

Schmidt, Ulrich; Birnbaum, Michael

Conference Paper

The Impact of Experience on Violations of Independence and Coalescing

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Risk and Ambiguity II, No. E06-V3

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Schmidt, Ulrich; Birnbaum, Michael (2014) : The Impact of Experience on Violations of Independence and Coalescing, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Risk and Ambiguity II, No. E06-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<http://hdl.handle.net/10419/100463>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The Impact of Experience on Violations of Independence and Coalescing*

Michael H. Birnbaum^a and Ulrich Schmidt^{b, c, †}

^a Department of Psychology, California State University, Fullerton, USA

^b Department of Economics, University of Kiel, Germany

^c Kiel Institute for the World Economy, Kiel, Germany

Abstract

The present paper reports a repeated experiment on decision making under risk where subjects have to tackle the same choice problems in several rounds. We fit a simple error model and investigate how behavior changes in the course of the experiment. Our analysis complements and extends Hey (2001) who analyzes for each subject the fit of several preference functionals from round to round. Instead, we focus on choice problems allowing for direct tests of independence and coalescing. We show that variability of responses as well as violations of independence and coalescing decrease from earlier to later rounds. Our results provide evidence in favor of expected utility in conjunction with the discovered preference hypothesis.

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

JEL classification: C91, D81

December 2012

* † 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

An important issue in the behavioral economics literature is the question whether observed behavioral biases and market anomalies are persistent or whether they tend to wash out with increasing experience of subjects. To analyze this question, many studies run repeated experiments and compare behavior in later rounds with that in earlier rounds (e.g. Cox & Grether, 1996; Loomes, Starmer & Sugden, 2003; Dufwenberg, Lindqvist, and Moore, 2005; Plott and Zeiler, 2005). Also this paper presents a repeated experimental study and investigates individual decision making under risk. In this context, two factors could contribute to differing choice behavior between the repetitions. First, it has been regularly observed that subjects often make errors when choosing between risky alternatives (early references include e.g. Camerer, 1989; Starmer and Sugden, 1989; Harless and Camerer, 1994; Hey and Orme, 1994). It may well be the case that the frequency of such errors decreases with experience of subjects. Second, the well-known discovered preference hypothesis (Plott, 1996; see also Smith, 1989; Harrison, 1994; Binmore, 1999) proposes that “individuals have a consistent set of preferences over states, but such preferences only become known to the individual with thought and experience” (Myagkov and Plott, 1997, p. 821). This implies that choice behavior in later rounds may reveal different preferences than in earlier rounds and, in particular, that observed irrationalities should decrease with experience.

In this paper we collect repeated binary choice data and employ a simple error model in order to disentangle “true” preferences from errors. This allows us to detect whether errors and/or true preferences change between repetitions. If true preferences change between repetitions, an obvious question is whether typical violations of expected utility (EU) decrease with experience, as being implied discovered preference hypothesis. In our study we focus on two important failures of EU, violations of independence and coalescing. The common consequence and common ratio effect of Allais are well-known experimental designs where substantial violations of independence have been observed. These violations motivated the development of alternative theories like rank-dependent utility (Quiggin, 1981; 1982; Luce 1991; Luce and Fishburn, 1991), cumulative prospect theory (Starmer and Sugden, 1989; Tversky and Kahneman, 1992; Wakker and Tversky, 1993) and configural weight models like

TAX and RAM (Birnbau and McIntosh, 1996) which imply weaker independence conditions. Our experimental design involves besides common consequence and common ratio effects also tests of these weaker conditions. Coalescing demands that if two branches in a gamble lead to the same consequence, they can be combined by adding their probabilities without altering the utility of the gamble. Violations of coalescing (also called splitting effects) have been observed in a number of studies (Starmer and Sugden 1993; Humphrey 1995, 2001). They are particularly troublesome as they can be utilized to generate violations of first-order stochastic dominance (Birnbau and Navarette, 1998). Therefore, it would be good news if these violations would decrease with experience.

