Netspar Network for Studies on Pensions, Aging and Retirement

Hans-Martin von Gaudecker

Arthur van Soest

Erik Wengström

Selection and Mode Effects in Risk Preference Elicitation Experiments

Discussion Paper 10/2008 - 042
October 28, 2008

Selection and Mode Effects in Risk Preference Elicitation Experiments *

Hans-Martin von Gaudecker
VU University Amsterdam and Netspar

Arthur van Soest
Tilburg University and Netspar

Erik Wengström †
University of Copenhagen

October 28, 2008

Abstract

We combine data from a risk preference elicitation experiment conducted on a representative sample via the Internet with laboratory data on student subjects for the same experiment in order to investigate effects of implementation mode and of subject pool selection. We find that the frequency of errors in the lab experiment is drastically below that of the representative sample in the Internet experiment. Average risk aversion is lower in the lab as well. Considering the student-like subsample of the Internet subjects and comparing a traditional lab design with an Internet-like design in the lab gives two ways to decompose these differences into differences due to subject pool selection and differences due to implementation mode. Both lead to the conclusion that the differences are due to selection and not to implementation mode. An analysis of the various steps leading to participation or non-participation in the Internet shows that these processes are selective in selecting subjects who make fewer errors, but do not lead to biased conclusions on risk preferences. These findings point at the usefulness of the Internet survey as an alternative to a student pool in the laboratory if the ambition is to draw inference on a broad population.

JEL Codes: C90, D81
Keywords: Risk aversion, Internet surveys, Laboratory experiments.

---

*Financial support from the Dutch National Science Foundation (NWO), the Swedish Institute for Banking Research (Bankforskningsinstitutet), and from the European Union under grant HPRN-CT-2002-00235 (RTN-AGE) is gratefully acknowledged. We thank the team of CentERdata, especially Marika Puumala, for their support with the experiments, as well as Morten Lau and Joachim Winter for very helpful comments on the experimental design. The analysis benefitted from comments received at presentations in Mannheim, Copenhagen, Gothenburg, the XIIth FUR conference at LUISS in Rome, and the ESA meetings in Nottingham and Tucson. Daniel Kemptner provided able research assistance.

†*Address for correspondence:* Department of Economics; University of Copenhagen; Studiestræde 6; DK-1455 Copenhagen K; Denmark. Email: erik.wengstrom@econ.ku.dk

1 Introduction

Recently, there has been an increased interest in eliciting economically important preference parameters by means of experimental methods (Harrison, Lau, and Williams (2002), Bleichrodt, Pinto, and Wakker (2001), among many others). In this context, researchers are generally interested in parameters that are valid for the general population and carry over to situations outside the laboratory setting. There are several reasons why it may be difficult or impossible to recover such parameters from standard experiments. First, the experimental design could differ too widely from real-world situations in terms of context, stakes, or similar features. The literature investigating this type of effect is reviewed in Harrison and List (2004) and Levitt and List (2007). Second, the subjects taking part in the experiment may not represent the population of interest. There has been a growing concern that the standard recruitment procedure – an experimenter inviting college students via emails or posters – may restrict socio-demographic variation too severely as to allow meaningful inference on the broad population of interest. Spurred by Harrison, Lau, and Williams (2002), this issue has been addressed in several recent field studies. In addition, there may be a selection effect that also applies if recruitment is broader than among students and is largely out of control of the experimenter: Participation in experiments is voluntary so that the participating subjects may differ from the population in relevant dimensions. In this paper, we address both types of selection effects.

Recent years have witnessed different approaches to enhance demographic variation in experimental situations. One rather laborious possibility is to take the laboratory from the university to the population of interest; for one of many examples along these lines see Harrison, List, and Towe (2007). Sample sizes usually do not exceed those typically encountered in the laboratory which may pose a problem in accounting for demographic heterogeneity. A similar strategy that has become available recently is to integrate experiments into existing household surveys; see for example the pioneering work by Fehr et al. (2003) and Dohmen et al. (2005) . Major advantages of this approach are that careful sampling frames are employed and that a lot of background information on participants is available. Until now, capacity constraints in the survey instruments and the relatively high costs of personal interviews have hindered a more widespread use of this method. The two cited studies were able to use moderately-sized subsamples (N=429 and N=450, respectively) from the much larger German Socio-Economic Panel. Third, experimenters have used convenience samples of Internet respondents recruited by means of newspaper advertising or email invitations. See, e.g., Lucking-Reiley (1999) and Güth, Schmidt, and Sutter (2007). While this approach facilitates conducting experiments with very large numbers of participants, there is typically no control over the recruitment process. Selection effects may arise since respondents have to read a particular newspaper, need to have access to the Internet, have to subscribe to a specific electronic mailing list, etc.

We combine the advantages of the last two approaches in conducting an experiment with a large sample (2,299 subjects) of respondents from a Dutch household survey, the CentERpanel. This is carried out over the Internet and avoids non-coverage of those without Internet access by providing them with a set-top box for their TV. In order to investigate the traditional experimenter-induced subject pool bias, we compare the Internet outcomes to those of parallel laboratory experiments with 178 students. However, replacing the laboratory by the Internet also changes the environment, unlike the case of comparisons based on a "mobile laboratory" approach (Andersen et al. 2007). Potential differences due to demographic variation can therefore be confounded with implementation mode effects. We address this issue from two angles. First, we introduce a treatment in the laboratory that replicates the Internet setting as closely as possible – no experimenter is present while subjects complete the experiment and subjects do not have to wait for the last person to finish before leaving the room. Second, our Internet sample is sufficiently large to analyse a subsample that resembles the student population in terms of age and education. If environmental factors play a role, they should lead to differences between the results for this Internet subsample and those for the laboratory experiment.

Section 2 describes our data and the experimental setup. The issue of experimenter-induced selection effects is taken up in section 3. We find that when moving from student samples to the general population, the most dramatic difference is a drastic rise in the number of violations of the most basic economic principles, namely choosing dominated options and non-monotonic behaviour. Risk aversion also turns out to be higher in the overall population. We cannot detect any differences arising from the environmental treatment for the young and educated, which leads us to conclude that the differences are driven by subject pool effects.

Selection effects stemming from voluntary participation have not received much attention until recently, see Harrison, Lau, and Rutström (2007) for a primer. The main reason for this is probably that there is usually little control over the recruitment process. Experimenters typically collect some demographic information about participating subjects, but the corresponding values of nonparticipants are not observed. A crucial feature of our setup is that we have access to rich background information for participants as well as nonparticipants. This allows us to estimate a model of selection into the experiment (Section 4). We find higher participation rates if incentives are provided. Participation is also larger for the more educated, the non-elderly, males, and for household members with large interest and expertise in financial matters. This induces some association of participation with inconsistent behaviour – participants typically have observed characteristics that make them less prone to making mistakes. On the other hand, we find that selective participation hardly affects the estimates of average risk preferences.

2 Data and Experimental Setup

This section provides a detailed description of our experimental design and subject pools. The starting point of the experiments is the multiple price list format, a well-established methodology for preference elicitation, which we modify in two ways. First, to help respondents understand their tasks, we include pie-chart representations of the probabilities in addition to the representations in numbers. Second, the experiment asks subjects to choose between risky choices involving gains and sometimes also losses, as in Holt and Laury (2002), but also in some cases involves a choice between lotteries with immediate resolution of uncertainty and late resolution of uncertainty, in the sense of (Kreps and Porteus 1978).

We first describe the multiple price list format and how we implement it. We then point out the aspects of the experiment that are specific to the Internet and laboratory settings, respectively. In particular, we highlight the features of our design aimed at disentangling subject pool ("selection") effects and implementation method ("mode") effects. The most important of these is the introduction of two environmental treatments in the laboratory. One replicates traditional experiments, the other mimics the Internet setting as much as possible. We term them "Lab-Lab" and "Lab-Internet" to avoid confusion with the CentERpanel experiment (also denoted as "Internet experiment"). The full experimental instructions, samples of choice problems, help screens, final questions, and the debriefing part are available at http://www.econ.ku.dk/wengstrom/.

