The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers

www.cesifo.org/wp

Pay Few Subjects but Pay Them Well: Cost-Effectiveness of Random Incentive Systems

Christoph March
Anthony Ziegelmeyer
Ben Greiner
René Cyranek

CESIFO WORKING PAPER NO. 5988
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
JULY 2016

An electronic version of the paper may be downloaded
- *from the SSRN website: www.SSRN.com*
- *from the RePEc website: www.RePEc.org*
- *from the CESifo website: www.CESifo-group.org/wp*

ISSN 2364-1428

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5988*

Pay Few Subjects but Pay Them Well: Cost-Effectiveness of Random Incentive Systems

Abstract

Most large-scale economic experiments use a between-subjects random incentive system-BRIS-which selects a subset of the participants at random and offers real payment only to the selected participants. We evaluate the relative impact of nominal payoffs and the selection probability on the incentive effectiveness of BRIS. High nominal payoffs and a tiny but unknown selection probability induce the monetary incentives in a large-scale Internet treatment. Two laboratory treatments complement the Internet treatment, one with high nominal payoffs and a moderate selection probability and the other with moderate nominal payoffs and a high selection probability. For the majority of our participants nominal payoffs have a bigger impact on the incentive effectiveness of BRIS than the probability of payment and the latter hardly dilutes the effects of monetary incentives. Our results suggest that cost-effective BRIS consist of high nominal payoffs and a small selection probability.

JEL-Codes: C910, C930, D120, D810.

Keywords: cost-effectiveness, random incentive system, large-scale experiments, risk aversion, decision error.

Christoph March
TUM School of Management
Technical University Munich
Arcisstrasse 21
Germany – 80333 Munich
christoph.march@wi.tum.de

Anthony Ziegelmeyer
Queen's Management School
Queen's University Belfast
185 Stranmillis Road
United Kingdom – Belfast BT9 5EE
email@anthonyziegelmeyer.com

Ben Greiner
University of New South Wales
School of Economics
Australia - UNSW Sydney NSW 2052
bgreiner@unsw.edu.au

René Cyranek
LMU Munich, Munich Experimental Laboratory for Economic and Social Sciences (mMlessa)
Giselastrasse 10
Germany – 80802 Munich
rene.cyranek@lmu.de

Financial support by the Max Planck Society and the German Science Foundation are gratefully acknowledged. We thank Nathaniel Wilcox for enlightening discussions, Martin Kocher, David Laibson, Sergey Popov, Klaus Schmidt, Joachim Winter, as well as audiences at seminars and conferences for helpful comments.

1 Introduction

Employee referral programmes are considered as among the most effective recruitment methods by half of the UK private sector (CIPD, 2013) and they have been adopted by about 70 percent of US firms. Employers often award a lottery ticket for every qualified applicant that an employee submits and recruitment platforms argue that lotteries constitute a cost-effective solution to incentivize participation in referral programmes (see e.g. CareerBuilder, 2012). Random payments also play a major role in fundraising for the private provision of public goods (Maeda, 2008), in health promotion plans to motivate beneficial behaviors (Haisley, Cryder, Loewenstein, and Volpp, 2008), and in surveys to increase contact, response and speed of response (Gajic, Cameron, and Hurley, 2012). But the type of behavioral intervention which relies almost exclusively on random payments as contingent financial incentives is the economic experiment with a large sample of participants.

Due to budget constraints, most large-scale economic experiments use a *between-subjects random incentive system*—BRIS—which selects a subset of the experimental subjects at random and offers real payment only to these selected subjects (e.g. Cohn, Engelmann, Fehr, and Maréchal, 2015; Dohmen, Falk, Huffman, and Sunde, 2010). BRIS seems like the inescapable payment protocol in large-scale economic experiments as it allows the experimenter to award sizeable nominal payoffs to a large sample of subjects at reasonable costs. But despite its increasing popularity the effectiveness of the (monetary) incentives induced by BRIS remains insufficiently explored, not least because incentive effects are rarely ascertained in an unequivocal way in economic experiments. In an attempt to fill this void, we compare the cost-effectiveness of three BRIS—each BRIS being used in a different experimental treatment—where subjects complete a risk elicitation task with clear incentive effects in each treatment. The three BRIS differ in their nominal payoffs or their selection probability of paid subjects (or both). Our findings could inform the design of random payments aiming at strong incentive effects in large-scale economic experiments but also in other behavioral interventions such as employee referral schemes and fundraising campaigns.

In each of the three incentive treatments, subjects complete the risk elicitation task developed by Holt and Laury (2002, HL hereafter). HL's (risk elicitation) task presents subjects with a menu of ten ordered choices between a safe and a risky lottery with the understanding that one of these choices is selected at random ex post for real payment. Extensive experimental evidence shows that scaling up lottery outcomes in HL's task significantly increases risk aversion (Harrison, Johnson, McInnes, and Rutström, 2005; Holt and Laury, 2002, 2005, 2008). Thanks to the presence of clear incentive effects in HL's task, the degree of risk aversion measured in a given incentive treatment is a sound indicator of the effectiveness of the incentives induced by the respective BRIS. Our core incentive treatment is $Scale50PrUnknown$ which is a large-scale treatment where lottery outcomes equal those in HL's $50\times$ payoff scale treatment (except that they are in Euros) and subjects are uninformed of the actual selection probability though they are most likely under the impression that the chance of real payment is tiny. The expected payment of a *randomly selected* risk neutral subject is (about) 120 Euros. To allow for the random selection of a few subjects from a large sample, we implemented $Scale50PrUnknown$ on the Internet where 3,582 subjects completed HL's task and only five of them were randomly selected to receive real payment. $Scale50PrUnknown$ is complemented with two incentive treatments conducted in the laboratory where the selection probability is public knowledge. In incentive treatment $Scale50Pr1/15$, 60 subjects choose between lotteries with the same outcomes as in $Scale50PrUnknown$ and four subjects receive real payment. In incentive treatment $Scale10Pr1/3$, 60 subjects choose between lotteries where outcomes of $Scale50PrUnknown$ are divided by 5, and twenty subjects are selected for payment. Thus, the experimenter's expected costs per paired

lottery choice are identical in the two laboratory treatments but the expected payment of a *randomly selected* risk neutral subject is only (about) 24 Euros in $Scale10Pr1/3$. Instructions are (almost) identical across treatments including the nominal amounts on the decision sheet, in the two laboratory treatments we simply employ different conversion rates from the experimental currency to Euros.[1]

Though it provides *indirect* evidence on the validity of BRIS, we did not design our experiment to check whether BRIS leads to biased estimates of risk aversion in HL's task (see subsection 2.2 for a discussion of the link between the incentive compatibility and the cost-effectiveness of BRIS). The primary goal of our experiment is to clarify the relative impact of the scale of nominal payoffs and the level of the selection probability on the effectiveness of incentives. First, by comparing risk aversion in $Scale50Pr1/15$ and $Scale10Pr1/3$ we investigate whether the scale of nominal payoffs is more effective in influencing subjects' behavior than the level of the selection probability when the experimenter's expected costs per choice are kept constant. Second, by comparing risk aversion in $Scale50Pr1/15$ and $Scale50PrUnknown$ we examine the impact of the selection probability on the effectiveness of incentives for a given scale of nominal payoffs and severely reduced experimenter's costs. Third, by comparing risk aversion in $Scale10Pr1/3$ and $Scale50PrUnknown$ we again investigate whether the scale of nominal payoffs is more incentive effective than the level of the selection probability but at reduced experimenter's costs. The presence of 637 non-students in the Internet sample also allows us to check whether incentive effects can be generalised to non-student pools.

Besides nominal payoffs or the selection probability, $Scale50PrUnknown$ differs from our laboratory treatments in terms of the demographic characteristics of subjects and the implementation mode. Though the laboratory implementation of BRIS with unknown and tiny selection probabilities seems hardly feasible, these two differences have the potential to confound our analysis of the relationship between BRIS and the effects of monetary incentives. We address this issue in several ways. We first collect in each incentive treatment a substantial amount of background information (e.g. age, gender, education level, employment and marital status) and we control for this observed heterogeneity in our statistical analysis. Moreover, subjects in $Scale50PrUnknown$ were mainly recruited from mailing lists composed of students and about three-quarters of our Internet sample consists of full-time students like in our laboratory samples. Only the implementation mode might play a role for the bulk of our samples. Lastly, we estimate a structural econometric model whose stochastic component combines two elements, a disturbance term according to which subjective values are prone to measurement error and a constant probability of a lapse of concentration according to which the choice is made completely at random. The characteristics of BRIS might influence errors in the formation of risk preferences. Errors in the execution of risk preferences are expected to pick up differences in the implementation mode as subjects are more likely to lose concentration in the Internet environment than in the controlled laboratory setting.

We find a strong tendency towards risk-averse behavior among subjects in each of our incentive treatments with at most 10 and 14 percent of the subjects being classified as risk-neutral and risk-loving. When the statistical analysis is carried out on the entire sample of subjects, risk aversion in $Scale50PrUnknown$ is lower than in $Scale50Pr1/15$ while it is larger than in $Scale10Pr1/3$, but none of these differences are statistically significant. A remarkable finding is that treatment effects on risk

[1] HL's task uses the *within-subjects random incentive system*—WRIS—where each subject performs a series of individual tasks knowing that only one of these tasks will be randomly selected for real payment. Thus, in each of our incentive treatments we use a combination of BRIS and WRIS which is referred to as a *hybrid* RIS by Baltussen, Post, van den Assem, and Wakker (2012). We simply refer to such a combination as BRIS since the characteristics of WRIS remain identical in each of our incentive treatments and we expect differences in the effectiveness of incentives to originate from the variation in the scale of nominal payoffs or the level of the between-subjects selection probability.

attitudes, as well as several demographic effects, significantly vary with the education status of subjects. For full-time students nominal payoffs have a bigger impact on the incentive effectiveness of BRIS than the probability of payment and that the latter hardly dilutes the effects of monetary incentives. For part-time students and for non-students none of the two components of BRIS has a significant impact on its incentive effectiveness. On the other hand, regardless of the education status of subjects, the estimated probability of a lapse of concentration is significantly larger in $Scale50PrUnknown$ than in each of the two laboratory treatments and there is no significant difference between the two laboratory treatments. On the whole, our findings show that the scale of nominal payoffs is more effective in influencing subjects' behavior than the probability of payment.

Related literature

Though there is a long experimental literature on the incentive compatibility of random incentive systems (see subsection 2.2 for references), we are unaware of any previous study that has examined the relative impact of nominal payoffs and the selection probability on the effectiveness of the incentives induced by BRIS. We review below the few experiments that have examined whether subjects reduce nominal payoffs to account for the probability of being among the paid subjects.[2]

Three studies conclude that the selection probability has limited impact on the effectiveness of the incentives. Harrison, Lau, and Rutström (2007, footnote 16) fail to reject the hypothesis that paying 26 subjects with a 1-in-10 probability generates the same responses as paying 51 subjects for certain in HL's task with 10× nominal payoffs. Harrison, Lau, and Williams (2002) elicit individual discount rates for a nationally representative sample of the Danish population where nominal payoffs range from US$450 to US$1,840 (depending on the payment date) and one out of either 5, 10 or 15 subjects receives actual payment (depending on the experimental session). The authors report that the level of the selection probability does not significantly impact predicted discount rates. In an Internet experiment on herding in financial markets, Drehmann, Oechssler, and Roider (2005) observe that large nominal payoffs combined with an unknown and tiny probability of payment influence subjects' behavior to the same extent as 10 times lower nominal payoffs combined with a 30 times higher and known probability of payment. Note however that incentive effects are less clear in the last two reviewed studies.

In contrast, Baltussen, Post, van den Assem, and Wakker (2012) find that a 10 percent chance of real payment significantly reduces risk aversion in their investigation of BRIS's validity in a dynamic choice experiment. Further research is needed to assess the incentive effectiveness of BRIS in complex and dynamic tasks.

The rest of the paper proceeds as follows. Section 2 describes and motivates the experimental design, and it details our experimental procedures and samples of subjects. Section 3 reports our main results and Section 4 concludes. The online supplementary material contains four appendices with the experimental instructions and complementary statistical analyses.

2 Design, Motivation, Procedures and Subjects

Our experimental design consists of three incentive treatments in which subjects complete HL's risk elicitation task and each of which uses a different BRIS to induce monetary incentives.

[2] Laury (2012) investigates the effectiveness of the incentives induced by WRIS in HL's risk elicitation task and fails to reject the hypothesis that paying all 10 decisions generates the same risk behavior as paying for 1 of 10 decisions.

Our core incentive treatment is $Scale50PrUnknown$. This large-scale treatment relies on an Internet sample of 3,582 subjects that offers a wider range of sociodemographic characteristics than usually found in convenience samples of students. Lottery outcomes in Euros equal those in HL's 50× payoff scale treatment and subjects only know that five of them are randomly selected to receive real payment. In most large-scale behavioral interventions experimenters don't know in advance how many subjects will participate and complete the choice tasks. Revealing the number of paid subjects rather than the percentage of paid subjects limits the experimenter's costs while allowing participation to grow unchecked. Though the actual (between-subjects) selection probability is unknown in $Scale50PrUnknown$, subjects are most likely under the impression that the sample size is large and that their chance of real payment is tiny.[3] $Scale50PrUnknown$ therefore constitutes a challenging environment for the effectiveness of the incentives induced by BRIS while allowing for the inference of risk preferences in a rather broad population at reasonable research costs.[4]

To assess the relative impact of the scale of nominal payoffs and the level of the selection probability, we complement $Scale50PrUnknown$ with two incentive treatments conducted in the laboratory. In the two laboratory treatments, the selection probability is public knowledge. In $Scale50Pr1/15$, lottery outcomes are identical to those in $Scale50PrUnknown$ and four out of 60 subjects receive real payment. In $Scale10Pr1/3$, lottery outcomes are one-fifth of those in $Scale50Pr1/15$ and twenty out of 60 subjects receive real payment. The experimenter's expected costs per choice are therefore identical in the two laboratory treatments but a risk neutral subject who is selected for payment receives in expected terms 120 and 24 Euros in $Scale50Pr1/15$ and $Scale10Pr1/3$ respectively (said differently, absent treatment effects, the expected costs are identical in the two laboratory treatments).

We cannot rule out that there is a selection bias in $Scale50PrUnknown$ due to voluntary participation on the Internet. As rightly noted by Harrison, Lau, and Rutström (2009) the use of random payments could attract subjects that are less risk averse than the population meaning that the selection bias would lower risk aversion in $Scale50PrUnknown$ compared to the two laboratory treatments. However, experimental evidence based on a related design suggests that selection effects are small to non-existent. Indeed, von Gaudecker, van Soest, and Wengström (2012) use a modified version of HL's task to estimate error propensities and risk parameters in the laboratory and over the Internet, and they find that self-selection slightly underestimates error propensities but has no effect on risk aversion. Using a different lottery choice task, Cleave, Nikiforakis, and Slonim (2013) also find that subjects' risk preferences are representative of those in the population from which they were recruited.

Below, we first describe HL's task, then we argue that the cost-effectiveness of BRIS is an empirical question worth investigating, and finally we detail our experimental procedures and samples of subjects.

2.1 HL's Risk Elicitation Task

We rely on an arguably transparent elicitation method for risk aversion introduced by Holt and Laury (2002). Each subject is presented with a menu of ten ordered decisions between a "safe" and a "risky"

[3] Indeed, subjects are aware that the underlying event for the research is the FIFA Soccer World Cup—the biggest single-event sporting competition in the world—which is hosted in Germany and that German institutions jointly conduct the research (see Section 2.3).

[4] The risk elicitation task was the first of three tasks that subjects had to complete. The two other tasks were related to the processing of information in parimutuel and double auction prediction markets and are not discussed here. Five and twenty subjects were selected for payment in the second and third task respectively. Instructions made clear to subjects that if they were randomly selected to receive real payment for having completed the risk elicitation task then they could not be selected to receive real payment for any other task.

Decision	Safe lottery	Risky lottery	Expected value difference
1	(0.1: €100, 0.9: €80)	(0.1: €192.50, 0.9: €5)	€58.25
2	(0.2: €100, 0.8: €80)	(0.2: €192.50, 0.8: €5)	€41.50
3	(0.3: €100, 0.7: €80)	(0.3: €192.50, 0.7: €5)	€24.75
4	(0.4: €100, 0.6: €80)	(0.4: €192.50, 0.6: €5)	€ 8.00
5	(0.5: €100, 0.5: €80)	(0.5: €192.50, 0.5: €5)	–€ 8.75
6	(0.6: €100, 0.4: €80)	(0.6: €192.50, 0.4: €5)	–€25.50
7	(0.7: €100, 0.3: €80)	(0.7: €192.50, 0.3: €5)	–€42.25
8	(0.8: €100, 0.2: €80)	(0.8: €192.50, 0.2: €5)	–€59.00
9	(0.9: €100, 0.1: €80)	(0.9: €192.50, 0.1: €5)	–€75.75
10	(1.0: €100, 0.0: €80)	(1.0: €192.50, 0.0: €5)	–€92.50

Notes: The third column reports the expected value difference between the safe and the risky lottery. In $Scale10Pr1/3$ the monetary outcomes of lotteries are divided by 5.

Table 1: Lottery-choice decisions in $Scale50PrUnknown$ and $Scale50Pr1/15$

one-stage lottery. The safe lottery offers less variable monetary outcomes than the risky lottery. The subject chooses either the safe or the risky lottery in each decision (we did not allow subjects to express indifference), and one decision is later selected at random for payout. Table 1 illustrates the payoff matrix presented to subjects in $Scale50PrUnknown$ and $Scale50Pr1/15$. This payoff matrix is identical to HL's matrix in their 50× payoff scale treatment except that lottery outcomes are in Euros (€1 ≃ US$1.25 at the time of the experiment). The first row shows that the safe lottery $(0.1 : €100, 0.9 : €80)$ offers a 10% chance of receiving €100 and a 90% chance of receiving €80 whereas the risky lottery $(0.1 : €192.50, 0.9 : €5)$ offers a 10% chance of receiving €192.50 and a 90% chance of receiving €5. Accordingly, the expected monetary value difference between the safe and the risky lottery equals €58.25 in the first row. As one proceeds down the matrix, the expected value of both one-stage lotteries increases, but the expected value of the risky lottery becomes greater than the expected value of the safe lottery. In the experimental instructions probabilities are explained in terms of throws of a ten-sided die and the expected value difference is not disclosed (see online Appendix A).

2.2 The Cost-Effectiveness of BRIS: A Practical Question Worth Investigating

We designed our experiment to assess the relative impact of the scale of nominal payoffs and the level of the selection probability on the effectiveness of the incentives induced by BRIS (henceforth the *incentive effectiveness* of BRIS). The relevance of such an inquiry may be questioned at first glance. Indeed, the emphasis of previous studies on the incentive compatibility of random incentive systems rather than on their cost-effectiveness suggests that incentive effectiveness has usually been considered an issue subsidiary to incentive compatibility by experimental economists. Here we first clarify the link between the incentive compatibility and the incentive effectiveness of BRIS. We then argue that, because of the extreme difficulty if not impossibility of establishing the incentive compatibility of random incentive systems, the more practical question of the cost-effectiveness of BRIS is also worth investigating.

Incentive Compatibility of BRIS

Incentive compatibility requires that subjects' preferences over safe and risky lotteries are truthfully elicited. Since randomizing devices are used to generate the payments awarded to subjects, the incentive

properties of BRIS depend critically on how subjects evaluate random payments over one-stage lotteries. Following Azrieli, Chambers, and Healy (2016), we specify below the restrictions on preferences over random payments which ensure that the BRIS we used are incentive compatible.

We start by describing formally the choices that subjects face in HL's task and how subjects are paid for their choices in our BRIS. Let X_{BRIS} denote the set of one-stage lotteries that subjects choose from in BRIS $\in \{Scale10Pr1/3, Scale50Pr1/15, Scale50PrUnknown\}$. We have that $X_{Scale50PrUnknown} = X_{Scale50Pr1/15} = \{(p : €100, 1-p : €80), (p : €192.50, 1-p : €5)$ with $p \in \{0.1, \ldots, 1\}\}$ and $X_{Scale10Pr1/3} = \{(p : €20, 1-p : €16), (p : €38.50, 1-p : €1)$ with $p \in \{0.1, \ldots, 1\}\}$. For each BRIS, we assume that the subject has a (complete and transitive) preference relation $\succeq$ on $\{X_{BRIS} \cup €0\}$ and we denote by $\mathbf{c}^* = (c_1^*, \ldots, c_{10}^*)$ a truthful vector of choices for $\succeq$ meaning that c_k^* is $\succeq$-*preferred* to the other one-stage lottery in decision $k \in \{1, \ldots, 10\}$. In each BRIS, two randomizing devices generate the possible payments awarded to the subject and these random payments assign an element in $\{X_{BRIS} \cup €0\}$ to each possible pair of realizations of the randomizing devices. In $Scale10Pr1/3$ and $Scale50Pr1/15$ random payments are two-stage lotteries over $\{X_{Scale10Pr1/3} \cup €0\}$ and $\{X_{Scale50Pr1/15} \cup €0\}$ respectively. Given a submitted list of 10 one-stage lotteries $\mathbf{c} = (c_1, \ldots, c_{10})$, the subject is awarded $(1/10 : c_1, \ldots, 1/10 : c_{10})$ with probability Pr and €0 with probability $1 - Pr$ where Pr equals 1/3 and 1/15 respectively. In $Scale50PrUnknown$ random payments are acts over one-stage lotteries over $\{X_{Scale50PrUnknown} \cup €0\}$. Given a submitted list $\mathbf{c} = (c_1, \ldots, c_{10})$, the subject is awarded $(1/10 : c_1, \ldots, 1/10 : c_{10})$ if she is among the paid subjects and €0 otherwise.