In order to compare our study with the related literature, one should distinguish – as Myagkov and Plott (1997) do – between “learning by thought” and “learning by experience”. In our study subjects have to tackle each choice problem altogether four times without any feedback given between the choices. This means that we can only observe learning by thought as subjects do not experience the consequences of their choices. A study of the common ratio effect with opportunities to learn by thought and experience has been performed by Barron and Erev (2003). Here subjects have to repeat a given choice problem 400 times and after each choice the lottery was played and paid out for real. For answering our question such a setup is problematic since it may involve wealth and portfolio effects. Although expected value maximization is clearly optimal for such a large number of independently played out lotteries, many subjects did not adhere to it. Also van de Kuilen & Wakker (2006) analyze learning by thought and learning by experience in a common ratio design but avoid possible wealth and portfolio effects by employing the random lottery incentive system. Subjects have to make choices in 15 common ratio problems and after each choice the preferred lottery is played out. At the end of the experiment one of the choices is randomly selected and the previously determined payoff is paid out for real. There is also a control group which did not receive any feedback after the single choices which means that subjects in the control group could only learn by thought. According to the results of van de Kuilen and Wakker (2006) learning by thought and experience leads to a significant increase of consistency with EU (the violation rate of independence decreases from 46.15% in the first round to 23.08% in the fifteenth round) while learning by thought only does not. This interesting study does, however, not disentangle between true preferences and errors. So it may well be the case that learning reduces errors and therefore erroneous violations of independence whereas true preferences remain unaffected.

There is indeed strong evidence that error rates decrease with repetition, see Hey (2001). His study is closely related to ours as it also considers repeated binary choice problems without feedback (learning by thought only). When analyzing the data, he considers each subject and each of the five repetitions separately. More precisely, he fits for each subject the parameters of a number of different preference functional (including EU and alternative theories like rank-dependent utility) repetition by repetition and compares their goodness of fit. He finds indeed some (limited) evidence that the majority of subjects is converging to EU as best fitting functional as required by the discovered preference hypothesis. As only learning by thought is involved, this conclusion is in contrast to the results by van de Kuilen & Wakker (2006).

Our study can be regarded as a complement to the analysis of Hey (2001). While our experimental setup is rather similar we analyze data choice problem by choice problem and not subject by subject. Our procedure has some disadvantages (in particular not taking into account the different degree of variability in subjects' responses) but does not incur the potential problems of fitting preference functional like prespecifying functional forms. By focusing on choice problems instead of subjects we can additionally analyze whether violation rates for some independence conditions decrease more than for others. Moreover, our analysis of violations of coalescing also requires distinguishing between choice problems instead of subjects.

Altogether, the goal of the present study is to analyze the impact of learning by thought on violations of independence and coalescing while controlling for errors. The next section presents our experimental design. Section 3 presents our error model. Note that this model is neutral with respect to independence and coalescing, i.e. true preferences may satisfy or violate these properties. Section 4 presents our results and Section 5 contains some concluding observations.

2 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 from lowest to highest. Explanation and playing out of lotteries involved a container containing numbered tickets from one to 100. Suppose a subject could for instance play out lottery A in Figure 2. Then she would win 20 Euro when drawing a ticket from 1 to 50, 30 Euro for a ticket between 51 and 80, and 40 Euro for a ticket between 81 and 100. All this was explained in the instructions which were given to the students in printed form 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.

<i>A:</i>	<i>50% to win 20 Euro</i>	<i>B:</i>	<i>33% to win 10 Euro</i>
	<i>30% to win 30 Euro</i>		<i>34% to win 15 Euro</i>
	<i>20% to win 40 Euro</i>		<i>33% to win 60 Euro</i>

Figure 1: 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. 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 were investigated in both, coalesced and split forms. All these 20 tests were replicated four times with counterbalanced left-right positioning. Additionally, in order to test the attentiveness of subjects, each booklet included 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 forms 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. Presenting subjects the lottery pairs in both coalesced and split form allows us to test whether coalescing is satisfied, i.e. whether choices in the split form do not differ systematically from those in the coalesced form.

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

The first six tests of independence in Table 3 are four common consequence effects (CCE1-4) and two common ratio effects (CRE1 and 2). Such tests have been widely used for testing the independence axiom of EU; 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 (α) 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 row 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 β 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' .

