Online Supplement to Accompany "Testing Independence Conditions in the Presence of Errors and Splitting Effects"

in *Journal of Risk and Uncertainty*, 2017, *54*(1), xxx-xxx.

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

Appendix A: Choice Problems of Experiment 1

The choice problems of Experiment 1 are presented in coalesced form in Table 3 of our main paper. The choice problems in canonical split form (corresponding to Table 3) are presented in Table A.1 of this Appendix. In canonical split form, the probabilities on corresponding ranked branches in the two lotteries within a choice are equal.

The choice problems were selected from previous research because they had produced large violations of expected utility (EU) or cumulative prospect theory (CPT). Our lottery pairs testing Upper Cumulative Independence (UCI) were taken from Birnbaum, Patton, & Lott (1999). For the Common Ratio Effect (CRE): CRE1 is from Birnbaum (2001) and CRE2 is from Starmer and Sugden (1989).

The choice problems used in our tests of Upper and Lower Cumulative Independence (UDI and LDI) are taken from Birnbaum (2005). That paper and Birnbaum and Chavez (1997) reported either minimal violations or violations contrary to CPT with inverse-S weighting function. Studies that reported evidence against EU and CPT based on choice problems of Experiment 2 include Birnbaum (2004b, 2005a, 2005b, 2006, 2007).

Our reanalysis of data from Experiment 1, investigating the effects of practice, is described in Birnbaum and Schmidt (2015).

		Safe				Risky			
Problem	No.	p_1 s_1	p_2 s_2	p_3 s_3	p_4 s_4	q_1 r_1	q_2 r_2	q_3 r_3	q_4 r_4
CCE1$_S$	7	*0.80* 0	*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
CCE2$_S$	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	
CCE3$_S$	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	
CCE4$_S$	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
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	

Table A.1. Choice problems of Experiment 1 in canonical split form. 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′*.

Appendix B: Supplemental Analyses of True and Error Models

Let p represent the proportion of people who truly prefer R in the choice problem comparing S and R, and let e represent the error rate for this problem. The probability of choosing gamble R on both presentations (i.e., both repetitions of the same choice problem) is represented as follows:

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

In other words, those people who truly prefer R over S have correctly detected and reported their preference twice and those who truly prefer S have made two errors. The probability of switching from R to S is given as follows:

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

This is the same as $P(SR)$, so the probability of reversals is $2e(1 - e)$. Finally, the probability of choosing S twice is $P(SS) = pe^2 + (1 - p)(1 - e)^2$. In theory, this model should reproduce the four observed frequencies corresponding to $P(RR)$, $P(RS)$, $P(SR)$, and $P(SS)$, which have 3 degrees of freedom (since they sum to the number of participants). The model uses two parameters (p and e), leaving 1 degree of freedom (df) to test the model. The parameters are estimated to minimize the $\chi^2(1)$ between observed and predicted frequencies.

Table B.1 presents the $\chi^2(1)$ tests of fit corresponding to the estimated values of p and e in Experiment 2 that are given in Tables 6 and 7 of the main paper. There are two $\chi^2(1)$ tests of fit for each set of data: one is a test of the true and error model and the other is a test of response independence [i.e., $P(RR^*) = p(R)p(R^*)$, $P(RS^*) = p(R)p(S^*)$, etc.], based on the same exact data entries, with the same number of parameters and using the same number of degrees of freedom. The true and error model does not imply independence.

All 52 tests of response independence in Table B.1 are significant (median Chi-Squares in Samples 1 and 2 were 37.43 and 45.25, respectively), but only 4 of the 52 Chi-Squares for the true and error model were significant ($p < .05$) with medians of 0.74 and 0.33, respectively. At the 0.05 level, we expect 2.6 tests to be significant by chance; the binomial probability to observe 4 or more "significant" (out of 52 with $p = 0.05$) is 0.26; we can retain the hypothesis that this true and error model is compatible with these data. Table B.2 shows the response frequencies for each choice problem of Experiment 2, aggregated over both samples. Aggregated data lead to the same conclusions as data for the two samples fit separately.