We now specify the necessary and sufficient conditions for the incentive compatibility of our BRIS. Consider $\mathbf{c}^* = (c_1^*, \ldots, c_{10}^*)$, a truthful vector of choices for $\succeq$, and $\mathbf{c}^\circ = (c_1^\circ, \ldots, c_{10}^\circ)$ where c_k° belongs to decision $k \in \{1, \ldots, 10\}$. Given the underlying preferences $\succeq$ on $\{X_{Scale10Pr1/3} \cup €0\}$, the two-stage lottery $2SL_1$ over $\{\mathbf{c}^* \cup €0\}$ dominates the two-stage lottery $2SL_2$ over $\{\mathbf{c}^\circ \cup €0\}$ in $Scale10Pr1/3$ since, for each possible pair of realizations of the randomizing devices, we have that the element in $\{X_{Scale10Pr1/3} \cup €0\}$ assigned by $2SL_1$ is $\succeq$-*preferred* to the element assigned by $2SL_2$.[5] Let $\succeq^e$ denote the (complete and transitive) extension of $\succeq$ over two-stage lotteries over $\{X_{Scale10Pr1/3} \cup €0\}$. $Scale10Pr1/3$ is incentive compatible if and only the extended preference relation $\succeq^e$ respects the dominance relation meaning that $2SL_1 \succeq^e 2SL_2$. Similarly, $Scale50Pr1/15$ (respectively $Scale50PrUnknown$) is incentive compatible if and only the extended preference relation over two-stage lotteries over $\{X_{Scale50Pr1/15} \cup €0\}$ (respectively acts over one-stage lotteries over $\{X_{Scale50PrUnknown} \cup €0\}$) respects the dominance relation. In words, our BRIS are incentive compatible if subjects never choose dominated random payments over one-stage lotteries.

Note that incentive compatibility imposes no structure on subjects' preferences over the set of one-stage lotteries they face in HL's task.

Incentive Effectiveness of BRIS

If subjects' extended preferences $\succeq^e$ respect the dominance relation then BRIS is incentive compatible and subjects systematically treat each of the ten decisions in HL's task as though it were in isolation. The incentive compatibility of BRIS implies that incentive effectiveness increases with the scale of nominal payoffs but is unaffected by the level of the selection probability (as long as it is positive). A cost-effective BRIS is therefore easily identified in large samples: it consists of sizable one-stage lottery outcomes and a tiny probability of real payment. Accordingly, $Scale50Pr1/15$ and $Scale50PrUnknown$ should be equally

[5] Indeed, €0 is $\succeq$-*preferred* to €0 and, for each decision $k \in \{1, \ldots, 10\}$, c_k^* is $\succeq$-*preferred* to c_k°.

more incentive effective than $Scale10Pr1/3$ and, of course, $Scale50PrUnknown$ is the least costly.

Based on the aforementioned arguments, we might conclude that incentive effectiveness is an issue subsidiary to incentive compatibility since among the class of BRIS whose empirical validity is firmly established the cost-effective ones are self-evident. This conclusion however depends on our ability to unequivocally identify the incentive compatible BRIS which is hardly possible in practice. Indeed, previous tests of the assumption that extended preferences respect the dominance relation have been inconclusive in various choice domains (see Azrieli, Chambers, and Healy, 2016, and the references therein), and Brown and Healy (2016) show that even the presentation format of a risk elicitation task might alter the conclusion of the test. The absence of clear experimental evidence on the validity of BRIS means that we would have to collect new evidence on the question by implementing a treatment where all subjects are paid to complete HL's task knowing before completion which of the ten decisions is payoff-relevant (all ten decisions would be shown to avoid framing effects). But even with such a costly experimental design the test of the incentive compatibility of BRIS relies on the assumption that subjects' preferences $\succeq$ are invariant to their experience with one-stage lottery *payoff-relevant* choices, an assumption that lacks empirical support (see, among others, Loomes, Moffatt, and Sugden, 2002).

The extreme difficulty in establishing the incentive compatibility of BRIS implies that its incentive effectiveness should be studied separately. At the present time, large-scale experiments claim (or at least assume) that BRIS is incentive compatible though they usually combine moderate nominal payoffs with a selection probability of at least 10 percent. By assessing the relative impact of nominal payoffs and the selection probability on the incentive effectiveness of BRIS, our experiment could help economists use more cost-effective random payments in large-scale studies.

Note that, though we did not design it to examine the validity of BRIS, our experiment provides *indirect* evidence on the incentive compatibility of BRIS as the latter implies that $Scale50Pr1/15$ and $Scale50PrUnknown$ are equally more incentive effective than $Scale10Pr1/3$.

2.3 Experimental Procedures and Subjects

To recruit the subjects of $Scale50PrUnknown$, we contacted various mailing lists almost exclusively composed of students. First, we contacted the mailing list of 8 experimental laboratories in Germany.[6] Second, we contacted mailing lists at the University of Cologne and posted links at the university's web pages. Email recipients could forward the invitation without invalidating the registration link. Individuals could also register by directly accessing the experiment website.[7] The homepage of the experiment website mentioned prominently that researchers from the Max Planck Institute of Economics in Jena and the University of Cologne were performing the study and provided contact details so that prospective subjects could verify the credibility of the experiment. A German and an English version of the experiment website were available.

Before completing the risk elicitation task, subjects had to register by filling in their name, email address, a chosen username and password. They then received an email with a link to complete their registration.[8] After having completed the risk elicitation task, subjects had to answer a questionnaire to

[6] We gratefully acknowledge the support of the experimental laboratory at the University of Bonn, the University of Cologne, the University of Erfurt, the Humboldt-University of Berlin, the Technical University of Berlin, the Max-Planck-Institute of Economics in Jena, the University of Magdeburg, and the University of Mannheim. All laboratories used the recruitment system ORSEE (Greiner, 2015) to invite their subjects.

[7] 82% of our subjects registered for $Scale50PrUnknown$ after receiving an invitation email, 16% registered following the recommendation of an acquaintance, and the remaining 2% registered via other means.

[8] Multiple registrations with the same email address were prevented. To avoid multiple registrations with different email

gather information on their demographic characteristics like gender, year of birth, education and marital status, etc. (see online Appendix A).[9]

All 8 sessions of the two laboratory treatments were conducted at the Cologne Laboratory for Economic Research (CLER) with 15 subjects in each session. Participants were recruited from the CLER subject pool with the restriction that they did not participate in $Scale50PrUnknown$. Sessions lasted approximately one hour. Laboratory procedures were identical to the Internet ones except that i) the ten lottery-choice decisions were made on a sheet of paper; ii) the lottery outcomes were in Experimental Currency Units and they were converted to Euros at pre-announced conversion rates of 1 ECU equals €1 and €0.2 in $Scale50Pr1/15$ and $Scale10Pr1/3$ respectively; and iii) to comply with the rules at CLER, each laboratory subject received a show-up fee of €2.50. Harrison, Lau, and Rutström (2009) show that the use of non-stochastic show-up fees can generate samples that are more risk averse than would otherwise have been observed. So, similar to the selection bias, the presence of a show-up fee might increase risk aversion in the two laboratory treatments compared to $Scale50PrUnknown$.

In each treatment, subjects randomly selected for payment were informed by email and their earnings were transferred to their bank account.

Samples of Subjects

Table 2 summarises the collected demographic characteristics of our subjects in each incentive treatment separately. For each demographic characteristic we report the number of observations available in the three treatments, except for gender, but for most demographics we only show the highly populated categories (for example, in addition to "married" or "single", subjects could choose either "divorced", or "widowed", or "other" to indicate their marital status).

Though women account for less than 40% of the subjects in $Scale50PrUnknown$, there is a larger fraction of women than men in the laboratory treatments (slightly above 60%). The main reason is that subjects in the laboratory sessions were recruited from a subset of the Internet subject pool with the restriction that they did not take part in $Scale50PrUnknown$. About three quarters of subjects are full-time students in each incentive treatment. Two-thirds of the remaining subjects are non-students in $Scale50PrUnknown$ whereas part-time students—who complete less than 12 credit hours per week of a semester—are the second largest group of subjects in the laboratory treatments (79% and 82% of the remaining subjects in $Scale50Pr1/15$ and $Scale10Pr1/3$ respectively). The presence of 637 non-students in the Internet sample compared to a total of 6 non-students in the laboratory samples implies that the former sample offers a wider range of demographic characteristics than the latter ones. For example, almost 100 subjects in the Internet sample are older than any subject in the laboratory samples, and almost 300 subjects in the Internet sample are employed full time compared to 1 subject in the laboratory samples.

3 Results

Numerous previous studies have established that increased incentives lead to greater risk aversion in HL's task (e.g. Holt and Laury, 2002, 2008). We therefore evaluate the incentive effectiveness of a given BRIS

addresses, we made clear that such attempts would be sanctioned by immediate exclusion from the experiment and all payments. We regularly conducted spot tests and we never had to exclude a subject because of multiple registrations.

[9] Most questions were optional. Though we encouraged subjects to skip questions they were uncomfortable with, we emphasised that answering the full questionnaire in a truthful way would strongly support our work as researchers.

	$Scale50PrUnknown$	$Scale50Pr1/15$	$Scale10Pr1/3$
Number of subjects	3,582	60	60
Females	0.37	0.63	0.63
Age	25.61	24.52	24.40
(N = 3474 | 60 | 60)	(14, 22, 24, 27, 69)	(20, 21, 24, 27, 44)	(19, 22, 24, 26, 43)
Education status ($N = 3,582 \mid 60 \mid 60$)			
Full-time students	0.73	0.72	0.78
Part-time students	0.09	0.22	0.18
Non-students	0.18	0.06	0.04
Marital status ($N = 3,476 \mid 60 \mid 60$)			
Married	0.06	0.02	0.05
Single	0.90	0.98	0.93
In charge of budget decisions ($N = 3,469 \mid 60 \mid 60$)			
Parent(s)	0.11	0.17	0.20
Self	0.82	0.82	0.75
Employment situation ($N = 3,499 \mid 60 \mid 60$)			
Full-time employment	0.08	0.00	0.02
Part-time employment	0.28	0.40	0.50
University employment	0.04	0.02	0.00
Only studying	0.55	0.45	0.38
	– – – – – – – – Students – – – – – – – –		
Major field of study ($N = 2,914 \mid 56 \mid 57$)			
Business administration	0.25	0.36	0.44
Economics	0.19	0.16	0.16
MNE	0.19	0.11	0.07
SSH	0.33	0.36	0.27
Other field	0.04	0.02	0.07
Pays for tuition and expenses ($N = 2,900 \mid 56 \mid 57$)			
Parent(s)	0.33	0.32	0.25
Self	0.20	0.14	0.25
Self and parent(s)	0.32	0.41	0.40

Notes: Except for the number of subjects and age, all entries are percentages. For the variable *Age*, the first row reports the average while the second row reports the minimum, the 1st quartile, the median, the 3rd quartile, and the maximum. For the variable *Employment situation*, the options were: full-time employed; part-time employed; self-employed; unemployed; employed at the university; only student; or other. In *Scale50PrUnknown*, subjects who indicated that they were either self-employed or unemployed were wrongly recorded as part-time employed. The 14 collected fields of study are grouped into 5 categories: business administration; economics; mathematics, natural sciences, and engineering (MNE); social sciences and humanities (SSH); and other field.

Table 2: Demographic characteristics of subjects

by inferring the subjects' degree of risk aversion from their choices between safe and risky one-stage lotteries in the respective incentive treatment. Our main statistical analysis compares the measured degrees of risk aversion in the three incentive treatments to clarify the relative impact of the scale of

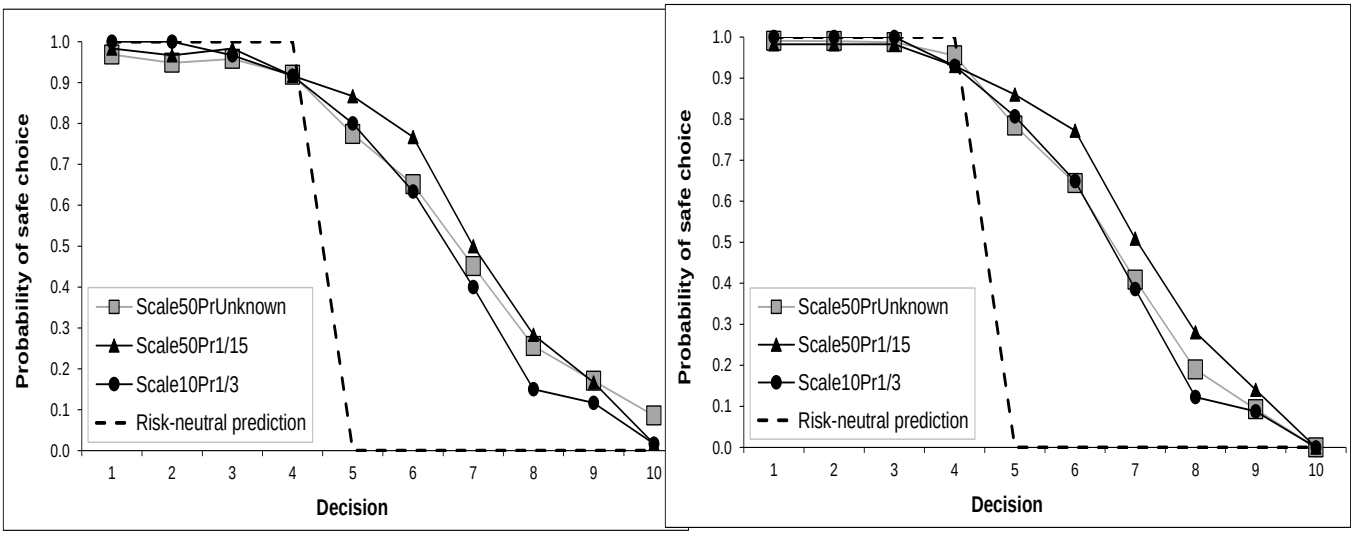
nominal payoffs and the level of the selection probability on the effectiveness of incentives.

As a complementary analysis, we report structural estimation results to investigate whether errors in the formation or the execution of risk preferences differ in the three incentive treatments.

3.1 Demographic and Treatment Effects on Risk Aversion

Following Holt and Laury (2002), we use in this subsection the total number of "safe choices", i.e. the number of times the safe lottery is chosen, as an indicator of risk aversion. We rely on this ordinal measure of risk aversion to avoid assuming a specific model of risk attitudes. Most deterministic theories of decision under risk predict that either the subject always chooses the risky lottery or she chooses the safe lottery in the first decision rows and then switches over to the risky lottery at the latest in the tenth decision row. The more risk-averse the individual the later she switches to the risky lottery with a risk-neutral individual choosing the safe lottery for the first four rows and the risky lottery thereafter. We observe that the large majority of choice sequences are consistent with the standard deterministic prediction especially in the laboratory treatments. The percentage of inconsistent sequences of choices is about 5% in each of the laboratory treatments and it is about 14% in $Scale50PrUnknown$.[10]

Figure 1 plots the proportion of safe choices for each of the ten decisions in the different treatments, separately for all sequences of choices (left panel) and for the subset of consistent sequences of choices (right panel). Each series of choice frequencies lies distinctly to the right of the risk-neutral prediction showing a clear tendency towards risk-averse behaviour among subjects in each incentive treatment. A comparison of the three series suggests that subjects are at least as risk-averse in $Scale50PrUnknown$ as in $Scale10Pr1/3$ and that risk aversion is largest in treatment $Scale50Pr1/15$. We also observe that the series of safe choices in the two panels are similar in the laboratory treatments whereas consistent choice sequences exhibit less risk aversion than all choice sequences in $Scale50PrUnknown$.



(a) All choice sequences

(b) Consistent choice sequences

Figure 1: Proportion of safe choices in each decision

[10] A probit regression shows that the frequency of inconsistent sequences of choices in each of the laboratory treatments is significantly smaller than in the Internet treatment (p-values = 0.035).

To study the effects of experimental conditions on risk attitudes we need to account for the observed heterogeneity in our statistical analysis as the Internet sample offers a wider range of demographic characteristics than the laboratory samples. We therefore estimate OLS models of the number of safe choices that condition on individual characteristics and treatments. Table 3 reports the results from three models of safe choices for all choice sequences (left panel) and for the subset of consistent choice sequences (right panel). We rely on the full sample of subjects in models 1 and 2 whereas model 3 relies on the restricted sample of students. Model 1 includes treatment dummies and controls for gender, age, employment and marital status, and whether the subject is in charge of budgeting or not. Model 2 enables us to distinguish between estimated numbers of safe choices for subjects with different education status (full-time students, part-time students and non-students). In addition to the explanatory variables included in model 1, model 3 controls for the duration (number of semesters) and level of education (undergraduate or graduate), the major field of study, and whether the student is primarily responsible for the payment of living expenses or not. For a given regression, subjects with missing values for the included variables are omitted. Table 1 in online Appendix B reports the estimates when the OLS model contains only an intercept and treatment dummies.[11]

When considering all choice sequences the number of safe choices estimated at mean demographic values equals at least 6 and is significantly larger than 4—the number of safe choices a risk-neutral individual makes—in each incentive treatment (p-values < 0.01). This observation confirms the strong tendency towards risk-averse behavior among subjects. For the full sample of subjects, risk aversion in $Scale50PrUnknown$ is larger than in $Scale10Pr1/3$ while it is lower than in $Scale50Pr1/15$ though none of the treatment differences is statistically significant. Interestingly enough, we find that treatment effects on risk aversion vary significantly with the education status of subjects. For full-time students, risk aversion in $Scale50PrUnknown$ is substantially lower than in $Scale50Pr1/15$ while it is larger than in $Scale10Pr1/3$, and treatment effects are statistically significant: i) the estimated number of safe choices is significantly higher in $Scale50Pr1/15$ than in $Scale50PrUnknown$ (resp. $Scale10Pr1/3$) at the 5% level (resp. 1% level); and ii) the estimated number of safe choices is significantly lower in $Scale10Pr1/3$ than in $Scale50PrUnknown$ at the 10 percent level. For part-time students, risk aversion in $Scale50PrUnknown$ is substantially larger than in $Scale50Pr1/15$ while it is lower than in $Scale10Pr1/3$, but we never reject the hypothesis that the estimated number of safe choices is the same in each incentive treatment at any conventional significance level. For non-students, risk aversion in $Scale50PrUnknown$ is lower than in $Scale50Pr1/15$ and it is substantially lower than in $Scale10Pr1/3$, but we never reject the hypothesis that the estimated number of safe choices is the same in each incentive treatment at any conventional significance level (note that there are few non-students in our laboratory samples).

We also find that demographic effects on risk attitudes vary significantly with the education status of subjects. First, female full-time students are significantly more risk-averse than male full-time students at the 10% level while we never reject the hypothesis that female and male make the same number of safe choices both for part-time students and non-students (p-values > 0.1). Second, we find a significant negative effect on risk aversion from age for non-students at the 5% level but no significant effect for part-time or full-time students (p-values > 0.1). Complementary results reported in Table 2 in online Appendix B show that the significant age effect is largely the consequence of less risk averse choices made by non-students older than most students.

Finally, we find that the regression results for consistent choice sequences contrast with those for

[11] Estimation results based on Ordered Probit regressions lead to the same qualitative conclusions (results are available upon request).