2.1 The Multiple Price List Format

Our experiments make use of an adapted version of the multiple price list format, introduced in economics by Binswanger (1980) and recently employed in the context of risk preferences by Holt and Laury (2002). An extensive description can be found in Andersen et al. (2006). In principle, multiple price lists work as follows: Each subject is presented a series of lotteries with identical payoffs but varying probabilities such as the one presented in Figure 1. In each of the four cases, the participant can choose between Option 'A' and Option 'B'. The table is designed such that the expected payoff of lottery 'A' starts out higher but moves up slower than the expected payoff of lottery 'B'. A participant with a monotonically increasing utility function switches at some point from the safer Option 'A' to the riskier Option 'B', or chooses 'B' throughout. This is because the last row is always a choice between two certain payoffs, with that of Option 'B' higher than that of Option 'A'.

A modification compared to previous studies is that we include pie-charts describing the probabilities of the outcomes, in addition to the verbal descriptions of the decision tasks. Using a design with low cognitive demands seems important when moving outside the traditional laboratory environment and using low educated subjects. Pilot experiments showed that this is appreciated by subjects who are not familiar with probability judgements. Because of the extra screen space needed for the graphical probability representations, we reduce the number of decision tasks per screen from the usual ten to four, avoiding that respondents need to scroll. To obtain precise responses despite this, subjects who are consistent in the sense that they do not switch back and forth between 'A' and 'B' and choose the higher certain payoff in the final question are routed to a second screen, containing lotteries with the same payoffs but a refined probability grid. The probability grid of the second screen involves 10%-steps located approximately between the respondent's highest choice of 'A' and lowest choice of 'B' on the previous screen. This is a version of the iterative multiple price list format described in Andersen et al. (2006).

Each subject faced the seven payoff configurations described in Table 1. For each configuration subjects make either four or eight decisions, depending on their answers on the first screen. Some of the riskier option 'B' lotteries involved negative payoffs, while payoffs from the safer option 'A' were all strictly positive. The actual payments were always made three months after the experiment and subjects were informed about this in the introduction. At the top of each screen, we indicated whether the outcome of the lottery was revealed immediately or in three months' time (see Figure 1).

Subjects were randomly assigned to one of three groups with different payoff treatments: one group with hypothetical and one group with real lotteries, both with the amounts shown in Table 1, and a third group with real payoffs but amounts divided by three. We refer to these as hypothetical, high, and low incentive treatments. Subjects in the high and low incentive groups received an upfront payment of 15 or 5 Euros, respectively. No payment at all was made to the hypothetical group of the CentERpanel experiment. The laboratory subjects in the hypothetical group received a participation fee of 5 Euros for recruitment reasons. In the incentives treatments, everyone received the participation fee, but only one in ten subjects got paid in addition for one of their chosen lotteries. The lottery to be paid out was selected at random to ensure incentive compatibility. In order to avoid negative payoffs, the highest possible loss did not exceed the fixed participation fee. We randomised the order in which the seven payoff configurations were presented. In an effort to remain close to earlier work, the first payoff configuration in the low incentive treatment is a scaled version of the payoff configuration in Table 1 of Holt and Laury (2002), multiplied by six and rounded to the next lowest integer. The other payoff configurations are derived in a similar way from those used by Holt and Laury.

2.2 The CentERpanel Experiment

The subjects in the Internet experiment are respondents in the CentERpanel,[1] aged 16 and older. The CentERpanel is managed by CentERdata, a survey research institute affiliated with Tilburg University. The panel contains more than 2,000 households and covers the complete Dutch population, excluding the institutionalised. Questionnaires and experiments are fielded over the Internet. To avoid selection bias, households without Internet access are provided with a set-top box for their TV (and with a TV if they do not have that either). Panel members get questions every weekend. They are reimbursed for their costs of participation (fees of dial-up Internet connections etc.) on a regular basis. We conducted our experiments in November and December of 2005. Our payments were included in one of the regular transactions three months after the experiments.

The welcome screen contained a brief introduction to the experiment followed by a non-participation option. See Figure 2 for the introductory screens of all treatments. For the treatments with real incentives, subjects were told the amount of the participation fee and that they had the chance to win substantially more or lose (part of) this money again. It was made clear that no payment would be made upon nonparticipation. In the hypothetical treatment, subjects were informed that the questionnaire consisted of choices under uncertainty in a hypothetical setting. In all treatments, subjects then had to indicate whether they wanted to participate or not. Respondents who opted for participation first went through two pages of online instructions before facing the seven price list configurations. The instructions and specially designed help screens could be accessed throughout the experiment. They were included to improve comparability with similar laboratory experiments, compensating for the absence of an experimenter.

In total, 2,299 persons logged into the system. About 12.7% opted for nonparticipation, leaving 2,008 respondents who started the experiment. 80 subjects dropped out before completing the questionnaire. Moreover, 138 respondents went through the experiment extremely rapidly. Those who took less than 5:20 minutes are treated as dropouts in the analysis below (see Section 4 for details about the choice of cut-off point). Our final sample thus consists of 1,790 subjects who made 91,808 choices.

[1]For related papers using data collected through the CentERpanel see, e.g., Donkers, Melenberg, and van Soest (2001) who analysed risk preferences using hypothetical questions, and Bellemare and Kröger (2007) for evidence from a trust game with real payoffs. More information about the CentERpanel can be found at http://www.uvt.nl/centerdata/.

The first three columns of Table 2 list descriptive statistics for the participants who completed the experiment ("final sample"), those who opted for nonparticipation, and those who dropped out in the course of the experiment or sped through it. As expected, the three groups differ in many respects. The variables in Table 2 can be broadly classified into six groups: Incentive treatment; education; sex and age; employment status and residential area; financial literacy and experience; income. Some of the questions, particularly those on assets and financial literacy and experience, are drawn from the DNB household survey (DHS), a survey focusing on financial issues held among CentERpanel respondents once a year. Not everybody in our sample took part in the DHS, implying that sample sizes fall if we include the corresponding variables in the analysis.

2.3 The Laboratory Experiment

In order to compare the answers in the Internet survey to those in the environment of a controlled laboratory experiment, we performed the same experiment in the economics laboratory at Tilburg University. In total, 178 students participated (8 sessions in September 2005 and 8 sessions in May 2006). The same treatments were carried out as in the Internet survey. The only difference was the above-mentioned payment of a show-up fee in the hypothetical treatment. The payment procedure for the incentives treatments was as in the CentERpanel experiment: The participation fee was transferred to participants' bank accounts three months after the experiment; one in ten subjects received the sum of the participation fee and the (possibly negative) payment from one randomly drawn lottery.

To distinguish effects due to different subject pools from effects due to replacing the controlled laboratory setting by the Internet environment, we also replicated this latter change in the lab. The first environmental treatment, labelled the "Lab-Lab" treatment, replicates the traditional setup used in laboratory experiments. In particular, an experimenter was present in the room to help the subjects and answer questions. In contrast to the CentERpanel experiment, no links to the instructions or help screens were shown in the core part of the experiment. Otherwise, the screens resembled the one in Figure 1. Participants also had to wait until everyone else in the session had finished before they could leave the room. In the second environmental treatment – termed the "Lab-Internet" treatment – the experimenter was not present. Instead subjects had access to the same help screens (including the introductory screens) as in the CentERpanel experiment. Moreover, subjects could leave directly after completing the experiment – they did not have to wait for everyone else.

