Monetary Incentives in Large-Scale Experiments: A Case Study of Risk Aversion*

Christoph March, Anthony Ziegelmeyer, Ben Greiner, and René Cyranek†

February 13, 2015

Abstract

Though lottery-based incentives are common in behavioral interventions such as fundraising campaigns or large-scale economic experiments, their effectiveness in motivating individuals remains largely unexplored. We rely on a large sample of participants with a diverse range of socioeconomic characteristics to compare the incentive effects of lotteries depending on the size of prizes and the payment likelihood. We find that the payment likelihood only mildly dilutes incentive effects which significantly vary with the education status of participants. For full-time students lotteries with large prizes and a small payment likelihood induce stronger incentive effects than lotteries of identical expected value but with moderate prizes and payment likelihood.

Keywords: Random incentive system, Incentives, Between-subjects design, Internet, Experimental methodology, Risk aversion, Stochastic choice, Decision error

JEL Classification: C90, C91, D12, D81

*Financial support by the Max Planck Society (MPG) and the German Science Foundation (DFG) are gratefully acknowledged. We thank Nathaniel Wilcox for enlightening discussions, Martin Kocher, David Laibson, Sergey Popov, Klaus Schmidt, Joachim Winter, participants of the Economics seminar at Queen's University Management School (Belfast) and other conferences for helpful comments.

†March: TUM School of Management, Technische Universität München, Arcisstrasse 21, 80333 Munich, Germany. Email: `christoph.march AT wi.tum.de`. Ziegelmeyer: Queen's University Management School, Queen's University Belfast, 185 Stranmillis Road, Belfast BT9 5EE, United Kingdom. Email: `email AT anthonyziegelmeyer.com`. Greiner: University of New South Wales, School of Economics, UNSW Sydney NSW 2052, Australia. Email: `bgreiner AT unsw.edu.au`. Cyranek: Ludwig-Maximilians-Universität München, Munich Experimental Laboratory for Economic and Social Sciences (mMlessa), Giselastrasse 10, D-80802 München, Germany. Email: `rene.cyranek AT lmu.de`.

1 Introduction

About 70 percent of US firms have an employee referral program as recruiting tool and, in addition to cash rewards, firms often award a lottery ticket for every qualified applicant that an employee submits (CareerBuilder, the global leader in human capital solutions, argues that lotteries constitute a cost-effective solution to incentivize employees to participate in referral programs; see CareerBuilder, 2012). Lottery-based incentives are also common in health promotion plans, fundraisers for charities, and experiments with large samples of participants. Indeed, 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).[1] For a given research budget and holding nominal payoffs constant, BRIS allows for larger samples which may come at the cost of lowering participants' task motivation because of reduced expected payouts per capita. Despite their increasing popularity, there is mixed empirical support for the effectiveness of lotteries in behavioral interventions and experimental research has provided little guidance as to which features of BRIS matter most for motivating participants.[2] In an attempt to fill this void, the present study clarifies the relative impact of the size of prizes and the payment likelihood of lottery-based incentives on the motivation of a large sample of participants with a diverse range of socioeconomic characteristics.

We report a large-scale experiment designed to compare the effects of monetary incentives in three BRISs which differ in their nominal payoffs and/or their selection probability of paid subjects. Subjects complete the risk elicitation task developed by Holt and Laury (2002, HL hereafter) in three incentive treatments, each of which uses a different BRIS to implement monetary incentives. 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 significant incentive effects in HL's task, the degree of risk aversion measured in our incentive treatments is a sound indicator of the effects of monetary incentives implemented by BRIS.

In incentive treatment *Scale50PrUnknown*, 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 payoff of a *randomly selected* risk neutral subject is (about) 120 Euros. To operationalize the random

[1] With the help of large-scale experiments, economists are able to learn about the extent to which results from student pools can be generalized to other subject pools (Bosch-Domènech, Montalvo, Nagel, and Satorra, 2002; Güth, Schmidt, and Sutter, 2007), infer the distribution of economically important preference parameters in a broad heterogeneous population (Andersen, Harrison, Lau, and Rutström, 2008; Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner, 2011), and characterize the variation in treatment effects across subpopulations which in turn enables policy makers to derive an optimal policy of which treatment should be given to whom (see Bellemare, Kröger, and Van Soest (2008) and Hermann, Thoni, and Gächter (2008) on the importance of variations of behavior across subgroups within a given society and across societies, respectively).

