American Economic Journal: Microeconomics 2022, 14(1): 143–163
https://doi.org/10.1257/mic.20190153

On the Experimental Robustness of the Allais Paradox[†]

By Pavlo Blavatskyy, Andreas Ortmann, and Valentyn Panchenko*

The Allais Paradox, or the common consequence effect, is a well-known behavioral regularity in individual decision-making under risk. Data from 81 experiments reported in 29 studies reveal that the Allais Paradox is a fragile empirical finding. The Allais Paradox is likely to be observed in experiments with high hypothetical payoffs, the medium outcome being close to the highest outcome and when lotteries are presented as a probability distribution (not in a compound form). The Allais Paradox is likely to be reversed in experiments when the probability mass is equally split between the lowest and highest outcomes in risky lotteries. (*JEL* D44, D81)

Initially proposed by Bernoulli (1738), expected utility theory (EUT) gained momentum in economics after von Neumann and Morgenstern (1947) provided its behavioral characterization. In fact, EUT became one of the cornerstones of the economic modeling edifice. Accordingly, EUT was subjected to thorough empirical scrutiny in numerous studies. Prominent among these were thought experiments proposed by Allais (1953, 527) and Ellsberg (1961) that challenged the descriptive validity of EUT. A considerable amount of work went, and continues to go, into the formulation of non-EUTs (Starmer 2000). In the present paper, we explore the vast, and sometimes contradicting, experimental literature on the Allais Paradox (AP). We argue that the AP is a fragile empirical finding. Specific choices of experimental design and implementation characteristics, and their parameterization, affect the likelihood of observing the AP (or the reverse thereof).

It is well known, and widely acknowledged (e.g., Hertwig and Ortmann 2001), that the way one conducts an experiment is "unbelievably important" (Camerer

*Blavatskyy: Montpellier Business School (email: p.blavatskyy@montpellier-bs.com); Ortmann: School of Economics, UNSW Business School, UNSW Australia (email: a.ortmann@unsw.edu.au); Panchenko: School of Economics, UNSW Business School, UNSW Australia (email: v.panchenko@unsw.edu.au). Michael Ostrovsky was coeditor for this article. We thank conference and seminar participants at ANZWEE 2013 in Brisbane, AP ESA 2013 in Tokyo, FUR 2014 in Rotterdam, CERGE-EI, Nanyang Technical University, the University of Passau, the University of Ulm, UNSW, and the University of Amsterdam for feedback. Special thanks to Ken Binmore, Michael Birnbaum, Gerd Gigerenzer, Simon Grant, Ganna Pogrebna, Johann Graf von Lambsdorff, and Peter Wakker. A special thank you also to the three referees for constructive criticism and expert guidance. Our study was approved by UNSW's HREAP Executive (HC200037). Pavlo Blavatskyy is a member of the Entrepreneurship and Innovation Chair, which is part of LabEx Entrepreneurship (University of Montpellier) and funded by the French government (Labex Entreprendre, ANR-10-Labex-11-01). The authors declare that they have no relevant or material financial interest that relate to the research described in this paper.

[†]Go to https://doi.org/10.1257/mic.20190153 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

2003, 34). Any test of a theory, such as EUT, is always a joint test of the theory and of the design and implementation choices the experimenter makes (Smith 2002, 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 and Müller 2012). Hence, any single study is worth only so much, and ultimately, it takes a body of evidence to establish the robustness of laboratory results and the reality of an alleged effect conditional on the various design and implementation choices made. The problem of how exactly a body of evidence is produced and evaluated has gained considerable attention and is at the heart of important methodological controversies and debates in both economics (e.g., Grether and Plott 1979; Harrison 1989, 1992; Plott and Zeiler 2005, 2011; Cason and Plott 2014) and psychology (e.g., Kahneman and Tversky 1996; Gigerenzer 1991, 1996).

One path increasingly taken by economists is metastudies. Metastudies sample the available evidence in a systematic, well-documented, and replicable manner. They allow us to quantify the impact of key design and implementation choices, which in turn allows the appropriate powering up of experimental studies, and to predict under what conditions behavioral regularities are likely to show up in the data. We provide a metastudy of experimental literature on the classic AP (also known as the common consequence effect). Strictly speaking, our methodology differs from a traditional meta-analysis (which uses statistics reported in previously published studies): this paper reanalyzes experimental data collected in previous studies.

The paper is organized as follows. In Section I, we describe the classic AP. Section II reviews the existing literature on the AP from a historical perspective and identifies six design and implementation details that might affect the AP. In Section III, we summarize our research methodology and present our results. Section IV concludes with a general discussion.

I. The Allais Paradox

Allais (1953, 527) designed a thought experiment to challenge the descriptive validity of EUT. This experiment was the starting point of what became known as the AP, or the common consequence effect. Allais (1953, 529–530) also designed a second thought experiment—in contemporary terminology, known as the common ratio effect—that is sometimes also referred to as the AP (e.g., van de Kuilen and Wakker 2006). In this paper, we discuss only the first Allais example (the common consequence effect for which at least one of the choice options is riskless).

The first Allais (1953) example consisted of two related decision problems, which we call Allais questions. In the first question, a decision-maker chooses between two options A and B:

- *Option A:* ₣100 million for certain
- *Option B:* ₣500 million with probability 0.1, ₣100 million with probability 0.89, nothing with probability 0.01

In the second question, a decision-maker chooses between another two options C and D:

- *Option C*: ₣100 million with probability 0.11, nothing with probability 0.89
- *Option D*: ₣500 million with probability 0.1, nothing with probability 0.9

The AP is illustrated in the probability triangle (Machina 1982) in Figure 1. 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. The left panel of Figure 1 shows a typical family of indifference curves for an expected utility maximizer—positively sloped parallel straight lines. Since AB is parallel to CD, option A is located on a higher indifference curve than option B (as shown in the left panel of Figure 1) if and only if option C is located on a higher indifference curve than option D. Thus, an expected utility maximizer weakly prefers A over B if and only if she weakly prefers C over D (e.g., footnote 4 in Huck and Müller 2012, 264).

A decision-maker choosing A over B and D over C violates EUT (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 horizontal 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 the right panel of Figure 1). 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 the right panel of Figure 1). Thus, when A is chosen over B and D is chosen over C, the map of indifference curves "fans out" along the horizontal axis of the probability triangle (cf. the right panel of Figure 1). Similarly, when B is chosen over A and C is chosen over D, the map of indifference curves "fans in" along the horizontal axis of the probability triangle and likewise violates EUT.

Typically, the majority of decision-makers display the horizontal-fanning-out choice pattern, and only a minority display the horizontal-fanning-in choice pattern. It is these two behavioral regularities (the violations and the asymmetry in fanning-out and fanning-in patterns) that together became widely known as the AP. In this paper, we argue that the AP is a fragile behavioral regularity and that specific choices of experimental design and implementation characteristics can systematically affect the likelihood of observing the AP (or the reverse thereof).

II. The Existing Literature

Allais (1953) originally designed his examples as a thought experiment. The advantages of thought experiments in research on individual choice are clear—the argument is more persuasive when a reader, who is as good as anybody else in the role of an 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 and biases program (e.g., Kahneman 2003; Tversky and Kahneman 1974).

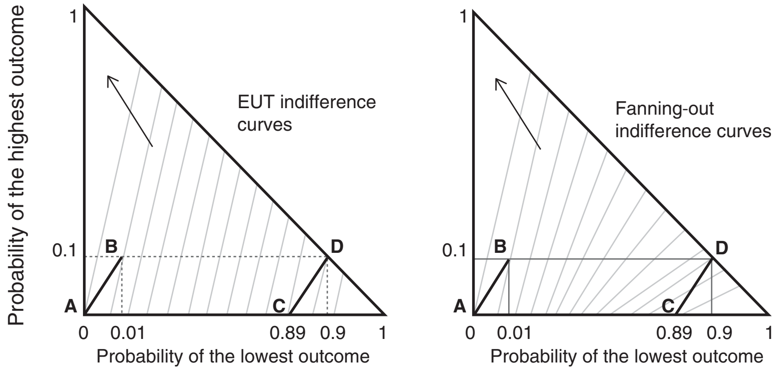

Figure 1. Illustration of the Allais Paradox in the Probability Triangle

Early experimental studies of the AP (e.g., Slovic and 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). Kahneman and Tversky (1979, 265) justified such nonincentivized experimental design as follows: "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." Whether this claim is correct is ultimately an empirical question. Laury and Holt (2008), for example, have demonstrated that the reflection effect documented in Kahnemann and 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 a sample drawn from a standard subject pool (a convenience sample of students), Huck and Müller (2012, 276, Figure 1) find that their participants exhibit the AP for large hypothetical outcomes but show significantly lower rates of EUT violations—about one-half for the representative sample and less than a third for the student sample—for low (real or hypothetical) outcomes for both their subject pools. Similar evidence was found in earlier between-subject experiments. The AP is found, for example, in the basic version of Allais questions with large hypothetical outcomes in Conlisk (1989, 395, Table 1). Yet, Conlisk (1989, 406–407, Appendix IV) finds almost no expected utility violations in a "pilot experiment" with small real outcomes. Camerer (1989, 92, Table 7) 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 1980s. By that time, a 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 motivated the development of many non-EUTs.