	All choice sequences			Consistent choice sequences		
	All participants		Students	All participants		Students
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Constant	6.583***	6.281***	6.167***	6.562***	6.108***	5.965***
	(0.204)	(0.338)	(0.418)	(0.203)	(0.342)	(0.416)
Scale50Pr1/15	0.215	0.529**	0.591**	0.335	0.568**	0.654**
	(0.242)	(0.265)	(0.269)	(0.247)	(0.271)	(0.278)
Scale10Pr1/3	−0.228	−0.396*	−0.416*	−0.125	−0.253	−0.241
	(0.219)	(0.237)	(0.237)	(0.214)	(0.238)	(0.239)
Female	0.120*	0.134*	0.149*	0.143**	0.129*	0.114
	(0.066)	(0.075)	(0.080)	(0.066)	(0.075)	(0.079)
Age	−0.015*	−0.002	−0.008	−0.021***	−0.002	−0.007
	(0.008)	(0.014)	(0.019)	(0.008)	(0.014)	(0.019)
Part-time student		0.573	0.353		−0.410	−0.118
		(0.901)	(1.012)		(0.956)	(1.002)
PT student x *Scale50Pr1/15*		−1.322**	−1.410*		−1.064	−1.298*
		(0.671)	(0.727)		(0.703)	(0.766)
PT student x *Scale10Pr1/3*		0.648	0.338		0.322	0.281
		(0.546)	(0.520)		(0.540)	(0.541)
PT student x *Female*		0.004	−0.059		0.064	0.040
		(0.235)	(0.243)		(0.237)	(0.240)
PT student x *Age*		−0.021	−0.004		0.021	0.023
		(0.034)	(0.041)		(0.036)	(0.041)
Non-student		0.288			0.780	
		(0.673)			(0.667)	
Non-student x *Scale50Pr1/15*		−0.533			−0.131	
		(0.529)			(0.508)	
Non-student x *Scale10Pr1/3*		1.606			1.608	
		(1.044)			(1.187)	
Non-student x *Female*		−0.112			0.133	
		(0.198)			(0.203)	
Non-student x *Age*		−0.021			−0.038**	
		(0.019)			(0.018)	
Controls for budgeting, marital status & employment status	Yes	Yes	Yes	Yes	Yes	Yes
Controls for duration and level of education, field of studies & payment of living expenses	No	No	Yes	No	No	Yes
R^2	0.009	0.016	0.028	0.013	0.020	0.027
Observations	3,478	3,478	2,750	2,985	2,985	2,381

Notes: Heteroskedasticity-robust standard errors are in parentheses. We systematically reject the null hypothesis that all the slope coefficients are zero (p-values < 0.01). *PT student* (resp. *Non-student*) x *MVI* refers to the interaction term between the dummy variable *Part-time student* (resp. *Non-student*) and *MVI*, one of the main variables of interest. * (10%); ** (5%); and *** (1%) significance level.

Table 3: OLS estimates of the number of safe choices

all choice sequences mainly in $Scale50PrUnknown$. First, for full-time students the difference in risk aversion between $Scale50Pr1/15$ and $Scale50PrUnknown$ increases whereas the difference between $Scale50PrUnknown$ and $Scale10Pr1/3$ decreases and becomes statistically insignificant. Second, part-time students are less (more) risk-averse than full-time students when considering only consistent choice sequences (all choice sequences) though the difference is never statistically significant at any conventional level.

In a nutshell, we cannot reject the hypothesis that the three BRIS are equally incentive effective

for part-time students and non-students. By contrast, we observe that the scale of nominal payoffs is significantly more incentive effective than the level of the selection probability for full-time students.

3.2 Structural Estimation Results

We report structural estimation results on risk preferences over one-stage lotteries with a particular focus on errors in the formation and the execution of risk preferences. For our inferences to be meaningful, we have to assume that BRIS is incentive compatible (though the descriptive results do not fully support the assumption as the impact of the selection probability on risk aversion is non-negligible).

The deterministic and stochastic components of our structural econometric model are as follows. First, we assume that subjects assess the relative value of the safe lottery relative to the risky lottery by comparing their expected utilities, i.e. the deterministic component is expected utility (EU). Indeed, experimental evidence shows that the list format of HL's task induces subjects to treat the ten decisions as one large decision which leads behavior to become more consistent with the reduction of compound lotteries (see Brown and Healy, 2016, and the references therein). And the reduction of compound lotteries combined with the incentive compatibility of BRIS implies that risk preferences must be consistent with EU.[12] Second, the stochastic component of choice under risk consists of two complementary elements. Fechner errors capture stochastic variations in the formation of risk preferences whereas trembles capture stochastic variations in the execution of preferences. Sensitivities to the difference in expected utilities might vary with the characteristics of BRIS, especially the scale of nominal payoffs. On the other hand, we expect trembles to pick up differences in the implementation mode as subjects are more likely to temporarily lose concentration in the Internet environment than in a controlled laboratory setting.

Online Appendix C presents the EU decision model in HL's task along with the specifics of the econometric model. The risk behavior of an EU maximizer is characterized by i) her ratio of utilities r such that, absent errors, if $r \in \left(\frac{d}{10}, \frac{d+1}{10}\right)$ then the safe lottery is chosen in the first $d \in \{1, \ldots, 9\}$ decisions and if $r \in \left(0, \frac{1}{10}\right)$ then the risky lottery is always chosen; ii) her sensitivity to the difference in the expected utilities of the two lotteries $k \in (0, \infty)$ such that, without trembles and for a given difference in expected utilities, the larger k the smaller the probability of mistakenly choosing the less preferred lottery; and iii) her tremble probability $w \in [0, 1]$ which corresponds to a constant probability of choosing completely at random. The most general specification of the econometric model allows for all three parameters to vary both with observable characteristics (*observed heterogeneity*) and with unobservable characteristics (*unobserved heterogeneity*). Though the general specification is identified in theory, the results of a simulation analysis indicate that allowing for unobserved heterogeneity in k or w hinders the empirical identification of the econometric model (see online Appendix C for details). Therefore, every specification of the econometric model that we consider allows for unobserved heterogeneity in r only.

For the sake of brevity, this section discusses the estimation results of restricted econometric models where k is homogeneous across subjects and treatments. Online Appendix D discusses the results of structural models where k varies with the experimental condition and the demographic characteristics of subjects. The fit of some of the regressions improves with heterogeneous k according to likelihood-ratio tests, but estimates of the ratio of utilities are (almost) identical and there are no significant demographic or treatment effects on k. The latter observation is likely to be driven by the fact that in any subsample of subjects the estimated k is large.

[12] We also agree with Abdellaoui, Driouchi, and L'Haridon (2011) that the simple frame provided by HL's task is unsuited to discriminate between probability weighting and outcome weighting and essentially provides an experimental measurement of risk attitudes under EU.

3.2.1 Parameter estimates at the treatment level

Table 4 shows the estimated parameters when the econometric model contains only an intercept and treatment dummies (again, k is assumed to be homogeneous across treatments). Table 1 in online Appendix D shows the estimated parameters for the full set of covariates.

	r	k	w
Constant	0.706***	17.689***	0.063***
	(0.004)	(1.015)	(0.002)
$Scale50Pr1/15$	0.023	-	-0.049***
	(0.029)		(0.008)
$Scale10Pr1/3$	-0.033	-	-0.049***
	(0.036)		(0.005)
Standard Deviation	1.021***	-	-
	(0.015)		

Notes: The number of observations is 3,702 and the log-likelihood is -11,974.40.
*** Significant at the 1% level.

Table 4: Estimated parameters for the model with treatment dummies

The results confirm a strong tendency towards risk-averse behavior among subjects since the estimated median ratio of utilities equals at least 0.67 in each treatment. Though treatment differences are not significant, the median subject is least risk averse in $Scale10Pr1/3$ and most risk averse in $Scale50Pr1/15$. In line with the OLS regression results, we observe that the scale of nominal payoffs has more impact on the incentive effectiveness of BRIS than the level of the selection probability.[13] Apparent from the results is also a considerable heterogeneity in risk preferences. While the overwhelming majority of subjects are risk averse in each treatment, 8 to 10 percent of subjects (depending on the treatment) are estimated to be risk neutral, and 9 to 14 percent are risk loving. Remarkably, the distribution of risk aversion has a thick right tail: at least 25 percent of the subjects in each treatment are inclined to pick the safe lottery in decision 8 or 9 where they forgo 59€ and 76€ in expected terms, respectively.

Regarding the stochastic component of choice under risk, the estimated (homogeneous) sensitivity to the difference in expected utilities is rather high.[14] When k is allowed to vary with the incentive treatment, complementary results show that treatment effects are non-significant (Table 4 in online Appendix D). Contrary to what we conjectured, the characteristics of BRIS do not significantly affect the stochastic variations in the formation of risk preferences. On the other hand, we confirm that the implementation mode has a significant effect on the stochastic variations in the execution of preferences. The median subject in $Scale50PrUnknown$ chooses completely at random with probability 6.3 percent compared to only 1.4 percent for the median laboratory subject, and this difference is statistically significant in a strong sense. Thus, the estimated trembles are significantly larger in the Internet environment than in the laboratory setting.

In sum, our structural estimation results show that the characteristics of BRIS do not significantly affect error rates and that more decision errors on the Internet ensue from a higher propensity of subjects to loose concentration.

[13] Our estimation approach follows von Gaudecker, van Soest, and Wengström (2012) who find no effect of the implementation mode on risk preferences when controlling for the difference in demographics between the Internet and the laboratory.

[14] A value of $k = 17.7$ implies that, absent trembles, the median subject picks the safe (risky) lottery with probability 0.54 (0.45) at decision 7, with probability 0.97 (0.95) at decision 6 (8), and with a probability larger than 0.99 at each decision $d < 6$ ($d > 8$).

3.2.2 Parameter estimates in subsamples of subjects

Table 5 presents the median predicted utility ratios and trembling probabilities stratified by the main demographics. For each sub-sample of subjects, the table shows the predicted medians, the 90% confidence intervals of the parameters, and the size of the sub-sample. The latter is determined by the number of observations included in models 2 and 3 of Section 3.1 respectively. The presentation conveys the total effect of varying the demographic variable, taking into account possible correlations between variables. Table 2 in online Appendix D contains the median predicted parameters stratified by *all* demographics.

Table 5: Median estimated parameters stratified by the main demographics

		Scale50PrUnknown		*Scale50Pr1/15*		*Scale10Pr1/3*	
		r	w	r	w	r	w
All	Full-time students	0.710	0.058	0.766	0.011	0.652	0.015
	$N = (2,474 \mid 43 \mid 47)$	(0.686,0.737)	(0.037,0.097)	(0.739,0.789)	(0.007,0.027)	(0.618,0.670)	(0.009,0.035)
	Part-time students	0.708	0.041	0.639	0.016	0.724	0.000
	$N = (319 \mid 13 \mid 11)$	(0.600,0.790)	(0.020,0.169)	(0.504,0.822)	(0.004,0.083)	(0.649,0.846)	(0.000,0.000)
	Non-students	0.692	0.091	0.683	0.000	0.849	0.000
	$N = (565 \mid 4 \mid 2)$	(0.625,0.749)	(0.000,0.145)	(0.637,0.695)	(0.000,0.000)	(0.845,0.854)	(0.000,0.000)
Male	Full-time students	0.710	0.037	0.766	0.007	0.642	0.015
	$N = (1,491 \mid 18 \mid 19)$	(0.666,0.723)	(0.036,0.063)	(0.747,0.777)	(0.007,0.017)	(0.600,0.666)	(0.007,0.023)
	Part-time students	0.699	0.032	0.618	0.006	0.715	0.000
	$N = (204 \mid 4 \mid 3)$	(0.537,0.766)	(0.019,0.153)	(0.504,0.628)	(0.004,0.009)	(0.688,0.724)	(0.000,0.000)
	Non-students	0.692	0.093				
	$N = (391 \mid 0 \mid 0)$	(0.623,0.748)	(0.000,0.146)				
Female	Full-time students	0.726	0.059	0.761	0.016	0.658	0.021
	$N = (983 \mid 25 \mid 28)$	(0.703,0.738)	(0.057,0.101)	(0.739,0.790)	(0.009,0.027)	(0.618,0.682)	(0.012,0.035)
	Part-time students	0.726	0.100	0.644	0.018	0.731	0.000
	$N = (115 \mid 9 \mid 8)$	(0.651,0.869)	(0.062,0.340)	(0.582,0.822)	(0.012,0.083)	(0.649,0.846)	(0.000,0.000)
	Non-students	0.693	0.088	0.683	0.000	0.849	0.000
	$N = (174 \mid 4 \mid 2)$	(0.625,0.752)	(0.000,0.142)	(0.637,0.695)	(0.000,0.000)	(0.845,0.854)	(0.000,0.000)
Youngest 25%	Full-time students	0.711	0.058	0.767	0.011	0.646	0.015
	$N = (829 \mid 15 \mid 23)$	(0.688,0.738)	(0.037,0.098)	(0.747,0.790)	(0.007,0.027)	(0.618,0.670)	(0.009,0.025)
	Part-time students	0.726	0.077	0.649	0.017	0.750	0.000
	$N = (101 \mid 6 \mid 3)$	(0.639,0.778)	(0.028,0.169)	(0.623,0.715)	(0.008,0.025)	(0.737,0.777)	(0.000,0.000)
	Non-students	0.716	0.078	0.695	0.000	0.845	0.000
	$N = (159 \mid 1 \mid 1)$	(0.670,0.760)	(0.000,0.104)	(0.695,0.695)	(0.000,0.000)	(0.845,0.845)	(0.000,0.000)
Middle 50%	Full-time students	0.710	0.058	0.766	0.011	0.669	0.014
	$N = (1,229 \mid 19 \mid 14)$	(0.679,0.726)	(0.037,0.097)	(0.723,0.780)	(0.007,0.027)	(0.600,0.682)	(0.007,0.025)
	Part-time students	0.703	0.036	0.642	0.019	0.724	0.000
	$N = (167 \mid 4 \mid 6)$	(0.543,0.767)	(0.022,0.153)	(0.613,0.822)	(0.004,0.083)	(0.715,0.846)	(0.000,0.000)
	Non-students	0.694	0.092	0.683	0.000	0.854	0.000
	$N = (272 \mid 2 \mid 1)$	(0.654,0.729)	(0.000,0.119)	(0.676,0.689)	(0.000,0.000)	(0.854,0.854)	(0.000,0.000)
Oldest 25%	Full-time students	0.709	0.057	0.760	0.009	0.651	0.024
	$N = (416 \mid 9 \mid 10)$	(0.665,0.725)	(0.036,0.096)	(0.739,0.789)	(0.007,0.018)	(0.623,0.665)	(0.013,0.035)
	Part-time students	0.680	0.030	0.582	0.012	0.669	0.000
	$N = (51 \mid 3 \mid 2)$	(0.604,0.850)	(0.016,0.230)	(0.504,0.619)	(0.004,0.018)	(0.649,0.688)	(0.000,0.000)
	Non-students	0.643	0.124	0.637	0.000		
	$N = (134 \mid 1 \mid 0)$	(0.605,0.679)	(0.077,0.164)	(0.637,0.637)	(0.000,0.000)		
Only studying	Full-time students	0.711	0.052	0.772	0.008	0.669	0.014
	$N = (1,727 \mid 22 \mid 21)$	(0.710,0.737)	(0.037,0.083)	(0.766,0.790)	(0.007,0.016)	(0.652,0.682)	(0.009,0.021)
	Part-time students	0.712	0.028	0.654	0.016	0.737	0.000
	$N = (117 \mid 5 \mid 2)$	(0.685,0.778)	(0.019,0.093)	(0.613,0.715)	(0.004,0.025)	(0.724,0.750)	(0.000,0.000)

Continued on next page

Table 5: Continued

		Scale50PrUnknown		*Scale50Pr1/15*		*Scale10Pr1/3*	
		r	w	r	w	r	w
	Non-students	0.660	0.000				
	$N = (5 \mid 0 \mid 0)$	(0.629,0.694)	(0.000,0.000)				
Full-time job	Full-time students	0.613	0.144				
	$N = (12 \mid 0 \mid 0)$	(0.589,0.631)	(0.141,0.213)				
	Part-time students	0.531	0.161				
	$N = (18 \mid 0 \mid 0)$	(0.471,0.771)	(0.090,0.432)				
	Non-students	0.698	0.094			0.845	0.000
	$N = (244 \mid 0 \mid 1)$	(0.621,0.722)	(0.072,0.136)			(0.845, 0.845)	(0.000, 0.000)
Part-time job	Full-time students	0.689	0.088	0.761	0.018	0.643	0.024
	$N = (634 \mid 17 \mid 22)$	(0.687,0.716)	(0.062,0.136)	(0.747,0.772)	(0.011,0.027)	(0.627,0.658)	(0.015,0.035)
	Part-time students	0.703	0.046	0.628	0.018	0.720	0.000
	$N = (165 \mid 7 \mid 8)$	(0.675,0.774)	(0.029,0.164)	(0.582,0.822)	(0.008,0.083)	(0.649,0.777)	(0.000,0.000)
	Non-students	0.692	0.110				
	$N = (145 \mid 0 \mid 0)$	(0.625,0.724)	(0.078,0.158)				
University job	Full-time students	0.644	0.027	0.723	0.008		
	$N = (48 \mid 1 \mid 0)$	(0.621,0.662)	(0.027,0.043)	(0.723,0.723)	(0.008,0.008)		
	Part-time students	0.794	0.000				
	$N = (8 \mid 0 \mid 0)$	(0.768,0.851)	(0.000,0.000)				
	Non-students	0.660	0.000				
	$N = (87 \mid 0 \mid 0)$	(0.636,0.673)	(0.000,0.000)				
Other job	Full-time students	0.676	0.044	0.739	0.009	0.618	0.012
	$N = (53 \mid 3 \mid 4)$	(0.662,0.692)	(0.031,0.070)	(0.739,0.739)	(0.009,0.009)	(0.600,0.619)	(0.007,0.012)
	Part-time students	0.654	0.075	0.504	0.004	0.846	0.000
	$N = (11 \mid 1 \mid 1)$	(0.604,0.732)	(0.029,0.132)	(0.504,0.504)	(0.004,0.004)	(0.846,0.846)	(0.000,0.000)
	Non-students	0.737	0.076	0.683	0.000	0.854	0.000
	$N = (84 \mid 4 \mid 1)$	(0.626,0.762)	(0.052,0.146)	(0.637,0.695)	(0.000,0.000)	(0.854,0.854)	(0.000,0.000)
Field of studies:	Full-time students	0.685	0.029	0.748	0.009	0.628	0.016
	$N = (591 \mid 16 \mid 18)$	(0.649,0.712)	(0.016,0.062)	(0.716,0.776)	(0.003,0.018)	(0.554,0.653)	(0.004,0.028)
Business	Part-time students	0.691	0.037	0.653	0.027	0.663	0.000
	$N = (75 \mid 4 \mid 5)$	(0.503,0.772)	(0.009,0.264)	(0.572,0.722)	(0.004,0.041)	(0.624,0.677)	(0.000,0.000)
Field of studies:	Full-time students	0.715	0.023	0.774	0.004	0.636	0.010
	$N = (435 \mid 8 \mid 8)$	(0.680,0.738)	(0.012,0.054)	(0.749,0.792)	(0.003,0.011)	(0.600,0.667)	(0.007,0.023)
Economics	Part-time students	0.608	0.021	0.556	0.006	0.551	0.000
	$N = (57 \mid 1 \mid 1)$	(0.432,0.704)	(0.006,0.284)	(0.556,0.556)	(0.006,0.006)	(0.551,0.551)	(0.000,0.000)
Field of studies:	Full-time students	0.728	0.044	0.773	0.008	0.664	0.018
	$N = (450 \mid 4 \mid 4)$	(0.687,0.752)	(0.019,0.084)	(0.715,0.804)	(0.005,0.012)	(0.636,0.684)	(0.012,0.024)
MNE	Part-time students	0.749	0.035	0.683	0.016		
	$N = (57 \mid 2 \mid 0)$	(0.546,0.839)	(0.000,0.225)	(0.661,0.704)	(0.015,0.017)		
Field of studies:	Full-time students	0.722	0.097	0.776	0.026	0.647	0.038
	$N = (769 \mid 15 \mid 11)$	(0.687,0.748)	(0.049,0.185)	(0.754,0.803)	(0.013,0.041)	(0.626,0.693)	(0.012,0.071)
SSH	Part-time students	0.759	0.065	0.694	0.037	0.723	0.000
	$N = (103 \mid 4 \mid 4)$	(0.557,0.837)	(0.014,0.241)	(0.406,0.736)	(0.013,0.045)	(0.705,0.763)	(0.000,0.000)
Field of studies:	Full-time students	0.706	0.034			0.649	0.010
	$N = (93 \mid 0 \mid 4)$	(0.664,0.727)	(0.017,0.069)			(0.643 0.655)	(0.008 0.016)
Other	Part-time students	0.664	0.015	0.576	0.002		
	$N = (10 \mid 1 \mid 0)$	(0.444,0.706)	(0.004,0.077)	(0.576,0.576)	(0.002,0.002)		

The results confirm the presence of treatment effects in the sample of full-time students. The estimated ratio of utilities is significantly larger in $Scale50Pr1/15$ ($r = 0.766$) than in $Scale10Pr1/3$ ($r = 0.652$) at the 5% level and it is significantly larger than in $Scale50PrUnknown$ ($r = 0.710$) at the 10% and 5% levels for models 2 and 3 respectively. Differences between $Scale50PrUnknown$ and $Scale10Pr1/3$ are not significant at any conventional level. For part-time students and non-students the estimated ratio

is largest in $Scale10Pr1/3$ and lowest in $Scale50Pr1/15$, but differences are non-significant. Regardless of the sample, the trembling probability is significantly larger in the Internet treatment than in the two laboratory treatments.

