Allais for All: Revisiting the Paradox*

Steffen Huck†
University College London & ELSE

Wieland Müller‡
Tilburg University, CentER & TILEC

December 14, 2007

Abstract

We administer the Allais paradox questions to, both, a representative sample of the Dutch population and to student subjects. Three treatments are implemented: one with the original high hypothetical payoffs, one with low hypothetical payoffs and a third with low real payoffs. Our key findings are—(i) a large bulk of violations stems from non-familiarity with large payoffs; (ii) we can identify groups of the general population that have much more difficulty to make consistent decisions and may, hence, be much more prone to making imprudent financial choices; this concerns mainly the lowly educated and unemployed; and (iii) the shares of violations of standard theory observed in the lab provide lower bounds for the shares of violations observed in the population at large.

Keywords: Expected utility theory; Allais paradox; common consequence effect; field experiments; representative sample.

JEL classification numbers: C93; D81.

*We thank Marcel Das and Marika Puumala of CentERdata (University of Tilburg) for their most efficient support in collecting the data. Furthermore, we thank Johannes Binswanger, Oliver Kirchkamp, Tobias Klein, Sabine Kröger, Imran Rasul, Jan van Ours, Anthony Ziegelmeyer and participants of the 3rd International Meeting on Experimental and Behavioral Economics and the IMPRS Uncertainty Summer School as well as seminar participants at Tilburg University and Humboldt University Berlin for helpful comments. We gratefully acknowledge financial help from the UK's Economic and Social Research Council via ELSE and a grant on 'Behavioral Mechanism Design'. The second author acknowledges financial help from the Netherlands Organisation for Scientific Research (NWO) through a VIDI grant.

†Department of Economics & ELSE, University College London, Gower Street, London WC1E 6BT, UK, Fax: +44 207 916 2774, Email: *s.huck@ucl.ac.uk*.

‡Department of Economics, Tilburg University, Postbus 90153, 5000 LE Tilburg, The Netherlands, Fax: +31 13 466 3042, Email: *w.mueller@uvt.nl*.

1 Introduction

This paper presents the first evidence on the consistency of risk preferences in a representative population sample. We find that consistency increases in task familiarity and is linked to several personal characteristics such as education, income and asset holdings. Moreover, we investigate the external validity of a laboratory experiment with a student population that implemented the same choice problems as our household panel study. We find that, in line with studies on other biases, deviations from rationality observed in the lab provide a lower bound for deviations in a population at large.

Recently, several studies have made significant progress in understanding risk preferences in populations, making use of innovative survey methods and field experiments (Harrison and List 2004) including game shows with large stakes (Andersen *et al.* 2007). However, estimating the curvature of an agent's utility function necessitates that this agent's risk preferences can be represented by a function with the expected utility property. As is known from countless laboratory studies (see Camerer 1995 and Starmer 2000 for surveys) this assumption can be problematic. From the perspective of these studies, the present paper takes one step back by focussing on consistency of risk preferences in a representative subject pool—well over 1,400 members of the CentER Panel, a representative sample of the Dutch population. We do this by falling back on the oldest consistency test of all—by reinvestigating the Allais paradox. Our results help to understand the reliability and robustness of investigations into the actual distribution of risk preferences in populations.

Our research strategy is threefold. First, we implement three different treatments in the main experiment with the panel. We analyze the original Allais question with payoffs of millions of Euros that, as when Allais asked Savage, were purely hypothetical. In our second treatment we scaled the payments down but kept them hypothetical. Our third treatment used the same downscaled payoffs but paid them out for real. Thus, we are able to examine to what extent violations are driven by lack of monetary incentives, on the one hand, and non-familiarity with large sums of money on the other.

Second, we are able to exploit the wide range of background information that is available for our subjects in order to study the roots of violations.[1] Which personal characteristics are

[1]Several other studies have also used the CentERpanel as a subject pool. Let us briefly mention some of these studies. Hey (2002) and Carbone (2005) analyze more complicated and sequential individual decision making tasks and do not find any background variable systematically influencing behavior. Bellemare and Kröger (2007) study a trust game and find "that heterogeneity in behavior is characterized by several asymmetries—men, the young

correlated with violations? Are violations a matter of insufficient education or limited experience with financial decision making? Can we identify 'problem groups' that are, perhaps, more likely to suffer (in particular late in life) from imprudent financial decision making?

Third, we conduct a laboratory experiment with the usual laboratory subject population (students) employing the same design that we used in the panel experiment. Thus, we are able to examine the external validity of a laboratory experiment in a clear and detailed manner. In particular, we can compare whether and how a lab study can tell us something about the population at large.

Pursuing our threefold research strategy we are, thus, able to present the most detailed evidence on the Allais paradox so far. Our results are useful for a number of practical issues. (1) Our results point to a number of conditions that make standard theoretical predictions more likely to hold.[2] (2) Our results identify certain parts of the population that, due to inconsistencies, may have difficulties in making prudent financial decisions. (3) Our results contribute to a better understanding of what can be reliably learned from laboratory experiments.

Along the first dimension of our research strategy we find that violations in the original paradox are to a large extent driven by very high payoffs with which, in real life, virtually nobody has any practical experience. Violations in the original Allais problem are twice as high as in both downscaled versions. This effect has been observed before with student samples (Conlisk 1989); we show that the pattern extends to the general population. We do not find this result surprising as it simply stresses that economic theory can be expected to work much better in environments where agents make repeatedly at least somewhat similar decisions and are, thus, well-adapted. On the other hand, we find no substantial difference between the two downscaled versions. Whether subjects are incentivized or not, violations are much lower in both cases.[3]

Along the second dimension, we are able to identify a whole array of personal characteristics that correlate with inconsistent decision making. Education, occupation, income and asset holdings

and elderly, and low educated individuals invest relatively less, but reward significantly more investments." (p.183) Gaudecker *et al.* (2007a) elicit risk preferences and report that older people, women, the relatively uneducated, and those with lower income are more risk averse. For another study on individual risk attitudes using a large and representative German sample, see Dohmen *et al.* (2005).

