UNSW Business School

Working Paper

Never Stand Still   Business School   Economics

UNSW Business School Research Paper No. 2015 ECON 14

Now you see it, now you don't: How to make the Allais Paradox appear, disappear, or reverse

P. Blavatskyy
A. Ortmann
V. Panchenko

Now you see it, now you don't: *How to make the Allais Paradox appear, disappear, or reverse*

Pavlo Blavatskyy[1], Andreas Ortmann[2] and Valentyn Panchenko[3] *

Abstract: The Allais Paradox, or Common Consequence Effect to be precise, is one of the most well-known behavioral regularities in individual decision making under risk. A common perception in the literature, which motivated the development of numerous generalized non-expected utility theories, is that the Allais Paradox is a robust empirical finding. We argue that such a perception does not accurately reflect the experimental evidence on the Allais Paradox and show how specific choices of parameters can make it appear, disappear, or reverse. For example, our results suggest that the Allais Paradox is likely to disappear when lotteries involve relatively small outcomes under real financial incentives and probability distributions are described as compound lotteries or in a frequency format (rather than as reduced-form simple lotteries). We also find that the Allais Paradox is likely to get reversed when lotteries are designed with an even division of the probability mass between the lowest and the highest outcomes.

JEL Classification Codes: D01; D81

Keywords: Decision under risk; the Allais Paradox; Common Consequence Effect; Expected Utility; Fanning-out; Experimental Practices

[1] School of Management and Governance, Murdoch University, 90 South Street, Murdoch, WA 6150, AUSTRALIA, Ph: +61 (0) 89360 2838, Fax: +61 (0) 89360 6966, Email: pavlo.blavatskyy@murdoch.edu.au, pavlo.blavatskyy@gmail.com

[2] Corresponding author, School of Economics, UNSW Business School, UNSW Australia, Sydney, NSW 2052, AUSTRALIA, Ph: +61 (0) 2 9385 3345, Email: a.ortmann@unsw.edu.au, aortmann@gmail.com

[3] School of Economics, UNSW Business School, UNSW Australia, Sydney, NSW 2052, AUSTRALIA, Ph: +61 (0) 2 9385 33 63, Email: v.panchenko@unsw.edu.au, valentyn.panchenko@gmail.com

*We thank conference and seminar participants at ANZWEE 2013 in Brisbane, AP ESA 2013 in Tokyo, FUR in Rotterdam, CERGE-EI, Nanyang Technical University, University of Passau, University of Ulm, and University of UNSW for feedback. Special thanks to Ganna Pogrebna.

1. Introduction

It is well known, and widely acknowledged (e.g., Hertwig & Ortmann 2001; Camerer 2003, p. 34), that the way one conducts an experiment is “unbelievably important”. Any test of a theory, such as Expected Utility Theory, is always a joint test of the theory and the design and implementation choices an experimenter makes (Smith 2002, p. 98). It is well-established that such choices can make a difference between the acceptance and rejection of a theory (e.g., Cherry et al. 2002; or of particular relevance here: Huck & Müller 2012). Hence any single study is only worth so much and ultimately it takes a body of evidence to establish the reliability of laboratory results. How exactly some such body of evidence gets produced and evaluated is a problem that has gained considerable attention and is at heart of important methodological controversies and debates both in economics (e.g., Grether & Plott 1979; Harrison 1989, 1992; Plott & Zeiler 2005, 2011; Cason & Plott 2014) and psychology (e.g., Kahneman & Tversky 1996; Gigerenzer 1991, 1996). In essence, to "emphasize what psychologists and experimental economists have learned about people, rather than *how* they have learned it" (Rabin 1998, p. 12) is a problematic strategy because the acceptance and rejection of a theory does depend on – sometimes subtle – details of design and implementation.

One path now increasingly trodded in economics are meta-studies, i.e., ways to sample the available evidence in a systematic, replicable, and well-documented manner (e.g, Engel 2011; Zhang & Ortmann 2014) that allows the quantification of the impact of key design and implementation characteristics, is important for the appropriate powering up of experimental studies, and allows us to predict under what conditions particular effects, or paradoxes, are likely to show up. Our study is a close relative to such undertakings.

Expected Utility Theory, arguably one of the cornerstones of the economic modeling edifice, has been tested in hundreds of studies. Prominent among these were tests proposed by Allais and Ellsberg which seemed to contradict EUT. Indeed, a widespread perception existed for decades that these paradoxes were robust empirical findings. Certainly the considerable amount of work that went, and continues to go, into the formulation of Non-Expected Utility theories suggests that much (Starmer 2000). In the present paper we explore the wide-spread perception that the Allais Paradox (AP) is a robust empirical finding [4].

[4] At a conference in Bratislava a few years back PB and AO got into an argument over what exactly the evidence is, PB maintaining that it was robust and in favor of the AP and AO contesting that claim. Rather than duelling each other, they decided to solve their differences in perception by something akin to an adversarial collaboration (e.g., Mellers et al. 2002), only without an arbiter. Later PB and AO asked VP to join forces since they realized that they were out of their depth once they got into serious estimation issues.

The paper is organized as follows. In section 2 we describe the AP. Section 3 reviews the existing literature on the AP. In section 4 we summarize our research methodology and present our results. In section 5 we examine experimental data collected by Loomes and Sugden (1998). Section 6 discusses the results. We conclude in section 7.

2. The Allais Paradox

Allais (1953, p.527) designed a thought experiment to challenge the descriptive validity of Expected Utility Theory. This thought experiment became widely known as the AP, or the Common Consequence Effect. Allais (1953, p. 529-530) also designed a second thought experiment, closely related to the first. This second Allais example -- in contemporary terminology known as the Common-Ratio Effect -- is sometimes also referred to as the AP (*e.g.*, van de Kuilen & Wakker 2006). In this paper we discuss only the first Allais example (the Common-Consequence Effect) and we do not deal with the second Allais example (the Common-Ratio Effect).

The Allais Paradox consists of two related hypothetical decision problems. In the following we call them Allais questions. First, a decision maker is asked to choose between two options A and B. Option A yields ₣100 million for certain. Option B yields ₣500 million with probability 0.1, ₣100 million with probability 0.89 and nothing with probability 0.01. Second, a decision maker is asked to choose between another two options C and D. Option C yields ₣100 million with probability 0.11 and nothing with probability 0.89. Option D yields ₣500 million with probability 0.1 and nothing with probability 0.9.

It is conventional to illustrate the AP in the Marschak-Machina probability triangle (Machina 1982). The horizontal (vertical) axis on Figure 1 shows the probability of the lowest (highest) outcome. The set of all probability distributions over three outcomes can be represented as a rectangular triangle with a side length of one. Choice option A is located at the origin (0,0), choice option B is located at the interior of the triangle at point (0.01,0.1) and so forth.

Choice options in Allais questions are constructed so that AB is parallel to CD and the length of AB equals the length of CD. Choice options in the Common Ratio Effect also involve two parallel lines AB′ and CD but choice option B′ is located on the hypotenuse (not in the interior of the triangle).

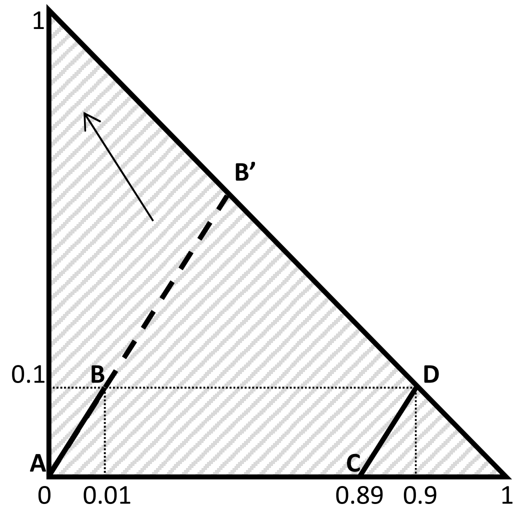

Figure 1 Illustration of the Allais Paradox in the Probability Triangle

It is straightforward to show (*e.g.*, footnote 4 in Huck & Müller, 2012, p. 264) that an Expected-Utility maximizer weakly prefers A over B if and only if she weakly prefers C over D. In the probability triangle, the indifference curves of an Expected Utility maximizer are positively-sloped parallel straight lines (one such family of indifference curves is shown as a map of grey lines in Figure 1). Since AB is parallel to CD then option B is located on a higher indifference curve than option A (as shown on Figure 1) if and only if option D is located on a higher indifference curve than option C.