Turning to demographic effects, our key results are as follows. First, females are significantly more risk-averse in the sample of full-time students only, and female students (full-time or part-time) commit significantly more decision errors than male students. Second, the negative effect of age on risk aversion is restricted to non-students. Third, a job at the university (full-time job) is associated with a significantly lower risk aversion for full-time (part-time) students. In addition, a full-time job outside the university is associated with a substantially higher propensity to commit errors for both full-time and part-time students (the estimated trembling probability in $Scale50PrUnknown$ is 14.4 and 16.1 percent respectively). Fourth, full-time students of business administration are the least risk averse, and full-time students of economics are the least error prone. Fifth, a few structural estimation results on risk aversion contrast with the OLS regression results. Marriage is associated with a significantly higher risk aversion for part-time students, but the trembling probability for this sub-group is extremely high (29.1 percent in $Scale50PrUnknown$). Full-time students with a part-time job are significantly more risk-averse, but also significantly more error prone. Sixth, we find that the trembling probability in $Scale50PrUnknown$ is significantly higher among non-students (9.1 percent), full-time students of the social sciences or humanities (9.7 percent), and full-time students not in charge of budget decisions (8.2 percent).[15]

Finally, we note that only a small part of the overall heterogeneity in risk preferences can be accounted for by observed covariates. For instance, in $Scale50PrUnknown$ 90 percent of the values of r predicted by a model which only considers observed heterogeneity account for less than 20 percent of the distribution of r when unobserved heterogeneity is also considered (see online Appendix D for details).

4 Conclusion

Our experimental evidence clarifies the relative impact of nominal payoffs and the between-subjects selection probability on the incentive effectiveness of BRIS. Both the descriptive and structural estimation results show that for the majority of student subjects (full-time students) nominal payoffs have a bigger impact on the incentive effectiveness of BRIS than the probability of payment and that the latter hardly dilutes the effects of monetary incentives. For the smaller share of student subjects (part-time students) and for non-students none of the two components of BRIS has a significant impact on its incentive effectiveness.

To further assess the dilution effect of the selection probability on monetary incentives, we compare the estimated CRRA indices of students in our incentive treatments along with those in the 1× and 10× treatments of Harrison, Johnson, McInnes, and Rutström (2005, HJMR hereafter) while controlling for available demographics (see online Appendix B.3 for details). HJMR recruited their subjects from a convenience pool of students and paid all of them to complete HL's risk elicitation task. We find that risk aversion in HJMR's 1×—where the scale of nominal payoffs is 1—is significantly lower than risk aversion in $Scale50PrUnknown$ and $Scale50Pr1/15$ at the 1% level and it is significantly lower than risk aversion in $Scale10Pr1/3$ at the 10% level. On the other hand, we cannot reject the null hypothesis that risk aversion in HJMR's 10× is identical to the one estimated in any of our incentive treatments at

[15] von Gaudecker, van Soest, and Wengström (2011) find that younger, more educated and more wealthy subjects, as well as males commit fewer errors.

any conventional significance level. At mean demographic values, risk aversion in $Scale10Pr1/3$ is only slightly lower than in HJMR's $10\times$ which confirms that the selection probability barely dilutes the effects of monetary incentives.[16]

With regard to decision errors, the key structural estimation results are that female students commit significantly more decision errors than male students, that the characteristics of BRIS do not significantly affect error rates, and that more decision errors on the Internet ensue from a higher propensity of subjects to loose concentration. Whenever we are willing to impose identifying assumptions, risky choices on the Internet and in the laboratory can be combined to make valuable inferences about the impact of random payments on the error and risk parameters.

Most importantly, our results suggest that in simple choice tasks cost-effective BRIS consist of large nominal payoffs and a small selection probability. Despite a potential recruitment bias and the absence of a participation fee, $Scale50PrUnknown$ is significantly more incentive effective than $Scale10Pr1/3$ for the bulk of our samples while the experimenter's expected costs are clearly lower in the Internet treatment than in the laboratory treatment. The fact that nominal payoffs are more salient than the selection probability *at the choice stage* is probably crucial for this methodological implication of our results. Experimental economists who aim at strong effects of monetary incentives in large-scale studies should use cost-effective BRIS and they should ensure the saliency of nominal payoffs in their setting.

References

Abdellaoui, M., A. Driouchi, and O. L'Haridon (2011): "Risk Aversion Elicitation: Reconciling Tractability and Bias Minimization," *Theory and Decision*, 71, 63–80.

Azrieli, Y., C. P. Chambers, and P. J. Healy (2016): "Incentives in Experiments: A Theoretical Analysis," The Ohio State University working paper.

Baltussen, G., T. Post, M. J. van den Assem, and P. P. Wakker (2012): "Random Incentive Systems in a Dynamic Choice Experiment," *Experimental Economics*, 15, 418–443.

Brown, A., and P. J. Healy (2016): "Separated Decisions," The Ohio State University working paper.

CareerBuilder (2012): "Referral Madness: How Employee Referral Programs Turn Good Employees Into Great Recruiters and Grow Your Bottom Line," CareerBuilder e-Book.

CIPD (2013): "Resourcing and Talent Planning," Annual Survey Report of the Chartered Institute of Personnel and Development in partnership with Hays.

Cleave, B. L., N. Nikiforakis, and R. Slonim (2013): "Is there selection bias in laboratory experiments? The case of social and risk preferences," *Experimental Economics*, 16, 372–82.

Cohn, A., J. Engelmann, E. Fehr, and M. Maréchal (2015): "Evidence for Countercyclical Risk Aversion: An Experiment with Financial Professionals," *American Economic Review*, 105, 860–85.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010): "Are Risk Aversion and Impatience Related to Cognitive Ability?," *American Economic Review*, 100, 1238–60.

[16]Though we acknowledge that there are differences in the procedures and samples of our experiment and HJMR, the comparison is still meaningful since i) both experiments use the list format of HL's task and ii) we focus on our sample of students while controlling for several observable characteristics (age, gender, level of education and major field of study).

Drehmann, M., J. Oechssler, and A. Roider (2005): "Herding and Contrarian Behavior in Financial Markets: An Internet Experiment," *American Economic Review*, 95, 1403–1426.

Gajic, A., D. Cameron, and J. Hurley (2012): "The Cost-Effectiveness of Cash Versus Lottery Incentives for a Web-Based, Stated-Preference Community Survey," *European Journal of Health Economics*, 13, 789–99.

Greiner, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1, 114–25.

Haisley, E., C. Cryder, G. Loewenstein, and K. Volpp (2008): "The Appeal of Lotteries and Their Use in Incentive Design," Society for Judgment and Decision Making Preconference: Using Human Nature to Improve Human Life, Chicago.

Harrison, G., E. Johnson, M. McInnes, and E. Rutström (2005): "Risk Aversion and Incentive Effects: Comment," *American Economic Review*, 95, 897–901.

Harrison, G., M. Lau, and E. Rutström (2007): "Estimating Risk Attitudes in Denmark: A Field Experiment," *Scandinavian Journal of Economics*, 109, 341–368.

Harrison, G., M. Lau, and M. Williams (2002): "Estimating Individual Discount Rates in Denmark: A Field Experiment," *American Economic Review*, 92, 1606–17.

Harrison, G. W., M. I. Lau, and E. E. Rutström (2009): "Risk attitudes, randomization to treatment, and self-selection into experiments," *Journal of Economic Behavior & Organization*, 70, 498–507.

Holt, C., and S. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), 1644–1655.

——— (2005): "Risk Aversion and Incentive Effects: New Data without Order Effects," *American Economic Review*, 95(3), 902–904.

——— (2008): "Further Reflections on the Reflection Effect," in *Risk Aversion in Experiments, Research in Experimental Economics, Volume 12*, ed. by J. Cox, and G. Harrison, pp. 405–440. Emerald Group Publishing Limited.

Laury, S. (2012): "Pay One or Pay All: Random Selection of One Choice for Payment," Discussion Paper, Andrew Young School, Georgia State University.

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.

Maeda, A. (2008): "Optimal Lottery Design for Public Financing," *Economic Journal*, 118, 1698–1718.

von Gaudecker, H., A. van Soest, and E. Wengström (2011): "Heterogeneity in Risky Choice Behavior in a Broad Population," *American Economic Review*, 101, 664–694.

——— (2012): "Experts in Experiments," *Journal of Risk and Uncertainty*, 45, 159–190.

Supplementary material for

Pay Few Subjects but Pay Them Well: Cost-Effectiveness of Random Incentive Systems

C. March, A. Ziegelmeyer, B. Greiner, and R. Cyranek

[For Online Publication]

The following appendices are for the convenience of the referees. Appendix A presents the experimental instructions along with the demographic questionnaire of the incentive treatment *Scale50PrUnknown*. Appendix B contains complementary descriptive regression results. Appendix C presents the expected utility decision model in HL's risk elicitation task along with its econometric implementation. Appendix D contains complementary structural estimation results.

Appendix A. Demographic Questionnaire and Instructions

The following sections present the English version of the registration screen, the experimental instructions and the demographic questionnaire of the incentive treatment $Scale50PrUnknown$. Instructions and the demographic questionnaire of the two laboratory treatments were adapted accordingly.

A.1 Registration

To register for the TorLabor markets please fill in the form below. You will receive a confirmation e-mail upon receipt of your completed form. This e-mail contains a link which you must follow to complete the registration process.

First name: [ ]
Last name: [ ]
E-mail: [ ]
Repeat e-mail: [ ]
Language: [∨]
- English
- German

For each match of the World Cup 2006, a new market will open a few days before the start of the match. **Would you like to be informed via e-mail each time a new market opens?**

◉ Yes. ○ No.

(It is always possible to enable/disable the delivery of e-mails.)

Please choose a username and a password. Both are necessary to participate in the markets. For your username, use only letters, numbers, dot ('.') or underscore ('_'). Your password must be at least 6 characters long.

Username: [ ]
Password: [ ]
Repeat password: [ ]

Please read carefully the rules and the terms and conditions (T&Cs). If after reading the rules and the terms and conditions, you understand and accept them, check the following box:

☐ I have read the rules and the terms and conditions and I accept them.

Important: Multiple registrations of the same person are NOT permitted and will lead to exclusion from participation and payment.

[Register]

A.2 Instructions

The bottom of the screen shows ten decisions. Each decision is a paired choice between "Option A" and "Option B". You are asked to make ten choices, and by completing this task you are offered a first possibility to earn some money at TorLabor soccer trading markets.

Later on, you will be offered more possibilities to earn money at TorLabor soccer trading markets. In particular, you may earn money by trading at TorLabor markets. Further details will be provided in due time. Please notice that if you are rewarded for completing the present task then you won't be rewarded for trading and that this information will be provided to you only after the World Cup 2006.

How your choices affect your payoffs

At the end of the World Cup 2006, 5 participants will be randomly selected. Imagine that you are one of these 5 randomly selected participants. A ten-sided die will be used to determine your payoffs. The faces of the die are numbered from 1 to 10. The ten-sided die will be thrown twice. The first throw will determine which of the 10 decisions you made will affect your payoffs. For this decision we will look at the option you have chosen. Then, for the chosen option, the die will be thrown a second time. You will receive the payoffs attached to the number of the die throw's result.

Thus, even though you will make ten decisions, only one of these will end up affecting your payoffs. You will not know in advance which decision will be used. Each decision has an equal chance of being used in the end.

Now, please look at Decision 1 at the top. Option A pays 100.00 euros if the throw of the ten-sided die yields 1, and it pays 80.00 euros if the throw yields 2, 3, 4, 5, 6, 7, 8, 9, or 10. Option B pays 192.50 euros if the throw of the die yields 1, and it pays 5.00 euros if the throw yields 2, 3, 4, 5, 6, 7, 8, 9, or 10. The other decisions are similar, except that as you move down the table, the chances of the higher payoff for each option increase. In fact, for Decision 10 in the bottom row, the die will not be needed since each option pays the highest payoff for sure, so your choice here is between 100 euros or 192.50 euros.

	Option A		I choose	**Option B**		I choose
Decision	Die throw yields	Option pays	option A	Die throw yields	Option pays	option B
1	1	100.00 euros	○	1	192.50 euros	○
	2,3,4,5,6,7,8,9 or 10	80.00 euros		2,3,4,5,6,7,8,9 or 10	5.00 euros	
2	1 or 2	100.00 euros	○	1 or 2	192.50 euros	○
	3,4,5,6,7,8,9 or 10	80.00 euros		3,4,5,6,7,8,9 or 10	5.00 euros	
3	1,2 or 3	100.00 euros	○	1, 2 or 3	192.50 euros	○
	4,5,6,7,8,9 or 10	80.00 euros		4,5,6,7,8,9 or 10	5.00 euros	
4	1,2,3 or 4	100.00 euros	○	1,2,3 or 4	192.50 euros	○
	5,6,7,8,9 or 10	80.00 euros		5,6,7,8,9 or 10	5.00 euros	
5	1,2,3,4 or 5	100.00 euros	○	1,2,3,4 or 5	192.50 euros	○
	6,7,8,9 or 10	80.00 euros		6,7,8,9 or 10	5.00 euros	
6	1,2,3,4,5 or 6	100.00 euros	○	1,2,3,4,5 or 6	192.50 euros	○
	7,8,9 or 10	80.00 euros		7,8,9 or 10	5.00 euros	
7	1,2,3,4,5,6 or 7	100.00 euros	○	1,2,3,4,5,6 or 7	192.50 euros	○
	8,9 or 10	80.00 euros		8,9 or 10	5.00 euros	
8	1,2,3,4,5,6,7 or 8	100.00 euros	○	1,2,3,4,5,6,7 or 8	192.50 euros	○
	9 or 10	80.00 euros		9 or 10	5.00 euros	
9	1,2,3,4,5,6,7,8 or 9	100.00 euros	○	1,2,3,4,5,6,7,8 or 9	192.50 euros	○
	10	80.00 euros		10	5.00 euros	
10	1,2,3,4,5,6,7,8,9 or 10	100.00 euros	○	1,2,3,4,5,6,7,8,9 or 10	192.50 euros	○
	- - -	80.00 euros		- - -	5.00 euros	

Continue

A.3 Demographic questionnaire

Please answer the questions below. Questions marked with a '*' are mandatory. In case you feel uncomfortable with answering some of the non-mandatory questions, you can skip them.

The Cologne Laboratory for Economic Research and the Max Planck Institute of Economics in Jena would be grateful if you would answer all questions truthfully. Your assistance is essential to the success of the research conducted by the Cologne Laboratory for Economic Research and the Max Planck Institute of Economics in Jena.

Our terms and conditions apply meaning that your data will be anonymised before the analysis.

Year of birth: [ ]

* Your home country: [ ]

* Your gender:
- ◯ female
- ◯ male

Your marital status:
- ◯ married
- ◯ single
- ◯ divorced
- ◯ widowed
- ◯ other

* Your degree of expertise in soccer is:

low	high
◯	◯

* Your degree of experience in trading markets is:

low	high
◯	◯

How have you heard about the TorLabor trading markets?:
- ◯ invitation email from an experimental lab
- ◯ friends
- ◯ newspaper
- ◯ coincidence
- ◯ another website
- ◯ other

Who in your household would you consider to be primarily in charge of expenses and budget decisions?:
- ◯ self
- ◯ spouse
- ◯ parent
- ◯ other
- ◯ don't know

How would you best describe your current employment situation?:
- ◯ full-time employed (not at university)
- ◯ part-time employed (not at university)
- ◯ self-employed (not at university)
- ◯ unemployed
- ◯ student only
- ◯ employed at university
- ◯ other

If you are a **student**, please answer these additional questions.

What describes your current situation best?:
- ◯ full-time student
- ◯ part-time student (less than 12 hours per week)

Your major field of studies: [ ]

Your current semester: [ ]

- ◯ undergraduate level
- ◯ graduate level

Who is primarily responsible for your tuition and living expenses while you are attending university?:
- ◯ self
- ◯ parent
- ◯ shared between parent and self
- ◯ scholarship / grant
- ◯ loans
- ◯ combination / other

Appendix B. Complementary Descriptive Regression Results

In this appendix we report a series of regression results which complement those reported in the main text. First, we compare the estimates of the number of safe choices in the different incentive treatments. Second, we show that the significant effect from age on risk attitudes is largely the consequence of less risk averse choices made by non-students older than most students. Third, we compare the estimates of the CRRA index in our incentive treatments with those in the 10× and 1×10× treatments of Harrison, Johnson, McInnes, and Rutström (2005, HJMR hereafter).

B.1 Estimates of the number of safe choices at the treatment level

Table 1 reports results from OLS models of the number of safe choices which only include treatment dummies. Results for all (resp. consistent) choice sequences are shown in the left (resp. right) panel, and in each panel estimates for all subjects, non-students, part-time and full-time students are shown separately.

We observe that differences in the estimated number of safe choices between treatments vary with the education status of participants. For non-students, risk aversion in $Scale50PrUnknown$ is lower than in the two laboratory treatments but there are no significant treatment effects (there are only 4 (resp. 2) non-students in the sample of $Scale50Pr1/15$ (resp. $Scale10Pr1/3$)). For part-time students, risk aversion in $Scale50PrUnknown$ is larger than in $Scale50Pr1/15$ while it is lower than in $Scale10Pr1/3$ but there are no significant treatment effects. For full-time students, we find that i) risk aversion in $Scale50PrUnknown$ is always significantly lower than in $Scale50Pr1/15$ at the 5% significance level; ii) risk aversion in $Scale50PrUnknown$ is always larger than in $Scale10Pr1/3$ and significantly so at the 10% level for all choice sequences; and iii) risk aversion in $Scale50Pr1/15$ is significantly larger than in $Scale10Pr1/3$ at the 1% level (resp. 5% level) for all choice sequences (resp. the consistent choice sequences).

	All choice sequences				Consistent choice sequences			
	All subjects	Non-students	Part-time students	Full-time students	All subjects	Non-students	Part-time students	Full-time students
Constant	6.181***	6.022***	6.260***	6.210***	6.043***	5.839***	6.082***	6.086***
	(0.031)	(0.079)	(0.108)	(0.036)	(0.031)	(0.080)	(0.109)	(0.035)
$Scale50Pr1/15$	0.269	0.228	−0.722	0.534**	0.395	0.411	−0.415	0.597**
	(0.239)	(0.423)	(0.614)	(0.262)	(0.244)	(0.423)	(0.651)	(0.268)
$Scale10Pr1/3$	−0.181	1.478	0.285	−0.401*	−0.061	1.661	0.118	−0.220
	(0.217)	(1.066)	(0.549)	(0.233)	(0.211)	(1.067)	(0.479)	(0.235)
R^2	0.001	0.002	0.005	0.002	0.001	0.004	0.002	0.003
Observations	3,702	643	358	2,701	3,183	534	302	2,347

Notes: Heteroskedasticity-robust standard errors are in parentheses. * (10%); ** (5%); and *** (1%) significance level.

Table 1: OLS estimates of the number of safe choices with treatment dummies only

We also note that, in line with the findings of Andersen, Harrison, Lau, and Rutström (2010), students exhibit higher risk aversion than non-students in $Scale50PrUnknown$ whether all or only consistent choice sequences are considered (the comparison is less compelling in the laboratory treatments due to the small number of non-students).

B.2 Age effect

Table 2 reports OLS regression results on the number of safe choices identical to those in the main text except that we control for age by partitioning subjects into three groups: (i) the 25% youngest subjects (dummy variable *Youngest*), the 25% oldest subjects (dummy variable *Oldest*), and (iii) the remaining middle-aged subjects who constitute the reference group. For the full sample of subjects, the 25% youngest are less than 22 years old whereas the 25% oldest are more than 27 years old. For the sample of full-time (resp. part-time) students, the 25% youngest are under the age of 22 (resp. 24) years whereas the 25% oldest are above the age of 26 (resp. 28) years. Finally, the 25% youngest non-students are less than 26 years old whereas the 25% oldest non-students are more than 36 years old.

We find no significant age effect for part-time and full-time students. On the contrary, the youngest non-students, under the age of 26, are significantly more risk averse than those aged between 26 and 36, whereas the oldest non-students, above the age of 36, are strongly significantly less risk averse than those from the reference group. These observations are well in line with the findings reported in Andersen, Harrison, Lau, and Rutström (2010) who compare estimates of risk (and time) preferences elicited from a convenience sample of university students in Copenhagen with estimates from a sample of the adult population in Denmark.