[2] Of course, some predictions derived under the assumption of expected utility theory might still hold when agents suffer from inconsistencies but this is rarely explicitly analysed.

[3] For early studies of the Allais paradox see, e.g., MacCrimmon (1968), Slovic and Tversky (1974), Allais and Hagen (1979), Kahneman and Tversky (1979). For the effect of downscaled payoffs see Conlisk (1989), Starmer and Sugden (1991), Harrison, G.W., (1994), Burke *et al.* (1996), and Fan (2002).

do all correlate with inconsistent decision making and in each case the direction of effects is as one would guess. The better educated are more consistent and so are those in employment, those who earn more and those who hold financial assets.

Finally, our methodological contribution reveals that the laboratory results are rather useful in predicting behavior in a general population. First, the relative treatment differences are precisely the same for both populations, panel and lab. Second, as demonstrated in a number of other studies (see Gächter, Huck, and Weizsacker 2007 for a survey) the violations of standard theory observed in the lab provide a *lower bound* for violations observed in the population at large.[4]

The remainder of the paper is organized as follows. In Section 2, we describe the main characteristics of the CentERpanel and introduce the experimental design. In Section 3 we present our results obtained with the panel. We first give a quick overview of the results and then present a more detailed analysis, based on regression results, that also accounts for the effect of sociodemographic characteristics. In Section 4 we introduce our lab results and compare them to those obtained in the panel. Section 5 concludes.

2 Design and data collection

We administer the original "Allais questions," which consist of two pairwise lottery choices. Consider the following two choice problems. First, a subject is asked to choose between lotteries A and A^* where

$$A = \text{ Certainty of €1 Million} \qquad \text{and} \qquad A^* = \begin{cases} 1/100 \text{ Chance of €0} \\ 89/100 \text{ Chance of €1 Million} \\ 10/100 \text{ Chance of €5 Million} \end{cases}$$

Second, a subject is asked to choose between lotteries B and B^* where

$$B = \begin{cases} 89/100 \text{ Chance of €0} \\ 11/100 \text{ Chance of €1 Million} \\ \end{cases} \qquad \text{and} \qquad B^* = \begin{cases} 90/100 \text{ Chance of €0} \\ \\ 10/100 \text{ Chance of €5 Million} \end{cases}$$

[4] Almost all of the experiments on the Allais paradox conducted so far have used students as their subjects. There are two notable exceptions. List and Haigh (2005) test the Allais paradox both with students and professional traders from the Chicago Board of Trade. They report that both students and professional traders show Allais paradox behavior, but find that traders do so to a smaller extent. Fatas et al. (in press) use students and politicians and report similar result with students being more prone to Allais paradox behavior.

Of the four possible answers[5] AB, A^*B^*, AB^*, and A^*B only the first two are consistent with expected utility theory (henceforth, EUT) whereas the last two are not.[6] Many laboratory experiments have shown that violations of EUT are frequent and that a larger share of subjects violating EUT chooses AB^* instead of A^*B.[7]

We have six simple treatments using a between-subjects design. To introduce these treatments, consider the following lotteries over three outcomes of monetary payoffs with probabilities as above, i.e., $A = (0, 1, 0)$, $A^* = (.01, .89, .10)$, $B = (.89, .11, 0)$, $B^* = (.90, 0, .10)$. Our three treatments were then as follows:

- Treatment HiHyp: Original Allais questions with high hypothetical payoffs of €0, €1 Mio, and €5 Mio.[8]
- Treatment LoHyp: Allais questions with low hypothetical payoffs of €0, €5, and €25.[9]
- Treatment LoReal: Allais questions with low real payoffs of €0, €5, and €25.

For all three treatments we had two sub treatments reversing the order of decisions. As we do not find any order effects in the data we pool the data throughout.[10]

We collected data from a representative sample of the Dutch population. The experiments were conducted by CentERdata—an institute for applied economic and survey research for the social sciences—that is affiliated with the University of Tilburg in the Netherlands. CentERdata carries out its survey research mainly by using its own panel called CentERpanel. This panel is internet based and consists of some 2000 households in the Netherlands which form a representative

[5] Note that despite Harrison's (1994) arguments, we did not allow subjects to express indifference between lotteries. The reason for this is that this option was rarely chosen in earlier studies. MacDonald and Wall (1989, p.48), in the context of an Allais paradox experiment, report that the option to indicate idifference was chosen in only three of the 378 [0.0079%] choices. Also, Battalio *et al.* (1990) report that in their Allais paradox treatments only few subjects indicated indifference among lotteries.

[6] To see this note that $u(A) = u(€1M) > 0.01u(€0) + 0.89u(€1M) + 0.1u(€5M) = u(A^*)$ implies (add $0.89u(€0) - 0.89u(€1M)$ to both sides of the last inequality) $u(B) = 0.89u(€0) + 0.11u(€1M) > 0.9u(€0) + 0.1u(€5M) + 0.1u(€5M) = u(B^*)$.

[7] See, e.g., MacCrimmon (1968), Slovic and Tversky (1974), Allais and Hagen (1979), Kahneman and Tversky (1979), Conlisk (1989), Starmer and Sugden (1991), Harrison, G.W., (1994), Burke *et al.* (1996), and Fan (2002).

[8] Note that $EV(A) = €1$ Mio, $EV(A^*) = €1.39$ Mio, $EV(B) = €0.11$ Mio, and $EV(B^*) = €0.5$ Mio.

[9] Note that in this case, $EV(A) = €5$, $EV(A^*) = €6.95$, $EV(B) = €0.55$, and $EV(B^*) = €2.50$.

[10] See Harrison *et al.* (2005a) for a recent study on the elicitation of risk preferences where order effects matter.

sample of the Dutch population.[11] One of the advantages of the CentERpanel is that the researcher has access to background information for each panel member such as demographic and financial data. Every weekend, the panel members complete a questionnaire on the Internet from their home. The experiment reported here was conducted with CentERpanel over the third weekend of January 2006.