A decision maker choosing A over B and D over C violates Expected Utility Theory (except for a special case when this decision maker happens to be exactly indifferent between A and B, which also implies indifference between C and D). This choice pattern is known, intuitively enough, as fanning-out. For A to be preferred over B the indifference curves must be relatively steep at the origin of the probability triangle (as shown in Figure 2). For D to be preferred over C the indifference curves must be relatively flat at the lower right corner of the probability triangle (as shown in Figure 2). Thus, when A is chosen over B and D is chosen over C, the map of indifference curves “fans out” in the probability triangle (see Figure 2). Similarly, when B is chosen over A and C is chosen over D, the map of indifference curves “fans in” and likewise violates Expected Utility Theory.

A general perception in the literature is that many people violate Expected Utility Theory in the two Allais questions. Moreover, these violations are highly asymmetric with the majority of people revealing the fanning-out choice pattern and only a minority revealing the fanning-in choice pattern. It is these two behavioural regularities that together became widely known as the Allais Paradox. In this paper we argue that the perception of the AP as a robust behavioral regularity does not accurately reflect actual experimental evidence, and that specific choices of parameters can make it appear, or disappear ad libitum. We discuss the implications of this finding in the Discussion and Conclusion sections

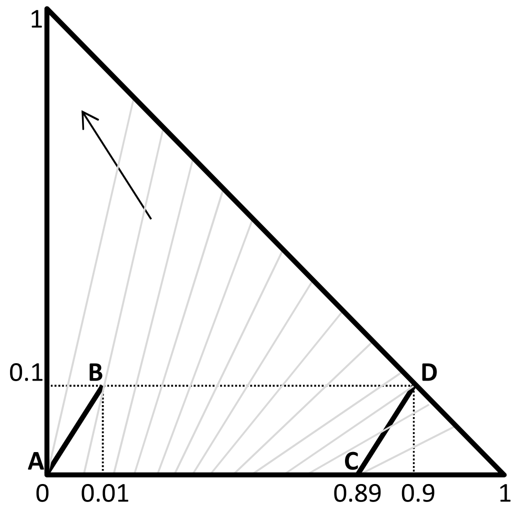

Figure 2 Fanning-out in the Probability Triangle

3. The Existing Literature

Allais (1953) originally designed his examples as a thought experiment. The tradition of thought experiments in individual decision making under risk can be traced back to the St. Petersburg Paradox (Bernoulli 1738). Arguably, no other field of economics saw such an extensive use of thought experiments as decision theory (other prominent examples are the Ellsberg, 1961, paradox and the recently proposed Machina, 2009, reflection example). The advantages of thought experiments in research on individual choice are eminent—the argument is more persuasive when a reader, who is as good as anybody else in the role of individual decision maker, finds herself with the incriminated choice pattern. This strategy has also been used to good effect by the proponents of the Heuristics & Biases program (e.g., , Kahneman 2003; Kahneman & Tversky 1979; Tversky & Kahneman 1974;).

Early experimental studies of the AP (e.g., Slovic & Tversky 1974) simply replicated the design of the Allais (1953) thought experiment (with the only substantial change apparently being a currency conversion of ₣100 million into $1 million and ₣500 million into $5 million). One of the most well-known studies Kahneman & Tversky (1979) justifies such non-incentivized experimental design as follows:

> “The method of hypothetical choices emerges as the simplest procedure by which a large number of theoretical questions can be investigated. The use of the method relies on the assumption that people often know how they would behave in actual situations of choice, and on the further assumption that the subjects have no special reason to disguise their true preferences. If people are reasonably accurate in predicting their choices, the presence of common and systematic violations of expected utility theory in hypothetical problems provides presumptive evidence against that theory.“ (Kahneman & Tversky 1979, p. 265).

Such an argument might be justifiable for behavioral regularities such as the Ellsberg (1961) paradox, which uses only a relatively small outcome ($100) that most subjects are likely to be familiar with. Yet, we believe that such an argument cannot be applied to the Allais paradox. Allais questions involve astronomically large monetary outcomes. Most individuals have little experience with such astronomical amounts in their daily life. Hence, they may find it difficult to imagine what their choices would be if the Allais questions were played out for real.

Whether the claim by Kahneman and Tversky is correct, is ultimately an empirical question. Laury & Holt (2008), for example, have demonstrated that the reflection effect documented in Kahnemann & Tversky (1979) fails to be the modal choice when this specific choice is properly incentivized. In a recent comprehensive study using a representative sample of the Dutch population as well as drawing on a standard subject pool, Huck & Müller (2012) find that their participants exhibit the AP for astronomically large hypothetical outcomes but show a significantly lower rate of Expected-Utility violations for low (real or hypothetical) outcomes. Similar evidence was found in earlier between-subject experiments. The AP is found, for example, in the basic version of Allais questions with astronomically large hypothetical outcomes in Conlisk (1989, Table 1, p. 395). Yet, Conlisk (1989, Appendix IV, p. 406-407) finds almost no Expected-Utility violations in a “pilot experiment” with small real outcomes. Camerer (1989, Table 7, p.92) finds that fanning-out choice patterns significantly outnumber fanning-in choice patterns when choice options have large hypothetical outcomes but not when choice options have small outcomes.

As documented, the first experimental studies of the AP with small real incentives appeared only at the end of 1980ies. By that time, a general consensus in the literature (coming from experiments with large hypothetical outcomes) had been established that the AP was a robust behavioral regularity and that in particular among those that violated EUT, the majority revealed a fanning-out choice pattern . This empirical finding motivated the development of numerous Non-Expected Utility theories.

The results of experimental studies with small real incentives that followed in the 1990s suggested that the AP was less wide-spread than the experiments with large hypothetical incentives seem to suggest (*e.g*., Harrison 1994, Section 1, pp. 226-231; Burke et al. 1996; Groes et al. 1999). In fact, several studies (*e.g*., Starmer 1992; Humphrey & Verschoor 2004; Blavatskyy 2013) even document a reversed AP when fanning-in choice patterns significantly outnumber fanning-out choice patterns. It has remained, until now, an open question how these findings could be reconciled.

The existing literature is focussed on the question whether there is a statistically significant asymmetry between fanning-out and fanning-in choice patterns. This pre-supposes that the reality of Expected-Utility violations is uncontested. We address both of these questions in this manuscript. There is tantalizing evidence from individual studies that suggest that the reality of the AP might be remarkably fragile. For example, Huck & Müller (2012) – in their recent and very comprehensive study -- find the AP in all treatments in that fanning-out choice patterns statistically significantly outnumber fanning-in choice patterns. Yet, in a laboratory experiment with low hypothetical (real) incentives only 4 (6) out of 79 subjects, *i.e*. only 5% (8%), reveal either a fanning-out or a fanning-in choice pattern. This seems hardly a threat for the validity of Expected Utility Theory; every theory that explains the behaviour of 9 out of 10, or even 19 out 20, is in our book remarkably successful.. (Yet, such a study might be cited as evidence of the AP contributing to the general perception that the paradox is a robust behavioural regularity.)

Apart from stakes and elicitation method, other design and implementation details are worth to look at. Several studies (*e.g*., Conlisk 1989; Bierman 1989; Carlin 1992) found that the AP is largely reduced when choice options in Allais questions are represented as compound lotteries rather than simple probability distributions. A similar effect was found when choice options are described in a frequency format (*e.g*., Carlin 1990). Arguably, frequency and compound lottery representations reduce cognitive load making both Allais questions an easier decision problem. This might decrease noise and imprecision in the revealed choice patterns and ultimately reduce the number of Expected-Utility violations. Huck & Müller (2012) have demonstrated that the choice of the subject pool also matters and interacts with stakes. In their high hypothetical treatment participants drawn from a representative sample of the population violate Expected-Utility Theory about 50 percent of the time while student subjects do so about 30 percent of the time. For the low-stakes treatments (both hypothetical and real) the violations of student subjects is about 10 percent for both conditions while that for participants drawn from a representative sample of the population it is about twice as high.