The last column of Table 2 contains the available demographic characteristics of the laboratory subjects. Much less information is available than for the CentERpanel experiment, and there is less variation in the basic demographic characteristics. Specifically, in terms of age and education, the laboratory population represents only a small fraction of the population representative sample in the first three columns.

3 Traditional Subject Pool Bias in the Laboratory

This section addresses "subject pool bias": the concern that the results of standard laboratory experiments are not representative for a broader, heterogeneous, population, since samples of students do not cover the population at large. Our design enables analysing subject pool bias in laboratory experiments controlling for implementation mode effects. With regard to elicitation and modelling of preferences, two issues are of special interest. The first is the structure and frequency of errors and violations of fundamental principles of choice. The second concerns the extent to which the distribution of preferences depends on the composition of the subject pool.

3.1 Errors and Inconsistencies

Several recent studies have highlighted the importance of accounting for errors in the decision-making process when modelling risky choice (cf. Hey (1995), Ballinger and Wilcox (1997), Loomes, Moffatt, and Sugden (2002), or Schmidt and Neugebauer (2007) for a few examples). We draw on this literature and recur to the revealed preference conditions of a choice model that puts very little structure on subjects' behaviour. The rationale behind this is to rule out confounding of errors and preference heterogeneity. This comes at the cost of "missing" some patterns that would be classified as errors under parametric structures. We use the following notation. A decision task j is characterised by the outcomes $\{A_j^{\mathrm{LOW}}, A_j^{\mathrm{HIGH}}, B_j^{\mathrm{LOW}}, B_j^{\mathrm{HIGH}}\}$ and the probability of the low outcome p_j^{LOW}. We assume that individual i's choice in task j, $Y_{i,j} \in \{A_j, B_j\}$ is:

$$
\begin{aligned}
(1) \qquad Y_{ij} = A_j \quad \text{if} \quad & w_i(p_j^{\mathrm{LOW}}) \cdot U_i(A_j^{\mathrm{LOW}}) + w_i(1-p_j^{\mathrm{LOW}}) \cdot U_i(A_j^{\mathrm{HIGH}}) \\
& > \\
& w_i(p_j^{\mathrm{LOW}}) \cdot U_i(B_j^{\mathrm{LOW}}) + w_i(1-p_j^{\mathrm{LOW}}) \cdot U_i(B_j^{\mathrm{HIGH}})
\end{aligned}
$$

and $Y_{ij} = B_j$ otherwise. The only restrictions we impose beyond the additive separability inherent in (1) are that the utility function U and the probability weighting function w are monotonically increasing for all subjects and that the latter is adding up to one: $U^{'} > 0$, $w^{'} > 0$, $w(p) + w(1-p) = 1$. This specification encompasses the major variants of generalised expected utility theory and prospect theory (with a fixed reference point) that we are aware of.[2] The individual subscripts on the utility function and probability weighting function allow for maximal heterogeneity across subjects.

We distinguish three choice patterns that are inconsistent with this model. First, a dominance violation occurs if somebody chooses option 'B' when the probability for the high outcome is zero or option 'A' when this probability is one.[3] The second type of inconsistency emerges when subjects switch back and forth from choosing 'A' and 'B' on the same screen. The third is when they switch back and forth between the initial screen and the follow-up screen (within a given payoff configuration). There was some overlap of probabilities on the two screens, so subjects could make a choice on the second that was inconsistent with their choice on the first screen.[4]

We consider the average number of violations of any of the three types as a summary statistic for error frequencies. Only one violation per subject is counted for each payoff configuration, limiting the maximum number to seven. In Figure 3, the average number of configurations with inconsistencies is presented by sample. The error frequency is much higher in the Internet experiment than in the Lab experiment: 2.43 (first bar) versus 1.21 (second bar). The null hypothesis that underlying population frequencies are the same is rejected using Mann-Whitney (MW) or Kolmogorov-Smirnov (KS) nonparametric tests (two-sided p-values< 0.01).

On the other hand, it is evident from Figure 3 that average error frequencies are very similar in the two Laboratory treatments: 1.29 in the "Lab-Internet" treatment (fourth bar) and 1.14 in the "Lab-Lab" treatment (fifth bar). We cannot reject the null hypothesis of identical underlying distributions using the MW or KS test (two-sided p-values> 0.3). This suggests that the disparity between the lab and the Internet is not due to the different environments under which the experiments were conducted. The higher frequency of violations observed in the Internet treatment is not driven by the presence of an experimenter or by the ability to leave immediately after finishing the experiment.

This already suggests that differences between Internet and Lab inconsistencies are driven by subject pool bias rather than implementation mode effects. In addition, we can also compare the inconsistency frequencies in the laboratory sample with a sub-sample of the Internet

[2]The Kreps and Porteus (1978) model mentioned above is not directly incorporated in this specification since it involves additional nonlinear transformations of (1). However, the three types of choice patterns we identify as inconsistent with (1) are also violating the axioms of Kreps and Porteus (1978).

[3]The former can only happen on a follow-up screen – if the respondent has already chosen option 'B' at probability 0.25

[4]For example, if on the first screen they switched from 'A' to 'B' when the high outcome probability went from 0.25 to 0.5, the second screen had high outcome probabilities $0.2, 0.3, 0.4$ and 0.5. An inconsistency then arises if they chose 'B' at probability 0.2 or if they chose 'A' at probability 0.5.

sample that has similar characteristics as the laboratory sample, thus largely controlling for subject pool selection. This sub-sample (labelled "Internet-Uni") has 96 observations on respondents between 18 and 34 years of age who hold a university degree or study at a university. Behaviour of the Internet participants in this sub-sample resembles that of the student sample: The average number of violations in the "Internet-Uni" sub-sample is 1.28 (third bar in Figure 3), close to the laboratory means of 1.29 and 1.14 – the differences are insignificant according to KS and MW tests. This confirms that the higher frequency of errors in the Internet treatment is exclusively driven by the different composition of the subject pool.[5]

Table 3 displays the frequencies of the different types of errors as a percentage of the number of possible violations. The pattern found in the laboratory is similar to results reported by Loomes, Moffatt, and Sugden (2002) on a different risky choice design: subjects make very few violations of dominance but make many more inconsistent choices when faced twice with the same decision problem. Inconsistencies between screens constitute more than 70% of all consistency violations in both the Lab-Lab and the Lab-Internet treatment. Our results indicate that this changes dramatically when the general population is considered, where dominance violations play a much larger role (more than 38% of all consistency violations), although the numbers of within and between screens inconsistencies are also larger than in the lab. As above, the figures suggest that the difference between the laboratory and the internet is mainly driven by subject pool effects: the error frequencies of the young and well educated in the "Internet-Uni" group resemble those of the laboratory samples. The only discrepancy concerns dominance violations, which appear to be slightly more common in the "Internet-Uni" sample than in the laboratory samples. However, using individual-level frequencies of dominance violations, we cannot reject the null hypothesis of identical underlying distributions (MW two-sided p-values>0.05; KS two-sided p-values>0.6). Taken together, the frequencies in Table 3 confirm that findings for student samples in the lab cannot always be generalised to the general population.

Finally, we checked whether providing monetary incentives makes subjects take more care in answering the questions so that they would make fewer errors. We find no evidence of this – differences between the treatments with different incentives are not significant according to the MW and KS tests.

[5] A concern with the Internet experiment that we cannot fully rule out is joint decision-making of household members. This said, more than 90% of those who completed the experiment reported that nobody else was present while going through the screens, and it seems unlikely that all of the remaining subjects came to the decisions jointly. Hence we think this issue is negligible.