Choice Problem	Sample 1 (n = 104) TE parameters		CHISQ		Sample 2 (n = 107) TE parameters		CHISQ	
	p	*e*	Indep	TE	*P*	*e*	Indep	TE
1	0.64	0.17	18.72	0.31	0.72	0.13	26.50	0.00
1a	0.40	0.11	38.91	1.77	0.37	0.11	37.28	0.05
2	0.60	0.13	31.69	1.08	0.65	0.08	49.56	0.00
2a	0.61	0.09	48.93	3.81	0.57	0.11	39.30	1.18
3	0.16	0.15	15.92	4.68	0.09	0.10	16.00	1.30
3a	0.69	0.23	7.10	0.67	0.46	0.14	30.58	0.99
4	0.80	0.13	20.97	1.08	0.88	0.19	7.30	3.82
4a	0.38	0.19	19.13	4.67	0.44	0.13	32.22	0.17
5	0.43	0.17	20.27	0.86	0.35	0.19	14.02	0.27
5a	0.42	0.21	10.78	0.03	0.38	0.09	47.98	0.53
6	0.09	0.15	7.26	2.42	0.13	0.15	9.96	0.04
6a	0.34	0.19	13.71	0.50	0.30	0.10	42.40	1.96
6c	0.18	0.08	35.95	0.07	0.09	0.13	8.94	0.00
7	0.34	0.06	62.53	2.18	0.35	0.05	67.48	0.40
7a	0.40	0.02	91.90	0.32	0.37	0.02	86.84	0.20
7b	0.39	0.05	69.89	0.11	0.38	0.02	90.75	0.00
7c	0.34	0.04	75.58	0.14	0.33	0.05	71.01	0.98
8	0.30	0.06	60.02	0.81	0.35	0.02	90.25	0.00
8a	0.40	0.03	80.57	0.00	0.37	0.02	90.59	0.00
8b	0.39	0.02	88.07	0.95	0.38	0.02	90.89	0.95
8c	0.34	0.06	71.02	7.59	0.31	0.04	73.89	0.49
9	0.59	0.02	91.99	0.32	0.58	0.02	87.55	0.20
9a	0.55	0.07	58.23	0.08	0.58	0.05	70.20	0.40
9b	0.69	0.05	67.68	0.98	0.65	0.08	64.40	7.36
10a	0.41	0.14	27.24	0.04	0.40	0.10	42.53	0.47
10d	0.34	0.18	14.91	0.03	0.34	0.15	21.79	0.00

Table B.1: Fit of the True and Error Model to Samples 1 and 2 of Experiment 2.

Choice	Response Pattern				Indep	Est. parameters		TE model
No.	*AA**	*AB**	*BA**	*BB**	CHISQ	*p*	*e*	CHISQ
1	52	28	25	106	45.2	0.68	0.15	0.17
1a	104	23	18	66	75.8	0.38	0.11	0.61
2	65	22	17	107	80.1	0.63	0.10	0.64
2a	71	12	25	103	88.5	0.59	0.10	4.46
3	144	25	19	23	30.6	0.12	0.12	0.82
3a	67	36	26	82	35.9	0.56	0.18	1.61
4	29	19	35	128	26.6	0.84	0.16	4.67
4a	92	34	19	66	52.3	0.41	0.15	4.18
5	89	30	32	60	34.0	0.39	0.18	0.06
5a	95	27	25	64	52.0	0.40	0.14	0.08
6	137	30	23	21	16.8	0.11	0.15	0.92
6a	108	24	26	53	51.0	0.32	0.14	0.08
6c	147	20	19	25	41.7	0.14	0.10	0.03
7	66	7	14	124	130.7	0.65	0.06	2.28
7a	78	4	4	125	178.7	0.62	0.02	0.00
7b	76	7	6	122	160.0	0.62	0.03	0.08
7c	65	10	6	130	146.5	0.67	0.04	0.99
8	64	6	9	132	149.5	0.67	0.04	0.59
8a	77	5	5	124	171.0	0.62	0.02	0.00
8b	78	4	4	125	178.7	0.62	0.02	0.00
8c	63	5	12	131	142.8	0.68	0.05	2.78
9	119	3	5	84	179.4	0.41	0.02	0.49
9a	106	12	11	82	128.1	0.44	0.06	0.04
9b	127	17	4	63	131.3	0.33	0.06	7.46
10a	68	20	24	98	68.9	0.59	0.12	0.36
10d	54	29	30	98	36.4	0.66	0.17	0.02