In addition, there are two “technical” design details that merit a closer look. Several studies reporting strong evidence of the AP designed Allais questions with the medium outcome being very close to the highest outcome (*e.g*., 2400 and 2500 Israeli pounds in Kahneman & Tversky 1979; 90 and 100 New Taiwanese dollars in treatments HR2 and CR2 in Fan 2002). Such design increases cognitive load making both Allais questions a harder decision problem. It is likely to increase imprecision and noise in the revealed choice patterns, which ultimately leads to a higher rate of

Expected-Utility violations. Blavatskyy (2010, experiment 2, pp. 232-235) found that the Common-Ratio Effect, not only disappears but is reversed when the medium outcome is moved away from the highest outcome. This finding suggests that a similar result exists for the Common-Consequence Effect.

The second noteworthy “technical” feature of the AP is an apparent similarity (or inconsequentiality) of probabilities in the second Allais question. In both questions, the riskier alternative can be obtained from the safer alternative by moving a probability mass of 0.11 away from the middle outcome (₣100 million) to the extreme outcomes. Allais divided this probability mass in uneven proportions between two extreme outcomes. Nearly all probability mass is allocated to the highest outcome (₣500 million). Specifically, a probability mass of 0.1 is allocated to the highest outcome and only a probability mass of 0.01 is allocated to the lowest outcome (zero). This uneven division of the probability mass creates a similarity (or inconsequentiality) of probabilities in the second Allais question.[5] In this question, decision maker faces a tradeoff between the middle outcome with a probability 0.11 and the highest outcome with a probability 0.1. Following a considerable literature on similarity considerations in these kind of problems (e.g., Leland 1994; Rubinstein 1998; see also the recent debate about the priority heuristic, Brandstaetter et al. 2008), one can argue that probability 0.11 is similar to (or approximately the same as) probability 0.1. This similarity (or inconsequentiality) can catalyze the AP. Indeed, experimental studies with an even division of the probability mass (*i.e.*, when lines AB and CD have a slope of one in the probability triangle) such as Starmer (1992), Humphrey & Verschoor (2004) and Blavatskyy (2013) all find the reversed Allais paradox when fanning-in choice patterns outnumber fanning-out choice patterns. Again, as of now, it was not clear how to reconcile these findings.

To summarize, the existing literature suggests that six design and implementation details might drive results of experimental studies on the Allais Paradox:

1) Outcome payoffs;
2) Whether incentives are hypothetical or real;
3) Framing of choice options;
4) Subject pool;
5) Ratio of the middle to the highest outcome;
6) Slope of lines AB and CD in the probability triangle.

Our meta-study suggests how the results in the literature can be reconciled.

[5] Allais (1953) writes that “Il y a lieu de noter que pour [la deuxième question] l'effet de complémentarité correspondant a une chance sur 100 de ne rien gagner est faible.” (Allais, 1953, p. 527)

Experiment	SS	SR	RS	RR	Conl-z	p-val	PH	PL	P	O	I	F	S
Cherry and Shogren (2007), no arbitrage	22	64	5	11	9.94	0.00	0.1	0.01	$5,257,62	0.2	0	1	0
Conlisk (1989), basic version	18	103	16	99	9.31	0.00	0.1	0.01	$8,787,34	0.2	0	1	0
Carlin (1992), experiment 1	16	42	4	27	6.92	0.00	0.1	0.01	$7,776,05	0.2	0	1	0
Carlin (1990), trial #1	12	27	2	24	5.63	0.00	0.1	0.01	$8,347,24	0.2	0	1	0
Huck and Müller (2012), HighHyp	82	136	62	121	5.44	0.00	0.1	0.01	$5,652,78	0.2	0	1	1
Huck and Müller (2012), LowReal	22	97	37	368	5.32	0.00	0.1	0.01	$28	0.2	1	1	1
Camerer (1989), large gains	3	17	1	9	5.11	0.00	0.1	0.1	$43,937	0.4	0	1	0
Cherry and Shogren (2007), pre cheap talk-arbitrage	16	34	7	4	4.97	0.00	0.1	0.01	$5,257,62	0.2	0	1	0
Cherry and Shogren (2007), pre real-arbitrage	11	33	7	3	4.91	0.00	0.10	0.01	$5,257,62	0.2	0	1	0
Carlin (1992), exp. 2, form AP8	9	40	9	50	4.87	0.00	0.1	0.01	$7,776,05	0.2	0	1	0
Huck and Müller (2012), LowHyp	22	77	29	373	4.76	0.00	0.1	0.01	$28	0.2	0	1	1
Da Silva, S., Baldo, D., & Matsushita, R. (2013)	38	42	13	15	4.20	0.00	0.33	0.01	$2,341	1.0	0	1	0
Fan (2002), CR2	15	55	21	111	4.05	0.00	0.1	0.01	$6	0.9	1	1	0
Cherry and Shogren (2007), post real-arbitrage	22	23	5	4	3.80	0.00	0.1	0.01	$5,257,62	0.2	0	1	0
Cherry and Shogren (2007), post cheap talk-arbitr	27	23	5	6	3.75	0.00	0.1	0.01	$5,257,62	0.2	0	1	0
Huck and Müller (2012), HighHyp lab	4	20	5	41	3.19	0.00	0.1	0.01	$5,652,78	0.2	0	1	0
Groes et al. (1999)	15	17	4	18	3.05	0.00	0.162	0.03	$1,700	0.9	0	0	0
Burke et al. (1996), fixed Allais	0	8	1	16	2.58	0.00	0.20	0.05	$14	0.5	0	1	0
Carlin (1992), exp. 2, form AP9	27	23	11	7	2.11	0.02	0.1	0.01	$7,776,05	0.2	0	0	0
Huck and Müller (2012), LowHyp lab	0	4	0	75	2.04	0.02	0.1	0.01	$28	0.2	0	1	0
Groes et al. (1999)	45	23	13	18	1.68	0.05	0.162	0.03	$1,700	0.9	1	0	0
Huck and Müller (2012), LowReal lab	1	5	1	67	1.65	0.05	0.1	0.01	$28	0.2	1	1	0
Burke et al. (1996), salient Allais	1	2	0	22	1.44	0.07	0.2	0.05	$14	0.5	1	1	0
Carlin (1990), trial #2	9	20	16	97	0.67	0.25	0.1	0.01	$8,347,24	0.2	0	0	0
Finkelshtain and Feinerman (1997)	26	22	20	112	0.31	0.38	0.1	0.01	$67,935	0.2	0	1	1
Camerer (1989), small gains, hypothetical	2	6	6	6	0.00	0.50	0.1	0.1	$18	0.5	0	1	0
Fan (2002), HR2	22	35	36	109	-0.12	0.45	0.1	0.01	$6	0.9	0	1	0
Conlisk (1989), pilot	0	2	3	44	-0.44	0.33	0.1	0.01	$44	0.2	1	1	0
Camerer (1989), small gains, real	4	1	2	3	-0.56	0.29	0.1	0.1	$18	0.5	1	1	0
Fan (2002), HR1	13	28	34	127	-0.76	0.22	0.1	0.01	$6	0.2	0	1	0
Humphrey & Verschoor (2004), Sironko	72	10	17	10	-1.35	0.09	0.25	0.25	$11	0.4	1	1	1
Humphrey & Verschoor (2004), Vepur	41	15	24	29	-1.45	0.07	0.25	0.25	$3	0.4	1	1	1
Fan (2002), CR1	2	8	15	177	-1.46	0.07	0.1	0.01	$6	0.2	1	1	0
Conlisk (1989), three-step version	49	23	36	104	-1.70	0.04	0.1	0.01	$8,787,34	0.2	0	0	0
Humphrey & Verschoor (2004), Ethiopia	43	12	25	20	-2.18	0.01	0.25	0.25	$11	0.4	1	1	1
Humphrey & Verschoor (2004), Guddimalakapura	45	18	34	21	-2.26	0.01	0.25	0.25	$3	0.4	1	1	1
Humphrey & Verschoor (2004), Bufumbo	35	11	27	23	-2.68	0.00	0.25	0.25	$11	0.4	1	1	1
Starmer (1992)	46	11	34	33	-3.59	0.00	0.1	0.1	$17	0.4	1	1	0
Blavatskyy (2013)	21	3	31	15	-5.82	0.00	0.25	0.25	$34	0.4	1	1	0

Table 1 Experimental data analyzed in this paper. Column "Experiment" lists experiments as labeled in the study from which they were taken. The relevant papers are asterisked in the References section. Column SS shows how many subjects chose A over B and C over D. Column SR (RS) shows how many subjects revealed a fanning-out (fanning-in) choice pattern. Column RR shows how many subjects chose B over A and D over C. Column O shows the ratio of the middle outcome to the highest outcome. Column Conl-z reports the Conlisk z-statistic and its p-value, respectively. The rows are ordered by the Conlisk z-statistic indicating fanning-out patterns in the top block, no paradox in the middle block (highlighted in grey-blue) and fanning-in patterns in the bottom block. The blocks are separated by thick black lines. Column PH (PL) shows the probability of the highest (lowest) outcome in lottery B in the first Allais question. Column P reports the highest payoff standardized to 2010 USD. Column I is a dummy variable that equals one if incentives are real and zero if they are hypothetical. Column F is a dummy variable that equals one if choice options are presented as lotteries (not in compound or frequency format). Column S is a dummy variable that equals one if subjects are not students.