After logging on to our experiment, panel members were randomly assigned to one of the six different treatments introduced above. After being informed about the nature of the experiment, subjects decided whether or not to participate—as common with many modules of the panel. For participating subjects, the next screen introduced an example of a pair of lotteries (which were referred to as "Options"). Subjects were told that their task would be to express preference for one of the two lotteries and, additionally, how the preferred lottery would be executed.[12] When subjects indicated that they were ready to start the experiment, they were, in two consecutive screens, presented with their two Allais questions. Only after answering both Allais questions, the two preferred lotteries were played out (by the computer) and subjects were informed about the outcome of their two preferred lotteries. In the treatments with real monetary payments, subjects were paid according to the outcomes in both of their preferred lotteries.[13]

In total 1676 members of the CentERpanel logged on to our experiment. Of the subjects logging on, 1426 (85.1%) subjects decided to participate in our experiment while 250 (14.9%) subjects decided not to participate. Table 1 shows descriptive statistics of our sample. The column labeled "Participation" in Table 1 shows descriptive statistics of participating subjects in each of the three main treatments. (The column labeled "Violation" shows statistics for participating subjects violating or not violating EUT, respectively, which we will analyze further below.) The data in Table 1 is grouped according to gender, age, education, occupation and income.

Concentrating on descriptive statistics in the Participation columns of Table 1, we note that by and large most variables are relatively identical across treatments. However, in some of the age

[11] For more information about the CentERpanel and the way it is administered see http://www.uvt.nl/centerdata/en/whatwedo/thecenterpanel/.

[12] For more details see Appendix A which contains a translation of the screens used in the treatments with low payoffs. Note that the experiment was administered in Dutch.

[13] Note the following about payments in treatment LoReal. CentERdata reimburses the telephone costs for filling in questionnaires by exchanging "CentERpoints" (1 CentERpoint = 0.01 Euro) to panel members' private bank accounts four times a year. Altough lotteries were described in Euro amounts, subjects in the treatments with real monetary earnings were informed that: "In this experiment you can earn real money that will be paid in the form of CentERpoints."

Category		Participation			Violation		
		HiHyp	LoHyp	LoReal	NO	YES	p-value, χ^2
Gender	Female	48.9	46.5	47.0	46.2	50.0	0.180
Age	Age 16-24	6.2	8.0	6.9	6.5	8.4	0.156
	Age 25-34	24.4	19.2	19.7	20.6	21.5	
	Age 35-44	18.7	19.4	16.6	19.8	14.4	
	Age 45-54	17.5	19.2	23.3	20.2	20.1	
	Age 55-64	18.2	16.2	17.2	16.3	19.0	
	Age 65+	15.0	18.2	16.4	16.6	16.7	
Education	Primary education	5.0	6.0	5.7	5.0	7.1	0.008
	Lower secondary education	28.5	26.6	25.2	25.0	30.2	
	Higher secondary education	12.0	13.4	13.7	13.9	11.4	
	Intermediate vocational training	20.0	20.2	20.2	19.7	21.3	
	Higher vocational training	24.2	23.2	23.5	23.9	22.9	
	University degree	10.2	10.6	11.6	12.6	7.1	
Occupation	Employed (contract)	54.4	49.1	52.3	54.7	45.2	0.003
	Freelance or self-employed	3.7	3.2	3.8	4.0	2.5	
	Unemployed	1.5	2.4	1.9	1.5	3.0	
	Student	5.5	6.6	6.7	5.9	7.3	
	Works in own household	10.7	13.0	12.0	11.4	13.2	
	Retired	17.2	17.8	16.8	16.5	19.0	
	Other occupation	7.0	8.0	6.5	6.0	9.8	
Houshold	HH gross income $\leq$ €2250	22.7	26.0	27.9	22.7	32.6	< 0.001
Income	HH gross income €2251−€3130	28.4	21.8	23.5	24.0	24.9	
	HH gross income €3131−€4350	25.2	24.4	25.8	26.8	21.2	
	HH gross income $\geq$ €4351	23.7	28.0	22.9	26.5	21.2	
Assets	Holds assets	18.4	16.2	15.5	18.0	13.2	0.025
Savings acct.	Has savings account	59.4	53.5	50.6	52.4	57.8	0.062
	Maximum no. of observations	401	501	524	988	438	

Notes: Except for the number of observations, numbers indicate column percentages. Since some of the members of the CentERpanel did not complete the Dutch Household Survey, some observations are missing. The column labeled "p-value, χ^2" shows p-levels of χ^2 tests for differences between proportions of violating and non-violating subjects in the category listed in column 1.

Table 1: Descriptive statistics of the samples

Treatment	Absolute Frequency of Choices					Relative Frequency of Choices				
	AB	A^*B^*	AB^*	A^*B	Σ	AB	A^*B^*	AB^*	A^*B	Violations
HiHyp	82	121	136	62	401	20.4	30.2	33.9	15.5	49.4
LoHyp	22	373	77	29	501	4.4	76.0	15.7	3.9	19.6
LoReal	22	368	97	37	524	4.1	70.2	18.5	7.1	25.6
Σ	126	862	310	128	1426	8. 8	60.4	21.7	9.0	30.7

Table 2: Summary of experimental results in the panel

and income brackets as well as in the category savings account, there is some more variation.

Of course, we were worried about a possible sample selection problem that could be due to the fact that some subjects explicitly decided not participate in the panel experiment. Therefore, for all regressions reported in the result section we ran Heckman (1976) selection models using variable "Ratio" as one of the exclusion variables. The variable "Ratio" measures the proportion of questionnaires completed by panel members in the three months proceeding our experiment. This variable can be assumed to affect the participation decision but not the decisions taken in the experiment. For none of the regressions we found evidence for a selection bias.[14]

3 Results

3.1 Overview of experimental results

A summary of the experimental results is given in Table 2. The table shows both, the absolute frequency of choices (left part) and the relative frequency of choices (right part).