[2] In regard to encouraging healthy behavior change, Volpp, John, Troxel, Norton, Fassbender, and Loewenstein (2008) document that lotteries which provide frequent small and infrequent large prizes are as effective as deposit contracts in significantly reducing short-term weight; in contrast, Niza, Rudisill, and Dolan (2014) find that offering participants a £5 voucher is much more likely to result in the screening of a sexually transmitted infection than offering participants a £200 lottery with unknown likelihood of payment. In regard to raising funds for public financing, Onderstal, Schram, and Soetevent (2013) conclude that an anonymous voluntary contribution mechanism is more effective for charitable giving than a lottery with multiple moderate prizes in a private value setting whereas Landry, Lange, List, Price, and Rupp (2006) find the opposite result in a common value setting where the lottery awards a single and large prize. Among others, Drehmann, Oechssler, and Roider (2005) and Harrison, Lau, and Rutström (2007) provide indicative evidence on the motivation effectiveness of BRIS (more details about the experimental literature are to be found in our conclusion).

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, expected payoffs per paired lottery choice are identical in the two laboratory treatments but the expected payoff 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. Section 2 describes the experimental design and Appendix A provides the demographic questionnaire along with the experimental instructions.[3]

Our experiment clarifies the relative impact of the two constituent elements of BRIS, the scale of nominal payoffs and the level of the selection probability, on the effects of monetary incentives. First, by comparing the degree of risk aversion in *Scale50Pr1/15* to the one in *Scale10Pr1/3* we investigate whether the scale of nominal payoffs is more effective in influencing subjects' behavior than the level of the selection probability since the two treatments have identical expected payoffs per lottery choice. Second, by comparing risk aversion in *Scale50Pr1/15* and *Scale50PrUnknown* we examine the impact of the selection probability on the effects of monetary incentives for a given scale of nominal payoffs. In *Scale50Pr1/15* the level of the publicly known selection probability is in line with earlier large-scale economic experiments whereas in *Scale50PrUnknown* subjects are most likely under the impression that the chance of real payment is much lower though they are uninformed of the actual selection probability. Third, by comparing risk aversion in *Scale50PrUnknown* and *Scale10Pr1/3* we explore whether a BRIS with large prizes and an unknown and (likely to be perceived as) tiny probability of payment induces monetary incentives as effective as those induced by a BRIS with moderate prizes and a known and large probability of payment. We further examine how the selection probability impacts on the effects of monetary incentives by comparing our risk aversion degrees to those in previous studies where subjects complete HL's task and all of them are paid. Finally, the presence of 637 non-students in the Internet sample allows us to check whether incentive effects can be generalized to non-student pools.

Besides nominal payoffs and/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 (age, gender, education level, employment and marital status, etc) and we control for this observed heterogeneity in our

[3]Like most large-scale economic experiments, we combine BRIS with the *within-subjects random incentive system* where each subject performs a series of individual tasks knowing that only one of these tasks will be randomly selected for real payment (see Laury, 2012, for evidence on the motivation effectiveness of the within-subjects random incentive system). Baltussen, Post, van den Assem, and Wakker (2012) refer to such a combination as a *hybrid* RIS. Our study compares the effects of monetary incentives in three hybrid RISs. Since the selection probability of the payoff-relevant decision is 10% in each treatment, we expect differences in the effects of monetary incentives to originate from the variation in the scale of nominal payoffs and the level of the between-subjects selection probability. Note also that HL's task is susceptible to induce cross-task contamination effects but investigating the extent of cross-task contamination effects in hybrid RISs is left for future research.

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 whereas errors in the execution of risk preferences are expected to pick up differences in the implementation mode as subjects are more likely to temporarily lose concentration in the Internet environment than in the controlled laboratory setting. Our estimation approach follows von Gaudecker, van Soest, and Wengström (2012) who find no effects of the implementation mode on risk preferences when controlling for the difference in demographics between the Internet and the laboratory.[4] Structural estimation results are also informative about the individual determinants of decision errors. Section 3 presents the model of decision under risk in HL's task along with its econometric implementation and Appendix B provides additional illustrations on the inference of risk aversion from choices.