4. Methodology and results

4.1. Data

We selected 39 experiments that were reported in 14 experimental studies and that all together contained 5035 observations. These studies are detailed in Table 1 and are preceded by an asterisk in the list of references. The studies were selected in April and May 2014 from the EconLit database with a string search "Allais paradox" OR "common consequence effect". The initial set of 75 references was whittled down by eliminating all non-experimental articles and working papers, i.e. only published papers reporting relevant experimental treatments were included.[6]

Note that columns SS and RR in Table 1 show how many subjects in each experiment revealed a choice pattern consistent with expected utility maximization. Column SR (RS) in Table 1 shows how many subjects revealed a fanning-out (fanning-in) choice pattern. Conlisk (1989) proposed a test statistic, the so-called Conlisk z-statistic, which takes values close to null under the null hypothesis of no Expected-Utility violation. Large positive values of the statistic indicate the AP (when fanning-out choice patterns SR outnumber fanning-in choice patterns RS). Large negative values of the statistic indicate the reverse AP (when fanning-in choice patterns RS outnumber fanning-out choice patterns SR). Experiments in Table 1 are listed in the decreasing order of the Conlisk z-statistic, *i.e*. experiments at the top of Table 1 document high rates of fanning-out choice patterns, experiments at the middle (highlighted in grey) show no systematic Expected-Utility violations, and experiments at the bottom document high rates of fanning-in choice pattern.

In addition, Table 1 reports the experimental design variables which might influence the results of the experimental study, as discussed in the previous section. Namely, column PH (PL) shows the probability of the highest (lowest) outcome in lottery B in the first Allais question. Column P reports the highest payoff standardized to 2010 USD.

---

[6] Our search identified several other experimental studies that we did not include for various reasons: Harless (1992) and Prelec (1990) considered lotteries inside the probability triangle, as does the displaced version in Conlisk (1989). L'Haridon & Placido (2008) did not respond to repeated requests for data. Li (2004) responded but could not retrieve the data. Mac Donald & Wall (1989) test the common ratio effect, as do Van Kuilen & Wakker (2006). Rao & Li (2011) is a study of intertemporal choice, as is Oliver (2003). Weber (2007) elicited indifferences in the Allais questions which is a different format from what we decided to study. Our search in EconLit did surprisingly not turn up studies such as Birnbaum (2007), Harrison (1994), List &Haigh (2005), and Starmer & Sugden (1991), for reasons that we understand only partially (e.g., the title of Birnbaum, 2007, mentions "Allais paradoxes"; it was probably the plural that had this paper not show up in our search). We are currently building an even more comprehensive database that includes these and additional studies; the results so far confirm the findings reported in the body of the text, as one might expect given the number of observations and independent studies already in our database. (See Table 5 in the Appendix).

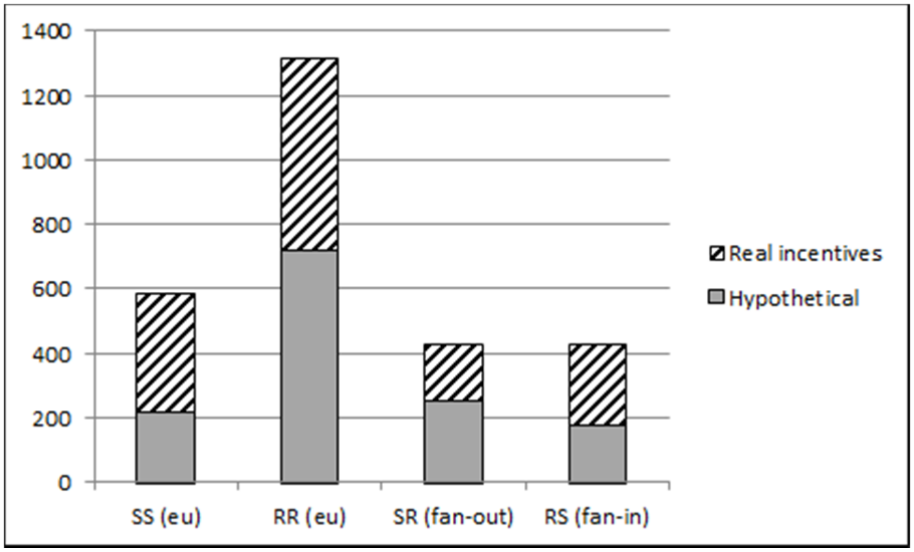

Figure 3 Observed outcomes. The numbers of the corresponding outcomes pooled across baseline dataset and reported separately for the experiments with real and hypothetical incentives.

In order to compare payoffs across different currencies and different years we first apply the PPP conversion factor[7] to all payoffs in foreign currencies to convert them to comparable USD payoffs and then use US CPI index (with 2010 as a base year) to bring all amount to 2010 USD. The PPP conversion factor and the US CPI index were sourced from the World Bank Database.

Column I is a dummy variable that equals one if financial incentives in the experiment are real and zero if they are hypothetical. Column F is a dummy variable that equals one if choice options are presented as lotteries (not in compound or frequency format). Column S is a dummy variable that equals one if subjects are not students.

Figure 3 shows the observed outcomes of choice patterns pooled across all the experiments in the baseline dataset conditional on whether incentives are real or hypothetical. Some regularity in the data is already apparent from a simple visual inspection of Figure 3 and/or Table 1. For example, the outcomes consistent with Expected-Utility Theory (no paradox) are prevalent across all the experiments with risky choice being a dominant outcome. However, the risky choice is less prevalent in the experiments with real incentives. Moreover, a great majority of studies that finds a systematic AP (fanning-out choice patterns outnumber fanning-in) use hypothetical incentives, as manifested by a high occurrence of a value of null in the I column at the top part of Table 1. The majority of studies, in contrast, that find a reverse AP (fanning-in choice patterns outnumber fanning-out) or no systematic violations of expected utility at all use real financial incentives, as manifested by a high occurrence of a value of one in the I column at the bottom part of Table 1.

[7] Purchasing power parity conversion factor is the number of units of a country's currency required to buy the same amount of goods and services in the domestic market as a U.S. dollar would buy in the US.

Another apparent regularity is that studies reporting a systematic Allais paradox (fanning-out choice pattern outnumbering fanning-in) typically design a pair of Allais questions with a very uneven division of the probability mass, as manifested by the fact that probability PH is often 10 times larger than probability PL at the top part of Table 1. On the other hand, studies reporting a reverse Allais paradox (fanning-in choice patterns outnumbering fanning-out) typically design a pair of Allais questions with an even division of the probability mass, as manifested by the fact that probability PH is often equal to probability PL at the bottom part of Table 1.

4.2. Econometric estimation

We use the reduced form regression to identifying possible relationships between the outcomes of the experiments and the experimental design variables. Data from all considered experiments is combined in one dataset. The weight of each experiment is the combined dataset depends on the number of observations in each experiment.

All experiments result in four discrete outcomes and hence multinomial logistic specification is a sensible model to use in this setting.[8] Logistic regression specifies that the log of the probabilities ratio has a linear structure. In particular, we consider the following model:

$$\log\left(\frac{P_i}{P_0}\right) = \beta_{i0} + \beta_{i1}\mathbf{P} + \beta_{i2}\mathbf{P}\times\mathbf{I} + \beta_{i3}\mathbf{F} + \beta_{i4}\mathbf{S} + \beta_{i5}\mathbf{O} + \beta_{i6}\mathbf{PH}/\mathbf{PL}$$

where P_i is the probability to observe a specific outcome, *i*=1,2,3 and P_0 is a baseline outcome.