Table B.2: Frequencies of response patterns in two replications, aggregated over Samples 1 and 2 of Experiment 2 ($n = 211$).

Previous research has shown that there are individual differences in the rate of preference reversals in the same choice problems, as if some participants have more “noise” in their data. Would apparent violations of EU be eliminated or diminished if the error model allowed for individual differences in error rates? Put another way: is it possible that different sub-groups of participants who differ in their error rates also differ with respect to their conformance to EU? To investigate this question, participants of Experiment 2 were separated into groups, according to the number of preference reversals between two replicates, and the true and error model was fit in each group separately.

There were 107 “noisy” participants, who had 5 to 7 preference reversals between repetitions (out of 26 choice problems, i.e., 73% to 81% self-agreement), and 104 “low noise” participants who had 4 or fewer reversals (85% self-agreement or better). Analysis of partitioned data is presented in Table B.3. Estimated true choice probabilities are similar between these groups (correlation was $r = 0.90$), and the error rates for different choice problems were correlated ($r = 0.82$), but estimated values of e were (as expected) much higher in the “noisy” than “low noise” group (medians = 0.15 versus 0.06, respectively). However, large violations of EU (compare Problems 1 and 3 and 4 and 6) appear in both groups, and the violations of EU, RDU and CPT (compare Problems 1 and 1a, and 3 and 3a) also appear in both groups. Therefore, allowing for individual differences in noise levels does not change the conclusions with respect to EU, RDU, and CPT.

Choice Problem No.	Noisy Group (*n* = 107)				Low Noise Group (*n* = 104)			
	TE parameters		CHISQ		TE parameters		CHISQ	
	p	*e*	Indep	TE	*P*	*e*	Indep	TE
1	0.65	0.20	14.74	2.42	0.71	0.11	34.47	1.77
1a	0.31	0.13	31.32	2.09	0.45	0.10	44.27	0.22
2	0.73	0.16	16.98	0.31	0.54	0.05	67.77	0.40
2a	0.68	0.15	26.28	3.75	0.53	0.06	64.82	0.81
3	0.07	0.16	4.27	0.14	0.16	0.09	32.36	0.99
3a	0.59	0.30	2.63	0.55	0.54	0.09	47.45	1.45
4	0.84	0.23	4.38	1.32	0.85	0.10	30.41	4.52
4a	0.37	0.24	9.53	5.05	0.44	0.08	52.20	0.07
5	0.33	0.23	7.63	0.11	0.43	0.13	29.93	0.66
5a	0.31	0.26	4.66	0.61	0.44	0.06	64.51	0.81
6	0.02	0.19	0.17	0.03	0.18	0.11	25.51	1.77
6a	0.21	0.20	8.31	0.12	0.40	0.08	48.49	0.00
6c	0.10	0.14	8.44	0.15	0.16	0.07	39.88	0.08
7	0.32	0.09	45.95	1.44	0.37	0.02	87.79	0.95
7a	0.42	0.04	76.85	0.00	0.35	0.00	104.00	0.00
7b	0.42	0.05	66.75	0.09	0.35	0.01	95.63	1.66
7c	0.31	0.07	57.13	1.86	0.36	0.02	91.44	0.32
8	0.30	0.07	53.30	1.12	0.35	0.01	99.70	0.83
8a	0.41	0.04	73.19	0.11	0.36	0.01	99.75	0.83
8b	0.42	0.03	80.31	0.14	0.35	0.01	99.70	0.83
8c	0.27	0.09	46.87	3.81	0.37	0.01	99.80	0.83
9	0.69	0.03	77.40	0.14	0.49	0.01	100.07	0.83
9a	0.68	0.09	42.93	0.22	0.46	0.03	84.91	0.20
9b	0.79	0.09	43.25	5.81	0.57	0.03	84.98	1.67
10a	0.45	0.15	26.08	0.33	0.37	0.09	44.09	0.06
10d	0.31	0.27	3.86	0.10	0.35	0.09	44.37	0.06