Violation of EUT. Note that the right-most column in Table 2 indicates that violations of EUT are observed in all treatments. In fact, we observe 49.5%, 19.6% and 25.6% violations of EUT in treatments HiHyp, LoHyp, and LoReal, respectively. Furthermore, in all treatments we observe that the fraction of EUT-violating AB^* answers is higher than the fraction of EUT-violating A^*B answers. The Z-statistic proposed in Conlisk (1989) indicates that the first fraction is significantly higher than the latter fraction at $p < 0.001$ in all treatments. We conclude that, as in earlier studies, violations of EUT are observed and that they are systematic in the sense that

[14] See Eckel and Grossman (2000), Bellemare and Kröger (2007), Gaudecker *et al.* (2007b) and Harrison *et al.* (2005b) for more evidence on selection issues.

AB^* is chosen more often than A^*B.

To facilitate comparison, note that Conlisk (1989) using a student sample for his "Basic Version" (which is comparable to our treatment HiHyp) reports the following relative frequencies of AB, A^*B^*, AB^*, and A^*B choices: 7.6%, 41.9%, 43.6%, and 6.8%. Thus, he observes EUT violation in 50.4% of the cases which compares to 45.5% in our panel treatment HiHyp.

The effect of high versus small hypothetical payoffs. Next consider the effect of high versus small hypothetical payoffs on the extent of EUT violation. For this purpose we compare the rates of EUT violations in treatments HiHyp and LoHyp. Table 2 shows that the rate of EUT violations drops from 49.4% in treatment HiHyp to 19.6% in treatment LoHyp. The D-statistic proposed in Conlisk (1989) indicates that this difference is highly significant at $p < 0.0001$ ($D = 9.115$).

Inspecting the relative frequencies of choices in Table 2 shows that moving from HiHyp to LoHyp sharply increases the fraction of choices consistent with expected value maximization (A^*B^*) at the expense of all other three possible responses. In particular, many more subjects prefer the payoff-maximizing choice A^* over A when (hypothetical) payoffs become small. A first possible explanation of this result is due to the fact that subjects in treatment LoHyp can be expected to be more familiar with the lower amounts of money leading them to make fewer mistakes.[15]

The effect of (small) real versus (small) hypothetical payoffs. Finally, consider the effect of (small) real versus (small) hypothetical payoffs on the extent of EUT violation. To analyze this, compare the rates of EUT violation in treatments LoHyp and LoReal. Table 2 shows that the rate of EUT violations is 19.6% in LoHyp whereas it is 25.6% in treatment LoReal. Thus, we see a slight increase in the share of EUT violations when we move from (small) hypothetical to (small) real payoffs. The D-statistic in Conlisk (1989) indicates that this difference is significant

[15] A second possible explanation for the reduction of Allais behavior (in particular the reduction of AB^* choices) in favor of an increase of expected value maximization (A^*B^* choices) when payoffs are drastically reduced has been put forward by Conlisk (1989). His explanation is based on two observations. First, Machina's (1982) generalization of EUT (especially the fanning-out model) does predict Allais behavior if payoffs are sufficiently large. However, when payoffs become small, the predictions of Machina's theory converge to the predictions made by EUT. Second, it is well-known that EUT converges to expected payoff maximization when payoffs become small. Taken together, these two observations provide an explanation for the above result. They predict Allais behavior when payoffs are sufficiently large and the disappearance of Allais behavior when payoffs become sufficiently small. The simpler explanation is, of course, that subjects have much more difficulties to make consistent decisions in domains they have no or little experience with.

($D = -1.6716$, $p = 0.047$). In contrast, Harrison (1994) and Burke *et al.* (1996) report that the use of low real instead of low hypothetical payoffs reduces the extent of EUT violation. For a broader overview on how incentives affect behavior in decisions under risk, see Camerer (1995, p. 634f).

3.2 Econometrics and the effect of socioeconomic characteristics

To test for across-treatment differences controlling for subjects' sociodemographic characteristics and to check whether any of theses characteristics influence behavior, we ran probit regressions with the variable "Violate" as the dependent variable. "Violate" is equal to 1 if a subject's answer to the Allais questions violates EUT (i.e., answers A^*B or AB^*), and is equal to 0 otherwise (i.e., answers AB or A^*B^*). The background variables we include in the regression are the ones shown in Table 1 above. The results are shown in Table 3 which reports marginal effects. Regression (1) includes all data whereas regressions (2) to (4) show results for each of the three treatments separately.

Let us first briefly reconsider across-treatment differences. For this purpose, refer to regression (1) in Table 3 which includes all data and controls for background variables. Importantly, note that in regression (1) the omitted treatment dummy is the one for LoHyp. Inspecting the treatment coefficients, we note that the coefficient for HiHyp is positive and big (0.302) and highly statistically significant whereas the coefficient of LoReal is also positive (0.053) but rather small and only borderline significant.

Let us now turn to providing some answers along the second dimension of our research strategy by analyzing the impact of socioeconomic background variables on subjects' behavioral responses. Refer first to Table 1 that under the heading "Violation" shows descriptive statistics of the subsamples violating and not violating EUT as well as p-levels of χ^2 tests. (For the latter, see the notes below Table 1.) Regarding gender, Table 1 reveals that women are slightly more likely to violate EUT than man. With respect to age, Table 1 does not suggest a clear effect although we note that the age bracket's [35-44] relative share is higher in the panel's subpopulation not violating EUT. Regarding education levels, those with lower secondary education and those subjects with a university degree stand somewhat out in the panel. The former because they violate EUT more often and the latter because they violate EUT less often. The most noticeable effect regarding occupation is that those employed on a contractual basis have a higher relative share in the sub sample not violating EUT. Finally, with respect to household income, Table 1 does not suggest a clear effect.