The highest outcome payoffs **P** and incentives dummy variable, **I**, are strongly correlated as studies with high stakes typically use no monetary incentives. The tetrachoric biserial correlation is -0.76. We do not include both of these variables in the specification, but instead we use interaction term $\mathbf{P}\times\mathbf{I}$ measuring the additional effect of payoffs when the incentives are real.

We start with three-variate logit with the following outcomes: no paradox (SS+RR) and fanning-out (SR) and fanning-in (RS). For better understanding we also consider all four outcomes. The same set of regressions is also performed using the extended dataset. Both results are reported in Appendix, Tables 7 and 8.

4.3. Results

Table 2 presents the results of the logistic regression. The relationship between the coefficient estimates and the probabilities of the outcomes is nonlinear. In order to simplify the interpretation of the results we report the average marginal (partial) effects, which are observation-specific marginal effects averaged over all observation. The original logit estimates are reported in Table 7 in

[8] An important assumption in the multinomial logistic model is that ratio P_i / P_0 is independent of the remaining probabilities, so called, the independence of irrelevant alternatives (IIA) assumption. Our model passes the Small-Hsiao test of the IIA assumption, see Table 7 in Appendix for details.

Appendix. Note that average marginal effects for each explanatory variable sum up to 1 over all possible outcomes.

Explanatory variables / Outcome probabilities	**P** (payoffs, in 2010 USD)	**P x I** (interact with I, I=1, real incent.)	**F** (=1, lottery)	**S** (=1, not student)	**O** (=mid/high)	**PH/PL** (slope)
P(SS+RR, eu)	**-3.86E-8**	**-7.6E-5**	**-0.160**	-0.029	**-0.317**	**0.005**
	0	3e-5	0.021	0.017	0.036	0.001
	0	0.021	0	0.091	0	0
P(SR, fan-out)	**3.44E-8**	**11.8E-5**	**0.171**	0.004	**0.213**	**0.003**
	0	3e-5	0.012	0.015	0.032	0.001
	0	0	0	0.781	0	0.009
P(RS, fan-in)	0.42E-8	**-4.2E-5**	**-0.011**	**0.025**	**0.104**	**-0.008**
	0	2e-5	0.018	0.012	0.025	0.001
	0.080	0.050	0.565	0.037	0	0

Table 2 Average marginal effects computed from three-outcome logit models. The first line alongside each outcome probability reports coefficient estimates, the second line their standard errors and the third line their p-values. Small numbers are reported in scientific format, where E-n is stands for $x10^{-n}$. Coefficients significant at 0.05 level are indicated with bold font.

A significant coefficient estimate on the highest outcome variable **P** suggests that we are more likely to observe a choice pattern inconsistent with expected utility theory when stakes are high, *ceteris paribus*. The magnitude of the coefficient is small as many studies use hypothetical payoffs in millions USD. Moreover, high stakes contribute to the increase in the occurrence of fanning-out pattern, but have no effect on the fanning-in pattern. The interaction term between **P** and **I** is significant: when participants are incentivized, the effect of high stakes is magnified considerably.

The significant coefficient on **F** dummy indicates that when choice options are presented as lotteries (as opposed to compounding or frequencies), we are likely to observe the fanning-out pattern in Allais questions. The coefficient on **S** dummy is significant only for a fanning-in choice pattern, which indicates that non-students are likely to exhibit a fanning-in pattern.

Allais paradox is more likely to be observed when the ratio of the middle to the highest outcome is higher (closer to 1) as indicated by a significant coefficient on variable **O**. When pairs of Allais questions are designed so that the middle outcome is close to the highest outcome (which, arguably, increases the cognitive burden of both questions), subjects tend to violate expected utility more frequently as a result of reduced risky (RR) choice. Both fanning-out and fanning-in choice patterns become more likely to be revealed; but instances of fanning-out happen nearly twice as frequently as instances of fanning-in. Thus, in the aggregate, subjects are more likely to reveal a systematic Allais paradox.

The significant coefficient on the **PH/PL** variable indicates that subjects are less likely to reveal a fanning-in choice pattern and more likely—to reveal a fanning-out choice pattern when Allais questions designed with an uneven division of the probability mass. Somewhat unexpectedly, in this case, subjects are also more likely to reveal a choice pattern consistent with Expected Utility Theory.

5. Additional insights from experimental data collected by Loomes & Sugden (1998)

Results from section 4 suggest that instances of violations of Expected Utility, that is, fanning-out and fanning-in choice patterns are more likely to be observed in decision problems with a high ratio of the middle outcome to the highest outcome while fanning-in choice patterns are more prevalent in problems with low slopes of lines AB and CD in the probability triangle. Loomes & Sugden (1998) collected experimental data that can be used to examine these findings within the same lab and subject population.

Loomes & Sugden (1998) asked two groups of 46 subjects to make 45 binary choices. Each decision problem was repeated twice in each group. Out of total 45 problems, there are four pairs of Allais questions. These are questions 5 and 8, 12 and 16, 20 and 24, 36 and 40 in Table 1a and Table 1b in Loomes et al. (2002, pp. 109-110). These questions are illustrated in the Marschak-Machina probability triangle on Figure 2 in Loomes & Sugden (1998, pp. 587-588). Since the slope **PH/PL** of lines AB and CD in the probability triangle is different in all four pairs we have an opportunity for examine the following hypothesis within the same subject population.

Hypothesis 1 Instances of the reverse Allais Paradox decrease with the ratio **PH/PL**.

Moreover, Loomes & Sugden (1998) used different lottery outcomes in two groups. While the lowest and the middle outcome were £0 and £10 in both groups, the highest outcome was £30 in group 1 and £20 in group 2. Hence, given results from section 4, we might expect more instances of the AP in group 2 (with a higher ratio of the middle outcome to the highest outcome).

Hypothesis 2 Violations of Expected Utility occur more often in group 2.

Tables 3 and 4 present the experimental data collected from Tables 2a and 2b in Loomes et al (2002, pp. 111-112). Tables 3 and 4 also show Conlisk z-statistic and its p-value. Recall that positive values of the statistic indicate the Allais Paradox (when fanning-out choice patterns **SR** outnumber fanning-in choice patterns **RS**). Negative values of the statistic indicate the reverse Allais Paradox (when fanning-in choice patterns **RS** outnumber fanning-out choice patterns **SR**). Zero values of the statistic indicate that there is no paradox.

Tables 3 and 4 show that the Conlisk z-statistic increases with the ratio **PH/PL** in both groups, which supports our Hypothesis 1. The evidence from group 1 is weak as all p-values for the Conlisk z-statistic are high. . Comparison across Tables 3 and 4 offers some support for Hypothesis 2. The evidence for reverse Allais Paradox stronger in group 2 as most of p-values for Conlisk z-statistic indicate that it is significantly different from zero.

In fact, in both groups we observe the reverse AP (in group 2 it is highly statistically significant). This is probably not surprising given that Loomes & Sugden (1998) designed their experiment with all factors that we identified in Section 4 as detrimental to the classic Allais Paradox: small payoffs with real incentives; probability distributions are presented as normalized frequencies (cf. Figure 3 in Loomes & Sugden 1998, p. 589); ratios **PH/PL** are relatively low.