The results of experimental studies with small real payoffs that followed in the 1990s suggested that the AP was less widespread than the experiments with large hypothetical outcomes seem to suggest (e.g., Harrison 1994, 226–231, Section 1; Burke et al. 1996; Groes et al. 1999). In fact, several studies (e.g., Starmer 1992; Humphrey and Verschoor 2004b; Blavatskyy 2013) even document a reversed AP where horizontal-fanning-in choice patterns significantly outnumber horizontal-fanning-out choice patterns. It has remained, until now, an open question of how these findings could be reconciled. This seems an undesirable state of affairs.

The existing literature tends to focus on the question of whether the asymmetry between horizontal-fanning-out and horizontal-fanning-in choice patterns is statistically significant. This presupposes that the frequency of EUT violations is of secondary importance. We address both of these issues in this paper. There is tantalizing evidence from individual studies that suggests that the frequency of EUT violations might be remarkably fragile. For example, Huck and Müller (2012)—in their very comprehensive study—find the AP in all treatments, in that horizontal-fanning-out choice patterns statistically significantly outnumber horizontal-fanning-in choice patterns. Yet, in their laboratory experiment with low hypothetical (real) payoffs, only 4 (6) out of 79 (74) subjects, i.e., only 5 percent (8 percent), reveal either a horizontal-fanning-out or horizontal-fanning-in choice pattern. This seems hardly a threat to the validity of EUT; every theory that explains the behavior of nine out of ten subjects 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 behavioral regularity.

Apart from payoff size and hypothetical versus real incentives, other design and implementation details are worth looking at. Several studies (e.g., Tversky and Kahneman 1981, problems 5–7; 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 EUT violations. Huck and Müller (2012) have demonstrated that the choice of the subject pool also matters: participants drawn from a representative sample of the population violate EUT more frequently than student subjects.

Besides, 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., 2,400 and 2,500 Israeli pounds in Kahneman and 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, which leads to a higher rate of EUT violations. Blavatskyy (2010, 232–235, experiment 2) 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 might exist 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 (0.1) is allocated to the highest outcome, and a probability mass of only 0.01 to the lowest outcome (0). This creates a similarity (or inconsequentiality) of probabilities in the second Allais question.[1] Following a considerable literature on similarity considerations in these kinds of problems (e.g., Leland 1994; Rubinstein 1998; see also the debate about the priority heuristic, Brandstätter, Gigerenzer, and Hertwig 2008), one can argue that probability 0.11 is similar in relative terms 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 and Verschoor (2004b); and Blavatskyy (2013) all find the reversed AP where fanning-in choice patterns outnumber fanning-out choice patterns. It was not clear how to reconcile these findings when we started our study.

To summarize, the existing literature suggests that six design and implementation details might drive results of experimental studies on the AP: (i) size of payoffs, (ii) whether incentives are hypothetical or real, (iii) presentation of choice options, (iv) subject pool, (v) ratio of the middle to the highest outcome, and (vi) slope of lines AB and CD in the probability triangle.

III. Methodology and Results

A. *Data*

A search in the Scopus database with the search line ((REF("Allais M" 1953)) OR TITLE-ABS-KEY ("Allais" OR "Common consequence")) AND TITLE-ABS-KEY ("experiment*") AND DOCTYPE ("ar") AND SUBJAREA ("ECON" OR "MULT") returned a list of 165 articles in October 2017. The vast majority of these articles are theoretical papers collecting no empirical data from human subjects. Only 22 of these articles collect new experimental data on the classic AP where a safer lottery in the first question yields the middle (positive) outcome with certainty. Several articles identified in the Scopus search collected new experimental data on the common ratio effect but referred to it as the AP, e.g., van Kuilen and Wakker (2006); Herrmann et al. (2017).

[1] 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, 527).

Going through the references cited in the 22 relevant articles, we identified another 11 articles that collected new experimental data on the classic AP but did not show up in the Scopus search. Not all relevant articles reported raw experimental data in their printed version. We collected missing information from various sources such as electronic supplementary materials, personal websites of the authors, and email exchanges with the authors.[2] This brings our sample to 29 relevant articles from which we were able to obtain experimental data on the classic AP in the format that we needed. These 29 articles are detailed in the notes to Table 1.

Our sample of 29 articles contains 81 experiments with different versions of the classic AP. We did not consider experiments with nonstandard modifications of the AP reported in the 29 sample articles, such as the displaced version in Conlisk (1989), where lotteries are located inside the probability triangle, or experiment 2 in Birnbaum, Schmidt, and Schneider (2017), where the lowest outcome is not zero. In summary, our dataset consists of 8,947 observations of the classic AP collected in 81 experiments and reported in 29 peer-reviewed published articles.

Our dataset of 81 experiments is presented in Table 1. Column "EUT-consistent choices, %" shows the percentage of subjects in each experiment who revealed a choice pattern consistent with EUT maximization. Column "Fanning-out-consistent choices, %" ("Fanning-in-consistent choices, %") in Table 1 shows the percentage of subjects revealing a horizontal-fanning-out (fanning-in) choice pattern.[3]

Conlisk (1989) proposed a test statistic, the so-called Conlisk z-statistic, which takes values close to null under the null hypothesis of no EUT violation. Large positive values of the statistic indicate the AP (when fanning-out choice patterns outnumber fanning-in choice patterns). Large negative values of the statistic indicate the reversed AP (when fanning-in choice patterns outnumber fanning-out choice patterns). 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 in the middle (highlighted in the shadowed area) show no systematic EUT violations, and experiments at the bottom document high rates of fanning-in choice pattern.

Besides, Table 1 reports the experimental design variables that might influence the results of the experimental study, as discussed in the previous section. Namely, column "Prob. of highest outcome" ("Prob. of lowest outcome") shows the probability of the highest (lowest) outcome PH (PL) in lottery B in the first

[2] The authors of two studies—Wu and Gonzalez (1996) and L'Haridon and Placido (2008)—did not respond to our requests for data. Li (2004) responded but could not retrieve the data. We used only those data that allowed us to readily identify variables listed in Table 1. Also note that Humphrey and Verschoor (2004b) use the same data as in Humphrey and Verschoor (2004a) and hence we exclude the latter.

[3] If subjects choose at random, we would observe a uniform distribution over the four outcomes: EUT-consistent safe (AC), EUT-consistent risky (BD), fanning out (AD), and fanning in (BC). We did Pearson's chi-squared tests for each experiment in our dataset and find that only 9 studies out of 83 fail to reject the null of uniform distribution. Six out of these 9 studies have 54 or fewer subjects, which is, arguably, a relatively small sample size. The nine studies (and their respective p-values) were Birnbaum (2007), experiment 2, condition A3, questions 6–12 (0.15); Bateman and Munro (2005), T1&T8 (0.45); Birnbaum, Schmidt, and Schneider (2017), exp1, CCE3 R4 (0.46); Butler and Loomes (2011), A$60 group (0.74); Bateman and Munro (2005), W3 and W7 (0.11); Camerer (1989), small gains, hypothetical (0.49); Camerer (1989), small gains, real (0.57); Loomes and Sugden (1998), group 1, Q12 and Q16 (0.37); and Birnbaum, Schmidt, and Schneider (2017), exp1, CCE3 R3 (0.11).