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 variants which are implied by rank-dependent utility, cumulative prospect theory, and configural weight models. A central property in this context is tail independence (TI) which was introduced by Green and Jullien (1988) using the term ordinal independence. Formally, TI demands that $S = (x_1, p_1; \dots; x_i, p_i; x_{i+1}, p_{i+1}; \dots; x_n, p_n) \succcurlyeq R = (x_1, p_1; \dots; x_i, p_i; x_{i+1}, q_{i+1}; \dots; x_n, q_n)$ if and only if $S' = (x_1, q_1; \dots; x_i, q_i; x_{i+1}, p_{i+1}; \dots; x_n, p_n) \succcurlyeq R' = (x_1, q_1; \dots; x_i, q_i;$

$x_{i+1}, q_{i+1}; \dots; x_n, q_n$) where $x_1 \geq x_2 \geq \dots \geq x_n$. This means 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 change 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 upper cumulative independence (UCI), which demands that decision weights depend only on cumulative probabilities. Formally, UCI demands that If $S = (s_1, p_1; s_2, p_2; \alpha, p_3) \prec R = (r_1, p_1; \gamma, p_2; \alpha, p_3)$ then $S' = (s_1, p_1 + p_2; \gamma, p_3) \prec R' = (r_1, p_1; \gamma, p_2 + p_3)$, where $\alpha > \gamma > s_2 > s_1 > r_1$. Substantial violations of UCI 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 the typical violating pattern is RS' .

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. 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 α is either the highest or the lowest outcome in both lotteries. If α 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.

3 The error model

Because the same person may make different decisions when faced with the same choice, it might happen that a person might show violations of certain properties simply due to error, or random variation in response. We use a model in which the probability that a person shows the observed preference pattern RS' is given as follows:

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

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

As noted by Birnbaum and Schmidt (2008), this model has been applied by Sopher and Gigliotti (1993) with the assumption that error rates are equal but transitivity is satisfied. The same data would refute transitivity if the error rates were assumed equal, as assumed by Harless and Camerer (1994). Thus, the conclusion one reaches depends on the specification of the errors. In order to reach stronger conclusions concerning Allais paradoxes, 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 error rates without assuming a utility model. This can be done by adding repetitions to presentations of a choice.

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

$$(2) \quad P(\text{RS}) = p(1 - e) + (1 - p)e = e(1 - e),$$

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

By adding one replication to both choices in a two-choice test of EU (R versus S and R' versus S'), we have four choices with 16 (4×4) possible response patterns, which have 15 degrees of freedom. But there are only 5 parameters to estimate from the data, two error terms and four probabilities of the four "true" response patterns. Because the four probabilities sum to 1, only three degrees of freedom are used in this estimation. We use the term *general model* to refer to the model in which all four true probabilities are allowed to be non-zero. This model is now over-determined, with 10 degrees of freedom left to test it. The EU theory is then a special case of this general model in which two of the true probabilities are fixed to zero ($p_{\text{SR}'} = p_{\text{SR}'} = 0$); therefore, the difference in chi-squares provides a chi-square test with two degrees of freedom. In sum, with replications we can estimate the error terms and test the applicability of EU model.

The data can be evaluated by Chi-Square statistics:

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

where f_i is the observed frequency and q_i is the predicted frequency of a particular response pattern. Parameters are selected to minimize this statistic. The difference in χ^2 between a fit of the model that allows all four true patterns to have non-zero probabilities and the special case in which $p_{\text{RS}'} = p_{\text{SR}'} = 0$ 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 might be caused by errors in the response of subjects.

Note that our experimental design involves not only two but four replications of each choice problem. Since fitting our error model requires only two replications, we can fit our

model for replications 1 and 2 and for replications 3 and 4. By comparing the results we can analyze how errors and true preferences change from the first two replications to the third and fourth replication.