We find a strong tendency toward 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. We compare our findings with those of a previous study where subjects complete HL's task and all of them are paid (Harrison, Johnson, McInnes, and Rutström, 2005, HJMR hereafter) and we conclude that risk aversion in *Scale50Pr1/15* and *Scale50PrUnknown* (resp. *Scale10Pr1/3*) is significantly higher than in HJMR's 1× payoff scale treatment in a strong (resp. weak) statistical sense. We also fail to reject the hypothesis that risk aversion in HJMR's 10× payoff scale treatment equals the one estimated in any of our incentive treatments at any conventional significance level. 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 attitudes, as well as several demographic effects, significantly vary with the education status of subjects. For non-students and part-time students, we never reject the hypothesis that risk aversion is identical in each incentive treatment at any standard significance level. For full-time students, risk aversion in *Scale50Pr1/15* is significantly larger than in *Scale10Pr1/3* at the 5 percent level, it is significantly larger than in *Scale50PrUnknown* at the 10 percent level, and there is no significant difference between the latter two treatments. 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. Section 4 reports our main results and Appendices C, D, and E complement the analysis presented in the main text.

Section 5 concludes and Appendix F contains a short review of the literature on random incentive systems.

[4] On the other hand, Hergueux and Jacquemet (2014) implement HL's 10× payoff scale treatment both online and in the laboratory with the same subject pool and they observe significantly less risk aversion and more inconsistencies online.

2 Experimental Design and Procedures

Our experimental design consists of three incentive treatments in which subjects complete HL's task and each of which uses a different BRIS to motivate them.

In the two laboratory treatments, the (between-subjects) selection probability is public knowledge. In *Scale50Pr1/15*, lottery outcomes in Euros equal those in HL's 50× payoff scale treatment 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. Expected payoffs per paired lottery 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.

Scale50PrUnknown relies on a large Internet sample of 3,582 subjects of more diverse demographic background than the two laboratory samples. Lottery outcomes are identical to those in *Scale50Pr1/15* and subjects know that five of them are randomly selected to receive real payment but they don't know the sample size. *Scale50PrUnknown* permits us, therefore, to assess heterogeneity of behavior at reasonable research costs, similarly to most large-scale economic experiments, and it also constitutes a challenging environment for the motivation effectiveness of BRIS. Though the actual selection probability is unknown, subjects are most likely under the impression that the chance of real payment is tiny. Indeed, subjects are fully made 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 2.2 below).[5]

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" 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 row (we did not allow subjects to express indifference), and one row 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 offers a 10% chance of receiving € 100 and a 90% chance of receiving € 80. Similarly, the risky lottery in the first row 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 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 Appendix A).

[5]The risk elicitation task was the first of three tasks that participants 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 participants 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	{€ 100, 0.1 ; € 80, 0.9}	{€ 192.50, 0.1 ; € 5, 0.9}	€ 58.25
2	{€ 100, 0.2 ; € 80, 0.8}	{€ 192.50, 0.2 ; € 5, 0.8}	€ 41.50
3	{€ 100, 0.3 ; € 80, 0.7}	{€ 192.50, 0.3 ; € 5, 0.7}	€ 24.75
4	{€ 100, 0.4 ; € 80, 0.6}	{€ 192.50, 0.4 ; € 5, 0.6}	€ 8.00
5	{€ 100, 0.5 ; € 80, 0.5}	{€ 192.50, 0.5 ; € 5, 0.5}	−€ 8.75
6	{€ 100, 0.6 ; € 80, 0.4}	{€ 192.50, 0.6 ; € 5, 0.4}	−€ 25.50
7	{€ 100, 0.7 ; € 80, 0.3}	{€ 192.50, 0.7 ; € 5, 0.3}	−€ 42.25
8	{€ 100, 0.8 ; € 80, 0.2}	{€ 192.50, 0.8 ; € 5, 0.2}	−€ 59.00
9	{€ 100, 0.9 ; € 80, 0.1}	{€ 192.50, 0.9 ; € 5, 0.1}	−€ 75.75
10	{€ 100, 1.0 ; € 80, 0.0}	{€ 192.50, 1.0 ; € 5, 0.0}	−€ 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*

2.2 Experimental Procedures and Participants

To recruit the participants 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 http://www.torlabor.de.[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 participants 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, participants 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, participants had to answer a questionnaire to gather information on their demographic characteristics like gender, year of birth, education and marital status, etc (see Appendix A for details).[9]

All 8 sessions of the two laboratory treatments were conducted at the Cologne Laboratory for Economic Research (CLER) with 15 participants 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

[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, 2004) to invite their participants.