3.2 Preferences

In order to compare preferences, we first consider the fraction of subjects choosing the safe option for a given probability of the high outcome. To get comparable data across probabilities we restrict attention to choices on the first screen. Subjects in the low incentive treatment are excluded, since they faced a different payoff scale and cannot be directly compared with subjects in the other treatments. We aggregate the data over the seven decision problems (looking at subgroups does not lead to additional insights). Comparing the answers in the laboratory and the Internet experiments, it is evident that there are considerable differences – see Figure 4. Except for the case of a 0.25 probability of the high outcome, the fractions of risky choices are higher in the laboratory than in the Internet experiment. The figure also suggests that the decisions in the lab are more sensitive to the probability of the high outcome than in the Internet experiment. There are two potential explanations for the smaller slope in the CentERpanel experiment: More heterogeneity in preferences or higher error rates. When deleting payoff configurations with dominance violations the pattern persists, suggesting that the larger number of dominance violations is not the only explanation. Still, disentangling the two explanations calls for a structural model of choice at the level of the individual, which is beyond the scope of this paper.[6] Andersen et al. (2007) also find a larger degree of preference heterogeneity in the field than in the lab.

We can check whether the differences are due to subject pool or implementation mode effects. First, Figure 5 shows that there are hardly any differences between the choice frequencies in the "Lab-Lab" and "Lab-Internet" treatments. Second, Figure 6 shows that the pattern in the "Internet-Uni" sub-sample of the Internet participants is similar to the pattern in the laboratory experiment. Both figures indicate that the main driving force for the difference between lab and Internet is the subject pool composition rather than the environment in which the experiments were conducted.

To obtain a simple measure of individual preferences we consider at which probabilities subjects switched from (the safer) option 'A' to (the riskier) option 'B' in each payoff configuration. Similar measures have been used in earlier studies, cf., e.g., Holt and Laury (2002). We can only compute bounds that will at best be a 5%-interval (e.g. between 75% as the highest 'A'-choice on the first screen and 80% as the lowest choice of 'B' on the second screen). In many cases, the bounds are substantially wider because of the inconsistencies discussed in section 3.1. We computed the bounds as follows: the lowest possible switch point is defined as the highest probability corresponding to an 'A' choice that is still lower than the minimum probability with a 'B' choice; the upper bound is the minimum probability with a 'B' choice

[6] Such a strategy is pursued in von Gaudecker, van Soest, and Wengström (2008). It confirms the finding of much higher heterogeneity of preferences in the CentERpanel experiment.

that is still higher than the maximum probability where option 'A' was chosen. If only choice 'A' ('B') was observed, both upper and lower bound were set to 100% (0%).

For each individual, we averaged the upper bounds and the lower bounds across the seven payoff configurations. This leaves us with two preference measures per individual – the higher the measure, the more averse the subject is to more risky choices. Coming back to model 1, we cannot distinguish between risk aversion arising from the shape of the utility function and from probability weighting. Doing so requires a structural model and a design that puts a higher demand on cognitive abilities than our design (Wakker and Deneffe 1996). To save space, we just report results using the midpoint of the two bounds. All results remain qualitatively the same if we use the upper or lower bounds or if we discard all payoff configurations with inconsistent choices.

The average switch point of 70.4% in the Internet experiment is considerably higher than the corresponding figure of 61.5% for the laboratory experiment.[7] Moreover, a similar difference between the two samples is found for all seven decision problems. Using the MW and KS tests we find that the differences between the laboratory and Internet samples are highly significant, both comparing averages across all questions or looking at each question separately (two-sided p-values < 0.01).[8] While our experiment supports Andersen et al.'s (2007) finding of greater preference heterogeneity in the field, more risk tolerance among students is in contrast with their finding of no significant differences in average risk aversion in a laboratory and a field experiment. There may be several explanations for this. First, it could simply be due to our much larger sample size in the CentERpanel than in their field experiment, yielding greater power of the test. Second, while it has been shown that there are no systematic effects of enforcing a switch point in the laboratory (Andersen et al. 2006), there could be such effects in a more heterogeneous population. Third, and somewhat along the same lines, we cannot exclude the possibility that there are mode effects for the general population, i.e. that although we cannot detect differences between the Internet and laboratory experiments for students, there may be differences for other population segments.

To disentangle subject pool bias and implementation mode effects as explanations for the observed differences in risk preferences, we also compared the average switch points in the "Lab-Lab" treatment and the "Lab-Internet" treatment. The MW and KS tests (using mean switch points across all payoff configurations or each payoff configuration separately), do not reject the null hypothesis that there is no difference (two-sided p-values > 0.10). The

[7] The findings in the laboratory are close to those in Holt and Laury (2002) for their high payoff treatments, which are closest to our treatments. Based on an extensive comparison to earlier results, they conclude that these findings are in line with previous estimates. Hence we do not repeat this exercise here.

[8] A structural analysis of the data confirms this finding, see von Gaudecker, van Soest, and Wengström (2008).

difference between average switch points in the "Lab-Lab" and "Lab-Internet" treatment is actually slightly positive, suggesting that the observed negative difference between the laboratory and the Internet experiments is not due to characteristics of the laboratory setting.

A similar conclusion emerges when we compare the average switch points in the "Internet-Uni" subsample to those of the student sample in the lab. The average switch point of 65.1 of the "Internet-Uni" sample is much closer to the laboratory mean of 61.5 than the overall Internet mean, and we cannot reject the null hypothesis of equality (two-sided p-values > 0.15). For each payoff configuration separately, MW and KS tests show no significant differences except for payoff configurations 3 and 5. Controlling for the composition of subject pools hence eliminates most of the differences between the Internet and laboratory findings. The disparity found between both errors and preferences in the Internet sample and the laboratory experiments is almost exclusively driven by the fact that the student population differs from the the general population.[9]

4 Self-Selection Bias in the CentERpanel Experiment

Conducting the experiment via an existing survey allows us to observe more features of the recruitment process than usual since we know many characteristics of all persons eligible for participation, regardless of whether they actually take part in the experiment or not. In order to get reliable population-wide estimates, sampling from a representative sub-population suffices only if non-response is perfectly random. Since people self-select into the experiment this condition may not hold. Indeed, the descriptive statistics in Table 2 suggest that there are some important differences between those who completed the experiment without rushing through it, those who chose not to participate, and those who started but rushed through or did not complete. In addition to this, Harrison, Lau, and Rutström (2007) have shown that self-selection effects may be important for the estimation of risk preferences. We look at the same issue, but we can exploit much more information about nonparticipants.

We first analyse the determinants of self-selection and then investigate their impact on observed choices. In order to structure the analysis, it is useful to divide the sampling process in the Internet experiment into three stages:

1. Dutch individuals are contacted at random and participate in the CentERpanel or not.

[9] Providing monetary incentives or not does not seem to affect the behaviour in any systematic way: the choices in the hypothetical and high incentive treatments are very similar. This is confirmed by the MW and KS tests on differences in average switch points. (Details not reported in the tables but available upon request from the authors.)

2. A random subsample of CentERpanel respondents is asked to take part in our experiment. After learning about the nature of the experiment, they decide to participate or decline participation.

3. Some of the subjects drop out during the experiment or click through it extremely rapidly.

Steps 2 and 3 are especially interesting for experimental economics because they resemble the recruitment process for laboratory experiments.

To see how Step 2 relates to typical recruitment procedures, note that some information on payoffs and the type of experiment is usually conveyed before subjects come to the lab. This is the typical form of communication in recruitment emails or on posters announcing the experiments. Such information is provided on our welcome screen (Figure 2). Subjects learn about the nature of the experiment and the possible payoffs, and then choose to participate or not.