Pairs of questions	**PH/PL**	**SS**	**SR**	**RS**	**RR**	**Conlisk z statistic**	**p-value**
5 and 8	2/3	53	11	16	11	-0.9618	0.1680
12 and 16	1	27	17	21	27	-0.64682	0.2588
20 and 24	1.5	22	16	18	36	-0.3413	0.3664
36 and 40	3	11	17	16	48	0.1731	0.4312

Table 3 Choice patterns revealed in group 1 (highest outcome £30) pooled over two repetitions

Pairs of questions	**PH/PL**	**SS**	**SR**	**RS**	**RR**	**Conlisk z statistic**	**p-value**
5 and 8	2/3	72	3	16	1	-3.1208	0.0009
12 and 16	1	56	8	21	7	-2.4807	0.0065
20 and 24	1.5	40	13	29	10	-2.5410	0.0055
36 and 40	4	26	12	15	39	-0.5752	0.2825

Table 4 Choice patterns revealed in group 2 (highest outcome £20) pooled over two repetitions

6. Discussion

Our results demonstrate that the Allais Paradox is by no means a robust behavioral regularity. The AP can be made to disappear, or even be reversed, when an experimenter makes specific choices for stakes, incentives, framing, and lottery design. Our result is in the spirit of Gigerenzer's deconstruction of well-known cognitive biases (Gigerenzer 1991); the allusion in our title to that paper is not coincidental. For example, our results indicate that people are more likely to violate Expected Utility Theory (in particular, in the direction consistent with fanning-out of indifference curves) when outcomes in the Allais questions are large. Indeed, Camerer (1989) finds that subjects tend to reveal fanning-out choice patterns when outcomes are large gains but finds no systematic violations of Expected Utility Theory when outcomes are small gains. As another example, our results indicate that people are more likely to violate expected utility theory (in particular, in the direction consistent with fanning-out of indifference curves) when probability distributions are presented as simple lotteries rather than compound lotteries or in a frequency format. Indeed, Conlisk (1989) finds that subjects tend to reveal fanning-out choice patterns when probability distributions are presented as simple lotteries but finds that violations of expected utility theory are more systematic in the direction of fanning-in choice patterns when probability distributions are presented as compound lotteries. In light of our results it cannot be claimed does that the Allais Paradox is a robust behavioral phenomenon. The interesting question is under what conditions it appears, disappears, or reverses.

It is important to get these empirical facts straight because empirical evidence ultimately affects the formation of economic theory. Decision theories are not descriptively accurate if they are built on the assumption that decision makers are prone to the kind of EUT violations captured by the AP independent of stakes, incentives, framing, and lottery design. A misleading perception of the AP as a robust behavioral regularity supports the existence of such theories and hinders the development of new decision theories that are more descriptively accurate. Thus, it is important to get experimental evidence straight to prompt the formation of relevant theories.

For example, our results suggests that we need a decision theory that could simultaneously rationalize a higher incidence of the fanning-out choice patterns in Allais questions with a high slope of lines AB and CD in the probability triangle as well as a higher incidence of fanning-in patterns in Allais questions with a low slope of lines AB and CD in the probability triangle. Blavatskyy (2015) has developed a generalization of classical models of disappointment aversion that can rationalize the AP results in classical Common-Consequence problems (as in Starmer & Sugden 1991) and the reverse

Allais Paradox—in Common Consequence problems with an even split of a probability mass (such as in Starmer 1992).

7. **Conclusion**

We started our investigation with divergent perceptions about the reality of the Allais Paradox. A key insight that emerges from our investigation is that the choice of specific realizations of design and implementation details matters and we demonstrated that the choices an experimenter makes can lead to the Allais Paradox to appear, or disappear, or even reverse. Our finding confirms that the way one conducts an experiment is unbelievably important (e.g., Hertwig & Ortmann 2001; Smith 2002; Camerer 2003). This is by no means a novel insight, at least to those working experimentally, but it has not yet been demonstrated for the Allais Paradox in a comprehensive, systematic, and tractable study. We have demonstrated that the choice of specific realizations of design and implementation details can make the difference between the acceptance and rejection of a theory. Our finding poses an interesting issue: Which of these design and implementation choices can be rationalized? We propose that representativeness / external validity may be candidates to guide our choice. These concepts themselves are of course subject to dispute, so for now a key insight of our study is that we can predict under what well-defined circumstances the Allais Paradox make an appearance, and when not. We note that our study is a close relative to meta-analyses and also to a model of evidence production and evaluation that we believe to be widely underused: Adversarial collaborations (Mellers, Hertwig, & Kahneman 2001). In the interest of a stabilization and consolidation of the evidence base we propose it as an important strategy. Our study of the Allais Paradox provides a viable proof of concept.

Appendix

Experiment	SS	SR	RS	RR	Conl-z	p-val	PH	PL	P	O	I	F	S
Birnbaum (2007), exp. 1, series A, questions 6-12	36	84	11	69	8.81	0.00	0.1	0.01	$2,103,04	0.5	0	1	0
Starmer and Sugden (1991)	37	57	14	52	5.56	0.00	0.2	0.05	$26	0.7	1	1	0
Birnbaum (2007), exp. 2, condition A2, questions 6-	40	66	29	64	3.93	0.00	0.1	0.1	$103	0.4	1	1	0
Harrison (1994), AP0	0	7	0	13	3.20	0.00	0.1	0.01	$29	0.2	0	1	0
List and Haigh (2005), students	4	13	3	10	2.76	0.00	0.2	0.05	$11	0.7	1	1	0
Harrison (1994), AP1	0	3	0	17	1.83	0.03	0.1	0.01	$29	0.2	1	1	0
Birnbaum (2007), exp. 2, condition A3, questions 6-	38	54	39	65	1.56	0.06	0.1	0.1	$103	0.4	1	1	0
Birnbaum (2007), exp. 2, condition A3, questions 6-	37	58	47	55	1.07	0.14	0.1	0.1	$103	0.4	1	1	0
List and Haigh (2005), traders	8	7	9	30	-0.50	0.31	0.2	0.05	$11	0.7	1	1	1

Table 5 Additional experimental data from studies that did not show up in the EconLit search but that were analyzed as a robustness check. Column "Experiment" lists experiments as labeled in the original study. Column SS shows how many subjects chose A over B and C over D. Column SR (RS) shows how many subjects revealed a fanning-out (fanning-in) choice pattern. Column RR shows how many subjects chose B over A and D over C. Column O shows the ratio of the middle outcome to the highest outcome. Column Conl-z report Conlisk z statistic and its p-value, respectively. The rows are orders by Conlisk z statistic indicating fanning-out patterns in the top block, no paradox in the middle block (highlighted in grey) and fanning-in patterns in the bottom block. The blocks are separated by thick black lines. Column PH (PL) shows the probability of the highest (lowest) outcome in lottery B in the first Allais question. Column P reports the highest payoff standardized to 2010 USD. Column I is a dummy variable that equals one if incentives are real and zero if they are hypothetical. Column F is a dummy variable that equals one if choice options are presented as lotteries (not in compound or frequency format). Column S is a dummy variable that equals one if subjects are not students.

Explanatory variables / Outcome probabilities	**P** (payoffs, in 2010 USD)	**P x I** (interact with I, I=1, real incent.)	**F** (=1, lottery)	**S** (=1, not student)	**O** (=mid/high)	**PH/PL** (slope)	**const**
Baseline dataset (Nobs = 5035)							
P(SR, fan-out)	**21.8E-8**	**7.39E-4**	**1.401**	0.104	**1.299**	**0.015**	**-3.547**
	1.4E-8	1.77E-4	0.143	0.090	0.198	0.007	0.192
	0	0	0	0.251	0	0.047	0
P(RS, fan-in)	**9.39E-8**	-2.8E-4	0.180	0.196	**1.457**	**-0.081**	**-1.941**
	1.96E-8	2.1E-4	0.170	0.112	0.244	0.011	0.244
	0	0.189	0.291	0.078	0	0	0
Diagnostics	lnL = -4179; SH test (Ho: IIA): omit SR pval=0.61, omit RS pval=0.59						
P(SS, eu)	**25.7E-8**	**6.83E-4**	**0.431**	**0.849**	**3.287**	**-0.118**	**-2.549**
	1.89E-8	1.93E-4	0.154	0.111	0.244	0.011	0.231
	0	0	0.005	0	0	0	0
P(SR, fan-out)	**27.5E-8**	**11.4E-4**	**1.529**	**0.305**	**2.119**	-0.011	**-3.589**
	1.51E-8	2.15E-4	0.152	0.095	0.210	0.008	0.202
	0	0	0	0.001	0	0.148	0
P(RS, fan-in)	**15.2E-8**	1.25E-4	**0.302**	**0.408**	**2.278**	**-0.110**	**-1.963**
	2.04E-8	2.43E-4	0.178	0.116	0.255	0.011	0.254
	0	0.608	0.09	0	0	0	0
Diagnostics	lnL = -5850; SH test (Ho: IIA): omit SS pval=0.17, omit SR pval=0.31, omit RS pval=0.7						
Extended dataset (Nobs = 6111)							
P(SS+RR, eu)	**19.1E-8**	**6.80E-4**	**1.536**	**-0.229**	**1.269**	**-0.011**	**-3.108**
	1.31E-8	1.70E-4	0.143	0.078	0.174	0.006	0.181
	0	0	0	0.003	0	0.075	0
P(RS, fan-in)	**8.73E-8**	-1.77E-4	0.156	0.148	**1.190**	**-0.090**	**-1.743**
	1.93E-8	2.09E-4	0.169	0.095	0.230	0.010	0.230
	0	0.396	0.357	0.119	0	0	0
Diagnostics	lnL = -5252; SH test (Ho: IIA): omit SR pval=0.47, omit RS pval=0.87						
P(SS, eu)	**24.4E-8**	**8.04E-4**	**0.481**	**0.676**	**3.051**	**-0.118**	**-2.392**
	1.8E-8	1.99E-4	0.151	0.092	0.221	0.009	0.211
	0	0	0.001	0	0	0	0
P(SR, fan-out)	**24.7E-8**	**11.5E-4**	**1.667**	-0.045	**2.061**	**-0.036**	**-3.145**
	1.41E-8	2.15E-4	0.151	0.083	0.186	0.007	0.191
	0	0	0	0.588	0	0	0
P(RS, fan-in)	**14.5E-8**	2.82E-4	0.283	**0.350**	**2.013**	**-0.117**	**-1.784**
	2.01E-8	2.47E-4	0.177	0.100	0.241	0.010	0.240
	0	0.253	0.11	0	0	0	0
Diagnostics	lnL = -7295; SH test (Ho: IIA): omit SS pval=0.09, omit SR pval=0.34, omit RS pval=0.83						