[7]82% of our participants 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 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 participant because of multiple registrations.

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

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 participant received a show-up fee of €2.50.

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

Samples of Participants

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

Though women account for less than 40% of the participants in *Scale50PrUnknown*, there is a larger fraction of women than men in the laboratory treatments (slightly above 60%). The main reason is that participants 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 participants are full-time students in each incentive treatment. Two-thirds of the remaining participants 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 participants in the laboratory treatments (79% and 82% of the remaining participants 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 participants in the Internet sample are older than any participant in the laboratory samples, and almost 300 participants in the Internet sample are employed full time compared to 1 participant in the laboratory samples.

3 Expected Utility in HL's Risk Elicitation Task

Our core theory of choice is expected utility (EU): Subjects assess the relative value of the safe lottery relative to the risky lottery by comparing their expected utilities. 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.

We restrict ourselves to EU decision-making for two main reasons: i) We share a common belief 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 (Abdellaoui, Driouchi, and L'Haridon, 2011); and ii) popular models that relax the independence axiom cannot be estimated consistently using BRIS (Harrison and Swarthout, 2014). In contrast, each of the ten choices made by an EU subject in our treatments coincides with the choice made if facing only the corresponding decision by itself.

	Scale50PrUnknown	*Scale50Pr1/15*	*Scale10Pr1/3*
Number of participants	3,582	60	60
Females	0.37	0.63	0.63
Age	25.61	24.52	24.40
($N = 3474 \mid 60 \mid 60$)	(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 participants and age, all entries are percentages. For the variable *Age*, the first row reports the average while the second row reports the minimum, the 1^{st} quartile, the median, the 3^{rd} quartile, and the maximum. To indicate their employment situation participants could choose one of the following options: full-time employed; part-time employed; self-employed; unemployed; employed at the university; only student; or other. In *Scale50PrUnknown*, participants 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 participants

3.1 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.

Formally, for any given sequence of choices, we denote by $d_{last\,S} \in \{1, \ldots, 10\}$ the largest decision in which the safe lottery is chosen and such that all previous choices are safe, and we denote by $d_{first\,R} \in \{1, \ldots, 10\}$ the smallest decision in which the risky lottery is chosen and such that all subsequent choices are risky. As a convention, we set $d_{last\,S}$ equal to zero if the risky lottery is chosen in the first decision and we set $d_{first\,R}$ equal to zero if the safe lottery is chosen in the last decision. Given a pair $(d_{last\,S}, d_{first\,R})$, we infer that the ratio belongs to the interval $(\min\{d_{last\,S}, 9\}/10, d_{first\,R}/10)$ if $d_{first\,R} \geqslant 1$, and that the ratio is strictly greater than $\min\{d_{last\,S}, 9\}/10$ otherwise. Thus, if all choices are safe the ratio is strictly greater than 9/10, and if all choices are risky but the first and third ones the ratio belongs to the interval (1/10, 4/10). Additional illustrations of the inferred bounds on the ratio are provided in Appendix B.

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 in a stochastic specification of choice under risk.

3.2 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.[10]

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

[10] 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.

[11] 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).

3.3 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

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

Our econometric model assumes that all parameters vary with observable characteristics (*observed heterogeneity*), and that the ratio of utilities, r, additionally varies with unobservable characteristics (*unobserved heterogeneity*). In Appendix E we present the general structural model which allows for observed and unobserved heterogeneity in all three parameters and we justify the use of the restricted model. Taking into account the interval restriction of the parameters we therefore have

$$r_i = \Lambda \left(\mathbf{x}_i\, \boldsymbol{\beta}^r + \tilde{\zeta}_i^r \right) = 1 / \left[1 + \exp \left(- \left(\mathbf{x}_i\, \boldsymbol{\beta}^r + \tilde{\zeta}_i^r \right) \right) \right] \tag{5}$$