Table B.3: Fit of the True and Error Model for two groups of participants who differed in number of preference reversals between replicates.

Appendix C: Expanded True and Error Model (TE-4 Model)

Although the model in which each choice problem has a different error rate fits the data well (Appendix B), this appendix considers a more complex model (Birnbaum 2012) in which the probability of making an error in a given choice problem might also depend on a person's true preference state. This assumption doubles the number of error rate parameters, but it might provide a means by which the EU model might be saved from data that would otherwise refute it. Thus, one would be adopting a more complex error model in hopes of retaining a simpler or more "rational" model of decision-making.

Let e and f represent the probabilities of making an error given that a person truly prefers S or R in the SR choice problem, and let e' and f' represent the probabilities of error given that a person truly prefers S' or R' in the $S'R'$ problem, respectively. Since there are two choice problems and two error rates for each problem, there are now four error rates in each test of EU. Hence, with two choice problems, this model is termed TE-4 and the model used in Appendix B (where $e = f$ and $e' = f'$) is a special case labeled TE-2.

According to TE-4, the probability to show the RS' response pattern on both replications is as follows:

(C.1) $P(RS', RS') = p_{RR'}(1-f)^2(f')^2 + p_{RS'}(1-f)^2(1 - e')^2 + p_{SR'}(e)^2(f')^2 + p_{SS'}(e)^2(1 - e')^2$

where $P(RS', RS')$ is the probability to observe RS' response pattern on both replications, and the other terms are as defined in Equation (1). There are 16 such equations for the 16 possible response patterns; the 16 corresponding observed frequencies have 15 degrees of freedom. The model uses four true probabilities of the response patterns (which sum to 1, using 3 df) and 4 error terms, so it uses 7 df, leaving 15 – 7 = 8 df to test the model.

Because the predicted frequencies of some cells are small in Experiment 2, cells representing the same combinations of response patterns were pooled prior to analysis, leaving 10 cell frequencies to be fit. (For example, frequencies of $SS'SR'$ and $SR'SS'$ were combined). Because these 10 frequencies sum to the number of participants, there are 9 df remaining in the data. There are at most 8 parameters to estimate from the data, which use 7 df (because $p_{SS'}$ + $p_{SR'}$ + $p_{RS'}$ + $p_{RR'}$ = 1), leaving at least 2 degrees of freedom in the tests.

The EU-4 model is the special case of TE-4 in which $p_{RS'} = p_{SR'} = 0$. The difference in fit between TE-4 and EU-4 is theoretically Chi-Square distributed with 2 df. The TE-2 model is the special case of TE-4 in which $e = f$ and $e' = f'$; within TE-2, EU-2 is a further special case in which $p_{RS'} = p_{SR'} = 0$. It is also possible to constrain the error terms to fit the replication data only (as in Table B.1); this constrained case of TE-2 is denoted TE-2c. The advantage of the constrained case is that the errors are estimated in a neutral way (from preference reversals in replications) that is independent of other assumptions, including EU.