No.	risky choices				difference of risky choices			inconsistencies		
	Rep 1	Rep 2	Rep 3	Rep 4	Rep 2 - 1	Rep 3 - 2	Rep 4 - 3	1 & 2	2 & 3	3 & 4
1	94	96	96	98	2	0	2	6	4	2
2	83	85	81	83	2	4	2	9	11	2
3	87	85	89	89	2	4	0	17	11	7
4	85	80	89	85	5	9	4	20	17	4
5	67	57	43	56	10	14	13	31	22	24
6	63	61	52	57	2	9	5	24	13	20
7	28	33	31	28	5	2	3	17	17	15
8	46	57	52	56	11	5	4	26	24	22
9	50	30	35	33	20	5	2	31	20	9
10	7	7	13	15	0	6	2	11	13	13
11	17	26	28	24	9	2	4	28	28	15
12	22	19	22	20	3	3	2	30	11	9
13	22	31	33	31	9	2	2	20	17	20
14	56	56	41	48	0	15	7	41	26	12
15	61	78	74	70	17	4	4	28	11	7
16	11	19	13	17	8	6	4	7	13	7
17	54	52	59	61	2	7	2	28	22	17
18	13	11	11	13	2	0	2	6	7	2
19	35	31	26	24	4	5	2	19	20	9
20	52	59	41	48	7	18	7	19	33	11
21	22	17	22	24	5	5	2	20	17	6
22	22	33	26	28	11	7	2	22	30	13
23	7	6	4	6	1	2	2	6	2	2
24	13	7	4	11	6	3	7	13	4	7
25	91	81	76	72	10	5	4	17	13	11
26	81	81	78	72	0	3	6	11	19	20
27	41	43	44	43	2	1	1	28	28	17
28	33	33	26	33	0	7	7	41	26	15
29	76	85	85	83	9	0	2	20	22	17
30	41	37	50	46	4	13	4	37	31	26
31	67	70	70	81	3	0	11	22	30	11
32	56	72	80	81	16	8	1	31	22	20
33	89	89	94	93	0	5	1	15	9	6
34	70	80	80	78	10	0	2	28	26	17
35	89	89	91	94	0	2	3	11	13	11
36	80	89	91	91	9	2	0	17	9	11
37	78	80	72	65	2	8	7	17	15	15
38	59	57	61	72	2	4	11	31	30	33
39	81	80	81	80	1	1	1	20	13	13
40	67	57	56	57	10	1	1	35	28	24
41	59	59	63	54	0	4	9	30	37	31
42	31	37	37	33	6	0	4	20	22	19
mean	53	54	53	54	5.02	4.79	3.83	21.67	18.71	13.62

Table 2: Descriptive results

4 Results

(i) descriptive analysis

A first glance at our results can be taken from Table 2 which presents some descriptive results. Columns two to five give for each lottery the percentage of subjects choosing the risky lottery in each of the four repetitions. The last row of the table shows that there is on average not much difference between the replications. In columns 6, 7, and 8 we calculated the absolute value of differences of these percentages between replications. If behavior becomes more consistent by learning one would expect that differences between replications 3 and 4 are lower than differences between the other replications. The last row of the Table shows that this is indeed true but this result is significant only at the 10%-level. Probably more informative is the analysis of inconsistencies in columns 9-11. An inconsistency occurs if a subject in a given choice problem chooses the risky lottery in one repetition and the safe lottery in another repetition. The columns report the relative frequency of such inconsistencies between repetitions 1 & 2, 2 & 3, and 3 & 4, respectively. As the last row reveals, inconsistencies decrease on average substantially (from 21.67% to 13.62%) in the course of the experiment. All the differences of inconsistencies between repetitions 1 & 2, 2 & 3, and 3 & 4 are significant at the 1%-level.

(ii) the error model

As explained in the preceding section, we need two replications to fit our error model. We use replications 1 & 2 and 3 & 4 to analyze how behavior changes in the course of the experiment. The results are presented in Table 3. For each lottery pair columns 2, 4, and 6 report respectively the estimated fraction of people preferring the risky lottery, the error rate, and the test statistic for the fit of the true and error estimated for the data from replications 1 and 2. Columns 3, 5, and 7 report corresponding information for the error model fitted with data from replications 3 and 4. From the last two columns we can infer that the true and error model fits the data quite well. Out of the 84 tests, the model can be rejected in only four cases at a significance level of 5%. Since four is also the number of expected violations in 84 test at this significance level we can conclude that our error model fits the data very well and that there is no need to consider more complex error models which are e.g. discussed in Birnbaum and Schmidt (2012).