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 into account these specifications the likelihood function is given by

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

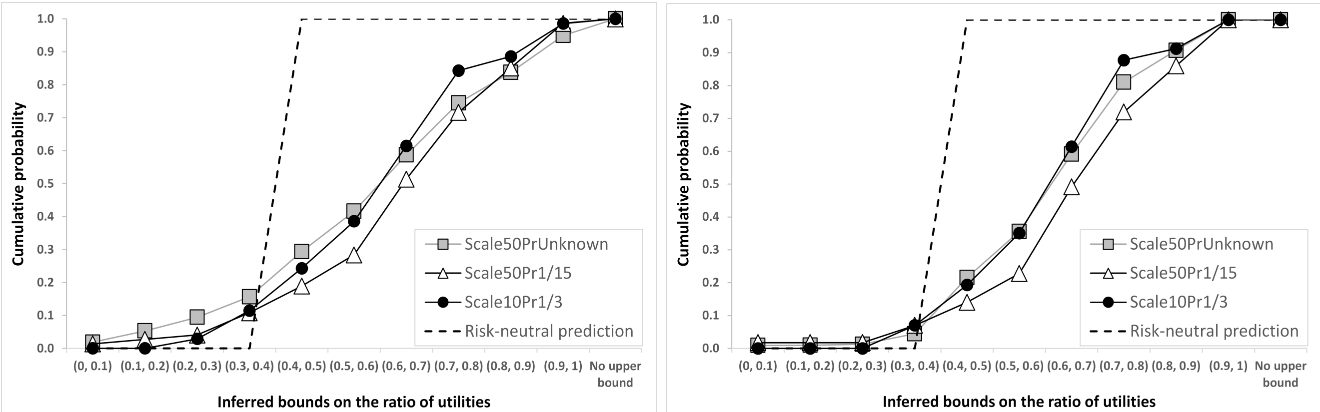
where $\phi_{\sigma_r}(\cdot)$ is the density of the (unidimensional) normal distribution with mean zero and standard deviation σ_r. The integral in (6) does not possess an analytical solution and we approximate it using standard simulation techniques. Concretely, we construct a sequence of $J = 1{,}000$ shuffled Halton draws per individual and we maximize 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.[12]

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

4 Results

4.1 Descriptive Evidence

Figure 1 plots the cumulative distribution of inferred bounds on the ratio of utilities in the different treatments, separately for all sequences of choices (left panel) and for the subset of consistent sequences of choices (right panel). If the sequence of choices is consistent, the ratio belongs to exactly one of the intervals $\left(\frac{i}{10}, \frac{i+1}{10}\right)$ with $i \in \{0, \ldots, 9\}$. If the sequence of choices is inconsistent, the ratio belongs to several intervals and whenever the safe lottery is chosen in the last decision no upper bound is inferred.



(a) All choice sequences

(b) Consistent choice sequences

Figure 1: Cumulative distributions of inferred bounds on the ratio of utilities

Figure 1 conveys several observations. First, cumulative distributions of inferred bounds on the ratio of utilities indicate that there is substantial heterogeneity in attitudes towards risk. About 9% (resp. 4% and 3%) of the ratios are smaller than 0.3 while about 25% (resp. 28% and 16%) of the ratios are larger than 0.8 in *Scale50PrUnknown* (resp. *Scale50Pr1/15* and *Scale10Pr1/3*) when all choice sequences are considered. Second, most inferred intervals of the ratio lie to the right of the risk-neutral interval $(0.4, 0.5)$ showing a clear tendency toward risk-averse behavior among subjects in each incentive treatment. Third, the (cumulative) distribution of bounds on the ratio in treatment *Scale50Pr1/15* (almost) first-order stochastic dominates the two other distributions whether all or only consistent sequences of choices are considered. This observation suggests that the degree of risk aversion is the largest in treatment *Scale50Pr1/15*. Fourth, the distributions of ratio bounds for all sequences of choices and for the consistent sequences of choices are very similar in treatments *Scale50Pr1/15* and *Scale10Pr1/3*. This observation suggests that the stochastic component of choice under risk is rather small in laboratory treatments. Finally, the distributions of ratio bounds in treatments *Scale50PrUnknown* and *Scale10Pr1/3* are very similar for consistent sequences of choices but when considering all sequences of choices the latter is more concentrated on intervals $(0.5, 0.6)$, $(0.6, 0.7)$ and $(0.7, 0.8)$. These observations suggest that the degree of risk aversion is comparable in treatments *Scale50PrUnknown* and *Scale10Pr1/3* and that decision errors are larger in the Internet than in the laboratory treatment.