The TE-4 model could, in principle, provide an error theory in which errors might be able to save EU. As it turns out empirically, however, even this model requires rejection of EU for the present data, as illustrated in Tables C.1 and C.2 for Problems 1 and 3. Table C.1 presents estimated parameter values for each model and indices of fit. Table C.2 presents observed and predicted frequencies for four models: TE-2c, EU-2c, TE-4, and EU-4.

Note that the *G* statistics (Chi-Squares) are acceptable for all three general versions of TE in both studies (Table C.1), so by the usual principle to prefer simpler models to more complex ones, there is no reason to adopt TE-4 over TE-2. The most parsimonious, acceptable model is TE-2c in both samples. According to this model, 49% and 66% of the subjects violated EU by having *RS′* as their true preference patterns in Samples 1 and 2, respectively.

However, our goal in this case is not to choose the most parsimonious error model, but rather to ask if by choosing a more complex error model, EU-4 might still be saved. The data give a clear answer to that question: No. The difference in *G* between a TE model and its special case, EU model is theoretically $\chi^2(2)$. All six differences exceed 9.21, which is the critical value of $\chi^2(2)$ with $\alpha = 0.01$, so EU can be rejected in all cases. Table C.2 shows that best-fit predictions of EU-4 deviate systematically from the observed frequencies.

Another statistical test helps pinpoint where the data refute EU. We can use a binomial test of the observed frequency of a *repeated* response pattern violating EU (*SR'SR'* or *RS'RS'*): does it exceed predicted frequency according to EU with this error model? In Samples 1 and 2, the observed frequencies of *RS'RS'* are 26 and 43 out of 104 and 107 participants, respectively. In all six cases (two samples by three models), observed violation patterns are much higher than the predictions of the best-fit EU model. EU-4 in Samples 1 and 2 implies predicted probabilities for the *RS'RS'* pattern of only 0.161 and 0.210. The binomial probabilities to observe 26 or 43 (or more) out of 104 and 107 are 0.013 and .000005, respectively. Thus, even with four error terms that depend on a person's true preference ("safe" or "risky") and the choice problem, EU can be rejected because the *RS'RS'* pattern is too frequent.

Sample 1									
Model	f	f'	e	e'	$p_{SS'}$	$p_{SR'}$	$p_{RS'}$	$p_{RR'}$	G
TE-4	0	0.10	0.34	0.15	0.58	0.04	0.27	0.11	0.50
EU-4	0	0.50	0.43	0.17	0.71	(0)	(0)	0.29	13.91
TE-2	**(= *e*)**	**(= *e*')**	0.17	0.14	0.35	0.02	0.49	0.14	2.46
EU-2	**(= *e*)**	**(= *e*')**	0.50	0.14	0.84	(0)	(0)	0.16	31.11
TE-2c	**(= *e*)**	**(= *e*')**	**0.17**	**0.15**	0.35	0.02	0.49	0.14	2.61
EU-2c	**(= *e*)**	**(= *e*')**	**0.17**	**0.15**	0.70	(0)	(0)	0.30	112.30

Sample 2									
Model	f	f'	e	e'	$p_{SS'}$	$p_{SR'}$	$p_{RS'}$	$p_{RR'}$	G
TE-4	0	0	0.33	0.11	0.46	0.05	0.47	0.03	1.16
EU-4	0.19	0.50	0.50	0.04	0.66	(0)	(0)	0.34	36.52
TE-2	**(= *e*)**	**(= *e*')**	0.13	0.10	0.25	0.03	0.66	0.06	1.58
EU-2	**(= *e*)**	**(= *e*')**	0.50	0.10	0.91	(0)	(0)	0.09	51.27
TE-2c	**(= *e*)**	**(= *e*')**	**0.13**	**0.10**	0.25	0.03	0.66	0.06	1.60
EU-2c	**(= *e*)**	**(= *e*')**	**0.13**	**0.10**	0.75	(0)	(0)	0.25	278.37

Table C.1: Estimated parameters and index of fit of true and error models to the frequencies of response combinations for Problems 1 and 3. Parameters that have been fixed or constrained are shown in parentheses, respectively.