No.	Rep 1&2	Rep 3&4	error 1&2	error 3&4	χ^2	χ^2
					Rep1&2	Rep 3&4
1	98	98	3	1	0,33	0,83
2	88	83	5	1	0,20	0,83
3	94	92	9	4	0,11	0,00
4	92	89	12	3	0,81	1,67
5	70	49	20	16	1,45	3,59
6	67	56	14	12	0,08	0,81
7	26	26	10	8	0,98	0,49
8	53	55	17	13	2,49	0,33
9	32	33	23	5	6,69	0,20
10	2	8	6	7	0,00	0,14
11	7	21	17	8	1,63	0,49
12	4	18	18	5	0,25	0,20
13	20	27	13	12	2,19	0,09
14	63	45	28	8	0,00	3,25
15	80	74	19	4	5,10	0,95
16	12	12	5	4	3,36	0,95
17	54	62	17	9	0,07	0,11
18	10	11	3	1	0,33	0,83
19	29	22	11	5	0,40	0,20
20	57	44	11	7	1,55	2,45
21	10	22	12	3	0,81	0,33
22	20	23	14	7	2,87	0,14
23	4	4	3	1	0,33	0,83
24	4	4	8	5	1,24	3,36
25	95	77	10	6	2,62	0,65
26	86	83	6	12	0,00	0,81
27	38	42	17	9	0,07	0,11
28	11	26	28	9	0,00	1,91
29	90	92	13	9	2,19	0,11
30	28	47	25	15	0,20	0,28
31	75	80	13	8	0,33	5,08
32	74	90	22	12	4,56	0,09
33	96	96	8	3	0,00	0,33
34	88	85	17	9	1,63	0,11
35	94	98	6	6	0,00	0,65
36	92	96	10	6	2,62	0,00
37	85	72	9	9	0,11	1,91
38	64	79	20	22	0,06	1,96
39	90	86	12	7	0,09	0,14
40	72	59	24	14	1,30	0,08
41	65	64	18	20	0,00	1,45
42	30	31	12	11	0,81	0,40
mean	54.02	54.31	13.52	8.00	1.19	0.93

Table 3: The error model

Our estimated preferences for the risky lottery do on average not change much in the course of the experiment. The correlation coefficient between values estimated from

repetitions 1 & 2 and 3 & 4 equals 0.96 and significant at the 1%-level. In line with the change of observed inconsistencies between replications, columns 4 and 5 of Table 3 show that estimated error rates are substantially decreasing, on average from 13.52% to 8.00%. This decrease is significant at the 1%-level.

Property	Observed		Estimated	
	Rep 1-2	Rep 3-4	Rep 1&2	Rep 3&4
CCE1 _C	17	17	10	15
CCE1 _S	17	7	5	5
CCE2 _C	39	33	22	25
CCE2 _S	34	33	22	27
CCE3 _C	38	26	31	24
CCE3 _S	25	10	0	4
CCE4 _C	38	32	41	18
CCE4 _S	36	21	27	12
CRE1 _C	55	57	62	60
CRE1 _S	44	48	43	49
CRE2 _C	35	27	26	21
CRE2 _S	23	15	4	7
LDI _C	7	5	0	0
UDI _C	14	14	5	0
UTI _C	49	45	54	42
UTI _S	29	18	5	6
LTI _C	25	20	12	11
LTI _S	14	9	1	0
UCI _C	41	33	25	10
UCI _S	22	25	13	14
LCI _C	40	33	17	7
LCI _S	39	32	32	23
mean	31	26	21	17
mean _C	33	29	25	19
mean _S	28	22	15	15
EU	33	27	24	22
EU _C	37	32	32	27
EU _S	30	23	17	17
NEU	28	24	16	11
NEU _C	29	25	19	12
NEU _S	26	21	13	11

Table 4: Violations of Independence

(iii) violations of independence

Violations of independence are analyzed in Table 4. In columns 2 and 3 we report for each independence test (subscript C denotes a test in coalesced form, subscript S in split form) the observed violation rate, for ease of comparison averaged over repetitions 1 & 2 and over repetitions 3 & 4. Columns 4 and 5 report the results of fitting our error model applied to the

two choice problems of each independence test. The table shows that true violation rates are in most cases lower than observed ones, indicating that a part of observed violations is caused by errors. The row “mean” reports the average violation rates. These are only slightly decreasing (5 percentage points for the observed and 4 percentage points for the estimated violations) but the decrease is significant at the 1%-level for observed violations and at the 5%-level for estimated ones. The two rows below distinguish between violations in coalesced and split form, indicating that the latter occur less frequently. The subsequent rows distinguish between independence conditions which are only implied by EU and the weaker conditions which are also implied by non-expected utility (NEU) models. This analysis shows that the weaker independence conditions are less frequently violated, at least according to the estimated true preferences. The lowest violation rates occur for weaker independence conditions presented in split form.