Table 1—Experimental Data Analyzed in This Paper

#	Obs.	EUT-consistent choices, %	Fanning-out-consistent choices, %	Fanning-in-consistent choices, %	Prob. of highest outcome	Prob. of lowest outcome	Highest outcome in 2010 USD	Middle/highest outcome	Real (1) or hypothetical (0) incentives	Lottery presentation (1) or not (0)	Students (1) or not (0)
1	186	41.4	55.4	3.2	0.1	0.01	$7.5m	0.2	0	1	1
2	75	41.3	58.7	0.0	0.33	0.01	$751	1.0	0	1	1
3	102	32.4	62.7	4.9	0.1	0.01	$5.3m	0.2	0	1	1
4	236	49.6	43.6	6.8	0.1	0.01	$8.8m	0.2	0	1	1
5	200	52.5	42.0	5.5	0.1	0.01	$2.2m	0.5	0	1	1
6	89	48.3	47.2	4.5	0.1	0.01	$7.8m	0.2	0	1	1
7	51	45.1	51.0	3.9	0.1	0.01	$7.4m	0.2	0	1	1
8	65	55.4	41.5	3.1	0.1	0.01	$8.3m	0.2	0	1	1
9	160	55.6	35.6	8.8	0.2	0.05	$26	0.7	1	1	1
10	206	57.3	33.0	9.7	0.33	0.01	$3.7k	1.0	0	1	1
11	401	50.6	33.9	15.5	0.1	0.01	$5.7m	0.2	0	1	0
12	524	74.4	18.5	7.1	0.1	0.01	$28	0.2	1	1	0
13	95	46.3	44.2	9.5	0.1	0.01	$7.5m	0.2	0	0	1
14	30	40.0	56.7	3.3	0.1	0.1	$43.9k	0.4	0	1	1
15	61	32.8	55.7	11.5	0.1	0.01	$5.3m	0.2	0	1	1
16	54	25.9	61.1	13.0	0.1	0.01	$5.3m	0.2	0	1	1
17	108	54.6	37.0	8.3	0.1	0.01	$7.8m	0.2	0	1	1
18	501	78.8	15.4	5.8	0.1	0.01	$28	0.2	0	1	0
19	108	49.1	38.9	12.0	0.33	0.01	$2.3k	0.96	0	1	0
20	202	62.4	27.2	10.4	0.1	0.01	$6	0.9	1	1	1
21	199	52.3	33.2	14.6	0.1	0.1	$108	0.4	1	1	1
22	54	48.1	42.6	9.3	0.1	0.01	$5.3m	0.2	0	1	1
23	61	54.1	37.7	8.2	0.1	0.01	$5.3m	0.2	0	1	1
24	54	83.3	16.7	0.0	0.1	0.01	$22	0.5	1	1	1
25	20	65.0	35.0	0.0	0.1	0.01	$43	0.2	0	1	1
26	70	64.3	28.6	7.1	0.1	0.01	$5.7m	0.2	0	1	1
27	54	61.1	31.5	7.4	0.16	0.03	$1.7k	0.9	0	0	1
28	80	85.0	13.8	1.3	0.1	0.11	$4	0.2	1	1	1
29	30	46.7	43.3	10.0	0.2	0.05	$12	0.7	1	1	1
30	25	64.0	32.0	4.0	0.2	0.05	$14	0.5	0	1	1
31	54	83.3	14.8	1.9	0.1	0.01	$22	0.5	1	1	1
32	54	83.3	14.8	1.9	0.1	0.01	$22	0.5	1	1	1
33	68	50.0	33.8	16.2	0.1	0.01	$7.8m	0.2	0	0	1
34	79	94.9	5.1	0.0	0.1	0.01	$28	0.2	0	1	1
35	20	85.0	15.0	0.0	0.1	0.01	$43	0.2	1	1	1
36	54	83.3	13.0	3.7	0.1	0.01	$22	0.5	1	1	1
37	99	63.6	23.2	13.1	0.16	0.03	$1.7k	0.9	1	0	1
38	74	91.9	6.8	1.4	0.1	0.01	$28	0.2	1	1	1
39	196	52.6	27.6	19.9	0.1	0.1	$108	0.4	1	1	1
40	25	92.0	8.0	0.0	0.2	0.05	$14	0.5	1	1	1
41	56	46.4	32.1	21.4	0.2	0.05	$39.8k	0.5	0	1	1

Notes: The rows are ordered by the Conlisk z-test statistic indicating fanning-out patterns in the top block (numbers 1–38), no paradox in the middle block (numbers 39–65), and fanning-in patterns in the bottom block (numbers 66–81). Row 1 Sopher and Gigliotti (1993), Treatment 1; Row 2 Kahneman and Tversky (1979); Row 3 Cherry and Shogren (2007), no arbitrage (pre- and post-merged); Row 4 Conlisk (1989), basic version; Row 5 Birnbaum (2007), experiment 1, series A, questions 6–12; Row 6 Carlin (1992), experiment 1; Row 7 Wu (1994), problem C7; Row 8 Carlin (1990), trial #1; Row 9 Starmer and Sugden (1991); Row 10 Wu (1994), problem C4; Row 11 Huck and Müller (2012), HighHyp; Row 12 Huck and Müller (2012), LowReal; Row 13 Sopher and Gigliotti (1993), Treatment 3; Row 14 Camerer (1989), large gains; Row 15 Cherry and Shogren (2007), pre–cheap talk arbitrage; Row 16 Cherry and Shogren (2007), pre–real arbitrage; Row 17 Carlin (1992), experiment 2, form AP8; Row 18 Huck and Müller (2012), LowHyp; Row 19 Da Silva, Baldo, and Matsushita (2013); Row 20 Fan (2002), CR2; Row 21 Birnbaum (2007), experiment 2, A2, Q6–12; Row 22 Cherry and Shogren (2007), post–real arbitrage; Row 23 Cherry and Shogren (2007), post–cheap talk arbitrage; Row 24 Birnbaum, Schmidt, and Schneider (2017), experiment 1, CCE2, repetition 4 ; Row 25 Harrison (1994), AP0; Row 26 Huck and Müller (2012), HighHyp lab; Row 27 Groes et al. (1999), hypothetical; Row 28 Agranov and Ortoleva (2017); Row 29 List and Haigh (2005), students; Row 30 Burke et al. (1996), fixed Allais; Row 31 Birnbaum, Schmidt, and Schneider (2017), experiment 1, CCE2, R1; Row 32 Birnbaum, Schmidt, and Schneider (2017), experiment 1, CCE2, R3; Row 33 Carlin (1992), experiment 2, form AP9; Row 34 Huck and Müller (2012) LowHyp lab; Row 35 Harrison (1994), AP1; Row 36 Birnbaum, Schmidt, and Schneider (2017), experiment 1, CCE2, R2; Row 37 Groes et al. (1999), real; Row 38 Huck and Müller (2012), LowReal lab; Row 39 Birnbaum (2007), experiment. 2, condition A3, questions 6–12; Row 40 Burke et al. (1996), salient Allais; Row 41 Chew and Waller (1986), experiment 2.

(*continued*)

Table 1—Experimental Data Analyzed in This Paper (*continued*)

#	Obs.	EUT-consistent choices, %	Fanning-out-consistent choices, %	Fanning-in-consistent choices, %	Prob. of highest outcome	Prob. of lowest outcome	Highest outcome in 2010 USD	Middle/highest outcome	Real (1) or hypothetical (0) incentives	Lottery presentation (1) or not (0)	Students (1) or not (0)
42	197	46.7	29.4	23.9	0.1	0.1	$108	0.4	1	1	1
43	76	56.6	25.0	18.4	0.3	0.2	$63	0.5	1	1	0
44	54	64.8	20.4	14.8	0.1	0.1	$31	0.4	1	1	1
45	142	74.6	14.1	11.3	0.1	0.01	$8.3m	0.2	0	0	1
46	54	59.3	22.2	18.5	0.1	0.1	$31	0.4	1	1	1
47	180	76.7	12.2	11.1	0.1	0.01	$67.9k	0.2	0	1	0
48	43	65.1	18.6	16.3	0.05	0.05	$147	0.4	0	1	1
49	44	47.7	27.3	25.0	0.2	0.2	$38	0.3	1	1	1
50	92	64.1	18.5	17.4	0.3	0.1	$56	0.3	1	1	1
51	34	70.6	14.7	14.7	0.3	0.2	$63	0.5	1	1	0
52	20	40.0	30.0	30.0	0.1	0.1	$18	0.5	0	1	1
53	202	64.9	17.3	17.8	0.1	0.01	$6	0.9	0	1	1
54	92	63.0	17.4	19.6	0.15	0.1	$56	0.3	1	1	1
55	49	89.8	4.1	6.1	0.1	0.01	$44	0.2	1	1	1
56	54	70.4	13.0	16.7	0.2	0.05	$12	0.7	1	1	0
57	10	70.0	10.0	20.0	0.1	0.1	$18	0.5	1	1	1
58	92	70.7	13.0	16.3	0.3	0.1	$37	0.3	1	1	1
59	92	58.7	18.5	22.8	0.1	0.1	$56	0.3	1	1	1
60	54	64.8	14.8	20.4	0.1	0.1	$28	0.4	1	1	1
61	202	69.3	13.9	16.8	0.1	0.01	$6	0.2	0	1	1
62	91	70.3	12.1	17.6	0.1	0.15	$56	0.3	1	1	1
63	109	75.2	9.2	15.6	0.25	0.25	$11	0.4	1	1	0
64	109	64.2	13.8	22.0	0.25	0.25	$3	0.4	1	1	0
65	202	88.6	4.0	7.4	0.1	0.01	$6	0.2	1	1	1
66	212	72.2	10.8	17.0	0.1	0.01	$8.8m	0.2	0	0	1
67	45	80.0	4.4	15.6	0.2	0.2	$26	0.5	1	1	1
68	54	63.0	11.1	25.9	0.1	0.1	$31	0.4	1	1	1
69	100	63.0	12.0	25.0	0.25	0.25	$11	0.4	1	1	0
70	56	75.0	5.4	19.6	0.05	0.05	$199	0.4	0	1	1
71	118	55.9	15.3	28.8	0.25	0.25	$3	0.4	1	1	0
72	92	68.5	8.7	22.8	0.1	0.1	$37	0.3	1	1	1
73	92	54.3	14.1	31.5	0.15	0.1	$37	0.3	1	1	1
74	96	60.4	11.5	28.1	0.25	0.25	$11	0.4	1	1	0
75	92	79.3	3.3	17.4	0.1	0.15	$37	0.3	1	1	1
76	124	63.7	8.9	27.4	0.1	0.1	$17	0.4	1	1	1
77	70	51.4	4.3	44.3	0.25	0.25	$34	0.4	1	1	1
78	156	52.6	6.4	41.0	0.2	0.25	$29	0.3	1	1	1
79	155	55.5	5.2	39.4	0.2	0.25	$35	0.3	1	1	1
80	156	50.6	5.8	43.6	0.2	0.2	$29	0.3	1	1	1
81	155	40.6	3.2	56.2	0.2	0.25	$23	0.4	1	1	1