Response	Study 1					Study 2				
Pattern	Data	TE-2c	EU-2c	TE-4	EU-4	Data	TE-2c	EU-2c	TE-4	EU-4
SS'SS'	18	19.5	36.2	19.2	16.8	18	17.4	49.2	17.3	16.7
SS'SR'	7	3.7	6.6	3.4	3.4	0	2.2	5.6	2.1	0.9
SS'RS'	11	8.9	7.4	9.9	12.5	10	8.8	7.3	8.7	17.8
SS'RR'	2	1.9	1.9	1.8	2.6	1	1.1	1.1	1.0	2.0
SR'SS'	1	3.7	6.6	3.4	3.4	4	2.2	5.6	2.1	0.9
SR'SR'	2	1.7	1.8	2.0	0.7	2	2.2	1.0	2.5	0.3
SR'RS'	1	1.9	1.9	1.8	2.6	0	1.1	1.1	1.0	2.0
SR'RR'	2	1.9	3.4	1.0	0.5	1	1.0	2.5	1.2	1.4
RS'SS'	10	8.9	7.4	9.9	12.5	8	8.8	7.3	8.7	17.8
RS'SR'	2	1.9	1.9	1.8	2.6	2	1.1	1.1	1.0	2.0
RS'RS'	26	26.4	2.0	26.0	16.8	43	43.5	1.3	44.3	22.4
RS'RR'	7	6.0	3.1	5.5	9.4	4	5.4	2.0	5.3	6.6
RR'SS'	1	1.9	1.9	1.8	2.6	0	1.1	1.1	1.0	2.0
RR'SR'	0	1.9	3.4	1.0	0.5	2	1.0	2.5	1.2	1.4
RR'RS'	4	6.0	3.1	5.5	9.4	8	5.4	2.0	5.3	6.6
RR'RR'	10	8.1	15.7	10.0	7.8	4	4.7	16.4	4.0	6.1

Table C.2: Observed frequencies of response patterns (Data) and predicted frequencies for Problems 1 and 3, according to best-fit true and error models.

Tables C.3 and C.4 show the analysis of Problems 1a and 3a, which are the canonical split forms of Problems 1 and 3. According to EU and CPT, how branches are split in a presentation should have no effect on how people choose between gambles. Instead, Problems 1a and 3a give opposite modal choices compared to Problems 1 and 3. Table C.2 shows that for Problems 1 and 3, there were 58 and 73 people who showed *RS*' pattern at least once in Samples 1 and 2 compared with 12 and 9 who showed the *SR*' pattern (not counting the cases who showed both patterns, *SR*'*RS*'), so the pattern *RS*' was significantly more frequent, $z = 5.50$ and 7.07, respectively. In contrast, Table C.4 shows that pattern *SR*' was more frequent than *RS*' in Problems 1a and 3a of Samples 1 and 2 (49 to 16 and 35 to 22; $z = -4.09$ and -1.72, respectively).

This effect due to splitting is reflected in the best-fit parameters in Tables C.1 and C.3. Whereas Table C.1 shows that the estimated true percentage (TE-2c) of subjects with *RS*' patterns is 49% and 66% in Samples 1 and 2, respectively, corresponding percentages in Table C.3 are only 5% and 10%, respectively. Whereas estimated percentages in Table C.1 of *SR*' are only 2% and 3%, they are 29% and 18% in Table C.3, respectively.

As in Table C.1, EU can be rejected in all six tests of Table C.3; in addition, observed frequencies of repeated patterns violating EU are significantly greater than predicted by EU for one or both repeated violation patterns, via binomial tests. Again, EU can be rejected with any of the error models, and CPT can be rejected because of the effects of splitting.

Tables C.5 and C.6 show response pattern frequencies for other pairwise tests of EU and CPT, which provide additional evidence that neither EU nor CPT can be saved for these data by this more complex error model. These cases have been fit to the true and error models with similar conclusions to those illustrated for Problems 1 and 3 in coalesced and canonical split forms.