Step 3 seems typical for the Internet environment – it usually plays no role in the laboratory. One may argue, however, that the participation decisions for laboratory experiments combine features of steps 2 and 3. Part of the nonparticipation in laboratory experiments may be similar to dropping out of the CentERpanel experiment, because of the negligible fixed costs of participation in the latter. Showing up at the laboratory at a specific time and date entails a significant cost – and subjects can be expected to have made the trade-off between the costs and benefits of participation beforehand. This is probably not the case in the Internet setting, where the experiment can be accessed within seconds of notification. Hence the cost-benefit analysis may well be postponed and carried out during the experiment. This may explain why subjects hardly ever leave the economics laboratory prematurely (and nobody left our laboratory sessions), whereas 4% of subjects did not finish the CentERpanel experiment. Similarly, rushing through the experiment can be considered as a form of nonparticipation, since there is a lower bound on the time needed to digest the instructions and to give serious answers. This minimum time certainly seems higher than the 1:43 minutes which is the minimum time observed in the Internet experiment. We checked several cut-off points between 3 and 7 minutes and finally chose the minimum duration observed in the laboratory (5:20 minutes). Results were robust to the precise value chosen. With this threshold, about seven percent of the Internet subjects fall into the category of "speeders."[10]

[10] The combined response rate for steps 2 and 3 in our Internet experiment is 78%. This seems to compare favourably to Harrison, Lau, and Rutström (2007), who employed more standard recruitment procedures in mailing out a letter to a random subsample of the Danish population and achieved a response rate of 38% (253 of 664 subjects), but it should be noted that our response rate is within a preselected sample that has shown a general inclination to fill out survey questionnaires.

An alternative explanation why Step 3 features prominently in the Internet experiment and not in laboratory experiments is the interaction with the experimenter and the typical rules in the laboratory. One difference is the possibility to ask questions. Internet participants who do not understand a task and cannot ask questions may more easily opt for randomly ticking options or drop out entirely. Another difference is that in typical laboratory experiments everybody is expected to stay until the last subject has finished, so that there is no point in rapid completion. We can analyse the consequences of these differences by comparing the "Lab-Lab" and "Lab-Internet" treatments (see Section 2.3). The distributions of the completion times are similar, with mean durations of about 12.5 minutes in both cases. Surprisingly, in the traditional "Lab-Lab" treatment the dispersion is higher and the left tail of the distribution has more mass. If rapid completion were due to the two factors mentioned above, we would expect the opposite. Completion times in the "Internet-Uni"- subgroup of the Internet subjects are lower than those of the laboratory subjects. This is consistent with our preferred interpretation of step 3 of the selection process, since the "Internet-Uni" group will contain more respondents who rush through the experiment.

4.1 The Determinants of Self-Selection

To analyse the factors that drive participation in the Internet experiment, we estimated a multinomial logit model with three possible outcomes: non-participation, rushing through or dropping out, and regular ("full") participation. Results are presented in Table 5, with full participation as the baseline category. Columns 1 and 2 contain the coefficients and standard errors of our basic specification, for nonparticipation and dropouts/speeders, respectively. Only basic covariates that are available for almost everybody are included – dummies for the incentive treatments, education, gender, age, occupational status, and residential area. In the extended specification (columns 3 and 4) the number of observations is lower since we use covariates from the DNB Household Survey questionnaires which were not filled out by all subjects. This specification adds household income (in four categories) and variables measuring financial expertise and preferences: Whether the respondent manages the household's finances, whether the employer offers a save-as-you-earn deduction arrangement,[11] whether the respondent holds such a plan, whether the household's financial portfolio includes savings

[11] This is an employer provided savings plan that is heavily subsidised by the state through tax deductions, see Alessie, Hochguertel, and van Soest (2006).These subsidies make net returns much higher than on any other safe asset. While it is easy to sign up for these plans and the employer does most of the paperwork, the default is not to participate. This may explain why employees with little financial knowledge or interest often do not sign up, cf., e.g., the work of Madrian and Shea (2001) on non-take up of 401(k) plans.

accounts, and whether it includes risky assets like mutual funds or stocks.[12]

For the variables included in both specifications, results for the two specifications are very similar. Nonparticipation is significantly less likely in the incentive treatments than in the hypothetical treatment (the benchmark). Translated into marginal effects, the coefficients indicate response rates that are almost ten percentage points higher than in the hypothetical treatment (all marginal effects are evaluated at a baseline with all dummy variables set to zero). The point estimates on not finishing the experiment are much smaller in magnitude and not significant. While incentives increase participation, they do not seem to attract a different subject pool: we estimated models that included a wide variety of interaction effects and none of them was significant. The coefficients on the high and low incentive treatments are not significantly different from each other.

Persons in the top two education categories are both significantly more likely to participate in the experiment and to finish it. Women's nonparticipation rates are four to five percentage points higher than those of men. Women also are slightly more prone to quit during the experiment or to finish it very rapidly. Age effects start to matter at age 45, beyond which participation rates decline. Those beyond 65 years of age are only half as likely to start the experiment as those younger than 35. At the same time, however, non-completion rates decrease significantly with age, mainly due to the fact that older participants are less likely to rush through the experiment.[13] The combined effects of age on full participation are small and insignificant in almost all cases.

Working respondents have higher participation rates than non-workers according to the parsimonious specification, but the effect becomes insignificant in the richer specification. Labour market status does not affect quitting or speeding. Living in an urban area has no significant impact at all.

Point estimates of the effects of income and wealth on participation are generally small and insignificant, and a joint test does not reject the null hypothesis that they play no role. The other financial variables are proxies for preferences and financial knowledge. Being the financial administrator of the household for example may reflect a preference for spending one's time with problems involving financial choices and taking financial risks. It significantly increases the propensities to participate and to finish the experiment in more than 5:20 minutes. A dummy for whether the employer offers a savings plan is a control variable necessary to avoid confounding the effects of holding such a plan with employment type. The vari-

[12]Other specifications of the selection model did not yield additional insights; they also included subjective income measures, wealth, and interactions between the covariates.

[13]We estimated models that treated the two components of step 3 (speeding through and dropping out) as separate outcomes. This is the only case where the distinction mattered, which is why we only report the results from the more parsimonious specification in the table.

able of interest, taking part in a save-as-you-earn savings arrangement, is associated with a (significant) six percentage points higher propensity to begin the experiment. This supports the interpretation that participation in the experiment is associated with financial expertise and interest. This interpretation is strengthened by the other two portfolio variables: On the one hand, having an ordinary savings account does not have any predictive power for taking part in the experiment, since saving accounts are commonly known and do not require much expertise or effort; on the other hand, the ownership of mutual funds, stocks, etc., is significantly associated with higher participation. These are much more sophisticated products and investing in them requires some financial knowledge. The results thus suggest that interested and knowledgeable individuals have a higher probability of participating.

4.2 The Impact of Self-Selection on Outcomes

To test whether selection effects matter for the outcomes considered in Section 3, we compare the observed (unweighted) sample distribution of full participants with a weighted distribution that corrects for the various steps of the selection process.

For Step 1, CentERdata provides standard survey weights based upon comparing with a much larger household survey drawn by Statistics Netherlands. We will assume that selection into the CentERpanel is independent of the variables of interest, conditional on the basic background variables used to construct these weights (age, sex, education, home ownership, region). This is a missing at random (MAR) assumption, see e.g. Little and Rubin (2002)). It implies that the standard weights can be used to correct for the selection in Step 1.

We make similar MAR assumptions for the other two steps, but then conditioning on the much larger set of background variables used in the previous subsection. We construct weights from a probit model that jointly explains the selection in Steps 2 and 3; each weight is the inverse of the predicted probability of being in the final sample. We multiply these weights with the weights for Step 1 to get weights that correct for all three steps of the selection process. Due to sample size considerations we opt for the parsimonious specification in the probit regression. We then test whether weighted sample statistics on the outcomes are significantly different from unweighted statistics. This can be seen as a test whether the selection process is selective, under the maintained assumption that selection in each step is MAR given all the covariates used to construct the weights.