Notes: The rows are ordered by the Conlisk z-test statistic indicating fanning-out patterns in the top block (numbers 1–38), no paradox in the middle block (numbers 39–65), and fanning-in patterns in the bottom block (numbers 66–81). Row 42 Birnbaum (2007), experiment 2, condition A3, questions 6–12; Row 43 Bateman and Munro (2005), T1 & T8; Row 44 Birnbaum, Schmidt, and Schneider (2017), experiment 1, CCE3, R2; Row 45 Carlin (1990), trial #2; Row 46 Birnbaum, Schmidt, and Schneider (2017), experiment 1, CCE3, R4; Row 47 Finkelshtain and Feinerman (1997); Row 48 Harless and Camerer (1994), extra to Chew and Waller (1986), experiment 1, c 1a; Row 49 Butler and Loomes (2011), A$60 group; Row 50 Loomes and Sugden (1998), group 1, Q36 & Q40; Row 51 Bateman and Munro (2005), W3 & W7; Row 52 Camerer (1989), small gains, hypothetical; Row 53 Fan (2002), HR2; Row 54 Loomes and Sugden (1998), group 1, Q20 & Q24; Row 55 Conlisk (1989), pilot; Row 56 List and Haigh (2005), traders; Row 57 Camerer (1989), small gains, real; Row 58 Loomes and Sugden (1998), group 2, Q36 & Q40; Row 59 Loomes and Sugden (1998), group 1, Q12 & Q16; Row 60 Birnbaum, Schmidt, and Schneider (2017), experiment 1, CCE3, R1; Row 61 Fan (2002), HR1; Row 62 Loomes and Sugden (1998), group 1, Q5 & Q8; Row 63 Humphrey and Verschoor (2004), Sironko; Row 64 Humphrey and Verschoor (2004), Vepur; Row 65 Fan (2002), CR1; Row 66 Conlisk (1989), three-step version; Row 67 Butler and Loomes (2011), A$40 group; Row 68 Birnbaum, Schmidt, and Schneider(2017), experiment 1, CCE3, R3; Row 69 Humphrey and Verschoor (2004b), Ethiopia; Row 70 Chew and Waller (1986), experiment 1; Row 71 Humphrey and Verschoor (2004b), Guddimalakapura; Row 72 Loomes and Sugden (1998), group 2, Q12 & Q16; Row 73 Loomes and Sugden (1998), group 2, Q20 & Q24; Row 74 Humphrey and Verschoor (2004), Bufumbo; Row 75 Loomes and Sugden (1998), group 2, Q5 & Q8; Row 76 Starmer (1992); Row 77 Blavatskyy (2013); Row 78 Baillon et al. (2016), CC2, stage 1; Row 79 Baillon et al. (2016), CC3, stage 1; Row 80 Baillon et al. (2016), CC1, stage 1; Row 81 Baillon et al. (2016), CC4, stage 1.

Allais question. Column "Highest outcome in 2010 USD" reports the highest payoff P standardized to 2010 US dollars (USD). To compare payoffs across different currencies and different years, we first apply the purchasing power parity conversion factor[4] to all payoffs in foreign currencies to convert them to comparable USD payoffs and then use the US CPI index (with 2010 as a base year) to express the outcomes in 2010 USD. The purchasing power parity conversion factor and the US CPI index were sourced from the World Bank Database. Column "Middle/highest outcome" shows the ratio O of the middle outcome to the highest outcome.

Column "Real (1) or hypothetical (0) incentives" is a dummy variable I that equals 1 if incentives, i.e., monetary outcomes in the experiment, were real and 0 if they were hypothetical. Column "Lottery presentation (1) or not (0)" is a dummy variable L that equals 1 if choice options were presented as lotteries (not in a compound or frequency format). Column "Students (1) or not (0)" is a dummy variable S that equals 1 if subjects were students.

The choice counts for Kahneman and Tversky (1979) are reconstructed from the number of participants and the frequencies of their choices reported in the paper. The reconstruction leads to the number of BC choices being –3, which may be due to rounding or reporting error. We set it to zero. Even if we exclude Kahneman and Tversky (1979) data from the analysis, the estimation results do not change substantially. We consider only stage 1 experiments in Baillon et al. (2016) to avoid any confounding with learning effects.

Figure 2 shows the fractions of the observed outcomes of choice patterns pooled across all the experiments in the dataset conditional on whether incentives are real or hypothetical. Some regularity in the data is apparent from a visual inspection of Figure 2 and/or Table 1. For example, the outcomes consistent with EUT (no paradox, labeled EUT in Figure 2) are prevalent across all the experiments, with choices not involving the riskless outcome being the clear modal choice for both hypothetical and real outcomes. (The risky choice is slightly less prevalent in the experiments with real incentives.) Moreover, fanning-out choice patterns clearly outnumber fanning-in choice patterns, by a factor of about three under hypothetical incentives. This pattern is reversed under real incentives, where fanning-in choice patterns outnumber modestly fanning-out choice patterns. Also of note, fanning-out choice patterns under hypothetical incentives are about twice as frequent as those under real incentives, as also suggested by a high (low) occurrence of a value of null (1) in column "Real (1) or hypothetical (0) incentives" at the top (bottom) part of Table 1.

Another apparent regularity is that studies reporting a classic AP (fanning-out choice pattern outnumbering fanning-in) typically use pairs of Allais questions with very uneven divisions of the probability mass, as manifested by the fact that probability PH is often ten times larger than probability PL at the top part of Table 1. On the other hand, studies reporting a reversed AP (fanning-in choice patterns outnumbering fanning-out) typically design pairs of Allais questions with an even division

[4] 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 US dollar would buy in the United States.

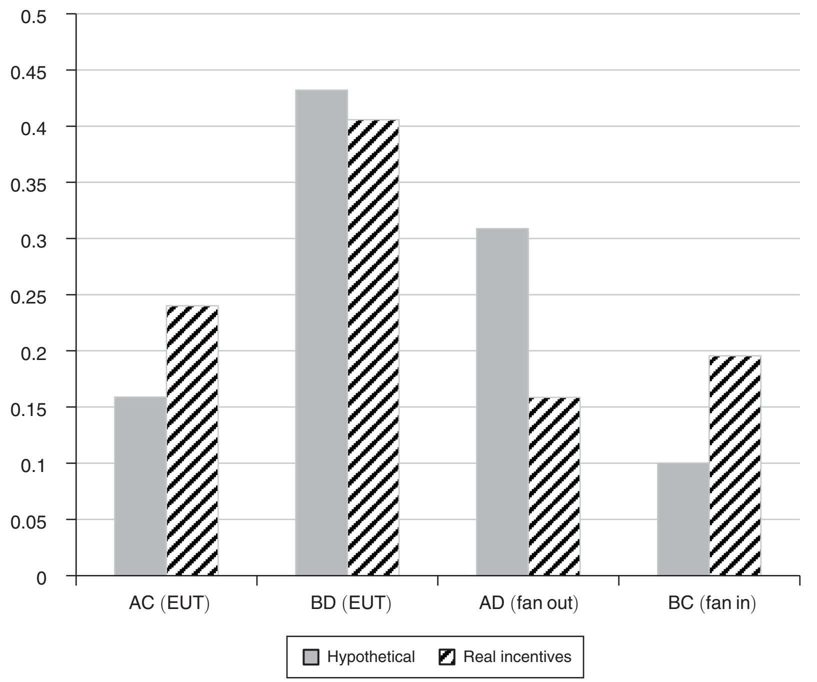

Figure 2. Observed Outcomes

Note: The fractions of the corresponding outcomes pooled across all data and reported separately for the experiments with real (4,835 observations) and hypothetical incentives (4,112 observations).

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.

B. *Regression Analysis*

We use the reduced form regression to describe statistical relationships between the outcomes of the experiments and the experimental design and implementation choices identified as relevant determinants of outcomes. Data from all considered experiments are combined in one dataset. Our unit of observation is an individual participant. We reconstruct individual choices from the frequencies of choice patterns reported in the 81 experiments in our dataset. The regressors stay the same for all participants in the same experimental treatment. The weight of each experiment in the combined dataset is given by the number of individual participants in each experiment.