Based on the principle of parsimony, one would prefer TE-2 or TE-2c over TE-4 for our data, since these models use fewer parameters and achieve fairly good fits. However, our purpose in this appendix is to ask if the more complex error models would allow us to retain the EU or CPT models. The results show that even when we allow these additional parameters, neither EU nor CPT can account for the systematic patterns of repeated violations of common consequence independence and coalescing.

Sample 1	Estimated Parameters of True and Error Model								Fit
Model	*f*	*f'*	*e*	*e'*	$p_{SS'}$	$p_{SR'}$	$p_{RS'}$	$p_{RR'}$	*G*
TE-4	0.21	0.00	0.02	0.44	0.48	0.00	0.22	0.29	2.82
EU-4	0.17	0.24	0.06	0.45	0.52	(0)	(0)	0.48	14.98
TE-2	**(= *e*)**	**(= *e'*)**	0.11	0.23	0.30	0.29	0.05	0.36	17.33
EU-2	**(= *e*)**	**(= *e'*)**	0.11	0.37	0.57	(0)	(0)	0.43	30.14
TE-2c	**(= *e*)**	**(= *e'*)**	**0.11**	**0.23**	0.30	0.29	0.05	0.36	17.36
EU-2c	**(= *e*)**	**(= *e'*)**	**0.11**	**0.23**	0.54	(0)	(0)	0.46	43.35

Sample 2	Estimated Parameters of True and Error Model								Fit
Model	*f*	*f'*	*e*	*e'*	*SS'*	*SR'*	*RS'*	*RR'*	*G*
TE-4	0.21	0.09	0.04	0.17	0.42	0.11	0.17	0.30	1.17
EU-4	0.20	0.31	0.07	0.30	0.55	(0)	(0)	0.45	24.09
TE-2	**(= *e*)**	**(= *e'*)**	0.11	0.14	0.44	0.18	0.10	0.28	2.95
EU-2	**(= *e*)**	**(= *e'*)**	0.12	0.31	0.61	(0)	(0)	0.39	26.21
TE-2c	**(= *e*)**	**(= *e'*)**	**0.11**	**0.14**	0.44	0.18	0.10	0.28	2.96
EU-2c	**(= *e*)**	**(= *e'*)**	**0.11**	**0.14**	0.60	(0)	(0)	0.40	52.71

Table C.3: Estimated parameters and fit of true and error models to the frequencies of response combinations for Choice Problems 1a and 3a, as in Table C.1.

Response Pattern	Sample 1 Data	TE-2c	EU-2c	TE-4	EU-4	Sample 2 Data	TE-2c	EU-2c	TE-4	EU-4
SS'SS'	13	16.0	26.3	15.6	14.6	29	28.2	37.9	29.1	25.6
SS'SR'	16	8.8	8.1	12.0	12.3	8	6.3	6.1	6.9	11.1
SS'RS'	2	2.5	3.6	1.6	1.3	2	4.3	4.8	3.3	2.6
SS'RR'	1	1.9	1.9	1.2	2.0	0	1.2	1.2	1.1	2.4
SR'SS'	11	8.8	8.1	12.0	12.3	6	6.3	6.1	6.9	11.1
SR'SR'	10	15.4	2.8	10.5	10.8	11	12.3	1.3	11.0	5.4
SR'RS'	2	1.9	1.9	1.2	2.0	0	1.2	1.2	1.1	2.4
SR'RR'	8	4.1	3.1	5.9	4.8	8	3.7	3.2	4.9	3.9
RS'SS'	2	2.5	3.6	1.6	1.3	5	4.3	4.8	3.3	2.6
RS'SR'	1	1.9	1.9	1.2	2.0	3	1.2	1.2	1.1	2.4
RS'RS'	6	4.3	2.5	4.7	1.9	8	7.3	1.2	8.2	3.1
RS'RR'	3	6.2	6.8	3.6	6.1	4	3.9	4.1	3.3	6.7
RR'SS'	0	1.9	1.9	1.2	2.0	1	1.2	1.2	1.1	2.4
RR'SR'	4	4.1	3.1	5.9	4.8	2	3.7	3.2	4.9	3.9
RR'RS'	3	6.2	6.8	3.6	6.1	3	3.9	4.1	3.3	6.7
RR'RR'	22	17.8	22.0	22.1	19.7	17	17.7	25.1	17.2	14.7