Table 7. Logit regression coefficients and diagnostics for three- and four-outcome models for baseline and extended datasets. The first line alongside each outcome probability reports coefficient estimates, the second line their standard errors and the third line their p-values. Small numbers are reported in scientific format, where E-n stands for $x10^{-n}$. Coefficients significant at 0.05 level are indicated with bold font. The results of the Small-Hsiao (SH) test for IIA are reported in diagnostics. The Hausman test frequently produced negative values of the test statistics and could not be used.

Explanatory variables / Outcome probabilities	**P** (payoffs, in 2010 USD)	**P x I** (interact with I, I=1, real incent.)	**F** (=1, lottery)	**S** (=1, not student)	**O** (=mid/high)	**PH/PL** (slope)
			Baseline dataset			
P(SS+RR, eu)	**-3.86E-8**	**-7.6E-5**	**-0.160**	-0.029	**-0.317**	**0.005**
	0	3e-5	0.021	0.017	0.036	0.001
	0	0.021	0	0.091	0	0
P(SR, fan-out)	**3.44E-8**	**11.8E-5**	**0.171**	0.004	**0.213**	**0.003**
	0	3e-5	0.012	0.015	0.032	0.001
	0	0	0	0.781	0	0.009
P(RS, fan-in)	0.42E-8	**-4.2E-5**	**-0.011**	**0.025**	**0.104**	**-0.008**
	0	2e-5	0.018	0.012	0.025	0.001
	0.080	0.050	0.565	0.037	0	0
P(SS, eu)	**2.18E-8**	**5.07E-5**	0.014	**0.102**	**0.319**	**-0.013**
	0	2E-05	0.018	0.015	0.030	0.001
	0	0.017	0.440	0	0	0
P(RR, eu)	**-5.97E-8**	**-18.20E-5**	**-0.184**	**-0.126**	**-0.643**	**0.018**
	0	4E-05	0.027	0.018	0.041	0.002
	0	0	0	0	0	0
P(SR, fan-out)	**3.43E-8**	**15.73E-5**	**0.180**	0.000	**0.217**	**0.004**
	0	3E-05	0.012	0.015	0.033	0.001
	0	0	0	0.996	0	0.002
P(RS, fan-in)	0.38E-8	-2.63E-5	-0.010	**0.024**	**0.106**	**-0.008**
	0	2E-05	0.019	0.012	0.026	0.001
	0.160	0.257	0.610	0.049	0	0
			Extended dataset			
P(SS+RR, eu)	**-3.56E-8**	**-8.87E-5**	**-0.177**	0.022	**-0.285**	**0.009**
	0	3e-5	0.021	0.015	0.035	0.001
	0	0.008	0	0.146	0	0
P(SR, fan-out)	**3.17E-8**	**12.7E-5**	**0.188**	**-0.044**	**0.194**	0.001
	0	3e-5	0.011	0.013	0.030	0.001
	0	0	0	0.001	0	0.646
P(RS, fan-in)	**0.39E-8**	**-3.82E-05**	**-0.012**	**0.023**	**0.091**	**-0.009**
	0	2e-5	0.019	0.010	0.024	0.001
	0.091	0.078	0.532	0.029	0	0
P(SS, eu)	**2.08E-8**	**5.89E-5**	0.017	**0.093**	**0.297**	**-0.012**
	0	2E-05	0.018	0.013	0.028	0.001
	0	0.007	0.357	0	0	0
P(RR, eu)	**-5.53E-8**	**-20.90E-05**	**-0.201**	**-0.071**	**-0.586**	**0.020**
	0	5E-05	0.027	0.016	0.038	0.001
	0	0	0	0	0	0
P(SR, fan-out)	**3.12E-8**	**17.06E-05**	**0.196**	**-0.047**	**0.195**	0.002
	0	3E-05	0.012	0.013	0.031	0.001
	0	0	0	0	0	0.168
P(RS, fan-in)	0.33E-8	-2.10E-05	-0.012	**0.025**	**0.095**	**-0.009**
	0	2E-05	0.019	0.011	0.025	0.001
	0.169	0.369	0.542	0.020	0	0

Table 8. Average marginal effects of three- and four-outcome logit for the baseline and extended dataset. The first line alongside each outcome probability reports coefficient estimates, the second line their standard errors and the third line their p-values. Small numbers are reported in scientific format, where E-n stands for x10-n. Coefficients significant at 0.05 level are indicated with bold font.

References[9]

Allais, Maurice (1953) “Le Comportement de l’Homme Rationnel devant le Risque: Critique des Postulates et Axiomes de l’Ecole Américaine” *Econometrica* **21**, 503-546.

Bernoulli, Daniel. (1738) “Specimen theoriae novae de mensura sortis” Commentarii Academiae Scientiarum Imperialis Petropolitanae

Bierman, Harold (1989) “The Allais Paradox: A Framing Perspective” *Behavioral Science* **34**, 46-52.

Birnbaum, Michael (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

Blavatskyy, Pavlo (2015) “A Theory of Decision-Making Under Risk as a Tradeoff between Expected Utility, Expected Utility Deviation and Expected Utility Skewness” available at http://ssrn.com/abstract=2505828

* Blavatskyy, Pavlo (2013) “Reverse Allais Paradox” *Economics Letters* **119(1)**, 60-64.

Blavatskyy, Pavlo (2012) “Probabilistic Subjective Expected Utility” *Journal of Mathematical Economics* **48** (**1**), 47-50.

Blavatskyy, Pavlo (2010) “Reverse Common Ratio Effect” *Journal of Risk and Uncertainty* **40**, 219-241

* Burke, Michael S., Carter, John R., Gominiak, Robert D., Ohl, Daniel F. (1996) “An experimental note on the Allais paradox and monetary incentives” *Empirical Economics* **21**, 617–632.

Brandtstaetter, Eduard, Gigerenzer, Gerd, Hertwig, Ralph (2008) “Risky Choice with Heuristics: Reply to Birnbaum (2008), Johnson, Schulte-Mecklenbeck, and Willlemsen (2008), and Rieger and Wang (2008)” *Psychological Review* **115**(1), 281 – 290.

Camerer, Colin F. (2003), *Behavioral Game Theory.* Princeton: Princeton University Press.

* Camerer, Colin F. (1989) “An Experimental Test of Several Generalized Utility Theories” *Journal of Risk and Uncertainty* **2**(1), 61–104.

* Carlin, Paul S. (1990) “Is the Allais paradox robust to a seemingly trivial change of frame?” *Economics Letters* **34**, 241-244.

* Carlin, Paul S. (1992) “Violations of the reduction and independence axioms in Allais-type and common-ratio effect experiments” *Journal of Economic Behavior and Organization* **19**, 213-235.

Cason, Timothy N, Plott, Charles R. (2014), "Misconceptions and Game Form Recognition: Challenges to Theories of Revealed Preference and Framing"*Journal of Political Economy* **122** (6), 1379-1381.