	(1)		(2)		(3)		(4)	
	All data		Only HiHyp		Only LoHyp		Only LoReal	
HiHyp	0.307***	(9.22)						
LoReal	0.058*	(1.87)						
Female	0.007	(0.25)	−0.074	(1.25)	0.056	(1.33)	−0.002	(0.05)
Age 25-34	−0.062	(0.77)	0.005	(0.03)	0.017	(0.15)	−0.200*	(1.77)
Age 35-44	−0.115	(1.45)	−0.186	(1.07)	−0.027	(0.24)	−0.175	(1.46)
Age 45-54	−0.050	(0.61)	−0.082	(0.46)	−0.011	(0.09)	−0.125	(0.96)
Age 55-64	−0.072	(0.87)	−0.162	(0.89)	0.031	(0.25)	−0.152	(1.21)
Age 65+	−0.145	(1.64)	−0.157	(0.75)	−0.134	(1.21)	−0.125	(0.86)
Lower second. edu.	−0.035	(0.63)	0.177	(1.35)	−0.147**	(2.23)	0.045	(0.51)
Higher second. edu.	−0.086	(1.50)	0.150	(1.07)	−0.152**	(2.40)	−0.027	(0.30)
Intermed. voc. training	−0.025	(0.43)	0.172	(1.26)	−0.134**	(1.97)	0.116	(1.19)
Higher vocat. training	−0.034	(0.59)	0.288**	(2.18)	−0.139**	(1.98)	−0.010	(0.11)
University degree	−0.134**	(2.21)	0.186	(1.28)	−0.188***	(2.85)	−0.152*	(1.69)
Employed (contract)	−0.130**	(2.53)	−0.234**	(2.04)	−0.127*	(1.82)	−0.118	(1.50)
Freelance or self-empl.	−0.149**	(2.11)	−0.435***	(3.08)	−0.081	(0.74)	−0.025	(0.21)
Unemployed	0.065	(0.65)	−0.023	(0.09)	0.043	(0.34)	0.046	(0.30)
Student	−0.101	(1.16)	−0.156	(0.79)	−0.033	(0.26)	−0.169	(1.43)
Works in own household	−0.061	(1.10)	−0.224	(1.78)	−0.015	(0.19)	−0.068	(0.82)
Retired	0.013	(0.19)	−0.063	(0.43)	0.145	(1.45)	−0.145	(1.56)
HH gr. inc. €2251-€3130	−0.077**	(2.26)	−0.042	(0.54)	−0.047	(0.95)	−0.129***	(2.64)
HH gr. inc. €3131-€4350	−0.110***	(3.23)	−0.167**	(2.10)	−0.033	(0.67)	−0.138***	(2.84)
HH gr. inc. ≥ €4351	−0.083**	(2.28)	−0.073	(0.88)	−0.076	(1.49)	−0.089*	(1.69)
Assets	−0.081**	(2.29)	−0.118	(1.59)	0.001	(0.02)	−0.112**	(2.02)
Savings account	0.054**	(2.05)	0.093	(1.64)	0.043	(1.13)	0.021	(0.53)
No. of observations	1424		400		500		524	

Note: Absolute value of z statistics in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Omitted categories are treatment LoHyp; age interval [16-24]; primary education; "other" occupation; household gross income smaller or equal to €2250. Two occupation observations missing.

Table 3: Results of probit regressions on violation of EUT. Marginal effects are reported.

To analyze the effect of socioeconomic background variables more formally, refer again to regression (1) in Table 3. We make the following observations.

- Gender and age have no significant influence on the extent of EUT violation. Note also that age does not appear to have a systematic effect in the sense that both subjects aged 35-44 and subjects aged 65 and above have a particularly low probability of violating EUT.[16]

- Regarding education, we find a strong tendency for violations to be reduced with further education. Overall, there is a strong effect of higher education that also shows in the separate specifications for both treatments with low payoffs. In LoHyp everything that improves on primary education reduces violations. Only in HiHyp there is no effect of education. This suggests an interesting interaction effect of experience with a decision domain and education. In the absence of any experience (as in HiHyp) education on its own does little to improve performance.

- Of the various occupational affiliations listed in Table 3, we find that the unemployed and 'others' do much worse than the employed, self-employed and free-lancers.[17] This is more pronounced in treatments with hypothetical payoffs.

- Regarding income, we notice that having a higher gross monthly household income (vis-à-vis the control group with the lowest gross monthly household income) significantly reduces EUT violation.[18] Interestingly, this is particularly pronounced in the treatment LoReal when actual money is at stake. (One could have conjectured that it would be the other way round as the marginal utility of making some money and, hence, the incentive to think a little harder might be higher for those on low incomes. Alas, it does not work this way.)

- Finally, holding assets significantly reduces EUT violation (by about 8%) whereas having a savings account significantly increases EUT violation (by about 5%). Maybe not surprisingly, those subjects holding assets tend to be expected value maximizers (mainly choosing A^*B^*) while those subjects who have a savings account display "Allais" behavior by tending to choose AB^*.[19]

[16] In the light of recent findings about sharply declining numeracy skills in the (British) population above 55 (Banks 2006) this is perhaps slightly surprising.

[17] A Wald test indicates that the effect of the these two occupations is not statistically different.

[18] Wald tests indicate that the effects of the three income variables are not statistically different.

[19] To look at the effect of holding assets or a savings account more closely, we defined the variable "only assets"

Moreover, refer to the rightmost column labeled "p-value, χ^2" in Table 1 that shows p-levels of χ^2 tests for differences between proportions of violating and non-violating subjects in the category listed in column 1. In accordance with the results we just derived from regression (1) in Table 3, the χ^2 tests also reveal the strongest differences in violation behavior in the categories education, occupation and household income.

In all a picture emerges that is reminiscent of recent studies by Benjamin *et al.* (2006), Burks *et al.* (2007), and Dohmen *et al.* (2007) who show that a range of behavioral biases might stem from cognitive limitations and low IQ. We find that violations are more prevalent in those who are lowly educated, unemplyed, on low income, and who have no significant asset holdings. This is, of course, particularly worrying as imprudent financial decision making and bad planning for retirement has the worst consequences in that group.