Table C.4: Observed and predicted frequencies of response patterns for Choice Problems 1a and 3a, as in Table C.2.

	Choice Problems			
Pattern	4 X 6	4a X 6a	2 X 5	2a X 5a
SS'SS'	18	75	34	55
SS'SR'	5	6	10	3
SS'RS'	17	12	12	6
SS'RR'	1	9	0	3
SR'SS'	5	8	10	10
SR'SR'	**1**	**2**	**11**	**3**
SR'RS'	1	7	1	1
SR'RR'	0	6	9	2
RS'SS'	29	12	10	15
RS'SR'	5	1	3	3
RS'RS'	**73**	**9**	**33**	**19**
RS'RR'	19	7	17	18
RR'SS'	1	2	3	4
RR'SR'	0	4	1	3
RR'RS'	16	9	18	10
RR'RR'	20	41	39	56
Total	211	210	211	211
G (EU-4)	18.85	1.79	28.06	11.88
SR'	11	26	41	21
RS'	154	49	90	68
Z	11.13	2.66	4.28	4.98
Binomial	0.222	0.028	0.108	0.05
	$p(x >=73)$	$p(x >=9)$	$p(x >= 33)$	$p(x >=19)$
	0.000026	0.138	0.019	0.010

Table C.5: Frequencies of response patterns in other tests of EU from the main design.

Response Pattern	Choice Problems					
	1 X 1a	3 X 3a	4 X 4a	6 X 6a	6a X 6c	5 X 5a
SS'SS'	41	61	24	86	92	61
SS'SR'	4	27	4	18	10	7
SS'RS'	20	4	9	11	15	9
SS'RR'	1	3	5	0	4	6
SR'SS'	4	14	1	17	5	11
SR'SR'	3	**42**	0	**16**	1	**10**
SR'RS'	2	4	3	4	2	6
SR'RR'	5	14	2	15	3	9
RS'SS'	17	2	20	10	22	14
RS'SR'	2	4	4	1	1	5
RS'RS'	**26**	0	**38**	1	**18**	**11**
RS'RR'	16	2	21	5	5	9
RR'SS'	3	6	4	5	2	2
RR'SR'	3	7	7	7	1	11
RR'RS'	9	2	11	0	10	6
RR'RR'	55	19	57	15	20	34
Total	211	211	210	211	211	211
G (EU-4)	12.93	12.79	9.32	1.40	7.40	8.69
SR'	19	104	14	73	20	48
RS'	88	10	99	27	70	49
z	6.67	-8.80	8.00	-4.60	5.27	0.10
Binomial	0.0360	0.0036	0.0096	0.2780	0.0442	0.0054

Table C.6: Response patterns in tests of coalescing (tests of EU and CPT).

Additional References in Online Supplement

Birnbaum, M. H. (2001). A Web-based program of research on decision making. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet Science* (pp. 23–55). Lengerich, Germany: Pabst Science Publishers.

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

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

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

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

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

Birnbaum, M. H. (2012). A true and error model with error rates that depend on the true preference state. Manuscript.

Birnbaum, M. H., & Chavez, A. (1997). Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes*, 71, 161–194.

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

Birnbaum, M. H., & Schmidt, U. (2015). The impact of learning by thought on violations of independence and coalescing. *Decision Analysis, 12*, 144-152. http://dx.doi.org/10.1287/deca.2015.0316.