Appendix C shows the distributions of inferred bounds on the ratio for subsamples of the different treatments. Distributions of ratio bounds for students are very similar to those for the entire sample in the laboratory treatments (where the proportion of non-students is at most 7%). On the other hand, distributions of ratio bounds shift slightly to the right in *Scale50PrUnknown* when the sample is restricted to students which suggests that the degree of risk aversion is larger for students than for non-students in the Internet treatment (where non-students constitute about 18% of the entire sample).

4.2 Demographic and Treatment Effects on Risk Aversion

Since the Internet sample offers a wider range of demographic characteristics than the laboratory samples, we need to account for the observed preference heterogeneity in our statistical analysis. To determine the joint role of demographics and experimental treatments on risk attitudes simultaneously, we estimate interval regression models of the ratio of utilities that condition on individual characteristics and treatments. Table 3 reports results from three interval regression models of ratio values for all choice sequences (left panel) and for the subset of consistent choice sequences (right panel) where coefficients are marginal effects. We rely on the full sample of participants 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 participant is in charge of budgeting or not. Model 2 enables us to distinguish between estimated ratios for participants 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, participants with missing values for the included variables are omitted. Table 2 in Appendix D reports the estimates of all explanatory variables.

At mean demographic values, the estimated ratio in model 1 is much larger than the ratio of a risk-neutral subject in each incentive treatment—the estimated ratio equals at least 0.65 whether all or only consistent sequences of choices are considered (see also Table 1 in Appendix D)—which confirms the strong tendency toward risk-averse behavior among subjects. Regression results of models 2 and 3 indicate that female full-time students are significantly more risk-averse than male full-time students when all choice sequences are considered. The gender effect weakens for consistent choice sequences which suggests more decision errors among female than male full-time students. On the other hand, whether all or only consistent choice sequences are considered, we do not reject the hypothesis that female and male have the same ratio both for part-time students and non-students (p-values > 0.1). We also find a strongly significant negative effect on risk aversion from age for non-students whether all or only consistent choice sequences are considered (p-values < 0.01). Given the low range of age values among students, it is not surprising that we do not find a significant effect from age on risk aversion both for part-time and full-time students (p-values > 0.1). Complementary results reported in Table 3 in Appendix D show that the significant age effect is largely the consequence of less risk averse choices made by non-students older than most students. Along with age and gender, employment status is a demographic characteristic whose marginal effects significantly depend on the participants' education status. We find that the employment situation of non-students does not significantly impact their risk attitudes. Full-time students who are employed at the university, most likely as research or teaching assistant, are significantly less risk-averse than full-time students who only study when all choice sequences are considered but the effect weakens for consistent choice sequences. Part-time students with a full-time job are significantly less risk-averse than part-time students who only study whether all or only consistent choice sequences are considered. Finally, whether participants are primarily in charge of budget decisions or not as well as participants' marital status don't significantly affect risk attitudes (when all choice sequences are considered, married part-time students are significantly more risk-averse than non-married ones according to model 2 but the effect becomes insignificant with the additional controls in model 3).