B.3 Comparing Our CRRA Estimates with Those from Previous Studies

We compare the estimates of the CRRA index in our incentive treatments with those in the 10× and 1×10× treatments of HJMR. In treatment 1×10× participants complete HL's risk elicitation task with outcomes of the safe (resp. risky) lottery equal to US$2.00 and US$1.60 (resp. US$3.85 and US$0.10), and then they have the possibility to give up their earnings in return for the chance to complete the task with lottery outcomes scaled up by 10. In treatment 10× participants complete the task only once with lottery outcomes scaled up by 10. We exclude the choices made in treatment 1×10× with lottery outcomes scaled up by 10 from our statistical analysis.

Table 3 reports interval regression estimates of the CRRA index in the different treatments where *HJMR1x* takes value 1 if choices have been made in treatment 1×10× with low payment (1×) and *HJMR10x* takes value 1 if choices have been made in treatment 10×. Results for all (resp. consistent) choice sequences are shown in the left (resp. right) panel, and in each panel estimates for all participants and students are shown separately. In regressions with the entire sample of participants we control for age and gender and in regressions with students we additionally control for the level of education (undergraduate or graduate) and the major field of study. For a given regression, participants with missing values for the included variables are omitted.

At mean demographic values, the CRRA estimate in treatment 1×10× with low payment and in treatment 10× equals 0.372 and 0.574 respectively. Compared to HJMR's treatments, we observe more risk aversion in *Scale50PrUnknown* and *Scale50Pr1/15* with CRRA estimates for all participants equal to 0.642 and 0.710 respectively (0.650 and 0.715 for students). Risk aversion in *Scale10Pr1/3* is of similar magnitude than in treatment 10× with CRRA estimate for all participants equal to 0.573 (0.554 for students) and it is larger than in treatment 1×10× with low payment. We now evaluate the statistical significance of the observed differences.

HJMR recruited 178 students from the University of South Carolina to complete HL's task in their two treatments (55 participated in treatment 10× and 123 in treatment 1×10×). Therefore, the most appropriate comparison of CRRA estimates is between estimates in HJMR's treatments and estimates

	All choice sequences			Consistent choice sequences		
	All participants		Students	All participants		Students
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Constant	6.255***	6.266***	5.986***	6.081***	6.082***	5.770***
	(0.058)	(0.065)	(0.134)	(0.057)	(0.062)	(0.130)
$Scale50Pr1/15$	0.220	0.542**	0.602**	0.340	0.575**	0.661**
	(0.240)	(0.264)	(0.269)	(0.245)	(0.271)	(0.279)
$Scale10Pr1/3$	−0.218	−0.388	−0.412*	−0.115	−0.251	−0.244
	(0.219)	(0.239)	(0.238)	(0.213)	(0.240)	(0.240)
Female	0.124*	0.132*	0.146*	0.152**	0.127*	0.108
	(0.066)	(0.075)	(0.080)	(0.066)	(0.075)	(0.079)
Youngest	−0.027	−0.050	−0.009	−0.020	−0.020	0.042
	(0.075)	(0.081)	(0.098)	(0.075)	(0.080)	(0.097)
Oldest	−0.168*	−0.121	−0.123	−0.154	−0.079	−0.076
	(0.095)	(0.106)	(0.119)	(0.096)	(0.107)	(0.119)
Part-time student		−0.045	0.298		0.059	0.492
		(0.225)	(0.438)		(0.230)	(0.443)
PT student x $Scale50Pr1/15$		−1.359**	−1.453**		−1.102*	−1.315*
		(0.666)	(0.711)		(0.711)	(0.756)
PT student x $Scale10Pr1/3$		0.612	0.303		0.319	0.283
		(0.563)	(0.525)		(0.545)	(0.542)
PT student x *Female*		0.005	−0.036		0.053	0.061
		(0.241)	(0.250)		(0.242)	(0.247)
PT student x *Youngest*		0.121	−0.048		0.078	−0.084
		(0.253)	(0.282)		(0.258)	(0.281)
PT student x *Oldest*		0.067	−0.017		0.268	0.147
		(0.333)	(0.356)		(0.344)	(0.361)
Non-student		−0.428			−0.344	
		(0.496)			(0.485)	
Non-student x $Scale50Pr1/15$		−0.432			−0.068	
		(0.473)			(0.465)	
Non-student x $Scale10Pr1/3$		1.615			1.670	
		(1.161)			(1.286)	
Non-student x *Female*		−0.133			0.108	
		(0.199)			(0.204)	
Non-student x *Youngest*		0.400**			0.450**	
		(0.200)			(0.198)	
Non-student x *Oldest*		−0.616**			−0.774***	
		(0.267)			(0.259)	
Controls for budgeting, marital status & employment status	Yes	Yes	Yes	Yes	Yes	Yes
Controls for duration and level of education, field of studies & payment of living expenses	No	No	Yes	No	No	Yes
R^2	0.010	0.020	0.029	0.012	0.023	0.027
Observations	3,478	3,478	2,750	2,985	2,985	2,381

Notes: Heteroskedasticity-robust standard errors are in parentheses. *PT student* x *MVI* refers to the interaction term between the dummy variable *Part-time student* and *MVI*, one of the main variables of interest. Likewise, *Non-student* x *MVI* refers to the interaction term between the dummy variable *Non-student* and a main variable of interest. * (10%); ** (5%); and *** (1%) significance level.

Table 2: OLS estimates of the number of safe choices for different groups of age

derived from the choices of our students. We find that risk aversion in treatment $1\times10\times$ with low payment is significantly lower than risk aversion in treatments *Scale50PrUnknown* and *Scale50Pr1/15* at the 1% level of significance but it is significantly lower than risk aversion in treatment *Scale10Pr1/3*

	All choice sequences		Consistent choice sequences	
	All participants	Students	All participants	Students
Constant	0.771***	0.632***	0.745***	0.616***
	(0.048)	(0.091)	(0.046)	(0.087)
Scale50Pr1/15	0.048	0.055	0.105	0.120*
	(0.076)	(0.077)	(0.070)	(0.072)
Scale10Pr1/3	−0.090	−0.120	−0.043	−0.049
	(0.075)	(0.075)	(0.070)	(0.070)
HJMR1x	−0.305***	−0.291***	−0.232***	−0.211***
	(0.055)	(0.054)	(0.052)	(0.052)
HJMR10x	−0.104	−0.078	−0.011	0.019
	(0.079)	(0.077)	(0.076)	(0.075)
Controls for age and gender	Yes	Yes	Yes	Yes
Controls for level of education & major field of studies	No	Yes	No	Yes
Log-likelihood	−6,744.68	−5,447.31	−6,214.15	−5,041.58
Observations	3,771	3,067	3,238	2,653
Left-censored obs.	2	2	0	0
Uncensored obs.	0	0	0	0
Right-censored obs.	684	533	289	229
Interval obs.	3,085	2,532	2,949	2,424

Note: * (10%); ** (5%); and *** (1%) significance level.

Table 3: Interval regression estimates of CRRA in our treatments and in HJMR

only at the 10% level of significance. On the other hand, we cannot reject the null hypothesis that risk aversion in treatment 10× equals the one estimated in any of our incentive treatments at any conventional significance level. These findings hold whether all or only consistent choice sequences are considered.

Appendix C. Expected Utility in HL's Risk Elicitation Task

In this appendix we present the expected utility (EU) decision model in HL's risk elicitation task. First, we expose the EU model in its deterministic form. Second, we assume that subjects' choices contain some random element and we embed the EU model into a model of stochastic choice. Finally, we derive our structural econometric model. Note that each of the ten choices made by an EU subject in our incentive treatments coincides with the choice made if facing only the corresponding decision by itself.

C.1 Deterministic EU Decision-Making

Let $i \in \{1, \ldots, I\}$ index subjects. In decision $d \in \{1, \ldots, 10\}$ the safe lottery $\tilde{S}_d = \left(0, \frac{10-d}{10}, \frac{d}{10}, 0\right)$ and the risky lottery $\tilde{R}_d = \left(\frac{10-d}{10}, 0, 0, \frac{d}{10}\right)$ are discrete probability distributions on the vector of monetary outcomes (l^R, l^S, h^S, h^R) where h^S and l^S (respectively h^R and l^R) denote the high and low monetary outcomes of the safe (respectively risky) lottery. In all decisions between paired lotteries, we have that $h^S = 2.00 \times scale$, $l^S = 1.60 \times scale$, $h^R = 3.85 \times scale$, and $l^R = 0.10 \times scale$ where $scale \in \{10, 50\}$.

Let $u_i(\cdot)$ denote the utility function of subject i. We refrain from making a functional form assumption about u_i. However, we postulate that, for each subject i, $u_i\left(h^R\right) > u_i\left(h^S\right) > u_i\left(l^S\right) > u_i\left(l^R\right)$ and we adopt the normalization $u_i\left(l^R\right) = 0$ and $u_i\left(h^R\right) = 1$. Hence, $0 < u_i\left(l^S\right) < u_i\left(h^S\right) < 1$. According to the deterministic EU model, subject i chooses the safe rather than the risky lottery in decision d if and only if

$$EU_i\left(\tilde{S}_d\right) > EU_i\left(\tilde{R}_d\right) \Leftrightarrow r_i \equiv \frac{u_i\left(l^S\right)}{1 - u_i\left(h^S\right) + u_i\left(l^S\right)} > \frac{d}{10}.$$

The smaller the ratio r_i of utilities the earlier subject i switches to the risky lottery, and once the subject chooses the risky lottery all subsequent choices consist of the risky lottery. Such a sequence of choices is called consistent. The ratio of a risk-neutral subject equals approximately 0.45 which implies that the safe lottery is chosen in the first four decisions and then the subject switches to the risky lottery.

Apart from consistent sequences of choices, we expect to observe *inconsistent* sequences of choices since the latter are quite common in experimental measurements of risk attitudes which rely on HL's task. Under the restriction that subjects' choices are governed by the EU model, inconsistent sequences of choices are assumed to provide less precise information concerning the ratio of utilities than consistent sequences of choices. In particular, any inconsistent sequence of choices in which the safe lottery is chosen in the last decision prevents the inference of an upper bound on the ratio of utilities. Though inconsistent sequences of choices might be the result of subjects being indifferent between some pairs of lotteries (indifference can even account for the choice of the safe lottery in decision 10 if utility is only weakly increasing in money), we favor the alternative interpretation that (binary discrete) choice under risk has a large stochastic component. To account for the random part in choice under risk, we embed the EU model into a stochastic specification of choice under risk.

C.2 Stochastic EU Decision-Making

We assume that subject i's choices are governed by a stochastic choice function $\mathbf{P}_i(\cdot, \cdot)$ which assigns a real-valued choice probability in the interval $[0, 1]$ to every ordered pair $(\tilde{S}_d, \tilde{R}_d)$. Without loss of generality $\mathbf{P}_i(\tilde{S}_d, \tilde{R}_d)$ denotes the probability that subject i chooses the safe lottery in decision d. Following the terminology coined by Wilcox (2008), we distinguish between *considered* choice probabilities and *overall* choice probabilities that subject i chooses the safe lottery in decision d. Considered choice probabilities

are linked to the expected utilities of the two lotteries, and overall choice probabilities are deduced from considered choice probabilities by adding constant probabilities that choices are made completely at random.

The considered component of the stochastic EU model is the Fechner strong utility model (Hey and Orme, 1994) according to which subject i attempts to choose the safe lottery in decision d if

$$\hat{k}_i \left[EU_i \left(\tilde{S}_d \right) - EU_i \left(\tilde{R}_d \right) \right] + \tilde{\epsilon}_i > 0$$

where $\hat{k}_i$ is subject i's sensitivity to the difference in expected utilities and $\tilde{\epsilon}_i$ follows the standard normal distribution. Since

$$\begin{aligned} EU_i \left(\tilde{S}_d \right) - EU_i \left(\tilde{R}_d \right) &= \left(\frac{10-d}{10} \right) u_i \left(l^S \right) - \left(\frac{d}{10} \right) \left[1 - u_i \left(h^S \right) \right] \\ &= \left[1 - u_i \left(h^S \right) + u_i \left(l^S \right) \right] \cdot \left(r_i - \frac{d}{10} \right), \end{aligned}$$

the strong utility probability that subject i chooses the safe lottery in decision d is given by

$$\mathbf{P}_i^{SU} \left(\tilde{S}_d, \tilde{R}_d \right) = \Phi \left(k_i \left(r_i - \frac{d}{10} \right) \right) \tag{1}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function and $k_i = \hat{k}_i \cdot \left[1 - u_i \left(h^S \right) + u_i \left(l^S \right) \right]$. Strong utility probabilities derive from lotteries being evaluated according to the EU model and calculations of expected utilities being subject to measurement error. For a given k_i, the larger the difference in expected utilities the smaller the probability of mistakenly choosing the less preferred lottery. Note that absent further restrictions on the utility function the payoff sensitivity parameter k_i is determined only up to a multiplicative factor. Since we are mainly interested in comparing the estimated value of this parameter across treatments (and individuals) we abstain from restricting the utility function.[1]

We add to the considered choice probabilities a trembling mechanism i.e. constant probabilities that subjects choose completely at random. Accordingly, the overall choice probability that subject i chooses the safe lottery in decision d, given (r_i, k_i, w_i), is

$$\mathbf{P}_i \left(\tilde{S}_d, \tilde{R}_d \right) = (1 - w_i) \, \Phi \left(k_i \left(r_i - \frac{d}{10} \right) \right) + \frac{w_i}{2}, \tag{2}$$

where $w_i \in [0, 1]$ is subject i's tremble probability. Trembles are clearly unconnected with the nature of the paired lottery choice itself and they are meant to capture subjects' momentary inattention or lapses of concentration.

Our stochastic component of choice under risk combines two complementary elements.[2] Fechner errors capture stochastic variations in the formation of risk preferences whereas trembles are unconnected with risk preferences and capture stochastic variations in the execution of preferences. We expect trembles to pick up differences in the implementation mode as subjects are more likely to temporarily lose concentration in the Internet environment than in a controlled laboratory setting. On the other hand,

[1] Note that the factor $\left[1 - u_i \left(h^S \right) + u_i \left(l^S \right) \right]$ does not imply that k_i and r_i are dependent. Indeed, it can be shown that $0 < u_i \left(l^S \right) < u_i \left(h^S \right) < 1$ if and only if $0 < r_i < 1$ and $0 < k_i < \hat{k}_i$, i.e. any pair (r_i, k_i) with these properties is admissible.

[2] Though the trembling mechanism is not viable as the principal stochastic component of choice under risk, the explanatory power of stochastic choice models can be significantly increased by the addition of trembles (Loomes, Moffatt, and Sugden, 2002).

sensitivities to the difference in expected utilities might vary with the characteristics of BRIS, especially the scale of nominal payoffs.

C.3 Structural Econometric Model

Equation (2) states the probability with which subject i chooses the safe lottery in decision $d \in \{1, \ldots, 10\}$ given her risk and error parameters (r_i, k_i, w_i). The stochastic EU model is capable of rationalizing all possible sequences of observed choices and constitutes the basis of our econometric model. We now detail the procedure to estimate the distribution of structural parameters in our samples.

Let $c_i^d = 1$ if subject i chooses the safe lottery in decision $d \in \{1, \ldots, 10\}$, and $c_i^d = -1$ otherwise. Given the subject-specific parameters (r_i, k_i, w_i), the likelihood of observing the choice c_i^d of subject i in decision d is given by

$$\ell_i^d \left(c_i^d \mid r_i, k_i, w_i \right) = (1 - w_i)\, \Phi \left(c_i^d\, k_i \left(r_i - \frac{d}{10} \right) \right) + \frac{w_i}{2}. \tag{3}$$

Assuming that errors are independent across decisions implies that the likelihood of observing the choice sequence $\mathbf{c}_i = \left(c_i^1, \ldots, c_i^{10} \right)$ of subject i given parameters (r_i, k_i, w_i) equals

$$\ell_i \left(\mathbf{c}_i \mid r_i, k_i, w_i \right) = \prod_{d=1}^{10} \ell_i^d \left(c_i^d \mid r_i, k_i, w_i \right). \tag{4}$$

In the most general econometric specification, all parameters vary both with observable characteristics (*observed heterogeneity*) and with unobservable characteristics (*unobserved heterogeneity*). Allowing for both observed and unobserved heterogeneity and taking into account the interval restriction of the parameters we have

$$z_i = g_z \left(\mathbf{x}_i\, \boldsymbol{\beta}^z + \tilde{\zeta}_i^z \right) \tag{5}$$

where $z \in \{r, k, w\}$, $g_k(x) = \exp(x)$ and $g_z(x) = \Lambda(x) = 1/\left(1 + \exp(-x)\right)$ for $z \in \{r, w\}$, $\mathbf{x}_i$ is a $1 \times K$ vector of regressors, $\boldsymbol{\beta}^z$ is a vector of coefficients of z, and $\tilde{\zeta}_i^z$ is the unobserved heterogeneity component of z. Assuming that $\tilde{\boldsymbol{\zeta}}_i = \left(\tilde{\zeta}_i^r, \tilde{\zeta}_i^k, \tilde{\zeta}_i^w \right)$ follows a joint normal distribution with mean $(0, 0, 0)$, covariance matrix $\Sigma' \Sigma$, and density $\phi_\Sigma(\cdot)$ implies that the likelihood contribution of subject i may be written as

$$\ell_i \left(\boldsymbol{\beta}^r, \boldsymbol{\beta}^k, \boldsymbol{\beta}^w, \Sigma \right) = \int_{\mathbb{R}^3} \left[\prod_{d=1}^{10} \ell_i^d \left(c_i^d \mid \Lambda \left(\mathbf{x}_i\, \boldsymbol{\beta}^r + \zeta^r \right), \exp \left(\mathbf{x}_i\, \boldsymbol{\beta}^k + \zeta^k \right), \Lambda \left(\mathbf{x}_i\, \boldsymbol{\beta}^w + \zeta^w \right) \right) \right] \phi_\Sigma \left(\boldsymbol{\zeta} \right) d\boldsymbol{\zeta}, \tag{6}$$

and the overall log-likelihood is given by

$$L \left(\boldsymbol{\beta}^r, \boldsymbol{\beta}^k, \boldsymbol{\beta}^w, \Sigma \right) = \sum_{i=1}^{I} \log \left(\ell_i \left(\boldsymbol{\beta}^r, \boldsymbol{\beta}^k, \boldsymbol{\beta}^w, \Sigma \right) \right). \tag{7}$$

Since the integral in (6) does not possess an analytical solution it has to be approximated. A standard simulation technique is to calculate a value of the functional at a series of randomly generated values of $\left(\tilde{\zeta}_i^r, \tilde{\zeta}_i^k, \tilde{\zeta}_i^w \right)$ and approximate the integral by the average across the function values (Train, 2003). We therefore construct a sequence of $J = 1{,}000$ shuffled Halton draws per parameter and individual. Letting H_i denote the $J \times 3$-matrix of Halton draws for subject i, and L the Choleski factor of Σ,[3] we then

[3] Thus, L is a lower-triangular matrix such that $L\, L' = \Sigma$.

generate J simulated values of $\tilde{\zeta}_i$ via

$$\begin{aligned}
\zeta^r_{i,j} &= L(1,1) \cdot \Phi^{-1}\left(H_i(j,1)\right), \\
\zeta^k_{i,j} &= L(2,1) \cdot \Phi^{-1}\left(H_i(j,1)\right) \,+\, L(2,2) \cdot \Phi^{-1}\left(H_i(j,2)\right), \\
\zeta^w_{i,j} &= L(3,1) \cdot \Phi^{-1}\left(H_i(j,1)\right) \,+\, L(3,2) \cdot \Phi^{-1}\left(H_i(j,2)\right) \,+\, L(3,3) \cdot \Phi^{-1}\left(H_i(j,3)\right),
\end{aligned}$$

where $\Phi^{-1}(\cdot)$ denotes the inverse cumulative distribution function of the standard normal distribution.

To maximise the (simulated) log-likelihood function (7) we employ the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with numerical derivatives which we restart multiple times with random initial values to rule out local maxima. Finally, the variance-covariance matrix of the parameter estimates is based on the outer product of gradients, and standard errors for transformed parameters are calculated using the delta method.[4]

While we are very much in favour of a specification with full heterogeneity, we acknowledge that practical considerations may force us to rely on a restricted specification. Indeed, the difficulty of estimating heterogeneity in both k and w has been acknowledged (von Gaudecker, van Soest, and Wengström, 2011). Below, we report the results of a simulation analysis conducted to assess the empirical identification of various specifications of the econometric model which assume different degrees of heterogeneity for the modelling parameters.