Property	Problem	Observed		Estimated	
		Rep 1-2	Rep 3-4	Rep 1&2	Rep 3&4
1-3	CCE1	15	8	4	4
5-7	CCE2	46	27	43	14
9-11	CCE3	39	12	18	2
10-12	CCE3	26	17	7	9
13-14	CCE4	42	24	31	17
15-17	CRE1	30	19	12	11
19-21	CRE2	31	17	20	11
20-22	CRE2	39	29	33	21
29-31	UTI	31	23	20	14
30-32	UTI	42	42	36	34
33-35	LTI	15	10	2	2
34-36	LTI	28	23	12	12
38-39	UCI	39	25	21	3
41-42	-	44	44	37	33
mean		33	23	21	13

Table 5: violations of coalescing

(iv) violations of coalescing

Violations of coalescing are reported in Table 5. Observed violation rates are always higher than estimated ones, indicating that also for coalescing violations are partly caused by errors. The last row reports average violation rates which are substantially decreasing for both observed and estimated values. In both cases the decrease is significant at the 1%-level.

5 Conclusions

This paper presented a repeated experiment on decision making under risk where people tackle the same choice problems on four occasions. As lotteries are only played out at the very end any differences in behavior in the course the experiment can be only attributed to learning by thought but not to learning by experience. When analyzing data we control for errors by the true and error model and statistical analysis shows that this model fits responses quite well.

Our results show that learning by thought has a strong impact on risk preferences as errors, violations of independence, and violations of coalescing are significantly decreasing from earlier to later rounds. In view of the contradictory results on the impact of learning by thought in the studies of Hey (2001) and van de Kuilen and Wakker (2006), our study supports the results of Hey (2001). Altogether, as Hey (2001) we provide evidence that EU in conjunction with the discovered preference hypothesis seems to be a reasonable characterization of individual behavior towards risk.

References

- Barron, G., and I. Erev (2003). Small Feedback-Based Decisions and Their Limited Correspondence to Description-Based Decisions, *Journal of Behavioral Decision Making* 16, 215–233.
- Binmore, K.G. (1999). Why Experiments in Economics? *Economic Journal* 109, F16–F24.
- Birnbaum, M.H. (2001). Decision and Choice: Paradoxes of Choice. In N.J. Smelser and P.B. Baltes, (eds.), *International Encyclopedia of the Social and Behavioral Sciences*. Elsevier: Oxford.
- Birnbaum, M.H. (2005). Three New Tests of Independence that Differentiate Models of Risky Decision Making. *Management Science* 51, 1346-1358.
- Birnbaum, M.H., and Chavez, A. (1997). Tests of Theories of Decision Making: Violations of Branch Independence and Distribution Independence. *Organizational Behavior and Human Decision Processes* 71, 161-194.
- Birnbaum, M.H., and McIntosh, W.R. (1996). Violations of Branch Independence in Choices Between Gambles. *Organizational Behavior and Human Decision Processes*, 67, 91-110.
- Birnbaum, M. H., and Navarrete, J. B. (1998). Testing Descriptive Utility Theories: Violations of Stochastic Dominance and Cumulative Independence. *Journal of Risk and Uncertainty* 17, 17-49.