In Table 4 the average number of inconsistencies and average mean switch points for the weighted and unweighted data are presented together with p-values of t-tests comparing mean values.[14] Taking the full selection process (steps 1, 2 and 3) into account the estimated

[14]The test works as follows: Let y denote our variable of interest (i.e. the average number of violations

population average number of inconsistencies is 2.60, compared to 2.43 for the unweighted sample. The difference is statistically significant (two-sided p-value=0.0001). This result is consistent with the findings in Section 4.1. The only variable for which we had a clear prior was education. We expect better educated people to make fewer errors. Indeed, the unweighted mean points at fewer inconsistencies than the weighted mean, in line with the fact that the better educated are overrepresented in the sample of participants. Controlling only for the selection in steps 2 and 3, the average number of inconsistencies is 2.55, which is significantly different from the unweighted number 2.43 (two-sided p-value:=0.001) but not from the 2.60, which also corrects for Step 1 (two-sided p-value=0.1). Hence, the selection bias reported for the full process originates mostly from steps 2 and 3, while step 1 has less of an impact.

For the estimates of risk preferences, we find a small underestimation of the mean switch points when using the unweighted sample compared to both weighted samples, but the difference is insignificant (irrespective of which switch points we use). Thus, although the participation decision is correlated with demographics, selection on observables does not fundamentally alter the results for our summary measure of risk preferences.

Using the approach developed byHeckman (1976, 1979) , Harrison, Lau, and Rutström (2007) find that not controlling for selection effects leads to an overestimation of average risk aversion in the population. Using the same methodology and similar exclusion restrictions,[15] our findings are different. Whether we use the full or parsimonious set of control variables and whether we estimate models jointly for all treatment groups or separately by incentive treatments, we never find a significant impact of self-selection on outcomes. Furthermore, we do not find any differences between incentive treatments, implying that we cannot establish a link between the size of the show-up fee and self-selection into the experiment. A possible explanation for the difference with Harrison, Lau, and Rutström's (2007) may be the completely different sampling frames. Future research will be needed to come to a final verdict on this issue.

or mean switch point) and $w(x)$ the weight. Our null hypothesis of no difference between the weighted and unweighted observations can then be stated as $E[w(x)y] = E[y]$ or $E[z] = 0$, with $z = (w(x) - 1)y$. Since we have a large sample size and few explanatory variables, we neglect the estimation error in w. The null hypothesis can then be tested with a standard t-test on whether the mean of z is zero or not.

[15]Exclusion restrictions refer to variables that affect participation but not risk aversion preferences, are needed if we do not want to rely on distributional assumptions for identification of the Heckman selection model. Following Harrison, Lau, and Rutström (2007), we used regional dummies.

5 Conclusions

We have first analysed the differences between a laboratory experiment among university students and a similar Internet experiment eliciting risk preferences among a sample covering the adult Dutch population, focusing both on estimated risk attitudes and on the tendency to make choices that violate economic theory ("errors"). The main result is that decomposing the differences into environmental ("mode") effects (laboratory vs. Internet) and subject pool bias ("selection effects") shows that selection effects dominate and mode effects hardly play a role. This result seems very robust and is found both when the laboratory experiment is changed so as to resemble the Internet experiment, and when the laboratory experiment is compared with the subsample of Internet subjects that are similar to the lab subjects.

The first implication of this is that results in laboratory experiments cannot simply be extrapolated to the general population. Students make a much smaller number of errors and also exhibit a higher degree of risk tolerance than the general population. The second implication is that embedding the experiment in a representative Internet survey has the potential to overcome this selection problem, since the conjecture that the experiment would be too difficult for respondents without the help of an experimenter assisting in person is not confirmed, at least not for the experiment at hand, with the help screens and graphical aids that were implemented.

This then leads to the second issue: it is obvious that the Internet sample provides better population coverage than the student sample in the lab, but there are still several stages in the process that lead to selection and non-response and which may bias the Internet sample based estimates for the population at large. We address this in the second part of the paper where we discuss and analyze the three stages in the process that drives voluntary participation in the Internet experiment. We find that higher education, being male, as well as interest and expertise in financial matters are predictors of participating in the experiment, and this type of selection leads to a sample with less prevalence of violations of monotonicity and dominance than in the general population. On the other hand, we found that selective participation in the Internet experiment did not induce selection bias on risk preferences, irrespective of whether or not we provide real monetary incentives or not.

We can therefore conclude that the Internet experiment is an appropriate way to estimate risk preferences for the general population, which is substantially less inclined to make risky choices than the subpopulation of students. Whether the conclusion that implementation mode does not matter also applies to more complicated experiments or experiments measuring different sorts of preferences, remains a topic for future research.

A Tables

Table 1: Characteristics of the Seven Payoff Configurations

Payoff Configuration	**Uncertainty Resolution, A**	**Payoff Low, A**	**Payoff High, A**	**Uncertainty Resolution, B**	**Payoff Low, B**	**Payoff High, B**
1	early	27	33	early	0	69
2	early	39	48	early	9	87
3	early	12	15	early	-15	48
4	early	33	36	late	6	69
5	early	18	21	late	-9	54
6	early	24	27	early	-3	60
7	late	15	18	late	-12	51

Note: These values were shown in the high incentive and hypothetical treatments. For the low incentive treatment they were divided by three.

Table 2: Selected Characteristics of Participants

Variable	Final Sample	CentERpanel Non-Participants	Dropouts Speeders	Laboratory Final Sample
Hypothetical treatment	0.31	0.50	0.37	0.37
Low incentive treatment	0.37	0.23	0.35	0.27
High incentive treatment	0.32	0.27	0.28	0.37
Primary / lower sec. education	0.31	0.44	0.34	.
Higher sec. / interm. voc. train.	0.33	0.30	0.41	.
Higher vocational training	0.25	0.20	0.16	.
University degree / univ. student	0.12	0.06	0.09	1.00
Female	0.45	0.54	0.56	0.46
Age 16-34 years	0.24	0.14	0.46	1.00
Age 35-44 years	0.19	0.13	0.21	.
Age 45-54 years	0.23	0.22	0.14	.
Age 55-64 years	0.18	0.18	0.09	.
Age 65+ years	0.16	0.33	0.10	.
Working	0.56	0.36	0.55	.
Unemployed, Looking for Job	0.02	0.03	0.03	.
Student, Pensioner, Housework	0.42	0.62	0.42	1.00
Lives in Urban Area	0.60	0.63	0.58	.
HH financial administrator	0.66	0.56	0.48	.
Employer offers Savings Plan	0.44	0.25	0.32	.
Has Sav. Plan via Employer	0.36	0.17	0.26	.
Has Sav. Acc. or similar	0.87	0.85	0.90	.
Holds Stocks, or similar	0.31	0.25	0.29	.
HH income below 22k Euros	0.34	0.35	0.33	.
HH income $\in$ [22k, 40k Euros)	0.49	0.51	0.49	.
HH income at least 40k Euros	0.17	0.14	0.18	.
Max. Number of Observations	1790	291	218	178

Note: The numbers shown indicate fractions in the final sample. Some households did not complete the questionnaires of the DHS from which some of the variables are drawn. Hence the number of observations is lower for some of the variables in question. This is particularly true for the last two sections of the table.

Table 3: Frequency of inconsistencies by type of error and subsample

	Dominance	Within	Between
Internet	11.0%	3.9%	21.1%
Laboratory	1.5%	1.8%	14.4%
Internet - Uni	3.9%	1.5%	13.1%
Lab - Internet	1.4%	1.8%	15.1%
Lab - Lab	1.5%	1.7%	13.7%

Note: The figures represent frequencies of the different types of errors as a percentage of the number of possible violations. The fractions of violations for the dominance category were obtained by dividing the total number of dominance violations in each category by the total number of screens shown to subjects on which dominance violations could be made. The numbers for the within category are calculated as the number of within violations, divided by the total number of screens shown to subjects in each group. The figures of the last column were obtained by dividing the number of between errors by the number of times the second screen was displayed to subjects.