C.3.1 Identification of the Econometric Model

To see that the model presented above is identified in theory, consider first the model without heterogeneity. Let i denote a consistent subject who picks the safe (risky) lottery n $(10-n)$ times where $0 \leqslant n \leqslant 9$. Clearly, i's decision sequence is perfectly captured by individual parameters $r_i \in \left(\frac{n}{10}, \frac{n+1}{10}\right)$, $w_i = 0$, and $k_i \to \infty$.[5] Assume next that subject j differs from subject i by a single inconsistent decision $\hat{d}$ such that either $\hat{d} < n$ and $c_j^{\hat{d}} = -1$, or $\hat{d} > n+1$ and $c_j^{\hat{d}}$. Note that decreasing k increases the probability of choices not in line with the EU model more strongly for decisions close to the switch point $d = n+1$, while increasing w increases the probability of such deviations uniformly for all decisions. Accordingly, the decision sequence of the inconsistent subject j is best explained by $k_j \ll k_i$, $w_j = 0$ if and only if $\left|\hat{d} - (n+1)\right|$ is small, and by $k_j \to \infty$, $w_j > 0$ otherwise. Hence, k and w are both identified in theory since they capture different forms of deviation from the EU model. On the other hand r becomes identified once we consider a group of (consistent) subjects with differing frequencies of choosing the safe lottery. Allowing for observed heterogeneity does not change these results, as long as the number of explanatory variables considered is reasonably small, since it merely enables the model to distinguish between different subgroups of subjects. Unobserved heterogeneity enables the model to distinguish subjects *within* a given subgroup. Since a continuous distribution is assumed (as opposed to a finite mixture model) identification is not affected by this assumption either.

While the model is identified in theory, it generates a likelihood function which is likely to be flat in k and w in a neighborhood of the likelihood-maximizing values. Concretely, since the true k (w) is likely to be large (small), changes in k and w in the neighborhood of the optimal values will lead to small changes in the likelihood function. This may pose a problem in terms of the *empirical identification* of

[4] The estimation procedure is programmed in Stata and the code is available from the authors upon request.

[5] Incidentally, this implies that the model is not identified at the individual level.

the model. Indeed, von Gaudecker, van Soest, and Wengström (2011) argue that "in practice it appears to be difficult to estimate heterogeneity in k and w separately". We therefore conducted a simulation analysis to assess the empirical identification of different versions of the model. Concretely, for each of four different versions of the model (with and without unobserved heterogeneity in k and w), two different group sizes ($I = 10$ and $I = 20$), and various collections $(\mu_r, \sigma_r, \mu_k, (\sigma_k), \mu_w, (\sigma_w))$ of means and standard deviations of the parameters, we constructed 15 simulated datasets as follows: For each subject $i = 1, \ldots, I$

(i) a vector of untransformed parameters $\left(\hat{r}_i, \hat{k}_i, \hat{w}_i\right)$ is determined by drawing for each parameter $z \in \{r, k, w\}$ a random value from the normal distribution with mean μ_z and standard deviation σ_z;

(ii) the transformed parameters are calculated via $r_i = \Lambda\left(\hat{r}_i\right)$, $k_i = \exp\left(\hat{k}_i\right)$, and $w_i = \Lambda\left(\hat{w}_i\right)$; and

(iii) a decision sequence $\left(c_i^1, \ldots, c_i^{10}\right)$ is generated using 10 Bernoulli random draws with probabilities $\ell_i^d\left(c_i^d \mid r_i, k_i, w_i\right)$, $d = 1, \ldots, 10$.

For each simulated dataset and each parameter z we then performed a grid search for the likelihood-maximizing mean and standard deviation, holding fixed the mean and standard deviation of the other parameters at their true values in order to give best chances to identifying the true values. The likelihood values in the grid were simulated using the same fixed $1,000 \times 3$ shuffled Halton draws per individual.

The results of the simulation analysis are presented in Tables 1 to 4. Each table summarises the results for a different version of the model. For a given number of subjects I, and a given vector of true means and standard deviations we report (i) the median likelihood-maximizing values of the means and standard deviations across the 15 datasets, and (ii) the differences between the true values and the median likelihood-maximizing values, normalised to lie between -1 (median at the lower bound of the grid) and +1 (median at the upper bound of the grid). There are three main findings. First, the mean and the standard deviation of the ratio of utilities r are usually well captured by the likelihood-maximizing values, regardless of the model. Second, the means of k and w are well captured, when no unobserved heterogeneity is present in either of these parameters. Third, when allowing for unobserved heterogeneity in either k or w the standard deviations of the two parameters are badly captured, which is likely to affect the estimation of the distribution means.

From the simulation analysis we conclude that allowing for unobserved heterogeneity in k or w might be asking too much from our data. We therefore report in the main text the estimation results of a restricted econometric model which allows for unobserved heterogeneity in r only. For the sake of completeness, we now detail this restricted structural econometric model.

C.3.2 Restricted Structural Econometric Model

Allowing for unobserved heterogeneity in r only and taking into account the interval restriction of the parameters we have

$$r_i = \Lambda\left(\mathbf{x}_i \boldsymbol{\beta}^r + \tilde{\zeta}_i^r\right) = 1/\left[1 + \exp\left(-\left(\mathbf{x}_i \boldsymbol{\beta}^r + \tilde{\zeta}_i^r\right)\right)\right] \tag{8}$$

where $\mathbf{x}_i$ is a $1 \times K$ vector of regressors, $\boldsymbol{\beta}^r$ is a vector of coefficients of r, and $\tilde{\zeta}_i^r$ is the unobserved heterogeneity component of r which we assume to be normally distributed. The vector of regressors contains 1, treatment dummies, and dummies related to (a subset of) the *collected* demographics. Similarly, $k_i = \exp\left(\mathbf{x}_i \boldsymbol{\beta}^k\right)$ and $w_i = \Lambda\left(\mathbf{x}_i \boldsymbol{\beta}^w\right)$ where the unobserved heterogeneity component is omitted. Taking

10 subjects												
	Simulation 1			Simulation 2			Simulation 3			Simulation 4		
	True	Median	Error	True	Median	Error	True	Median	Error	True	Median	Error
μ_r	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.5	-0.2
σ_r	-1.0	-0.7	0.2	-1.0	-1.0	0.0	-1.0	-0.7	0.2	-1.0	-1.6	-0.4
μ_k	2.3	1.9	-0.2	2.3	2.3	0.0	1.4	1.4	0.0	1.4	1.4	0.0
μ_w	-1.5	-2.0	-0.2	-3.0	-4.5	-0.6	-3.0	-3.0	0.0	-1.5	-1.5	0.0
20 subjects												
	Simulation 1			Simulation 2			Simulation 3			Simulation 4		
	True	Median	Error	True	Median	Error	True	Median	Error	True	Median	Error
μ_r	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
σ_r	-1.0	-1.0	0.0	-1.0	-1.3	-0.2	-1.0	-0.7	0.2	-1.0	-1.0	0.0
μ_k	2.3	2.3	0.0	2.3	2.3	0.0	1.4	1.4	0.0	1.4	1.4	0.0
μ_w	-3.0	-3.5	-0.2	-1.5	-2.0	-0.2	-3.0	-3.0	0.0	-1.5	-1.5	0.0

Table 1: Simulation Results: Model with (unobserved) heterogeneity in r only.

10 subjects												
	Simulation 1			Simulation 2			Simulation 3			Simulation 4		
	True	Median	Error	True	Median	Error	True	Median	Error	True	Median	Error
μ_r	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
σ_r	-1.0	-1.0	0.0	**-1.0**	**-2.5**	**-1.0**	-1.0	-1.0	0.0	-1.0	-1.0	0.0
μ_k	2.3	2.3	0.0	2.3	2.3	0.0	1.4	1.4	0.0	2.3	2.3	0.0
μ_w	-3.0	-3.5	-0.2	-1.5	-2.5	-0.4	-3.0	-3.0	0.0	-3.0	-4.5	-0.6
σ_w	**-1.0**	**-2.5**	**-1.0**	**-1.0**	**0.5**	**1.0**	**-1.0**	**0.5**	**1.0**	**-2.5**	**-4.0**	**-1.0**
20 subjects												
	Simulation 1			Simulation 2			Simulation 3			Simulation 4		
	True	Median	Error	True	Median	Error	True	Median	Error	True	Median	Error
μ_r	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
σ_r	-1.0	-1.0	0.0	-1.0	-1.3	-0.2	-1.0	-0.7	0.2	-1.0	-1.0	0.0
μ_k	2.3	2.3	0.0	2.3	3.0	0.4	1.4	1.4	0.0	2.3	2.3	0.0
μ_w	-3.0	-3.5	-0.2	-1.5	-2.0	-0.2	-3.0	-3.0	0.0	**-3.0**	**-5.5**	**-1.0**
σ_w	**-1.0**	**0.5**	**1.0**	**-1.0**	**0.5**	**1.0**	-1.0	-0.7	0.2	**-2.5**	**-4.0**	**-1.0**

Table 2: Simulation Results: Model with (unobserved) heterogeneity in r and w.

into account these specifications the likelihood function is given by

$$L\left(\boldsymbol{\beta}^r, \boldsymbol{\beta}^k, \boldsymbol{\beta}^w, \sigma_r\right) = \sum_{i=1}^{I} \log\left(\int_{\mathbb{R}} \left[\prod_{d=1}^{10} \ell_i^d\left(c_i^d \mid \Lambda\left(\mathbf{x}_i \boldsymbol{\beta}^r + \zeta^r\right), \exp\left(\mathbf{x}_i \boldsymbol{\beta}^k\right), \Lambda\left(\mathbf{x}_i \boldsymbol{\beta}^w\right)\right)\right] \phi_{\sigma_r}\left(\zeta_r\right) d\zeta_r\right) \tag{9}$$

where $\phi_{\sigma_r}(\cdot)$ is the density of the (unidimensional) normal distribution with mean zero and standard deviation σ_r. The integral in (9) does not possess an analytical solution and we approximate it using standard simulation techniques. Concretely, we construct a sequence of $J = 1,000$ shuffled Halton draws per individual and we maximise the (simulated) log-likelihood function via the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with numerical derivatives. Finally, the variance-covariance matrix of the parameter estimates is based on the outer product of gradients, and standard errors for transformed parameters are calculated using the delta method.

10 subjects

	Simulation 1			Simulation 2			Simulation 3			Simulation 4		
	True	Median	Error	True	Median	Error	True	Median	Error	True	Median	Error
μ_r	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
σ_r	-1.0	-1.0	0.0	-1.0	-1.3	-0.2	-1.0	-0.7	0.2	-1.0	-1.0	0.0
μ_k	2.3	2.3	0.0	2.3	2.3	0.0	1.4	1.4	0.0	2.3	2.3	0.0
σ_k	**-1.0**	**-2.2**	**-0.8**	**-1.0**	**-2.5**	**-1.0**	**-1.0**	**-2.5**	**-1.0**	-2.5	-2.8	-0.2
μ_w	-3.0	-3.5	-0.2	-1.5	-2.0	-0.2	-3.0	-2.5	0.2	-3.0	-4.0	-0.4

20 subjects

	Simulation 1			Simulation 2			Simulation 3			Simulation 4		
	True	Median	Error	True	Median	Error	True	Median	Error	True	Median	Error
μ_r	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
σ_r	-1.0	-1.0	0.0	-1.0	-1.0	0.0	-1.0	-1.0	0.0	-1.0	-1.0	0.0
μ_k	2.3	2.3	0.0	2.3	2.3	0.0	1.4	1.4	0.0	2.3	2.3	0.0
σ_k	-1.0	-1.3	-0.2	**-1.0**	**-2.5**	**-1.0**	**-1.0**	**-2.5**	**-1.0**	**-2.5**	**-4.0**	**-1.0**
μ_w	-3.0	-4.0	-0.4	-1.5	-2.0	-0.2	-3.0	-2.0	0.4	**-3.0**	**-5.5**	**-1.0**

Table 3: Simulation Results: Model with (unobserved) heterogeneity in r and k.

10 subjects

	Simulation 1			Simulation 2			Simulation 3			Simulation 4			Simulation 5		
	True	Median	Error	True	Median	Error	True	Median	Error	True	Median	Error	True	Median	Error
μ_r	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
σ_r	-1.0	-0.7	0.2	-1.0	-1.0	0.0	-1.0	-1.3	-0.2	-1.0	-1.3	-0.2	-1.0	-1.0	0.0
μ_k	2.3	2.3	0.0	1.4	1.4	0.0	2.3	3.4	0.6	2.3	2.7	0.2	2.3	2.3	0.0
σ_k	**-1.5**	**-2.5**	**-1.0**	**-1.0**	**-2.5**	**-1.0**	**-1.0**	**-2.5**	**-1.0**	-1.0	-1.3	-0.2	**-2.5**	**-4.0**	**-1.0**
μ_w	**-3.0**	**-5.0**	**-0.8**	-3.0	-2.5	0.2	-1.5	-2.5	-0.4	-3.0	-4.0	-0.4	-3.0	-3.5	-0.2
σ_w	**-1.0**	**0.2**	**0.8**	-1.0	-0.7	0.2	**-1.0**	**-2.5**	**-1.0**	-2.5	-2.8	-0.2	**-1.0**	**0.5**	**1.0**

20 subjects

	Simulation 1			Simulation 2			Simulation 3			Simulation 4			Simulation 5		
	True	Median	Error	True	Median	Error	True	Median	Error	True	Median	Error	True	Median	Error
μ_r	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
σ_r	-1.0	-1.0	0.0	-1.0	-0.7	0.2	**-1.0**	**-2.5**	**-1.0**	-1.0	-1.0	0.0	-1.0	-1.0	0.0
μ_k	2.3	2.3	0.0	1.4	1.4	0.0	2.3	1.9	-0.2	2.3	2.3	0.0	2.3	2.3	0.0
σ_k	-1.0	-0.7	0.2	-1.0	-0.7	0.2	**-1.0**	**-2.5**	**-1.0**	**-1.0**	**-2.5**	**-1.0**	-2.5	-2.2	0.2
μ_w	-3.0	-3.5	-0.2	-3.0	-2.0	0.4	-1.5	-2.0	-0.2	**-3.0**	**-5.0**	**-0.8**	-3.0	-4.5	-0.6
σ_w	**-1.0**	**0.2**	**0.8**	-1.0	-0.1	0.6	-1.0	-1.9	-0.6	**-2.5**	**-1.0**	**1.0**	**-1.0**	**0.5**	**1.0**

Table 4: Simulation Results: Model with (unobserved) heterogeneity in r, k and w.

Appendix D. Complementary Structural Estimation Results

In this appendix we report additional structural estimation results to complement those present in the main text.

D.1 Complete Results with a Homogeneous Sensitivity to Payoff Differences

Here we complement the estimation results shown in Table 4 and Table 5 in the main text. Table 1 reports the estimated parameters when the model contains the full set of covariates. Like for Table 4 in the main text, the results are presented on the original parameter scale meaning that the constant terms are given by $g_z\left(\beta^z_{\text{Constant}}\right)$, $z = r, k, w$, where $g_z(\cdot) = \Lambda(\cdot)$ for $z \in \{r, w\}$, and $g_k(\cdot) = \exp(\cdot)$, and the treatment effects are given by $g_z\left(\beta^z_{\text{Constant}} + \beta^z_{\text{Treatment}}\right) - g_z\left(\beta^z_{\text{Constant}}\right)$, $z = r, k, w$. Table 2 shows the corresponding median parameters stratified by the demographics.

Note that standard errors of the coefficients of the trembling probability cannot always be calculated for models 2 and 3. This is due to the fact that no inconsistent choice sequences are observed for several sub-samples of subjects. In such cases the trembling probability is zero for the sub-sample, and the maximization algorithm converges to a boundary solution. Since the likelihood function is very flat near the boundary, the algorithm is likely to encounter convergence problems. This usually results in maximum likelihood estimates with standard errors huge or lacking. To improve convergence we tried to estimate constrained versions of the models. Concretely, for model 2 we set the coefficients of *Non-student* and the interaction terms *Part-time Student* × *Scale10Pr1/3*, *Part-time Student* × *Unijob*, *Non-student* × *Scale50Pr1/15*, *Non-student* × *Scale10Pr1/3*, and *Non-student* × *Unijob* such that the transformed trembling probability equals zero for the corresponding sub-samples.[6] The constrained model 2 converged properly to the same value of the likelihood-function as the unconstrained version. Unfortunately, a similar technique could not be successfully implemented for model 3.

Table 1: Estimated parameters for models with the full set of covariates

	All Participants				Students	
	Model 1		Model 2		Model 3	
	r	w	r	w	r	w
Constant	0.749***	0.031***	0.714***	0.039***	0.697***	0.013***
	(0.017)	(0.003)	(0.040)	(0.008)	(0.048)	(0.004)
Scale10Pr1/15	0.017	-0.025***	0.056*	-0.032***	0.065**	-0.011***
	(0.027)	(0.004)	(0.033)	(0.008)	(0.032)	(0.004)
Scale10Pr1/3	-0.038	-0.026***	-0.058	-0.030***	-0.064	-0.009***
	(0.033)	(0.003)	(0.044)	(0.007)	(0.046)	(0.003)
Female	0.012	0.017***	0.015	0.022***	0.017*	0.004***
	(0.007)	(0.002)	(0.009)	(0.005)	(0.010)	(0.001)
Age	-0.002**	2.7E-04***	-1.7E-04	-1.1E-04	-0.001	0.001***
	(0.001)	(9.5E-5)	(0.002)	(3.4E-04)	(0.002)	(7.5E-05)
Married	0.003	0.005	-0.020	0.009	-0.026	0.001
	(0.017)	(0.003)	(0.035)	(0.008)	(0.042)	(0.003)

Continued on next page

[6] The concrete values we employed were -15 for the coefficient of *Non-student*, and -20 for the coefficients of the interaction terms.

Table 1: Continued

	All Participants				Students	
	Model 1		Model 2		Model 3	
	r	*w*	*r*	*w*	*r*	*w*
Not budgeting	0.013	0.010***	0.012	0.017***	0.014	0.005**
	(0.009)	(0.002)	(0.012)	(0.004)	(0.012)	(0.002)
Full-time job	-0.023	0.038***	-0.097	0.113***	-0.066	0.078**
	(0.016)	(0.006)	(0.145)	(0.038)	(0.160)	(0.034)
Part-time job	-0.017**	0.023***	-0.022**	0.027***	-0.026**	0.007***
	(0.009)	(0.003)	(0.011)	(0.006)	(0.012)	(0.003)
University job	-0.044**	-0.024***	-0.066	-0.011	-0.097*	-0.007**
	(0.021)	(0.003)	(0.043)	(0.007)	(0.052)	(0.003)
Other job status	-0.013	0.005	-0.047	-0.007	-0.054	-0.001
	(0.017)	(0.004)	(0.039)	(0.007)	(0.040)	(0.003)
Payment expenses: Self					0.021	0.003*
					(0.014)	(0.002)
Payment expenses: Shared					0.001	2.2E-04
					(0.012)	(0.001)
Payment expenses: Other					-0.008	-0.004**
					(0.014)	(0.002)
Semester					-0.001	-0.001***
					(0.002)	(3.4E-04)
Graduate					0.012	-0.002
					(0.013)	(0.001)
Economics					0.030**	-0.003*
					(0.014)	(0.002)
MNE					0.041***	0.007***
					(0.014)	(0.002)
SSH					0.032**	0.027***
					(0.013)	(0.007)
Other field of study					0.012	0.001
					(0.024)	(0.003)
Part-time student			0.093	0.069	0.068	0.231
			(0.074)	(0.065)	(0.104)	(0.192)
PT student × Scale50Pr1/15			-0.118	-0.053	-0.136	-0.181
			(0.073)	(0.054)	(0.085)	(0.167)
PT student × Scale10Pr1/3			0.070	-0.079	0.027	*-0.235*
			(0.075)	(0.065)	(0.116)	-
PT student × Female			-0.002	0.133**	0.004	0.211***
			(0.022)	(0.061)	(0.029)	(0.072)
PT student × Age			-0.003	-0.006	-0.001	-0.011
			(0.002)	(0.005)	(0.004)	(0.014)

Continued on next page

Table 1: Continued

	All Participants				Students	
	Model 1		Model 2		Model 3	
	r	*w*	*r*	*w*	*r*	*w*
PT student × Married			0.132***	0.230*	0.148**	0.287**
			(0.049)	(0.118)	(0.067)	(0.125)
PT student × Not budgeting			0.020	0.011	0.027	0.001
			(0.027)	(0.027)	(0.035)	(0.056)
PT student × Full-time job			-0.049	0.244**	-0.114	0.421***
			(0.166)	(0.124)	(0.195)	(0.081)
PT student × Part-time job			0.015	0.018	0.010	0.056
			(0.022)	(0.032)	(0.028)	(0.058)
PT student × University job			0.130**	-0.098	0.169**	*-0.238*
			(0.059)	(0.065)	(0.076)	-
PT student ×			-0.010	0.036	-0.181	0.208
Other job status			(0.065)	(0.063)	(0.111)	(0.170)
PT student ×					-0.023	0.021
Payment expenses: Self					(0.037)	(0.063)
PT student ×					-0.015	-0.119
Payment expenses: Shared					(0.036)	(0.093)
PT student ×					0.030	-0.126
Payment expenses: Other					(0.039)	(0.093)
PT student × Semester					3.7E-04	-0.010
					(0.004)	(0.007)
PT student × Graduate					-0.028	-0.049
					(0.035)	(0.044)
PT student × Economics					-0.095**	-0.079
					(0.048)	(0.077)
PT student × MNE					-0.006	-0.054
					(0.042)	(0.054)
PT student × SSH					0.024	0.051
					(0.037)	(0.074)
PT student ×					-0.029	-0.142
Other field of study					(0.065)	(0.206)
Non-student			0.026	-0.039***		
			(0.206)	(0.008)		
Non-student × Scale50Pr1/15			-0.096	0.032***		
			(0.197)	(0.008)		
Non-student × Scale10Pr1/3			0.176	0.030***		
			(0.130)	(0.007)		
Non-student × Female			-0.019	*-0.022*		
			(0.021)	-		

Continued on next page

Table 1: Continued

	All Participants				Students	
	Model 1		Model 2		Model 3	
	r	w	r	w	r	w
Non-student × Age			-0.003	*1.1E-04*		
			(0.002)	-		
Non-student × Married			0.023	*-0.009*		
			(0.043)	-		
Non-student × Not budgeting			-0.010	*-0.017*		
			(0.025)	-		
Non-student × Full-time job			0.139	-0.058		
			(0.248)	(0.040)		
Non-student × Part-time job			0.064	0.038***		
			(0.202)	(0.013)		
Non-student × University job			0.074	0.011		
			(0.206)	(0.007)		
Non-student ×			0.114	0.055***		
Other job status			(0.205)	(0.012)		
Standard deviation	1.019***		1.016***		0.984***	
	(0.016)		(0.016)		(0.019)	
k	17.937***		18.092***		18.857***	
	(1.103)		(1.143)		(1.411)	
Log-likelihood	-11,138.0		-11,087.7		-8,439.8	
Observations	3,478		3,478		2,750	

Regression coefficients are transformed back to the original scale, i.e. the constant is $g_z\left(\beta^z_{Constant}\right)$, coefficients of (standard) covariates are $g_z\left(\beta^z_{Constant} + \beta^z_{Covariate}\right) - g_z\left(\beta^z_{Constant}\right)$, the partial effect of setting the dummy variable to one or increasing age (or semester) by one, and coefficients of interaction terms are $g_z\left(\beta^z_{Constant} + \beta^z_{Cov.1} + \beta^z_{Cov.2} + \beta^z_{Cov.1\times Cov.2}\right) - g_z\left(\beta^z_{Constant} + \beta^z_{Cov.1}\right) - \left[g_z\left(\beta^z_{Constant} + \beta^z_{Cov.2}\right) - g_z\left(\beta^z_{Constant}\right)\right]$.