All experiments result in four revealed choice patterns from the two questions: two choice patterns consistent with EUT—AC (subjects choose A over B and C over D) and BD (subjects choose B over A and D over C)—fanning-out choice pattern AD (subjects choose A over B and D over C), and fanning-in choice pattern BC (subjects choose B over A and C over D). Hence, the multinomial logistic specification is a sensible model to use in this setting; see also Huck and Müller (2012). Logistic regression specifies that the natural log of the probability ratios has a linear structure. In particular, we consider the following model:

$$\ln\left(\frac{\mathrm{Pr}_i}{\mathrm{Pr}_{\mathrm{AC}}}\right) = \beta_{i0} + \beta_{i1}\ln P \times I + \beta_{i2}\ln P \times \left(1 - I\right) + \beta_{i3}I + \beta_{i4}L + \beta_{i5}S$$
$$+ \beta_{i6}O + \beta_{i7}PH/PL,$$

where Pr_i is the probability to observe a specific choice pattern, $i = \{\mathrm{BD, BC, AD}\}$, and AC is set as the baseline outcome.[5]

The highest payoffs P are natural logged to reconcile a wide range of P values starting from US\$(2010)3–US\$(2010)8.8 million and reflect saturation. There is a strong negative correlation between $\ln P$ and the real incentives dummy variable, I, as studies with high payoffs typically use no real monetary incentives. We use the interaction terms $\ln P \times I$ and $\ln P \times (1 - I)$ to allow for different slopes for $\ln P$ for the cases of real and hypothetical payoffs, respectively.

C. *Results*

Table 2 presents the results of the four-outcome logistic regression. The relationship between the coefficient estimates and the probabilities of the revealed choice patterns is nonlinear. To simplify the interpretation of the results, we report the average marginal effects, which are observation-specific marginal effects averaged over all observations.[6] Note that average marginal effects for each explanatory variable sum up to zero over all possible choice patterns.

We report regular standard errors as well as cluster-robust standard errors. The cluster-robust method allows for correlated residuals within clusters but not across clusters. Correlations may be induced by some unobserved conditions specific to a cluster. We cluster at the level of the research team (proxied by published articles) and hence have 29 clusters.

Two explanatory variables affect mostly risk preferences: dummy variables for incentives (not in interaction with the size of payoffs) and students. In particular, having real incentives increases the probability of risk-averse EUT-consistent choices by 0.128, and having student subjects reduces this probability by 0.054, yet

[5] We also considered several alternative model specifications (as suggested by the referees) such as binary logit EUT-consistent versus non-EUT-consistent outcomes, linear probability model with the same two outcomes, binary logit EUT-consistent versus fanning-out (dropping fanning-in) outcomes, three-outcome logit EUT consistent versus fanning out versus fanning in, and ordered three-outcome logit with the following order: fanning in, EUT consistent, and fanning out. The results for these alternative model specifications as well as additional regression information and diagnostics are reported in Appendix Table A1 (logit average marginal effects and linear probability model) and Appendix Table A2 (logit regression coefficients for log odds ratios). The four-outcome logit specification had one of the highest pseudo R^2 values, and that is why we report and discuss this specification in the main text.

[6] In logit regressions, the coefficients are estimated for the odds ratios. For ease of interpretation, these coefficients are transformed into marginal effects of independent variable on the probability of specific choice for each choice category (see Appendix for the details). The marginal effects can be added to obtain the marginal effects of combined outcomes; e.g., the marginal effect for the EUT-consistent AC and BD is the sum of the marginal effects for the AC and BD outcomes for each explanatory variable. The logit coefficient estimates for the log odds ratios are reported in Table A2 in the Appendix. Also note that the computed marginal effects for the binary logit model nearly coincide with the coefficients of the analogous linear probability model reported in Appendix Table A1.

Table 2—Average Marginal Effects Computed from the Logit Model

Explanatory variables	$\ln P \times I$	$\ln P \times (1-I)$	I	L	S	O	PH/PL
Prob. of choice	(ln payoffs, real)	(ln payoffs, hypothet.)	(=1, real incentives)	(=1, lottery)	(=1, student)	(=mid/high)	(slope in the prob. triangle)
Pr(AC, EUT safe)	**0.013**	**0.015**	**0.128**	−0.008	**−0.054**	**0.336**	**−0.005**
Stand. errors	−0.005	−0.002	−0.025	−0.017	−0.011	−0.024	−0.001
Cl. stand. errors	−0.027	−0.004	−0.115	−0.056	−0.061	−0.13	−0.005
Pr(BD, EUT risky)	**−0.056**	**−0.032**	−0.025	**−0.121**	**0.039**	**−0.588**	**0.015**
Stand. errors	−0.007	−0.001	−0.03	−0.021	−0.013	−0.028	−0.001
Cl. stand. errors	−0.039	−0.005	−0.163	−0.089	−0.063	−0.165	−0.006
Pr(AD, fan out)	**0.047**	**0.019**	**−0.105**	**0.155**	0.005	**0.207**	0.001
Stand. errors	−0.006	−0.001	−0.028	−0.017	−0.011	−0.026	−0.001
Cl. stand. errors	−0.017	−0.004	−0.065	−0.055	−0.027	−0.064	−0.003
Pr(BC, fan in)	−0.004	−0.002	0.002	−0.026	0.011	**0.045**	**−0.011**
Stand. errors	−0.004	−0.001	−0.021	−0.017	−0.01	−0.022	−0.001
Cl. stand. errors	−0.011	−0.005	−0.06	−0.037	−0.03	−0.077	−0.005

Notes: We report regular standard errors (in parentheses) and the cluster-robust standard errors (clustered at the article level). Coefficients significant at the 0.05 level for both the regular and cluster-robust methods are highlighted with **bold black** font. Coefficients significant at the 0.05 level for the regular but not the cluster-robust method are highlighted with **bold red** font.

both are not significant when using cluster-robust standard errors. At the same time, having student subjects is not significant (statistically and economically) for the probabilities of the fanning-out and fanning-in choice patterns.

The variables that increase the probability of the fanning-out choice pattern the most are having hypothetical incentives $I = 0$; higher payoffs P, especially when they are real; presentation of choice options as lotteries L (not in a frequency or compound lottery form); and the higher ratio of middle to highest payoff O. Having hypothetical incentives increases, on average, the probability of the fanning-out pattern by 0.105. Real payoffs contribute to an average 0.047 increase in the probability of the occurrence of the fanning-out pattern per 1 percent increase in P. Hypothetical payoffs have a similar but somewhat smaller effect (0.019). When choice options are presented as lotteries, we are much more likely to observe the fanning-out pattern; i.e., the increase in the corresponding probability is 0.155. The closer the ratio of middle to highest payoff O is to 1, the higher the probability of selecting the sure-choice option A in the first question. This leads to the higher probability of the EUT-consistent safer AC pattern (average increase in the probability is 0.0336 per 0.1 increase in the ratio) and the higher probability of the fanning-out pattern (average increase in the probability is 0.0207 per 0.1 increase in the ratio).

The variables that have a statistically significant effect on the probability of fanning-in choice pattern are the ratio of middle to highest payoff O and the slope of lines AB and CD in the probability triangle PH/PL. The division of the probability mass captured by the PH/PL is an important predictor for the fanning-in pattern. One unit increase in PH/PL leads to an average 0.011 decrease in the probability of the fanning-in pattern. Even though this effect might appear small, note that the ratio PH/PL can be as high as 33 in Kahneman and Tversky (1979).

To summarize, we find the probability of observing the classic AP—that is, the fanning-out pattern—can be significantly increased by having hypothetical incentives, increasing payoffs, presenting the questions in the lottery format, and setting the mid payoff closer to the highest payoff. At the same time, the probability mass distribution is the significant predictor for the reversed AP—that is, decreases in PH/PL lead to increases in the fanning-in pattern.

D. *Discussion*

Our results demonstrate that the AP is by no means a robust behavioral regularity. The instances of the AP are affected by specific experimenters' choices for the size of payoffs, incentives, lottery presentation, and design. Our result is in the spirit of Gigerenzer's (1991) deconstruction of well-known alleged cognitive biases. For example, our results indicate that people are more likely to violate EUT (in particular, in the direction consistent with the fanning out of indifference curves) when outcomes in the Allais questions are large and hypothetical. Indeed, Camerer (1989) finds that subjects tend to reveal fanning-out choice patterns when outcomes are large gains but finds no systematic violations of EUT when outcomes are small gains. If high payoffs increase the likelihood of observing fanning out, then—in those rare real-world situations where decision-makers have to decide over large stakes—they are more likely to exhibit the AP.

As another example, our results indicate that people are more likely to violate EUT (in particular, in the direction consistent with the 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 EUT are more systematic in the direction of fanning-in choice patterns when probability distributions are presented as compound lotteries. In light of our results, the claim that the AP is a robust behavioral phenomenon is incorrect: we identify the experimental conditions that affect the likelihood of observing the AP (or the reverse thereof).

It is important to get these empirical facts straight because empirical evidence ultimately affects the development 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 payoff size, incentives, lottery presentation, and 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 development of relevant theories.