4 The lab experiment

As mentioned in the introduction, the third dimension of our research strategy is concerned with the external validity of laboratory experiments that are typically carried out with rather homogenous subject pools. Of course, the preceding section has shown that there are important sources of heterogeneity in the population at large that simply cannot be detected when the subject pool is restricted to students. The same is, of course, true for any highly selected convenience sample. But what about the questions we analyzed first—the effects of different treatments, the differences between high and low and real and hypothetical payoffs? Would a lab experiment give us reliable results to analyze such questions (as it has been implicitly assumed for a long time in the experimental community, perhaps negligently without much testing)? To shed more light on these issues

which equals 1 if a subject holds assets but has no savings account (otherwise it equals 0), the variable "only savings account" which equals 1 if a subject has a savings account but holds no assets (otherwise it equals 0), and the variable "assets & savings account" which equals 1 if a subject holds assets and has a savings account (otherwise it equals 0). Hence, the reference group consists of those subjects who neither hold assets nor have a savings account. Replacing the variables "assets" and "savings account" in regression (1) in Table 3 by the new variables "only assets," "only savings account," and "assets & savings account," leaves the other variables of regression (1) almost unchanged (including significance levels) and shows that while the coefficients of the variables "only assets" and "assets & savings account" are negative (-0.073 and -0.032) but insignificant, the coefficient of the variable "only savings account" is positive (0.055) and significant at the 5% level. So it is not only the financially savvy who hold assets who do comparatively well but also people without any savings—perhaps because, having no financial cushion, they cannot afford making many mistakes.

Treatment	Absolute Frequency of Choices					Relative Frequency of Choices				
	AB	A^*B^*	AB^*	A^*B	Σ	AB	A^*B^*	AB^*	A^*B	Violations
HiHyp	4	41	20	5	70	5.7	58.6	28.6	7.1	35.7
LoHyp	0	75	4	0	79	0	94.9	5.1	0	5.1
LoReal	1	67	5	1	74	1.4	90.5	6.8	1.4	8.2
Σ	5	183	29	6	223	2. 2	82.1	13.0	2. 7	15.7

Table 4: Summary of experimental results in the lab

we conducted an additional lab experiment in the laboratory of Tilburg University using Dutch speaking student subjects drawn from the normal subject pool.

The lab experiment was conducted in the same way as the experiment using the CentER-panel. That is, student subjects did the experiment using a web browser in the lab and using the same screens as the subjects in the panel. However, there were two small exceptions. First, lab subjects received a 10 Euro show-up fee. (Potential participants were informed about this in the invitation E-mail.) But of course, mirroring the panel design again, only subjects assigned to treatments with real payment had the chance to earn additional money during the experiment. This was not announced prior to the experiment. Second, lab subjects were not offered the choice of not participating in the experiment once they had reported to the lab and the experiment was started. This was done in an effort to mimic the normal procedures in lab experiments where by reporting to the lab, a subject usually confirms his or her decision to participate.

After the experiment we asked subjects to fill in a questionnaire in which we elicited some basic background information. Naturally, the information we collected from lab subjects is very limited and can not be compared in scope and quality to the background information available from members of CentERpanel. The lab experiments were conducted in December 2006 using 223 subjects in total.

As in the panel experiment we did not observe any order effects of presenting the Allais questions such that we present only pooled data in Table 4 which shows the same information for the lab data that Table 2 showed for the panel. We make the following observations. First, as in the panel experiments, we observe EUT violations in all treatments, albeit to a much lesser degree.[20] This mirrors the main result in Gächter, Huck, and Weizsacker's meta-study: Violations from

[20] Again, we observe that the fraction of EUT-violating AB^* answers is significantly higher than the fraction of EUT-violating A^*B answers in all lab treatments ($p < 0.001$, Conlisk's (1989) Z-statistic).

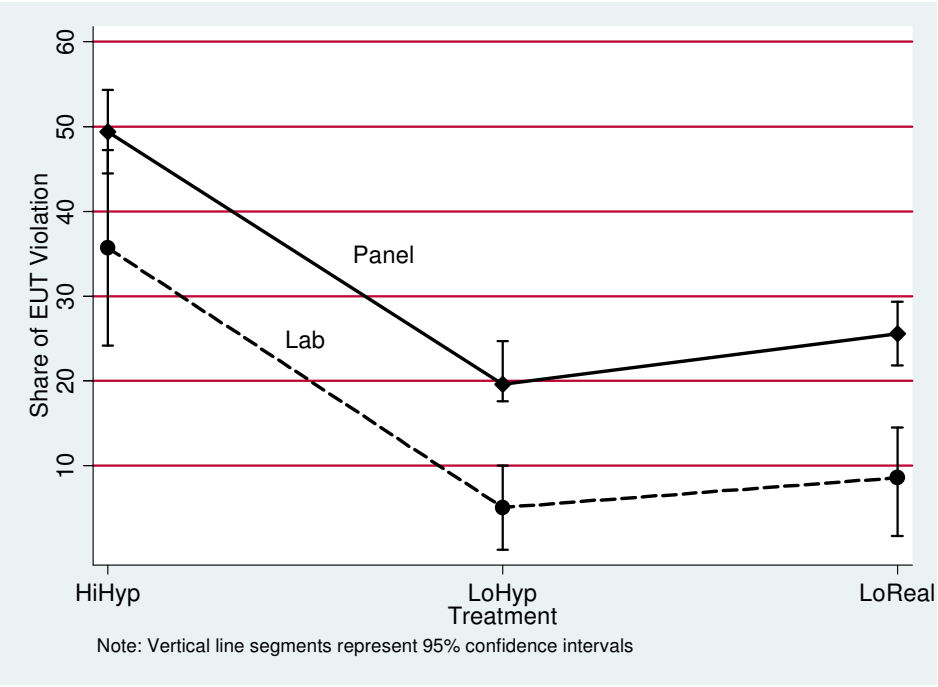

Figure 1: The share of choices violating EUT in the panel and the lab

orthodox theoretical predictions and biases observed in the lab form a lower bound for violations and biases observed in the population at large. Second, as in the panel, moving from high hypothetical payoffs to low hypothetical payoffs reduces the extent of EUT violation significantly ($p < 0.001$, $D = 4.881$). Third, moving from low hypothetical payoffs to low real payoffs increases the extent of EUT violation slightly but insignificantly ($p < 0.226$, $D = -0.7525$). The similarities between the observations in the panel and in the lab are evident.