Significance level: 1% (***), 5% (**), 10% (*)

Table 2: Median estimated parameters stratified by demographics

		Scale50PrUnknown		*Scale50Pr1/15*		*Scale10Pr1/3*	
		r	w	r	w	r	w
All	Full-time students $N = (2,474 \mid 43 \mid 47)$	0.710 (0.686,0.737)	0.058 (0.037,0.097)	0.766 (0.739,0.789)	0.011 (0.007,0.027)	0.652 (0.618,0.670)	0.015 (0.009,0.035)
	Part-time students $N = (319 \mid 13 \mid 11)$	0.708 (0.600,0.790)	0.041 (0.020,0.169)	0.639 (0.504,0.822)	0.016 (0.004,0.083)	0.724 (0.649,0.846)	0.000 (0.000,0.000)
	Non-students $N = (565 \mid 4 \mid 2)$	0.692 (0.625,0.749)	0.091 (0.000,0.145)	0.683 (0.637,0.695)	0.000 (0.000,0.000)	0.849 (0.845,0.854)	0.000 (0.000,0.000)
Male	Full-time students $N = (1,491 \mid 18 \mid 19)$	0.710 (0.666,0.723)	0.037 (0.036,0.063)	0.766 (0.747,0.777)	0.007 (0.007,0.017)	0.642 (0.600,0.666)	0.015 (0.007,0.023)
	Part-time students $N = (204 \mid 4 \mid 3)$	0.699 (0.537,0.766)	0.032 (0.019,0.153)	0.618 (0.504,0.628)	0.006 (0.004,0.009)	0.715 (0.688,0.724)	0.000 (0.000,0.000)
	Non-students $N = (391 \mid 0 \mid 0)$	0.692 (0.623,0.748)	0.093 (0.000,0.146)				
Female	Full-time students $N = (983 \mid 25 \mid 28)$	0.726 (0.703,0.738)	0.059 (0.057,0.101)	0.761 (0.739,0.790)	0.016 (0.009,0.027)	0.658 (0.618,0.682)	0.021 (0.012,0.035)
	Part-time students $N = (115 \mid 9 \mid 8)$	0.726 (0.651,0.869)	0.100 (0.062,0.340)	0.644 (0.582,0.822)	0.018 (0.012,0.083)	0.731 (0.649,0.846)	0.000 (0.000,0.000)
	Non-students $N = (174 \mid 4 \mid 2)$	0.693 (0.625,0.752)	0.088 (0.000,0.142)	0.683 (0.637,0.695)	0.000 (0.000,0.000)	0.849 (0.845,0.854)	0.000 (0.000,0.000)
Youngest 25%	Full-time students $N = (829 \mid 15 \mid 23)$	0.711 (0.688,0.738)	0.058 (0.037,0.098)	0.767 (0.747,0.790)	0.011 (0.007,0.027)	0.646 (0.618,0.670)	0.015 (0.009,0.025)
	Part-time students $N = (101 \mid 6 \mid 3)$	0.726 (0.639,0.778)	0.077 (0.028,0.169)	0.649 (0.623,0.715)	0.017 (0.008,0.025)	0.750 (0.737,0.777)	0.000 (0.000,0.000)
	Non-students $N = (159 \mid 1 \mid 1)$	0.716 (0.670,0.760)	0.078 (0.000,0.104)	0.695 (0.695,0.695)	0.000 (0.000,0.000)	0.845 (0.845,0.845)	0.000 (0.000,0.000)
Middle 50%	Full-time students $N = (1,229 \mid 19 \mid 14)$	0.710 (0.679,0.726)	0.058 (0.037,0.097)	0.766 (0.723,0.780)	0.011 (0.007,0.027)	0.669 (0.600,0.682)	0.014 (0.007,0.025)
	Part-time students $N = (167 \mid 4 \mid 6)$	0.703 (0.543,0.767)	0.036 (0.022,0.153)	0.642 (0.613,0.822)	0.019 (0.004,0.083)	0.724 (0.715,0.846)	0.000 (0.000,0.000)
	Non-students $N = (272 \mid 2 \mid 1)$	0.694 (0.654,0.729)	0.092 (0.000,0.119)	0.683 (0.676,0.689)	0.000 (0.000,0.000)	0.854 (0.854,0.854)	0.000 (0.000,0.000)
Oldest 25%	Full-time students $N = (416 \mid 9 \mid 10)$	0.709 (0.665,0.725)	0.057 (0.036,0.096)	0.760 (0.739,0.789)	0.009 (0.007,0.018)	0.651 (0.623,0.665)	0.024 (0.013,0.035)
	Part-time students $N = (51 \mid 3 \mid 2)$	0.680 (0.604,0.850)	0.030 (0.016,0.230)	0.582 (0.504,0.619)	0.012 (0.004,0.018)	0.669 (0.649,0.688)	0.000 (0.000,0.000)
	Non-students $N = (134 \mid 1 \mid 0)$	0.643 (0.605,0.679)	0.124 (0.077,0.164)	0.637 (0.637,0.637)	0.000 (0.000,0.000)		
Only studying	Full-time students $N = (1,727 \mid 22 \mid 21)$	0.711 (0.710,0.737)	0.052 (0.037,0.083)	0.772 (0.766,0.790)	0.008 (0.007,0.016)	0.669 (0.652,0.682)	0.014 (0.009,0.021)
	Part-time students $N = (117 \mid 5 \mid 2)$	0.712 (0.685,0.778)	0.028 (0.019,0.093)	0.654 (0.613,0.715)	0.016 (0.004,0.025)	0.737 (0.724,0.750)	0.000 (0.000,0.000)
	Non-students $N = (5 \mid 0 \mid 0)$	0.660 (0.629,0.694)	0.000 (0.000,0.000)				
Full-time job	Full-time students $N = (12 \mid 0 \mid 0)$	0.613 (0.589,0.631)	0.144 (0.141,0.213)				
	Part-time students $N = (18 \mid 0 \mid 0)$	0.531 (0.471,0.771)	0.161 (0.090,0.432)				
	Non-students $N = (244 \mid 0 \mid 1)$	0.698 (0.621,0.722)	0.094 (0.072,0.136)			0.845 (0.845, 0.845)	0.000 (0.000, 0.000)

Continued on next page

Table 2: Continued

		Scale50PrUnknown		*Scale50Pr1/15*		*Scale10Pr1/3*	
		r	w	r	w	r	w
Part-time job	Full-time students $N = (634 \mid 17 \mid 22)$	0.689 (0.687,0.716)	0.088 (0.062,0.136)	0.761 (0.747,0.772)	0.018 (0.011,0.027)	0.643 (0.627,0.658)	0.024 (0.015,0.035)
	Part-time students $N = (165 \mid 7 \mid 8)$	0.703 (0.675,0.774)	0.046 (0.029,0.164)	0.628 (0.582,0.822)	0.018 (0.008,0.083)	0.720 (0.649,0.777)	0.000 (0.000,0.000)
	Non-students $N = (145 \mid 0 \mid 0)$	0.692 (0.625,0.724)	0.110 (0.078,0.158)				
University job	Full-time students $N = (48 \mid 1 \mid 0)$	0.644 (0.621,0.662)	0.027 (0.027,0.043)	0.723 (0.723,0.723)	0.008 (0.008,0.008)		
	Part-time students $N = (8 \mid 0 \mid 0)$	0.794 (0.768,0.851)	0.000 (0.000,0.000)				
	Non-students $N = (87 \mid 0 \mid 0)$	0.660 (0.636,0.673)	0.000 (0.000,0.000)				
Other job	Full-time students $N = (53 \mid 3 \mid 4)$	0.676 (0.662,0.692)	0.044 (0.031,0.070)	0.739 (0.739,0.739)	0.009 (0.009,0.009)	0.618 (0.600,0.619)	0.012 (0.007,0.012)
	Part-time students $N = (11 \mid 1 \mid 1)$	0.654 (0.604,0.732)	0.075 (0.029,0.132)	0.504 (0.504,0.504)	0.004 (0.004,0.004)	0.846 (0.846,0.846)	0.000 (0.000,0.000)
	Non-students $N = (84 \mid 4 \mid 1)$	0.737 (0.626,0.762)	0.076 (0.052,0.146)	0.683 (0.637,0.695)	0.000 (0.000,0.000)	0.854 (0.854,0.854)	0.000 (0.000,0.000)
Not Married	Full-time students $N = (2,420 \mid 43 \mid 45)$	0.710 (0.687,0.737)	0.058 (0.037,0.097)	0.766 (0.739,0.789)	0.011 (0.007,0.027)	0.652 (0.618,0.670)	0.015 (0.009,0.035)
	Part-time students $N = (307 \mid 12 \mid 10)$	0.703 (0.600,0.770)	0.039 (0.020,0.153)	0.634 (0.504,0.715)	0.015 (0.004,0.025)	0.724 (0.649,0.777)	0.000 (0.000,0.000)
	Non-students $N = (426 \mid 4 \mid 2)$	0.700 (0.639,0.755)	0.090 (0.000,0.137)	0.683 (0.637,0.695)	0.000 (0.000,0.000)	0.849 (0.845,0.854)	0.000 (0.000,0.000)
Married	Full-time students $N = (54 \mid 0 \mid 2)$	0.689 (0.621,0.718)	0.072 (0.033,0.117)			0.641 (0.623 0.660)	0.028 (0.025 0.030)
	Part-time students $N = (12 \mid 1 \mid 1)$	0.868 (0.685,0.897)	0.291 (0.081,0.432)	0.822 (0.822,0.822)	0.083 (0.083,0.083)	0.846 (0.846,0.846)	0.000 (0.000,0.000)
	Non-students $N = (139 \mid 0 \mid 0)$	0.662 (0.608,0.708)	0.101 (0.000,0.152)				
Budgeting	Full-time students $N = (2,033 \mid 33 \mid 33)$	0.710 (0.679,0.726)	0.058 (0.036,0.097)	0.766 (0.739,0.780)	0.011 (0.007,0.018)	0.645 (0.618,0.670)	0.014 (0.009,0.025)
	Part-time students $N = (266 \mid 12 \mid 10)$	0.703 (0.543,0.743)	0.037 (0.020,0.153)	0.634 (0.504,0.822)	0.015 (0.004,0.083)	0.724 (0.649,0.846)	0.000 (0.000,0.000)
	Non-students $N = (447 \mid 4 \mid 2)$	0.692 (0.623,0.733)	0.095 (0.000,0.146)	0.683 (0.637,0.695)	0.000 (0.000,0.000)	0.849 (0.845,0.854)	0.000 (0.000,0.000)
Not Budgeting	Full-time students $N = (441 \mid 10 \mid 14)$	0.723 (0.700,0.738)	0.082 (0.053,0.136)	0.774 (0.758,0.790)	0.016 (0.010,0.027)	0.659 (0.641,0.682)	0.022 (0.013,0.036)
	Part-time students $N = (53 \mid 1 \mid 1)$	0.759 (0.707,0.868)	0.082 (0.027,0.373)	0.715 (0.715,0.715)	0.025 (0.025,0.025)	0.777 (0.777,0.777)	0.000 (0.000,0.000)
	Non-students $N = (118 \mid 0 \mid 0)$	0.700 (0.627,0.760)	0.076 (0.000,0.121)				
Expenses: Parents	Full-time students $N = (810 \mid 14 \mid 14)$	0.717 (0.677,0.744)	0.043 (0.016,0.139)	0.772 (0.739,0.804)	0.008 (0.003,0.040)	0.646 (0.587,0.684)	0.011 (0.005,0.028)
	Part-time students $N = (66 \mid 4 \mid 0)$	0.751 (0.607,0.818)	0.049 (0.015,0.241)	0.713 (0.576,0.736)	0.028 (0.002,0.044)		
Expenses: Self	Full-time students $N = (405 \mid 2 \mid 6)$	0.722 (0.674,0.756)	0.055 (0.019,0.198)	0.785 (0.784,0.785)	0.027 (0.019,0.034)	0.631 (0.626,0.672)	0.020 (0.012,0.038)
	Part-time students $N = (116 \mid 5 \mid 6)$	0.724 (0.533,0.841)	0.063 (0.013,0.298)	0.661 (0.406,0.702)	0.030 (0.013,0.045)	0.659 (0.551,0.721)	0.000 (0.000,0.000)

Continued on next page

Table 2: Continued

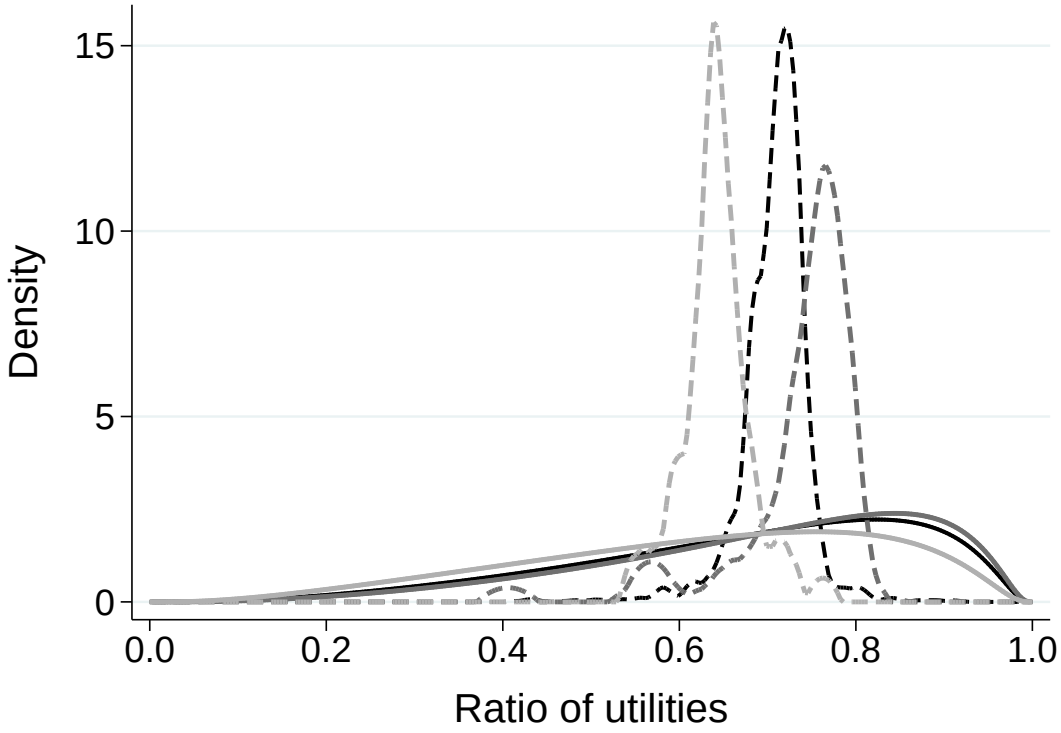
		Scale50PrUnknown		*Scale50Pr1/15*		*Scale10Pr1/3*	
		r	w	r	w	r	w
Expenses: Joint	Full-time students $N = (757 \mid 21 \mid 20)$	0.709 (0.655,0.740)	0.041 (0.016,0.147)	0.765 (0.716,0.782)	0.012 (0.004,0.035)	0.640 (0.581,0.684)	0.019 (0.009,0.063)
	Part-time students $N = (89 \mid 2 \mid 3)$	0.697 (0.519,0.775)	0.023 (0.005,0.115)	0.564 (0.556,0.572)	0.005 (0.004,0.006)	0.725 (0.663,0.763)	0.000 (0.000,0.000)
Expenses: Other	Full-time students $N = (366 \mid 6 \mid 5)$	0.705 (0.653,0.733)	0.031 (0.013,0.101)	0.763 (0.735,0.784)	0.005 (0.003,0.016)	0.602 (0.554,0.643)	0.008 (0.004,0.064)
	Part-time students $N = (31 \mid 1 \mid 1)$	0.767 (0.510,0.837)	0.023 (0.007,0.121)	0.671 (0.671,0.671)	0.016 (0.016,0.016)	0.676 (0.676,0.676)	0.000 (0.000,0.000)
Field of studies: Business	Full-time students $N = (591 \mid 16 \mid 18)$	0.685 (0.649,0.712)	0.029 (0.016,0.062)	0.748 (0.716,0.776)	0.009 (0.003,0.018)	0.628 (0.554,0.653)	0.016 (0.004,0.028)
	Part-time students $N = (75 \mid 4 \mid 5)$	0.691 (0.503,0.772)	0.037 (0.009,0.264)	0.653 (0.572,0.722)	0.027 (0.004,0.041)	0.663 (0.624,0.677)	0.000 (0.000,0.000)
Field of studies: Economics	Full-time students $N = (435 \mid 8 \mid 8)$	0.715 (0.680,0.738)	0.023 (0.012,0.054)	0.774 (0.749,0.792)	0.004 (0.003,0.011)	0.636 (0.600,0.667)	0.010 (0.007,0.023)
	Part-time students $N = (57 \mid 1 \mid 1)$	0.608 (0.432,0.704)	0.021 (0.006,0.284)	0.556 (0.556,0.556)	0.006 (0.006,0.006)	0.551 (0.551,0.551)	0.000 (0.000,0.000)
Field of studies: MNE	Full-time students $N = (450 \mid 4 \mid 4)$	0.728 (0.687,0.752)	0.044 (0.019,0.084)	0.773 (0.715,0.804)	0.008 (0.005,0.012)	0.664 (0.636,0.684)	0.018 (0.012,0.024)
	Part-time students $N = (57 \mid 2 \mid 0)$	0.749 (0.546,0.839)	0.035 (0.000,0.225)	0.683 (0.661,0.704)	0.016 (0.015,0.017)		
Field of studies: SSH	Full-time students $N = (769 \mid 15 \mid 11)$	0.722 (0.687,0.748)	0.097 (0.049,0.185)	0.776 (0.754,0.803)	0.026 (0.013,0.041)	0.647 (0.626,0.693)	0.038 (0.012,0.071)
	Part-time students $N = (103 \mid 4 \mid 4)$	0.759 (0.557,0.837)	0.065 (0.014,0.241)	0.694 (0.406,0.736)	0.037 (0.013,0.045)	0.723 (0.705,0.763)	0.000 (0.000,0.000)
Field of studies: Other	Full-time students $N = (93 \mid 0 \mid 4)$	0.706 (0.664,0.727)	0.034 (0.017,0.069)			0.649 (0.643 0.655)	0.010 (0.008 0.016)
	Part-time students $N = (10 \mid 1 \mid 0)$	0.664 (0.444,0.706)	0.015 (0.004,0.077)	0.576 (0.576,0.576)	0.002 (0.002,0.002)		
Under-graduate	Full-time students $N = (1,015 \mid 28 \mid 28)$	0.712 (0.669,0.742)	0.059 (0.024,0.170)	0.770 (0.723,0.791)	0.014 (0.004,0.040)	0.640 (0.587,0.672)	0.019 (0.008,0.064)
	Part-time students $N = (88 \mid 9 \mid 4)$	0.748 (0.597,0.808)	0.065 (0.018,0.264)	0.686 (0.556,0.736)	0.030 (0.006,0.045)	0.691 (0.551,0.725)	0.000 (0.000,0.000)
Graduate	Full-time students $N = (1,323 \mid 15 \mid 17)$	0.715 (0.660,0.747)	0.033 (0.014,0.117)	0.767 (0.716,0.804)	0.007 (0.003,0.019)	0.636 (0.554,0.693)	0.009 (0.004,0.032)
	Part-time students $N = (214 \mid 3 \mid 6)$	0.707 (0.522,0.824)	0.033 (0.006,0.225)	0.572 (0.406,0.576)	0.004 (0.002,0.013)	0.669 (0.624,0.763)	0.000 (0.000,0.000)

The table shows predicted sub-sample medians and 90% confidence intervals of the utility ratio and the trembling probability along with the size of the respective sub-sample. The total number of observations is determined by the number of observations included in, respectively, model 2 (results for gender, age, employment status, marital status, and budgeting), and model 3 (results for tuition, field of studies, and graduate level for full-time students and part-time students).