IV. Conclusion

Allais (1953) proposed a textbook example of a possible violation of EUT. Yet, as a test of the descriptive validity of EUT, the original Allais (1953) example is of limited value. The example involves very large outcomes that are typically not implementable in laboratory experiments with real incentives. Moreover, differences in probability values in the Allais (1953) example are relatively small. It is challenging to find a reasonable scenario where people are faced with decisions resembling the original Allais (1953) example. In other words, the environment of the Allais (1953) example is unfamiliar to most experimental subjects. Allais (1953, 526) designed his example to maximize the advantage (or disadvantage) of extreme complementarity between lottery outcomes in the choice between a sure payoff and a risky lottery when outcomes are very large.[7] In a certain sense, this example is a stress test of EUT: if the theory were to hold in such an extreme example, one could reasonably expect it to hold in less extreme situations. Arguably, a more practical test of the descriptive validity of EUT would avoid astronomically large outcomes and tiny probability differences. Indeed, the experimental evidence suggests that the independence axiom of the EUT is less frequently violated in the interior of the probability triangle (cf. Camerer 1995).

A perception frequently found in the literature, which motivated the development of numerous generalized non-EUTs, is that the AP is a robust empirical finding. Above, we have brought this perception to the data in a meta-analysis. Specifically, we have demonstrated how specific choices of design and implementation characteristics and parameters affect the likelihood of observing the AP (or the reverse thereof). The AP is likely to be observed in experiments with high hypothetical payoffs, the medium outcome being close to the highest outcome (which makes a harder choice) and when lotteries are presented as a probability distribution (not in a compound form). The AP is likely to be reversed in experiments when the probability mass is equally split between the lowest and highest outcomes in risky lotteries (which makes an easier choice).

Our findings confirm that the way one designs and conducts an experiment may have a substantial effect on the outcomes. This is by no means a novel insight, but it had not yet been demonstrated for the AP in a comprehensive, systematic, and tractable way.

Appendix

Average Marginal Effects.—Marginal effect of continuous explanatory variable x on probability $\Pr(Y_i = k)$ that individual i chooses outcome k is $dP(Y_i = k)/dx = \hat{\Pr}(Y_i = k)\left(\beta_k - \sum_{j=1}^{K}\hat{\Pr}(Y_i = j)\,\beta_j\right)$, where $\hat{\Pr}(Y_i)$ are the predicted probabilities of corresponding outcomes and βs are the coefficient estimates on explanatory variable x from the corresponding logit (relative to the baseline outcome). Note that the computed marginal effect is individual specific. To compute the overall marginal effect, we average all individual marginal effects. For a discrete explanatory variable, the marginal effect is computed by calculating the average predicted probabilities for each value of the discrete variable and then taking differences. Estimation and transformations to the average marginal effects were performed in Stata 16.

[7] Allais (1953, 526) considers "des cas extrêmes où l'avantage (ou l'inconvénient) de la complémentarité peut devenir particulièrement marqué. Tel est en particulier le cas des choix entre des gains certains et des gains aléatoires, lorsque les gains ont une grande valeur par rapport à la fortune du joueur."

Table A1—Logit Average Marginal Effects and Linear Probability Model

Explanatory variables / Prob. of choice	$\ln P \times I$ (ln payoffs, real)	$\ln P \times (1-I)$ (ln payoffs, hypothet.)	I (=1, real incentives)	L (=1, lottery)	S (=1, student)	O (=mid/high)	PH/PL (slope in the prob. triangle)
Logit two-outcome spec: EUT versus non-EUT; pseudo R^2= 0.032							
Pr(non-EUT)	**0.035**	**0.023**	−0.016	**0.185**	0.024	**0.309**	**−0.008**
SE	(0.006)	(0.002)	(0.029)	(0.021)	(0.013)	(0.029)	(0.001)
p-values	0	0	0.591	0	0.067	0	0
Cl. SE	(0.018)	(0.003)	(0.083)	(0.046)	(0.034)	(0.061)	(0.002)
Cl. *p*-value	0.058	0	0.85	0	0.477	0	0
Linear probability model with two outcomes: EUT (0) versus non-EUT (1); adj R^2 = 0.042							
Pr(non-EUT)	**0.035**	**0.023**	−0.023	**0.192**	0.024	**0.298**	**−0.008**
SE	(0.006)	(0.001)	(0.028)	(0.021)	(0.013)	(0.029)	(0.001)
p-values	0	0	0.409	0	0.065	0	0
Cl. SE	(0.018)	(0.003)	(0.075)	(0.046)	(0.033)	(0.058)	(0.002)
Cl. *p*-value	0.055	0	0.759	0	0.476	0	0.001
Logit two-outcome spec: EUT versus non-EUT excluding BC fan-in outcome; pseudo R^2 = 0.068							
P(AD, fan out)	**0.048**	**0.024**	**−0.072**	**0.188**	0.008	**0.265**	−0.002
SE	(0.007)	(0.001)	(0.030)	(0.019)	(0.013)	(0.029)	(0.001)
p-values	0	0	0.016	0	0.536	0	0.096
Cl. SE	(0.019)	(0.002)	(0.060)	(0.056)	(0.030)	(0.045)	(0.002)
Cl. *p*-value	0.01	0	0.23	0.001	0.788	0	0.406
Logit three-outcome spec: EUT, fan-out AD and fan-in BC; pseudo R^2 = 0.057							
Pr(AC and BD, EUT)	**−0.037**	**−0.020**	0.046	**−0.152**	−0.022	**−0.271**	**0.011**
SE	(0.007)	(0.002)	(0.030)	(0.021)	(0.013)	(0.030)	(0.001)
p-values	0	0	0.12	0	0.097	0	0
Cl. SE	(0.016)	(0.003)	(0.067)	(0.045)	(0.031)	(0.051)	(0.003)
Cl. *p*-value	0.023	0	0.492	0.001	0.487	0	0.002
Pr(AD, fan out)	**0.043**	**0.021**	**−0.070**	**0.167**	0.009	**0.218**	0.001
SE	(0.006)	(0.001)	(0.027)	(0.017)	(0.011)	(0.026)	(0.001)
p-values	0	0	0.009	0	0.447	0	0.295
Cl. SE	(0.017)	(0.003)	(0.058)	(0.052)	(0.026)	(0.054)	(0.003)
Cl. *p*-values	0.011	0	0.229	0.001	0.74	0	0.731
Pr(BC, fan in)	−0.006	0.000	0.024	−0.015	0.013	**0.053**	**−0.012**
SE	(0.004)	(0.001)	(0.021)	(0.017)	(0.010)	(0.022)	(0.001)
p-values	0.17	0.814	0.255	0.358	0.178	0.017	0
Cl. SE	(0.010)	(0.005)	(0.065)	(0.041)	(0.030)	(0.077)	(0.005)
Cl. *p*-values	0.535	0.956	0.718	0.707	0.661	0.487	0.024
Ordered logit three-outcome spec in the following order: fan-in BC, EUT, fan-out AD; pseudo R^2 = 0.04							
Pr(AC and BD, EUT)	**−0.017**	**−0.011**	0.029	**−0.091**	0.003	**−0.052**	**−0.004**
SE	(0.003)	(0.001)	(0.015)	(0.011)	(0.007)	(0.015)	(0.001)
p-values	0	0	0.055	0	0.633	0.001	0
Cl. SE	(0.010)	(0.003)	(0.041)	(0.041)	(0.021)	(0.049)	(0.002)
Cl. *p*-values	0.068	0.001	0.492	0.026	0.878	0.286	0.066
Pr(AD, fan out)	**−0.006**	**−0.004**	0.009	**−0.030**	0.001	**−0.017**	**−0.001**
SE	(0.001)	(0.000)	(0.005)	(0.004)	(0.002)	(0.005)	(0.000)
p-values	0	0	0.06	0	0.633	0.001	0
Cl. SE	(0.004)	(0.002)	(0.013)	(0.019)	(0.006)	(0.015)	(0.001)
Cl. *p*-values	0.158	0.034	0.483	0.124	0.87	0.254	0.112
Pr(BC, fan in)	**0.023**	**0.014**	−0.038	**0.121**	−0.004	**0.070**	**0.005**
SE	(0.005)	(0.001)	(0.020)	(0.015)	(0.009)	(0.020)	(0.001)
p-values	0	0	0.055	0	0.633	0.001	0
Cl. SE	(0.012)	(0.003)	(0.054)	(0.053)	(0.027)	(0.062)	(0.003)
Cl. *p*-value	0.063	0	0.483	0.024	0.876	0.261	0.05

(*continued*)

Table A1—Logit Average Marginal Effects and Linear Probability Model (*continued*)