- Birnbaum, M. H., and Schmidt, U. (2008). An Experimental Investigation of Violations of Transitivity in Choice under Uncertainty. *Journal of Risk and Uncertainty*, 37, 77-91.
- Birnbaum, M.H., and Schmidt, U. (2012). Allais Paradoxes: Error, Restricted Branch Independence, or Coalescing. Working Paper.
- Camerer, C. (1989), An Experimental Test of Several Generalized Utility Theories. *Journal of Risk and Uncertainty* 2, 61-104.
- Cox, J.C., and Grether, D.M. (1996). The Preference Reversal Phenomenon: Response Mode, Markets and Incentives. *Economic Theory* 7, 381–405.
- Dufwenberg, M., Lindqvist, T., and Moore, E. (2005). Bubbles and Experience: An Experiment. *American Economic Review* 95, 1731-1737.
- Green, J.R., and Jullien, B. (1988). Ordinal Independence in Nonlinear Utility Theory. *Journal of Risk and Uncertainty*. 1, 355-387.
- Harrison, G.W. (1994). Expected Utility and the Experimentalists. *Empirical Economics* 19, 223–53.
- Harless, D., and Camerer, C. (1994). The Predictive Utility of Generalized Expected Utility Theories. *Econometrica* 62, 1251-1289.
- Hey, J.D. (2001). Does Repetition Improve Consistency? *Experimental Economics* 4, 5-54.
- Hey, J.D., and Orme, C. (1994). Investigating Generalisations of Expected Utility Theory Using Experimental Data, *Econometrica* 62, 1291-1326.
- Humphrey, S.J. (1995). Regret Aversion or Event-Splitting Effects? More Evidence under Risk and Uncertainty. *Journal of Risk and Uncertainty* 11, 263-274.
- Humphrey, S.J. (2001). Non-Transitive Choice: Event-Splitting Effects or Framing Effects? *Economica* 68, 77-96.
- Loomes, G., Starmer, C., and Sugden, R. (2003). Do Anomalies Disappear in Repeated Markets? *Economic Journal* 113, C153–C166.
- Luce, R.D. (1991). Rank- and Sign-Dependent Linear Utility Models for Binary Gambles. *Journal of Economic Theory* 53, 75-100.
- Luce, R.D., and Fishburn, P.C. (1991). Rank- and Sign-Dependent Linear Utility Models for Finite First-Order Gambles. *Journal of Risk and Uncertainty* 4, 29-59.
- Myagkov, M., and Plott, C.R. (1997). Exchange Economies and Loss Exposure: Experiments Exploring Prospect Theory and Competitive Equilibria in Market Environments. *American Economic Review* 87, 801–28.
- Plott, C.R. (1996). Rational Individual Behaviour in Markets and Social Choice Processes: The Discovered Preference Hypothesis', In K.J. Arrow, E. Colombatto,

- M. Perlman and C. Schmidt (eds) *The Rational Foundations of Economic Behaviour*, Macmillan: Basingstoke, 225–50.
- Plott, C.R., and Zeiler, K. (2005). The Willingness to Pay-Willingness to Accept Gap, the ‘Endowment Effect’, Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. *American Economic Review* 95, 530-545.
- 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 Behaviour and Organization* 3, 323—343.
- Smith, V.L. (1989). Theory, Experiment and Economics. *Journal of Economic Perspectives* 3, 151–69.
- Sopher, B., and Gigliotti, G. (1993). Intransitive Cycles: Rational Choice or Random Errors? An Answer Based on Estimation of Error Rates with Experimental Data. *Theory and Decision* 35, 311-336.
- Starmer, C., and Sugden, R. (1989). Probability and Juxtaposition Effects: An Experimental Investigation of the Common Ratio Effect. *Journal of Risk and Uncertainty* 2, 159-178.
- Starmer, C., and Sugden, R. (1993). Testing for Juxtaposition and Event-Splitting Effects. *Journal of Risk and Uncertainty* 6, 235-254.
- Tversky, A., and Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* 5, 297-323.
- Van de Kuilen, G., and Wakker, P.P. (2006). Learning in the Allais Paradox. *Journal of Risk and Uncertainty* 33, 155-164.
- Wakker, P.P., and Tversky, A. (1993). An Axiomatization of Cumulative Prospect Theory. *Journal of Risk and Uncertainty* 7, 147-176.
- Wakker, P.P., Erev, I., and Weber, E. (1994). Comonotonic Independence: The Critical Test between Classical and Rank-Dependent Utility Theories. *Journal of Risk and Uncertainty* 9, 195-230.
- Wu, G. (1994). An Empirical Test of Ordinal Independence. *Journal of Risk and Uncertainty* 9, 39-60.

APPENDIX

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