Table 4: Weighted data

Weight	Average # inconsistencies	p-value steps 2,3	p-value no weight	Meanswitch	p-value steps 2,3	p-value no weight
Steps 1,2,3	2.6	0.1	0.0001	71.55	0.2	0.1
	(0.05)			(0.50)		
Steps 2,3	2.55	.	0.001	71.14	.	0.1
	(0.05)			(0.49)		
None	2.43	.	.	70.36	.	.
	(0.05)			(0.53)		

Note: Variables in category Steps 1,2,3 use weights for step 1,2 and 3; variables in category Steps 2,3 use weights for step 2 and 3. Average mean switch points are calculated using data from the hypothetical and high incentive treatments only. Standard errors are given in parenthesis. P-values comes from t-tests, described in footnote 14 in Section 4.2, with the null hypotheses of equal means.

Table 5: Self-Selection into the CentERpanel Experiment

Specification	NP (1)	DO/SP (2)	NP (3)	DO/SP (4)
Low incentive treatment	-1.047***	-.220	-1.053***	-.247
	(.163)	(.174)	(.183)	(.207)
High incentive treatment	-.699***	-.277	-.881***	-.342
	(.156)	(.183)	(.188)	(.226)
Higher sec. / interm. voc. train.	-.285*	-.120	-.180	-.290
	(.160)	(.177)	(.185)	(.215)
Higher vocational training	-.413**	-.677***	-.329	-.801***
	(.181)	(.228)	(.217)	(.282)
University degree / univ. student	-.881***	-.447	-.840**	-.622*
	(.281)	(.279)	(.356)	(.356)
Female	.383***	.275*	.357**	.253
	(.137)	(.152)	(.158)	(.184)
Age 35-44 years	.336	-.450**	.385	-.471*
	(.242)	(.196)	(.301)	(.253)
Age 45-54 years	.579***	-1.076***	.761***	-1.044***
	(.221)	(.222)	(.268)	(.275)
Age 55-64 years	.489**	-1.326***	.781***	-1.214***
	(.231)	(.262)	(.274)	(.301)
Age 65+ years	1.120***	-1.171***	1.349***	-1.297***
	(.233)	(.278)	(.284)	(.341)
Working	-.402**	-.104	-.185	-.054
	(.175)	(.179)	(.214)	(.229)
Unemployed, Looking for Job	.059	.198	-.011	.263
	(.416)	(.439)	(.515)	(.519)
Lives in Urban Area	.165	-.059	.235	-.051
	(.137)	(.151)	(.159)	(.183)
HH financial administrator			-.379**	-.328*
			(.164)	(.193)
Employer offers Savings Plan			.180	-.369
			(.281)	(.365)
Has Sav. Plan via Employer			-.845***	-.045
			(.310)	(.379)
Has Sav. Acc. or similar			.095	.228
			(.225)	(.292)
Holds Stocks, or similar			-.368**	.058
			(.177)	(.204)
HH income $\in$ [22k, 40k Euros)			.153	.191
			(.176)	(.208)
HH income at least 40k Euros			.040	.539*
			(.259)	(.280)
Constant	-1.693***	-1.147***	-1.735***	-1.244***
	(.253)	(.240)	(.371)	(.401)
No. of Observations	2296	2296	1802	1802

Note: Coefficient estimates and corresponding standard errors of multinomial logit regression. Columns indicate categories of the dependent variable by regression type. The reference category contains those respondents who completed the experiment in more than 5:20 minutes. Columns (1) and (3) list estimates for opting for nonparticipation on the first screen (NP); columns (2) and (4) those for dropping out before completion (DO) or finishing the experiment in less than 5:20 minutes (SP). Left-out categories of relevant variables are hypothetical treatment; primary and lower secondary education; ages 16-34; other type of occupation; and household income less than 22,000 Euro. Asterisks indicate significance at the 10%, 5%, and 1%-level.

B Figures

Figure 1: Screenshot of Payoff Configuration 5, First Screen

Figure 2: Translations of the Welcome Screens in the CentERpanel Experiment

High (Low) Incentive Treatment

Welcome to this economic experiment carried out by researchers of Tilburg University. The experiment is about making choices under uncertainty. Please read the instructions carefully in order to understand how the experiment works.

If you have questions after the beginning of the experiment, you can return to the instructions by clicking on a link at the top of the screen. If you have questions on the specific screen, you can click on 'Help' at the top right corner of the screen.

You will receive 15 (5) Euros for participating. Then you can, depending on the choices you make and on chance, earn more or lose part of the 15 (5) Euros. If completing the total experiment, you receive the reward for participating, possibly increased by your gain (or reduced by your loss) in one of the choices you have made. Whether the latter occurs and which choice then determines your payoff, will be determined by chance. **Your total reward will be added to your CentERpoints.**

The questions are not designed to test you. Answers are therefore not correct or incorrect; please give the answers that reflect your own preferences. Assume in each choice problem that this choice determines your actual payoff.

This questionnaire is about making choices, and your payoff depends on your choices and on chance. If you do not want to participate out of principle, you can indicate this below. In that case you will not continue with the questionnaire.

O Yes, I proceed with the questionnaire

O No, I do **not** want to complete this questionnaire

Hypothetical Treatment

Welcome to this economic experiment carried out by researchers of Tilburg University. The experiment is about making choices under uncertainty. Please read the instructions carefully in order to understand how the experiment works.

If you have questions after the beginning of the experiment, you can return to the instructions by clicking on a link at the top of the screen. If you have questions on the specific screen, you can click on 'Help' at the top right corner of the screen.

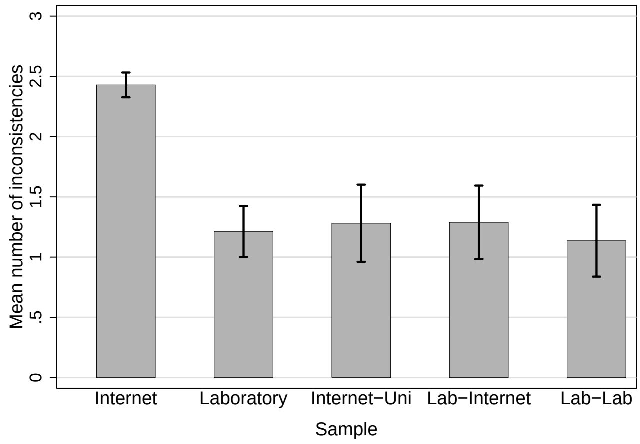
The questions are not designed to test you. Answers are therefore not correct or incorrect; please give the answers that reflect your own preferences.

This questionnaire is about making choices between several situations in which you can (hypothetically) gain or lose money. Your revenue depends on the choices you make and on chance. **What matters is what you would do in hypothetical situations, in reality, there is nothing at stake for you.** If you nevertheless do not want to participate out of principle, you can indicate this below. In that case you will not continue with the questionnaire.

O Yes, I proceed with the questionnaire

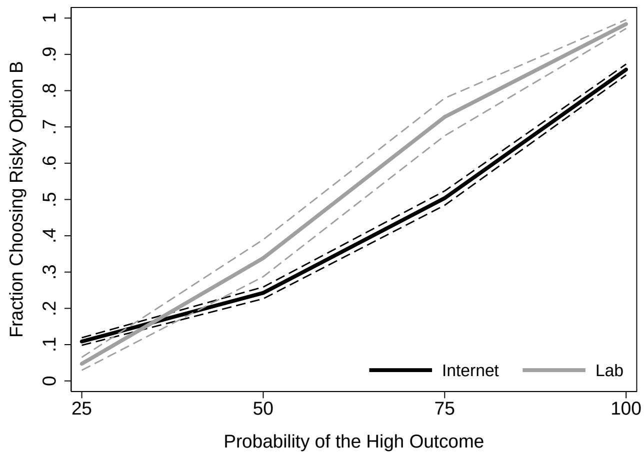
O No, I do **not** want to complete this questionnaire

Figure 3: Average Number of Answers that Violate Monotonicity or Dominance by Sample.