Explanatory variables Prob. of choice	$\ln P \times I$ (ln payoffs, real)	$\ln P \times (1-I)$ (ln payoffs, hypothet.)	I (=1, real incentives)	L (=1, lottery)	S (=1, student)	O (=mid/high)	PH/PL (slope in the prob. triangle)
Logit four-outcome spec: EUT safe AC, EUT risky BD, fan-out AD and fan-in BC; pseudo $R^2 = 0.065$							
Pr(AC, EUT safe)	**0.013**	**0.015**	**0.128**	−0.008	**−0.054**	**0.336**	**−0.005**
SE	(0.005)	(0.002)	(0.025)	(0.017)	(0.011)	(0.024)	(0.001)
p-values	0.007	0	0	0.617	0	0	0
Cl. SE	(0.027)	(0.004)	(0.115)	(0.056)	(0.061)	(0.130)	(0.005)
Cl. *p*-value	0.636	0	0.264	0.882	0.374	0.01	0.31
Pr(BD, EUT risky)	**−0.056**	**−0.032**	−0.025	**−0.121**	**0.039**	**−0.588**	**0.015**
SE	(0.007)	(0.001)	(0.030)	(0.021)	(0.013)	(0.028)	(0.001)
p-values	0	0	0.388	0	0.003	0	0
Cl. SE	(0.039)	(0.005)	(0.163)	(0.089)	(0.063)	(0.165)	(0.006)
Cl. *p*-values	0.153	0	0.876	0.175	0.542	0	0.021
Pr(AD, fan out)	**0.047**	**0.019**	**−0.105**	**0.155**	0.005	**0.207**	0.001
SE	(0.006)	(0.001)	(0.028)	(0.017)	(0.011)	(0.026)	(0.001)
p-values	0	0	0	0	0.68	0	0.31
Cl. SE	(0.017)	(0.004)	(0.065)	(0.055)	(0.027)	(0.064)	(0.003)
Cl. *p*-values	0.005	0	0.105	0.004	0.864	0.001	0.762
Pr(BC, fan in)	−0.004	−0.002	0.002	−0.026	0.011	**0.045**	**−0.011**
SE	(0.004)	(0.001)	(0.021)	(0.017)	(0.010)	(0.022)	(0.001)
p-values	0.35	0.09	0.917	0.115	0.258	0.037	0
Cl. SE	(0.011)	(0.005)	(0.060)	(0.037)	(0.030)	(0.077)	(0.005)
Cl. *p*-values	0.694	0.66	0.971	0.476	0.713	0.557	0.018

Notes: Coefficients significant at the 0.05 level for both the regular and cluster-robust methods are highlighted with **bold black** font. Coefficients significant at the 0.05 level for the regular but not the cluster-robust method are highlighted with **bold red** font.

Models.—In addition to the four-outcome logit presented and discussed in the main text, we also considered several alternative model specifications (as suggested by the referees) such as binary logit EUT-consistent versus non-EUT-consistent outcomes, linear probability model with the same two outcomes, binary logit EUT-consistent versus fanning-out (dropping fanning-in) outcomes, three-outcome logit EUT consistent versus fanning out versus fanning in, and ordered three-outcome logit with the following order: fanning in, EUT consistent, and fanning out. The ordered logit model assumes that outcomes can be ordered in a specific way and that the coefficients of the linear relationships for all the logs of "higher outcome" to "lower outcome" ratios are the same.

Discussion.—Four-outcome specification has one of the largest pseudo R^2 values; the average marginal effects of the combined outcomes generally agree with the marginal effects of the binary and three-outcome logit models. The average marginal effects of the binary EUT/non-EUT logit model are nearly the same as the coefficient estimates of the analogous linear probability model. The average marginal effects of the binary logit model of EUT against the fanning-out pattern (excluding the fanning-in pattern) are nearly the same as those of the fanning-in pattern in the three-outcome logit model. The pseudo R^2 and the marginal effects of the ordered three-outcome logit model suggest that this specification is not a good choice. Likely, the assumptions behind the ordered logit model do not hold in this case.

For completeness in Table A2, we present the logit regression coefficients for log odds ratios.

Table A2—Logit Regression Coefficients for log Odds Ratios

Explanatory variables / Odds ratios	$\ln P \times I$ (ln payoffs, real)	$\ln P \times (1 - I)$ (ln payoffs, hypothet.)	I (=1, real incentives)	L (=1, lottery)	S (=1, student)	O (=mid/high)	PH/PL (slope in the prob. triangle)
Logit two-outcome spec: EUT versus non-EUT (baseline EUT)							
Non-EUT/EUT	**0.155**	**0.100**	−0.070	**0.819**	0.106	**1.369**	**−0.034**
SE	(0.029)	(0.007)	(0.129)	(0.093)	(0.058)	(0.130)	(0.005)
p-values	0	0	0.591	0	0.067	0	0
Cl. SE	(0.083)	(0.013)	(0.367)	(0.205)	(0.148)	(0.278)	(0.009)
Cl. *p*-values	0.061	0	0.85	0	0.474	0	0
Logit two-outcome spec: EUT versus fan-out AD excluding BC fan-in outcome (baseline EUT)							
Fan-out AD/EUT	**0.265**	**0.131**	**−0.398**	**1.038**	0.044	**1.465**	−0.009
SE	(0.037)	(0.008)	(0.166)	(0.109)	(0.071)	(0.160)	(0.005)
p-values	0	0	0.016	0	0.537	0	0.096
Cl. SE	(0.101)	(0.013)	(0.330)	(0.318)	(0.163)	(0.278)	(0.011)
Cl. *p*-values	0.008	0	0.228	0.001	0.787	0	0.409
Logit three-outcome spec: EUT versus fan-out AD and fan-in BC (baseline EUT)							
fan-out AD/EUT	**0.269**	**0.133**	**−0.413**	**1.056**	0.076	**1.489**	**−0.012**
SE	(0.037)	(0.008)	(0.168)	(0.109)	(0.071)	(0.163)	(0.005)
p-values	0	0	0.014	0	0.282	0	0.022
Cl. SE	(0.103)	(0.014)	(0.345)	(0.315)	(0.155)	(0.296)	(0.012)
Cl. *p*-values	0.009	0	0.231	0.001	0.622	0	0.289
Fan-in BC/EUT	0.017	**0.030**	0.089	0.136	0.125	**0.796**	**−0.096**
SE	(0.037)	(0.010)	(0.170)	(0.139)	(0.080)	(0.184)	(0.009)
p-values	0.633	0.004	0.602	0.328	0.12	0	0
Cl. SE	(0.076)	(0.041)	(0.517)	(0.313)	(0.232)	(0.608)	(0.039)
Cl. *p*-values	0.818	0.46	0.863	0.664	0.592	0.191	0.014
Ordered logit in the order: fan-in BC, EUT, fan-out AD							
"Higher"/"Lower"	**0.140**	**0.085**	−0.229	**0.729**	−0.026	**0.421**	**0.032**
SE	(0.028)	(0.006)	(0.119)	(0.091)	(0.054)	(0.121)	(0.004)
p-values	0	0	0.055	0	0.633	0.001	0
Cl. SE	(0.073)	(0.020)	(0.326)	(0.316)	(0.166)	(0.372)	(0.017)
Cl. *p*-values	0.054	0	0.482	0.021	0.877	0.258	0.063
Logit four-outcome spec (baseline EUT, safe AC)							
EUT risky BD/AC	**−0.223**	**−0.165**	**−0.741**	**−0.289**	**0.392**	**−3.383**	**0.067**
SE	(0.038)	(0.011)	(0.180)	(0.124)	(0.077)	(0.179)	(0.006)
p-values	0	0	0	0.019	0	0	0
Cl. SE	(0.245)	(0.038)	(1.014)	(0.538)	(0.480)	(1.196)	(0.045)
Cl. *p*-values	0.363	0	0.465	0.591	0.414	0.005	0.134
Fan-out AD/AC	**0.153**	0.010	**−1.158**	**0.770**	**0.305**	**−0.807**	**0.032**
SE	(0.043)	(0.012)	(0.213)	(0.134)	(0.087)	(0.204)	(0.007)
p-values	0	0.394	0	0	0	0	0
Cl. SE	(0.154)	(0.041)	(0.679)	(0.399)	(0.361)	(0.956)	(0.039)
Cl. *p*-values	0.32	0.801	0.088	0.053	0.399	0.399	0.41
Fan-in BC/AC	**−0.106**	**−0.096**	**−0.634**	−0.164	**0.358**	**−1.501**	**−0.046**
SE	(0.043)	(0.014)	(0.216)	(0.159)	(0.095)	(0.221)	(0.010)
p-values	0.013	0	0.003	0.301	0	0	0
Cl. SE	(0.118)	(0.026)	(0.363)	(0.252)	(0.393)	(0.558)	(0.019)
Cl. *p*-values	0.369	0	0.081	0.514	0.362	0.007	0.015

Notes: Coefficients significant at 0.05 level for both the regular and cluster-robust methods are highlighted with **bold black** font. Coefficients significant at 0.05 level for the regular, but not the cluster-robust method are highlighted with **bold red** font.

REFERENCES

Agranov, Marina, and Pietro Ortoleva. 2017. "Stochastic Choice and Preferences for Randomization." *Journal of Political Economy* 125 (1): 40–68.