Figure 1 shows the shares of choices violating EUT in the two subsamples. It appears that the graph indicating the share of EUT violation in the panel can quite accurately be obtained by shifting the graph indicating the share of EUT violation in the lab upwards by about 15 percentage points. This means that although the share of EUT violations is consistently higher in the panel than in the lab, the "comparative statics results" of moving from one treatment to another could have been reliably predicted by the lab experiments. We consider this an important result of this study.

5 Conclusions

Using a representative sample of the Dutch population we revisit the Allais paradox. Our main results are threefold. First, violations are systematic but much lower when stakes are low. Second, there is considerable heterogeneity in the population and violations are particularly prevalent among the lowly educated, those poor in income and asset holdings, and the unemployed. Third, comparing the panel results with a laboratory experiment we find that the relative treatment differences are identical in panel and lab although violations in the lab are much lower than in the panel. The lab results seem to provide a lower bound for the extent of EUT violation in a general population.

Our findings appear to imply two general messages. First, laboratory experiments with convenience samples of students might, in general, be more useful to study *relative* effects rather than *absolute* levels (see also Levitt and List 2007 who make a similar point in the context of social preferences). When it comes to the absolute measurement of behavior, it appears that lab results will draw a too optimistic picture. The population at large, it turns out, is less rational than student samples are.

Second, our results suggest that the predictive power of EUT in a general population is limited. In particular, parts of the population that are potentially more likely to experience economic hardship might react less stringently to economic mechanisms as orthodox theory would suggest. When making predictions about responses to economic incentives one might want to differentiate carefully between different strata of a population.

References

[1] Allais, M. (1953): Le comportement de l'homme rationnel devant le risqué: critique des postulats et axioms de l'ecole americaine, *Econometrica* 21, 503–46.

[2] Allais, M. and Hagen, O., eds., (1979): *Expected Utility Hypotheses and the Allais Paradox*, Dordrecht, Holland: Reidel.

[3] Andersen, S., G.W. Harrison, M.I. Lau, and E.E. Rutstrom (2007): Risk aversion in game shows, in: J.C. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments,* Greenwich, CT: JAI Press, Research in Experimental Economics, Volume 12, forthcoming.

[4] Banks, J. (2006): Economic choices, capabilities and outcomes at older ages, *Fiscal Studies* 27, 281-311.

[5] Battalio R.C., J.H. Kagel, and K. Jiranyakul (1990): Testing between alternative models of choice under uncertainty: Some initial results, *Journal of Risk and Uncertainty* 3, 25-50.

[6] Bellemare, C. and S. Kröger (2007): On representative social capital, *European Economic Review* 51, 183–202.

[7] Benjamin, D.J., S.A. Brown, and J.M. Shapiro (2006): Who is "behavioral"? Cognitive ability and anomolous preferences, mimeo.

[8] Burke, M.S., J.C. Carter, R.D. Gominiak, and D:F: Ohl (1996): An experimental note on the Allais paradox and monetary incentives, *Empirical Economics* 21, 617–632.

[9] Burks. S.V., J.P. Carpenter, L. Götte, K.A. Monaco, A. Rustichini, and K. Porter (2007): Using behavioral economic field experiments at a large motor carrier: The context and design of the truckers and turnover project, NBER Working Paper No. W12976.

[10] Camerer, C. (1995): Individual decision making, in *The Handbook of Experimental Economics*, edited by J. H. Kagel and A. E. Roth, Princeton, N.J.: Princeton Univerity Press, Chapter 8, 587-703.

[11] Carbone, E. (2005): Demographics and behaviour, *Experimental Economics* 8, 217–232.

[12] Conlisk, J. (1989): Three variants on the Allais example, *American Economic Review* 79, 392–407.

[13] Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G.G. Wagner (2005): Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey, IZA Discussion Paper 1730.

[14] Dohmen, T., A. Falk, D. Huffman, U. Sunde (2007): Are Risk Aversion and Impatience Related to Cognitive Ability?, IZA Discussion Paper 2735.

[15] Eckel, C. and P. Grossman (2000): Volunteers and pseudo-volunteers: The effect of recruitment method on subjects' behavior in experiments, *Experimental Economics* 3, 107–120.

[16] Fan, C.-P. (2002): Allais paradox in the small, *Journal of Economic Behavior and Organization* 49, 411-421.

[17] Fatas, E., T. Neugebauer, and P. Tamborero (in press): How politicians make decisions: A political choice experiment, *Journal of Economics*, forthcoming.

[18] Gächter, S., S. Huck, and G. Weizsacker (2007): Socio-demographics and choice in experimental economics, mimeo.

[19] Gaudecker, H.-M. von, A. van Soest, and E. Wengström (2007a): Risk preferences in the small for a large population, mimeo.

[20] Gaudecker, H.-M. von, A. van Soest, and E. Wengström (2007b): Experimental elicitation of risk preferences: Some further steps towards representativeness, mimeo.

[21] Harrison, G.W., (1994): Expected utility and the experimentalists, *Empirical Economics* 19, 223–253.

[22] Harrison, G.W. and J.A. List (2004): Field experiments, *Journal of Economic Literarture* 42, 1013-1059.

[23] Harrison, G.W., E. Johnson, M.M. McInnes, and E.E. Rutstrom (2005a): Risk aversion and incentive effects: Comment, *American Economic Review* 95, 897-901.