[9] Studies marked by asterisk were used in the empirical analysis

* Cherry, Todd L., Shogren, Jason F. (2007) “Rationality Crossovers” *Journal of Economic Psychology* **28**(2), 261–277.

* Conlisk, John (1989) “Three Variants on the Allais Example” *American Economic Review* **79** (3), 392-407.

* Da Silva, Sergio, Baldo, Dinora, Matsushita, Raul (2013) “Biological Correlates of the Allais Paradox” *Applied Economics* **45**(4-6), 555–568.

Ellsberg, Daniel (1961) “Risk, Ambiguity, and the Savage Axioms” *Quarterly Journal of Economics* **75**, 643-669.

Engel, Christoph (2011) “Dictator Games: A Meta Study” *Experimental Economics* **14**, 583–610.

* Fan, C.-P. (2002) “Allais paradox in the small” *Journal of Economic Behavior and Organization* **49**, 411–421.

Fatas, Enrique, Neugebauer, Tibor., Tamborero, Pilar (2007) “How politicians make decisions: A political choice experiment” *Journal of Economics* **92**, 167–196.

* Finkelshtain, Israel. Feinerman, Eli (1997) “Framing the Allais Paradox as a Daily Farm Decision Problem: Tests and Explanations” *Agricultural Economics* **15**(3), 155–167.

Gigerenzer, Gerd (1991) "How to make cognitive illusions disappear: Beyond ‘heuristics and biases.’". *European Review of Social Psychology* **2** (1), 83–115.

Gigerenzer, Gerd (1996) “On narrow norms and vague heuristics: A reply to Kahneman and Tversky.” *Psychological Review* **103**, 592-596.

Grether, David M., Plott, Charles R. (1979) “Economic Theory of Choice and the Preference Reversal Phenomenon*” The American Economic Review* **69**(4), 623-663.

* Groes, Ebbe, Jacobsen, Hans Jorgen, Sloth, Birgitte, Tranaes, Torben (1999) “Testing the Intransitivity Explanation of the Allais Paradox” *Theory and Decision* **47,** 229–245.

Harless, David W. (1992) “Predictions about Indifference Curves Inside the Unit Triangle: A Test of Variants of Expected Utility Theory.” *Journal of Economic Behavior and Organization* **18**(3), 391–414.

Harrison, Glenn W. (1989),”Theory and Misbehavior of First-Price Auctions.” *American Economic Review* **79**(4), 749-762.

Harrison, Glenn W. (1992) “Theory and misbehavior of first-price auctions: reply*”. American Economic Review* **82**, 1426–1443.

Harrison, Glenn W. (1994) “Expected utility and the experimentalists” *Empirical Economics* **19**, 223–253.

Hertwig, Ralph,Ortmann, Andreas (2001) “Experimental Practices in Economics: A Challenge for Psychologists?” *Behavioral and Brain Sciences* **24**, 383 - 403.

* Huck, Steffen, Müller, Wieland (2012) "Allais for all: Revisiting the paradox in a large representative sample" *Journal of Risk and Uncertainty* **44**, 261–293.

* Humphrey, Steven, Verschoor, Arjan (2004) "The probability weighting function: experimental evidence from Uganda, India and Ethiopia" *Economics Letters* **84**, 419-425.

Kahneman, Daniel (2003) "Maps of Bounded Rationality: Psychology for Behavioral Economics."*American Economic Review* **93**(5), 1449-1475.

Kahneman, Daniel, Tversky, Amos (1979) "Prospect Theory: An Analysis of Decision Under Risk" *Econometrica* **47**(2): 263 – 292.

Kahneman, Daniel, Tversky, Amos (1996) "On the reality of cognitive illusions." *Psychological Review* **103**(3), 582-91.

Laury, Susan K, Holt, Charles A. (2008) Payoff Scale Effects and Risk Preference Under Real and Hypothetical Conditions", pp. 1047 – 1053 in Plott, Charles R., Smith, Vernon L. (2008) *Handbook Experimental Economics Results*. Amsterdam: North-Holland.

Leland, John W. (1994) "Generalized similiarity judgements: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty* **9**, 151-172.

L'Haridon, Oliver, Placido, Larida (2008) "An Allais Paradox for Generalized Expected Utility Theories? *Economics Bulletin* **4**(19), 1–6.

Li, Shu (2004) "An Alternative Way of Seeing the Allais-Type Violations of the Sure-Thing Principle." *Humanomics* **20**(1-2), 17–31.

List, John A., Haigh, Michael, S. (2005) "A simple test of expected utility theory using professional Traders" *Proceedings of the National Academy of Science* **102**, 945–948

Loomes, Graham, Sugden, Robert. (1998) "Testing different stochastic specifications of risky choice" *Economica* **65**: 581-598.

Loomes, Graham, Moffatt, Peter, Sugden, Robert (2002) "A microeconomic test of alternative stochastic theories of risky choice" *Journal of Risk and Uncertainty* **24**, 103-130.

MacDonald, Don N., Wall, Jerry L. (1989). An Experimental Study of the Allais Paradox over Losses: Some Preliminary Evidence. *Quarterly Journal of Business and Economics* **28**(4), 43–60.

Machina, Mark (1982) "Expected utility' analysis without the independence axiom" *Econometrica* **50:** 277–323.

Machina, Mark (2009) "Risk, Ambiguity, and the Rank-Dependence Axioms" *American Economic Review* **99**(1), 385-392.

Mellers, Barbara, Hertwig, Ralph, Kahneman, Daniel (2001). "Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration"*Psychological Science* **12**(4), 269–275.

Oliver, Adam (2003) “A quantitative and qualitative test of the Allais paradox using health outcomes” *Journal of Economic Psychology* **24** : 35–48

Plott, Charles R., Zeiler, Kathryn (2005) "The Willingness to Pay-Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations." *American Economic Review* **95**(3), 530-545.

Plott, Charles R., Zeiler, Kathryn (2011) "The Willingness to Pay—Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations: Reply." *American Economic Review* **101**(2), 1012-28.

Prelec, Dragan (1990) “’Pseudo-endowment” Effect, and its Implications for Some Recent Nonexpected Utility Models.” *Journal of Risk & Uncertainty* **3**(3), 247–259.

Quiggin, John (1981) “Risk perception and risk aversion among Australian farmers” *Australian Journal of Agricultural Recourse Economics* **25**, 160-169.

Rabin, Matthew (1998) “Psychology and Economics” *Journal of Economic Literature* **36**(1, March 1998), 1 – 46.

Rao, Li-Lin, Li, Shu. (2011). “New Paradoxes in Intertemporal Choice.” *Judgment and Decision Makin*g, **6**(2), 122–129.

Rubinstein, Ariel (1988) “Similiarity and decision making under risk (Is there a utility theory resolution to the Allais-paradox?) *Journal of Economic Theory* **46**, 145-153

Slovic, Paul, Tversky, Amos (1974) “Who accepts Savage’s axiom?” *Behavioral Sciences* ***19***, 368–373

Smith, Vernon L. (2002) “Method in Experiment: Rhetoric and Reality.” *Experimental Economics* 5(2), 91–110 .

Starmer, Chris (1992) "Testing New Theories of Choice under Uncertainty Using the Common Consequence Effect." *The Review of Economic Studies* **59**(4), 813-30.

Starmer, Chris (2000) "Developments in Non-expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk" *Journal of Economic Literature* **38**(2), 332-382.

Starmer, Chris, Sugden, Robin. (1991) “Does the random-lottery incentive system elicit true preferences? An experimental investigation” *American Economic Review* **81**, 971–978.

Tversky, Amos, Kahneman, Daniel (1974) “Judgement under Uncertainty: Heuristics and Biases. Science185 (4157, Sept 27), 1124-1131.

van de Kuilen, Gijs, Wakker, Peter (2006) “Learning in the Allais paradox” *Journal of Risk and Uncertainty* **33**, 155–164.

Weber, Bethany J. (2007) “The Effects of Losses and Event Splitting on the Allais Paradox.” *Judgment and Decision Making* **2**(2), 115–125.

Zhang, Le, Ortmann, Andreas (2014) “The effects of the take-option in dictator-game experiments: a comment on Engel’s (2011) meta-study” *Experimental Economics* **17**, 414- 420.