Allais, Maurice. 1953. "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulates et Axiomes de l'Ecole Américaine." *Econometrica* 21: 503–46.

Baillon, Aurélien, Han Bleichrodt, Ning Liu, and Peter P. Wakker. 2016. "Group Decision Rules and Group Rationality under Risk." *Journal of Risk and Uncertainty* 52 (2): 99–116.

Bateman, Ian, and Alistair Munro. 2005. "An Experiment on Risky Choice amongst Households." *Economic Journal* 115 (502): 176–89.

Bernoulli, Daniel. 1738. "Specimen Theoriae Novae de Mensura Sortis." *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 5: 175–92.

Bierman, Harold Jr. 1989. "The Allais Paradox: A Framing Perspective." *Behavioral Science* 34 (1): 46–52.

Birnbaum, Michael 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 (2): 154–73.

Birnbaum, Michael H., Ulrich Schmidt, and Miriam D. Schneider. 2017. "Testing Independence Conditions in the Presence of Errors and Splitting Effects." *Journal of Risk and Uncertainty* 54 (1): 61–85.

Blavatskyy, Pavlo R. 2010. "Reverse Common Ratio Effect." *Journal of Risk and Uncertainty* 40 (3): 219–41.

Blavatskyy, Pavlo R. 2013. "The Reverse Allais Paradox." *Economics Letters* 119 (1): 60–64.

Blavatskyy, Pavlo, Andreas Ortmann, and Valentyn Panchenko. 2022. "Replication data for: On the Experimental Robustness of the Allais Paradox." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.38886/E119324V2.

Brandstätter, Eduard, Gerd Gigerenzer, and Ralph Hertwig. 2008. "Risky Choice with Heuristics: Reply to Birnbaum (2008), Johnson, Schulte-Mecklenbeck, and Willemsen (2008), and Rieger and Wang (2008)." *Psychological Review* 115 (1): 281–90.

Burke, Michael S., John R. Carter, Robert D. Gominiak, and Daniel F. Ohl. 1996. "An Experimental Note on the Allais Paradox and Monetary Incentives." *Empirical Economics* 21 (4): 617–32.

Butler, David, and Graham Loomes. 2011. "Imprecision as an Account of Violations of Independence and Betweenness." *Journal of Economic Behavior and Organization* 80 (3): 511–22.

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

Camerer, Colin F. 1995. "Individual Decision Making." In *Handbook of Experimental Economics*, edited by John H. Kagel and Alvin E. Roth, 587–703. Princeton, NJ: Princeton University Press.

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

Carlin, Paul S. 1990. "Is the Allais Paradox Robust to a Seemingly Trivial Change of Frame?" *Economics Letters* 34 (3): 241–44.

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 (2): 213–35.

Cason, Timothy N., and Charles R. Plott. 2014. "Misconceptions and Game Form Recognition: Challenges to Theories of Revealed Preference and Framing." *Journal of Political Economy* 122 (6): 1235–70.

Cherry, Todd, L., Peter Frykblom, and Jason F. Shogren. 2002. "Hardnose the Dictator ." *American Economic Review*, 92 (4): 1218–21.

Cherry, Todd L., and Jason F. Shogren. 2007. "Rationality Crossovers." *Journal of Economic Psychology* 28 (2): 261–77.

Conlisk, John. 1989. "Three Variants on the Allais Example." *American Economic Review* 79 (3): 392–407.

Da Silva, Sergio, Dinorá Baldo, and Raul Matsushita. 2013. "Biological Correlates of the Allais Paradox." *Applied Economics* 45 (5): 555–68.

Ellsberg, Daniel. 1961. "Risk, Ambiguity, and the Savage Axioms." *Quarterly Journal of Economics* 75 (4): 643–69.

Fan, Chinn-Ping. 2002. "Allais Paradox in the Small." *Journal of Economic Behavior and Organization* 49 (3): 411–21.

Finkelshtain, Israel, and Eli Feinerman. 1997. "Framing the Allais Paradox as a Daily Farm Decision Problem: Tests and Explanations." *Agricultural Economics* 15 (3): 155–67.

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 (3): 592–96.

Grether, David M., and Charles R. Plott. 1979. "Economic Theory of Choice and the Preference Reversal Phenomenon." *American Economic Review* 69 (4): 623–38.

Groes, Ebbe, Hans Jørgen Jacobsen, Birgitte Sloth, and Torben Tranæs. 1999. "Testing the Intransitivity Explanation of the Allais Paradox." *Theory and Decision* 47: 229–45.

Harrison, Glenn W. 1989. "Theory and Misbehavior of First-Price Auctions." *American Economic Review* 79 (4): 749–62.

Harrison, Glenn W. 1992. "Theory and Misbehavior of First-Price Auctions: Reply." *American Economic Review* 82 (5): 1426–43.

Harrison, Glenn W. 1994. "Expected Utility and the Experimentalists." *Empirical Economics* 19 (2): 223–53.

Harless, David W., and Colin F. Camerer. 1994. "The Predictive Utility of Generalized Expected Utility Theories." *Econometrica* 62 (6): 1251–89.

Herrmann, Tabea, Olaf Hübler, Lukas Menkhoff, and Ulrich Schmidt. 2017. "Allais for the Poor: Relations to Ability, Information Processing, and Risk Attitudes." *Journal of Risk and Uncertainty* 54 (2): 129–56.

Hertwig, Ralph, and Andreas Ortmann. 2001. "Experimental Practices in Economics: A Challenge for Psychologists?" *Behavioral and Brain Sciences* 24 (3): 383–403.

Hong, Chew Soo, and William S. Waller. 1986. "Empirical Tests of Weighted Utility Theory." *Journal of Mathematical Psychology* 30 (1): 55–72.

Huck, Steffen, and Wieland Müller. 2012. "Allais for All: Revisiting the Paradox in a Large Representative Sample." *Journal of Risk and Uncertainty* 44 (2): 261–93.

Humphrey, Steven, and Arjan Verschoor. 2004a. "Decision-Making under Risk among Small Farmers in East Uganda." *Journal of African Economies* 13 (1): 44–101.

Humphrey, Steven, and Arjan Verschoor. 2004b. "The Probability Weighting Function: Experimental Evidence from Uganda, India and Ethiopia." *Economics Letters* 84 (3): 419–25.

Kahneman, Daniel. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review* 93 (5): 1449–75.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–92.

Kahneman, Daniel, and Amos Tversky. 1996. "On the Reality of Cognitive Illusions." *Psychological Review* 103 (3): 582–91.

Laury, Susan K., and Charles A. Holt. 2008. "Payoff Scale Effects and Risk Preference under Real and Hypothetical Conditions." In *Handbook of Experimental Economics Results,* edited by Charles R. Plott and Vernon L. Smith, 1047–53. Amsterdam: North-Holland.

Leland, John W. 1994. "Generalized Similarity Judgments: An Alternative Explanation for Choice Anomalies." *Journal of Risk and Uncertainty* 9: 151–72.

L'Haridon, Oliver, and Laetitia Placido. 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., and Michael S. Haigh. 2005. " A Simple Test of Expected Utility Theory Using Professional Traders." *Proceedings of the National Academy of Science* 102 (3): 945–948.

Loomes, Graham, and Robert Sugden. 1998. "Testing Different Stochastic Specifications of Risky Choice." *Economica* 65 (260): 581–98.

Machina, Mark J. 1982. "'Expected Utility' Analysis without the Independence Axiom." *Econometrica* 50 (2): 277–323.

Plott, Charles R., and Kathryn Zeiler. 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): 530–45.

Plott, Charles R., and Kathryn Zeiler. 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.

Rubinstein, Ariel. 1988. "Similarity and Decision Making under Risk (Is There a Utility Theory Resolution to the Allais-Paradox?)." *Journal of Economic Theory* 46 (1): 145–53.

Slovic, Paul, and Amos Tversky. 1974. "Who Accepts Savage's Axiom?" *Behavioral Sciences* 19 (6): 368–73.

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

Sopher, Barry, and Gary Gigliotti. 1993. "A Test of Generalized Expected Utility Theory." *Theory and Decision* 35 (1): 75–106.

Starmer, Chris. 1992. "Testing New Theories of Choice under Uncertainty Using the Common Consequence Effect." *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): 332–82.

Starmer, Chris, and Robin Sugden. 1991. "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation." *American Economic Review* 81 (4): 971–78.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–31.

Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481): 453–58.

van de Kuilen, Gijs, and Peter P. Wakker. 2006. "Learning in the Allais Paradox." *Journal of Risk and Uncertainty* 33 (2): 155–64.

von Neumann, John and Oscar Morgenstern. 1947. *Theory of Games and Economic Behavior*. 2nd ed. Princeton, NJ: Princeton University Press.

Wu, George. 1994. "An Empirical Test of Ordinal Independence." *Journal of Risk and Uncertainty* 9 (1): 39–60.

Wu, George, and Richard Gonzalez. 1996. "Curvature of the Probability Weighting Function." *Management Science* 42 (12): 1676–90.