[24] Harrison, G.W., M.I. Lau, and E.E. Rutström (2005b): Risk attitudes, randomization to treatment, and self-selection into experiments, Working Paper 05-01, Department of Economics, College of Business Administration, University of Central Florida.

[25] Harrison, G.W., J.A. List, and C. Towe (2007): Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion, *Econometrica* 75, 433-458.

[26] Heckman, J. (1976): The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models, *Annals of Economic and Social Measurement* 5, 475-492.

[27] Hey, J.D. (2002): Experimental economics and the theory of decision making under risk and uncertainty, *Working Paper*, University of York.

[28] Kahneman, D. and A. Tversky (1979): Prospect theory: An analysis of decision under risk, *Econometrica* 47, 263-291.

[29] Levitt, S.D. and J.A. List (2007): What do laboratory experiments measuring social preferences reveal about the real world, *Journal of Economic Perspectives* 21, 153-174

[30] List, J.A. and M.S. Haigh (2005): A simple test of expected utility theory using professional traders, *Proceedings of the National Academy of Sciene* 102, 945-948.

[31] MacCrimmon, K.R. (1968): Descriptive and normative implications of the decision-theory postulates, in: K. Borch and J. Mossin (eds.), *Risk and Uncertainty*, London: Macmillan, chapter 1.

[32] MacDonald, D.N., and J.L. Wall (1989): An experimental study of the Allais paradox over losses: Some preliminary evidence, *Quarterly Journal of Business Economics* 28, 43–60.

[33] Machina, M.J. (1982): Expected utility analysis without the independence axiom, *Econometrica* 50, 277–323.

[34] Slovic, P. and A. Tversky (1974): Who accepts Savage's axiom?, *Behavioral Sciences* 19, 368-373.

[35] Starmer, C. (2000): Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk, *Journal of Economic Literature* 38, 332-82.

[36] Starmer, C. and R. Sugden (1991): Does the random-lottery incentive system elicit true preferences? An experimental Investigation, *American Economic Review* 81(4), 971-978.

A Instructions (Translation)

The experiment was administered in Dutch. Here we give a translation of the screens presented in treatment LoHyp and [LoReal]

Screen 1:

This research is conducted by researchers of Tilburg University and University College London. The questionnaire consists of two choice problems in which you will be asked to make a choice between two situations. Based on your choices and luck you may win an amount of money. Please note: In this experiment all amounts are hypothetical, in reality you cannot win any money. [in LoReal: In this experiment you can earn real money that will be paid in the form of CentERpoints.]

If you do not want to participate as a matter of principle, you can indicate this below. You will then go directly to the end of the questionnaire.

◯ I continue with the questionnaire.

◯ No, I do not want to participate in this questionnaire.

Continue

Screen 2:

Instructions

You will shortly be presented with two questions. You will be asked to make a choice between two options which provide you with different chances to win something. Please see an example of such a situation here below. In the first option you have a chance of 80 percent to win nothing and a chance of 20 percent to win 10 Euro. The second option provides you with a chance of 20 percent of nothing and a chance of 80 percent of winning 20 Euro.

OPTION 1:	80 percent chance	nothing	(if number is between 1 and 80) and
	20 percent chance	10 euro	(if number is between 81 and 100)
OPTION 2:	20 percent chance	nothing	(if number is between 1 and 20) en
	80 percent chance	20 euro	(if number is between 21 and 100)

We would like to know whether you prefer Option 1 or Option 2 (in these instructions you don't have to choose yet). After you have made the choice, the computer will play out the option

you chose. The computer generates a random number that is between 1 and 100. The chance distribution of the chosen option then defines how much you win with this number.

For example: in the Option 1 above you get nothing if the computer generates a number between 1 and 80 (this is indicated above in Option 1 in brackets), but if the computer generates a number between 81 and 100 you will get 10 Euro. In Option 2 you get nothing if the computer generates a number between 1 and 20, but with a number between 21 and 100 you win 20 Euro. As already mentioned, it concerns hypothetical amounts here, in reality you cannot win any money. [in LoReal: If you win something then this amount will be added to your account of CentERpoints.]

If you are ready to start the experiment, press "Continue."

Continue

Screen 3:

Which of the following two options do you prefer?

OPTION A:	certainty	of 5 euro	(if number is between 1 and 100) and
OPTION B:	1 percent chance	of nothing	(if number is 1) and
	89 percent chance	of 5 euro	(if number is between 2 and 90) and
	10 percent chance	of 25 euro	(if number is between 91 and 100)

◯ Option A

◯ Option B

Continue

Screen 4:

Which of the following two options do you prefer?

OPTION C:	89 percent chance	of nothing	(if number is between 1 and 89) and
	11 percent chance	of 5 euro	(if number is between 90 and 100)
OPTION D:	90 percent chance	of nothing	(if number is between 1 and 90) and
	10 percent chance	of 25 million euro	(if number is between 91 and 100)

◯ Option C

◯ Option D

Continue

Screen 5:

You have now made the two decisions. Press "Continue" to see the results if the options you chose.

Continue

Screen 6:

In the first question (option A or B) you have chosen Option X ([description of the chosen option]). The computer generated the number [random number]. Thus, you have won [in treatment LoHyp: the hypothetical] amount of [...] euro with this option.

In the second question (option C or D) you have chosen Option Y ([description of the chosen option]). The computer generated the number [random number]. Thus, you have won the [in treatment LoHyp: the hypothetical] amount of [...] euro with this option.

In total you have won the [in treatment LoHyp: the hypothetical] amount of [...] euro in this experiment.

Continue

Screen 7:

Do you have any comments regarding the questionnaire?

◯ Yes

◯ No

Continue

Screen 8 [In case the answer to the question on Screen 7 was Yes.]**:**

You can type in your comments below.

Continue

Screen 9:

This is the end of the questionnaire. Thank you for your participation.