D.2 Parameter Heterogeneity and Observed Characteristics

The results from the structural econometric models establish that there are important differences between sociodemographic groups. Still, the standard deviation of r is considerably and significantly larger than zero for each model, and it is reduced only slightly when taking into account more demographic covariates. This suggests that only a small part of the overall heterogeneity can be accounted for by observed characteristics. In order to assess the extent to which this is possible, Figure 1 plots for each treatment the density of the overall distribution of r against the density of the distribution implied by the observable characteristics only. The figure is based on model 3, but similar results are obtained for model 1 and

2. The figure clearly confirms that unobserved heterogeneity is an important part of individuals' risk preferences. 90% of the utility ratios predicted for treatment *Scale50PrUnknown* (*Scale50Pr1/15* and *Scale10Pr1/3*, respectively) by a model which relies on observed heterogeneity only lie in an interval that accounts for less than 20% (42% and 25%, respectively) of the distribution of r when unobserved heterogeneity is also considered. Moreover based on observed heterogeneity alone risk-loving behaviour ($r < 0.4$) is predicted for less than 1% of the population in all three treatments.



Notes: Solid lines are the estimated parameter distributions taking observed and unobserved heterogeneity into account. Dashed lines are kernel density estimates over the predicted utility ratios when unobserved heterogeneity is neglected. The black, dark-gray, and light-gray lines picture, respectively, the distributions for treatment *Scale50PrUnknown*, *Scale50Pr1/15*, and *Scale10Pr1/3*.

Figure 1: Distributions of the ratio of utilities with and without unobserved heterogeneity

D.3 Robustness to Different Model Specifications

We here report robustness checks which account for the influence of (heterogeneity in) the stochastic choice parameters k and w. For simplicity we focus (mainly) on the specification where the vector of covariates contains only treatment dummies.

First, Table 3 presents the results of the restricted model without trembles. The model without trembles is rejected by a likelihood ratio test (the value of the test statistic is 2,575.9 which is distributed chi-square with 3 degrees of freedom) which suggests that differences in decision errors are not captured well by differences in the sensitivity to expected utility differences. Still, the estimates confirm that there are no significant differences between treatments with respect to risk aversion. Moreover, absent trembles, differences in the sensitivity to expected utility differences between the internet and the laboratory become significant. We also estimated the model with stochastic decision-making captured solely by trembles. However, the resulting discontinuity in the likelihood function renders this model specification unsuitable

for estimation with maximum simulated likelihood techniques. Indeed, from any (random) initial vector of coefficients, the model never converged and estimation steps were usually accompanied by warning messages regarding numerical derivatives and flatness of the likelihood function. Accordingly, allowing for errors in the *considered* choice probabilities is necessary to make the model amenable to maximum (simulated) likelihood techniques.

	r	k
Constant	0.691***	5.415***
	(0.003)	(0.023)
Scale50Pr1/15	0.024	4.444***
	(0.023)	(0.557)
Scale10Pr1/3	-0.030	4.147***
	(0.028)	(0.326)
Standard Deviation	0.804***	-
	(0.014)	

Number of observations is 3,702, and log-likelihood is -13,263.5. Regression coefficients are transformed back to the original scale.

Table 3: Estimated parameters for the model with treatment dummies and no trembles

Second, we consider the influence of allowing for (observed) heterogeneity in the stochastic choice parameters. Tables 4 and 5 contain the results of the 4 models (w/o and with demographic covariates, respectively) when allowing for observed heterogeneity in the sensitivity to payoff differences.

	r	k	w
Constant	0.706***	17.537***	0.063***
	(0.004)	(1.020)	(0.002)
Scale50Pr1/15	-0.001	1.4E+06	-0.042***
	(0.004)	(1.3E+11)	(0.007)
Scale10Pr1/3	-0.033	5.416	-0.048***
	(0.035)	(8.238)	(0.005)
Standard Deviation	1.024***	-	-
	(0.015)		
Log-likelihood		-11,975.8	
Observations		3,702	
p-value (LR-test)		1.000	

Regression coefficients are transformed back to the original scale. Significance level: 1% (***), 5% (**), 10% (*)

Table 4: Estimated parameters for the model with treatment dummies and observed heterogeneity in k

Likelihood-ratio tests reveal that allowing for observed heterogeneity in k significantly improves the fit of models 2 and 3, but not for the other two models.[7] In addition, few of the coefficients of the covariates of k are significantly different from zero, and the coefficients of the estimated ratio of utilities hardly change. Furthermore treatment differences in the trembling probability are barely affected.

[7] Indeed, the log-likelihood is *lower* for the model w/o demographic covariates when allowing for observed heterogeneity in k. However, small differences in the likelihood-ratio need to be treated with care, since they may be caused by the finite number of Halton draws, or a failure of the algorithm to converge fully when the likelihood-function is very flat in the neighborhood of the optimum.

Table 5: Estimated parameters for models with observed heterogeneity in k.

	All Participants						Students		
	Model 1			Model 2			Model 3		
	r	k	w	r	k	w	r	k	w
Constant	0.749***	22.309***	0.033***	0.714***	20.465*	0.038***	0.682**	30.446	0.014**
	(0.019)	(5.579)	(0.005)	(0.040)	(10.597)	(0.010)	(0.040)	(19.425)	(0.005)
Scale50Pr1/15	0.018	7.973	-0.027***	0.055	3.832	-0.031*	0.065*	3.825	-0.012
	(0.026)	(16.744)	(0.007)	(0.035)	(25.654)	(0.018)	(0.034)	(49.901)	(0.009)
Scale10Pr1/3	-0.038	9.935	-0.027***	-0.069***	2.8E+05	-0.027***	-0.069**	3.7E+04	-0.009**
	(0.032)	(16.106)	(0.004)	(0.015)	(3.8E+08)	(0.007)	(0.010)	(5.3E+06)	(0.004)
Female	0.013*	-1.931	0.017***	0.014	-2.406	0.020***	0.020**	-3.287	0.004**
	(0.007)	(2.409)	(0.003)	(0.010)	(2.841)	(0.005)	(0.008)	(4.386)	(0.002)
Age	-0.002**	-0.147	2.0E-04	-1.3E-04	-0.060	-3.9E-05	2.3E-05	-0.508	0.001**
	(0.001)	(0.247)	(1.5E-4)	(0.002)	(0.460)	(4.1E-04)	(0.002)	(1.162)	(9.1E-05)
Married	0.002	0.028	0.006	-0.029*	4.7E+02	0.018*	-0.014	12.867	0.002
	(0.019)	(6.551)	(0.005)	(0.016)	(7.2E+03)	(0.011)	(0.023)	(33.448)	(0.003)
Not budgeting	0.013	0.335	0.010***	0.013	1.080	0.017***	0.012	-0.414	0.005**
	(0.009)	(3.297)	(0.003)	(0.012)	(3.975)	(0.006)	(0.009)	(5.216)	(0.002)
Full-time job	-0.022	-1.393	0.040***	-0.089	1.4E+02	0.110**	-0.056	6.9E+02	0.088**
	(0.017)	(5.033)	(0.009)	(0.118)	(7.4E+03)	(0.043)	(0.126)	(4.7E+04)	(0.044)
Part-time job	-0.017*	-3.029	0.022***	-0.023**	-1.344	0.025***	-0.023**	-0.342	0.008**
	(0.009)	(2.547)	(0.004)	(0.011)	(3.007)	(0.007)	(0.009)	(4.479)	(0.003)
University job	-0.043**	3.443	-0.025***	-0.052	2.922	-0.010	-0.114**	6.3E+04	-0.005
	(0.021)	(7.390)	(0.004)	(0.044)	(31.487)	(0.025)	(0.013)	(9.1E+06)	(0.003)
Other job status	-0.013	-1.684	0.004	-0.034	22.518	-0.006	-0.055*	24.838	-4.2E-04
	(0.018)	(4.922)	(0.005)	(0.025)	(2.2E+02)	(0.007)	(0.032)	(1.9E+02)	(0.004)
Payment expenses: Self							0.012	-8.423	0.001
							(0.011)	(6.068)	(0.002)

Continued on next page

Table 5: Continued

	All Participants						Students		
	Model 1			Model 2			Model 3		
	r	*k*	*w*	*r*	*k*	*w*	*r*	*k*	*w*
Payment expenses: Shared							-2.0E-04	-3.207	-7.2E-05
							(0.009)	(5.171)	(0.001)
Payment expenses: Other							-0.018	0.299	-0.004**
							(0.014)	(6.465)	(0.002)
Semester							-0.001	0.208	-0.001**
							(0.002)	(0.880)	(4.6E-04)
Graduate							0.014	5.617	-0.001
							(0.012)	(7.845)	(0.001)
Economics							0.032**	-1.211	-0.003
							(0.012)	(5.679)	(0.002)
MNE							0.044**	-2.994	0.007**
							(0.012)	(5.768)	(0.003)
SSH							0.035**	-5.400	0.027**
							(0.011)	(6.338)	(0.009)
Other field of study							0.016	-6.829	-0.001
							(0.017)	(8.825)	(0.004)
Part-time student				0.096	-18.158*	-0.032***	-0.063	49.174	0.204
				(0.081)	(10.751)	(0.012)	(0.133)	(92.944)	(0.476)
PT student × Scale50Pr1/15				-0.157**	2.9E+04	0.027	-0.191	6.2E+02	-0.103
				(0.075)	(7.0E+07)	(0.019)	(0.117)	(2.6E+03)	(0.244)
PT student × Scale10Pr1/3				0.040	-2.8E+05	0.021*	0.060	5.9E+06	-0.209
				(0.066)	(3.8E+08)	(0.011)	(0.043)	(2.9E+10)	(0.615)
PT student × Female				0.002	1.533	-0.013	0.012	-62.496	-0.100
				(0.024)	(2.957)	(0.009)	(0.038)	(77.633)	(0.199)

Continued on next page

Table 5: Continued

	All Participants						Students		
	Model 1			Model 2			Model 3		
	r	k	w	r	k	w	r	k	w
PT student × Age				-0.004	0.344	0.001	0.005	1.230	-0.007
				(0.002)	(0.488)	(4.7E-04)	(0.006)	(3.001)	(0.029)
PT student × Married				0.139***	-4.8E+02	-0.022**	0.234**	-64.970	-0.219
				(0.043)	(7.2E+03)	(0.011)	(0.080)	(67.888)	(0.921)
PT student × Not budgeting				0.018	-2.143	-0.022***	0.046	-10.287	-0.110
				(0.031)	(4.084)	(0.008)	(0.043)	(28.289)	(0.246)
PT student × Full-time job				-0.037	-1.4E+02	-0.088*	-0.132	-7.6E+02	0.023
				(0.151)	(7.4E+03)	(0.049)	(0.165)	(4.7E+04)	(0.198)
PT student × Part-time job				0.019	-0.074	-0.028***	-0.010	-44.427	-0.092
				(0.022)	(3.273)	(0.008)	(0.035)	(56.886)	(0.158)
PT student × University job				0.117*	-4.074	0.004	0.188**	-6.3E+04	-0.213
				(0.067)	(31.524)	(0.029)	(0.080)	(9.1E+06)	(0.707)
PT student × Other job status				-0.050	-24.160	-3.5E-04	-0.085	7.7E+02	0.407
				(0.073)	(2.2E+02)	(0.012)	(0.118)	(1.3E+05)	(0.307)
PT student × Payment expenses: Self							-0.017	8.692	0.152
							(0.048)	(28.781)	(0.223)
PT student × Payment expenses: Shared							-0.010	37.298	-0.069
							(0.044)	(55.035)	(0.155)
PT student × Payment expenses: Other							0.086	-8.873	-0.144
							(0.055)	(32.768)	(0.371)
PT student × Semester							-0.002	13.222	0.005
							(0.005)	(16.764)	(0.011)
PT student × Graduate							-0.064	46.075	-0.012
							(0.043)	(93.963)	(0.093)

Continued on next page

Table 5: Continued

	All Participants						Students		
	Model 1			Model 2			Model 3		
	r	k	w	r	k	w	r	k	w
PT student × Economics							-0.079	-56.393	-0.185
							(0.049)	(72.354)	(0.394)
PT student × MNE							0.022	-55.492	-0.119
							(0.048)	(71.191)	(0.237)
PT student × SSH							0.052	-66.656	-0.179
							(0.042)	(84.395)	(0.291)
PT student × Other field of study							-0.039	1.1E+04	-0.040
							(0.062)	(2.5E+06)	(0.173)
Non-student				0.024	-17.488	-0.038			
				(0.232)	(2.4E+02)	(0.032)			
Non-student × Scale50Pr1/15				-0.060	7.1E+02	*0.031*			
				(0.061)	(2.3E+05)	-			
Non-student × Scale10Pr1/3				0.221	-2.1E+05	*0.027*			
				(0.157)	(2.8E+08)	-			
Non-student × Female				-0.018	5.932	-0.020*			
				(0.020)	(2.8E+02)	(0.011)			
Non-student × Age				-0.003	0.473	4.1E-05			
				(0.002)	(33.055)	(1.6E-03)			
Non-student × Married				0.036	-4.7E+02	-0.018			
				(0.027)	(7.2E+03)	(0.013)			
Non-student × Not budgeting				-0.006	11.413	-0.017*			
				(0.023)	(1.0E+03)	(0.010)			
Non-student × Full-time job				0.138	-1.4E+02	-0.088*			
				(0.257)	(7.4E+03)	(0.053)			

Continued on next page

Table 5: Continued

	All Participants						Students		
	Model 1			Model 2			Model 3		
	r	k	w	r	k	w	r	k	w
Non-student × Part-time job				0.071	-1.446	-0.009			
				(0.228)	(2.4E+02)	(0.031)			
Non-student × University job				0.073	-5.363	*0.010*			
				(0.233)	(2.4E+02)	-			
Non-student × Other job status				0.089	-25.212	0.019			
				(0.229)	(3.2E+02)	(0.031)			
Standard deviation	1.017***	-	-	1.026***	-	-	0.986**	-	-
	(0.016)			(0.016)			(0.016)		
Log-likelihood		-11,132.3			-11,061.9			-8,391.1	
Observations		3,478			3,478			2,750	
p-Value (H0: k=const.)		0.320			0.045			0.000	

Regression coefficients are transformed back to the original scale, i.e. the constants are given by $g_z\left(\beta^z_{\text{Constant}}\right)$, where $g_z(\cdot) = \Lambda(\cdot)$, $z = r, w$, and $g_k(\cdot) = \exp(\cdot)$, and represent the median parameters in treatment $Scale50PrUnknown$. Effects of standard covariates are $g_z\left(\beta^z_{\text{Constant}} + \beta^z_{\text{Covariate}}\right) - g_z\left(\beta^z_{\text{Constant}}\right)$. Interaction effects are $g_z\left(\beta^z_{Constant} + \beta^z_{Cov.1} + \beta^z_{Cov.2} + \beta^z_{Cov.1\times Cov.2}\right) - g_z\left(\beta^z_{Constant} + \beta^z_{Cov.1}\right) - \left[g_z\left(\beta^z_{Constant} + \beta^z_{Cov.2}\right) - g_z\left(\beta^z_{Constant}\right)\right]$.

Significance level: 1% (***), 5% (**), 10% (*)

	Homogeneous w		Homogeneous k and w
	r	k	r
Constant	0.706***	17.436***	0.706***
	(0.004)	(0.974)	(0.004)
Scale50Pr1/15	0.024	9.273	0.023
	(0.028)	(17.824)	(0.029)
Scale10Pr1/3	-0.073***	5.0E+05	-0.033
	(0.008)	(1.1E+09)	(0.036)
Standard Deviation	1.023***	-	1.021***
	(0.014)		(0.015)
Constant k	-		17.636***
			(1.008)
Constant w	0.061***		0.061***
	(0.002)		(0.002)
Log-likelihood	-11,986.2		-11,987.0
Observations	3,702		3,702
p-value (LR-test)	1.7E-4		3.0E-5

Regression coefficients are transformed back to the original scale.
Significance level: 1% (***), 5% (**), 10% (*)

Table 6: Estimated parameters for models with treatment dummies and no heterogeneity in errors

Still, allowing k to vary with demographics refines some results for the trembling probability. Accordingly, decision errors of some sub-samples of subjects are better captured by a smaller trembling probability and a lower sensitivity to payoff differences (see e.g. the coefficient of *Part-time Student* in model 2), or vice versa (e.g. *Part-time Student* $\times$ *Scale50Pr1/15* in models 2 and 3).

Finally, Table 6 presents the results of two model specifications where the first restricts w to be homogeneous across subjects and the second restricts both k and w to be homogeneous across subjects. Only treatment dummies are included in the vector of covariates. Both specifications confirm the results regarding treatment effects on risk preferences. On the other hand when only w is restricted to be homogeneous there are no treatment differences with respect to the sensitivity to payoff differences, and the model is rejected by a likelihood ratio test (the value of the test statistic is 20.83 which is distributed chi-square with 2 degrees of freedom). When both k and w are assumed homogeneous, the estimated trembling probability is heavily distorted towards the estimated trembling probability for treatment *Scale50PrUnknown* and the model is rejected by a likelihood ratio test (the test statistic is 22.35 which is distributed chi-square with 4 degrees of freedom).

The results strongly confirm our claim that treatment differences between the laboratory and the internet are best captured by differences in the trembling probabilities. On the other hand randomness in the considered choice probabilities is important for estimation purposes since it generates a reasonably smooth likelihood function. Our results therefore support the interpretation of Loomes, Moffatt, and Sugden (2002) that trembles are errors that individuals can learn to avoid whereas the insensitivity to differences in expected utilities results from the imprecision of preferences which is an inherent and stable property of preferences.[8]

[8]Notice however, that Loomes, Moffatt, and Sugden (2002) mainly argue about the random preferences model.

References

Andersen, S., G. Harrison, M. Lau, and E. Rutström (2010): "Preference Heterogeneity in Experiments: Comparing the Field and Laboratory," *Journal of Economic Behavior and Organization*, 73, 209–224.

Harrison, G., E. Johnson, M. McInnes, and E. Rutström (2005): "Risk Aversion and Incentive Effects: Comment," *American Economic Review*, 95, 897–901.

Hey, J., and C. Orme (1994): "Investigating Generalisations of Expected Utility Theory using Experimental Data," *Econometrica*, 62, 1291–1326.

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.

Train, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press, New York.

von Gaudecker, H., A. van Soest, and E. Wengström (2011): "Heterogeneity in Risky Choice Behavior in a Broad Population," *American Economic Review*, 101, 664–694.

Wilcox, N. T. (2008): "Stochastic Models for Binary Discrete Choice under Risk: A Critical Primer and Econometric Comparison," in *Risk Aversion in Experiments, Research in Experimental Economics, Volume 12*, ed. by J. Cox, and G. Harrison, pp. 197–292. Emerald Group Publishing Limited.