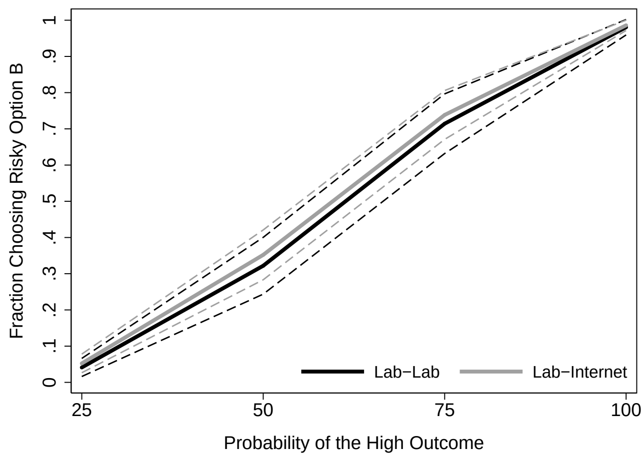
Note: Values shown are means over all screens per subject (minimum 0, maximum 7). "Internet" consists of unweighted numbers from CentERpanel respondents. "Lab" are averages for all laboratory subjects and "Internet-Uni" mean values for those respondents of the CentERpanel that are less than 35 years of age and hold a university degree or study to obtain one. "Lab-Lab" and "Lab-Internet" are averages for laboratory subjects in the"Lab-Lab" and "Lab-Internet" treatments respectively.

Figure 4: Risky Choices, Internet and Lab Subsamples



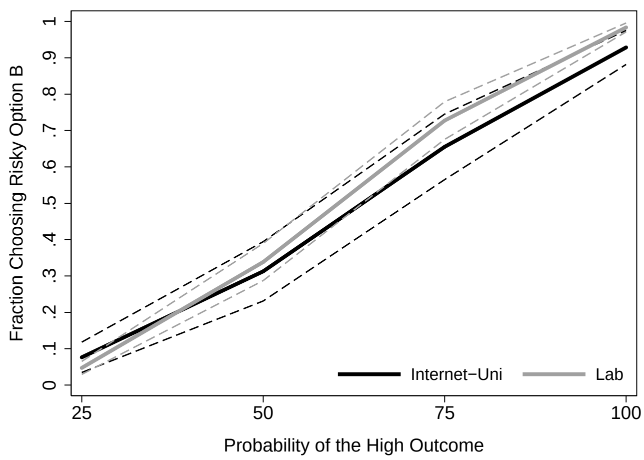
Note: Solid lines depict means over all first screen choices. Dashed lines display corresponding confidence intervals. Data from the low incentive treatment is excluded. "Internet" consists of the raw numbers from CentERpanel respondents. "Lab" are averages for laboratory subjects.

Figure 5: Risky Choices, Lab-Internet and Lab-Lab Subsamples



Note: Solid lines depict means over all first screen choices. Dashed lines display corresponding confidence intervals. Data from the low incentive treatment is excluded. "Lab-Lab" are averages for laboratory subjects in the "Lab-Lab" treatment and "Lab-Internet" mean values for subjects of the "Lab-Internet" treatment.

Figure 6: Risky Choices, Internet-Uni and Lab Subsamples



Note: Solid lines depict means over all first screen choices. Dashed lines display corresponding confidence intervals. Data from the low incentive treatment is excluded. "Lab" are averages for laboratory subjects and "Internet-Uni" mean values for those respondents of the CentERpanel that are less than 35 years of age and hold a university degree or study to obtain one.

References

Alessie, R., S. Hochguertel, and A. van Soest (2006): "Non-Take-Up of Tax-Favored Savings Plans: Evidence from Dutch Employees," *Journal of Economic Psychology*, 27(4), 483–501.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2006): "Elicitation Using Multiple Price List Formats," *Experimental Economics*, 9(4), 383–405.

——— (2007): "Preference Heterogeneity in Experiments: Comparing the Field and Laboratory," Centre for Economic and Business Research Discussion Paper No. 2005-03, Copenhagen.

Ballinger, T., and N. Wilcox (1997): "Decisions, Error and Heterogeneity," *The Economic Journal*, 107(443), 1090–1105.

Bellemare, C., and S. Kröger (2007): "On Representative Social Capital," *European Economic Review*, 51(1), 183–202.

Binswanger, H. P. (1980): "Attitudes Towards Risk: An Experimental Measurement in Rural India," *American Journal of Agricultural Economics*, 62, 395–407.

Bleichrodt, H., J. L. Pinto, and P. P. Wakker (2001): "Making Descriptive Use of Prospect Theory to Improve the Prescriptive Use of Expected Utility," *Management Science*, 47(11), 1498–1514.

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.

Donkers, B., B. Melenberg, and A. van Soest (2001): "Estimating Risk Attitudes Using Lotteries; A Large Sample Approach," *Journal of Risk and Uncertainty*, 22(2), 165–195.

Fehr, E., U. Fischbacher, B. von Rosenbladt, J. Schupp, and G. G. Wagner (2003): "A Nation-Wide Laboratory: Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys," IZA Discussion Paper 715.

Güth, W., C. Schmidt, and M. Sutter (2007): "Bargaining Outside the Lab – A Newspaper Experiment of a Three Person-Ultimatum Game," *Economic Journal*, 117(518), 449–469.

Harrison, G. W., M. I. Lau, and E. E. Rutström (2007): "Risk Attitudes, Randomization to Treatment, and Self-Selection Into Experiments," University of Central Florida Economics Working Paper No. 05-01.

Harrison, G. W., M. I. Lau, and M. B. Williams (2002): "Estimating Discount Rates in Denmark: A Field Experiment," *American Economic Review*, 92(5), 1606–1617.

Harrison, G. W., and J. A. List (2004): "Field Experiments," *Journal of Economic Literature*, 42(4), 1009–1055.

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(2), 433–458.

Heckman, J. 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(4), 475–492.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–61.

Hey, J. D. (1995): "Experimental Investigations of Errors in Decision-Making under Risk," *European Economic Review*, 39, 633–641.

Holt, C. A., and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

Kreps, D. M., and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

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(2), 153–174.

Little, R. J., and D. B. Rubin (2002): *Statistical Analysis with Missing Data*. John Wiley & Sons Inc., New York, 2nd edn.

Loomes, G., P. G. Moffatt, and R. Sugden (2002): "A Microeconometric Test of Alternative Stochastic Theories of Risky Choice," *Journal of Risk and Uncertainty*, 24(2), 103–130.

Lucking-Reiley, D. (1999): "Using Field Experiments to Test Equivalence between Auction Formats: Magic on the Internet," *American Economic Review*, 89(5), 1063–1080.

Madrian, B. C., and D. F. Shea (2001): "The Power of Suggestion: Inertia in 401(K) Participation and Savings Behavior," *Quarterly Journal of Economics*, 116(4), 1149–1187.

Schmidt, U., and T. Neugebauer (2007): "Testing Expected Utility in the Presence of Errors," *Economic Journal*, 117(518), 470–485.

von Gaudecker, H.-M., A. van Soest, and E. Wengström (2008): "Risk Preferences in the Small for a Large Population," Netspar Discussion Paper 2008-011.

Wakker, P., and D. Deneffe (1996): "Eliciting von Neumann-Morgenstern Utilities When Probabilities are Distorted or Unknown," *Management Science*, 42(8), 1131–1150.