	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	0.712***	0.671***	0.650***	0.706***	0.661***	0.647***
	(0.018)	(0.035)	(0.041)	(0.018)	(0.034)	(0.040)
Scale50Pr1/15	0.015	0.045	0.052*	0.033	0.057**	0.065**
	(0.024)	(0.028)	(0.028)	(0.023)	(0.027)	(0.026)
Scale10Pr1/3	−0.027	−0.044	−0.044	−0.012	−0.025	−0.024
	(0.024)	(0.027)	(0.027)	(0.023)	(0.026)	(0.026)
Female	0.015**	0.017**	0.016**	0.014**	0.013*	0.011
	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)
Age	−0.002**	7E-05	−8E-05	−0.002***	−2E-04	−7E-04
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Part-time student		0.051	0.032		−0.041	−0.012
		(0.084)	(0.096)		(0.086)	(0.095)
PT student x *Scale50Pr1/15*		−0.132**	−0.146**		−0.106*	−0.130**
		(0.060)	(0.063)		(0.058)	(0.061)
PT student x *Scale10Pr1/3*		0.057	0.028		0.032	0.028
		(0.063)	(0.065)		(0.061)	(0.062)
PT student x *Female*		0.004	0.002		0.006	0.004
		(0.024)	(0.025)		(0.023)	(0.025)
PT student x *Age*		−0.002	−1E-04		0.002	0.002
		(0.003)	(0.004)		(0.003)	(0.004)
Non-student		0.049			0.078	
		(0.093)			(0.088)	
Non-student x *Scale50Pr1/15*		−0.041			−0.013	
		(0.098)			(0.092)	
Non-student x *Scale10Pr1/3*		0.162			0.161	
		(0.133)			(0.124)	
Non-student x *Female*		−0.010			0.013	
		(0.019)			(0.019)	
Non-student x *Age*		−0.003*			−0.004**	
		(0.002)			(0.002)	
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
Log-likelihood	−6,315.83	−6,302.58	−4,953.32	−5,824.31	−5,814.31	−4,591.22
Observations	3,478	3,478	2,750	2,985	2,985	2,381
Left-censored obs.	0	0	0	0	0	0
Uncensored obs.	0	0	0	0	0	0
Right-censored obs.	294	294	223	0	0	0
Interval obs.	3,184	3,184	2,527	2,985	2,985	2,381

Notes: *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 3: Interval regression estimates of the ratio of utilities

Turning to treatment effects, we focus on the regression results of model 3 since non-students are basically absent from our laboratory samples. As for several demographics, we observe that treatment effects on risk attitudes significantly vary with the education status of participants. 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 ratio is the same in each incentive treatment at the 5 percent significance level (the estimated ratio in *Scale50Pr1/15* is significantly lower than in *Scale50PrUnknown* in a weak sense—p-value = 0.09—when considering all choice sequences but the effect vanishes when considering only consistent choice sequences). For full-time students, risk aversion in *Scale50PrUnknown* is substantially lower than in *Scale50Pr1/15* while it is larger than in *Scale10Pr1/3*, and there are statistically significant treatment effects: i) the estimated ratio is significantly higher in *Scale50Pr1/15* than in *Scale50PrUnknown* at the 5 (resp. 10) percent level for consistent (resp. all) choice sequences; and ii) the estimated ratio is significantly larger in *Scale50Pr1/15* than in *Scale10Pr1/3* at the 5 percent level whether all or only consistent choice sequences are considered. However, we never reject the hypothesis that the estimated ratio of full-time students is the same in *Scale50PrUnknown* and *Scale10Pr1/3* (p-values > 0.1).[13]

In a nutshell, our regression results show that we cannot reject the hypothesis that the two BRISs with identical expected payoffs per choice—*Scale50Pr1/15* and *Scale10Pr1/3*—induce equally strong effects of monetary incentives for part-time students whereas the scale of nominal payoffs is significantly more effective than the level of the selection probability in motivating full-time students. Moreover, though decreasing the selection probability for a given scale of nominal payoffs—*Scale50Pr1/15* versus *Scale50PrUnknown*—significantly reduces the effects of monetary incentives for full-time students, we cannot reject the hypothesis that *Scale50PrUnknown* and *Scale10Pr1/3* are equally effective in influencing the behavior of students. Thus, a BRIS with large prizes and an unknown and (likely to be perceived as) tiny probability of payment induces monetary incentives as effective as those induced by a BRIS with moderate prizes and a known and large probability of payment.

4.3 Differences in Risk Aversion With Certain Payment Treatments

We compare the effects of monetary incentives observed in our treatments to those observed in previous studies where all participants are paid. We restrict ourselves to certain payment treatments where participants complete HL's risk elicitation task under just one payment condition (to avoid order effects).

First, we estimate CRRA indices in our incentive treatments along with those in the 10× and 1×10× treatments of Harrison, Johnson, McInnes, and Rutström (2005, HJMR hereafter). In treatment 1×10× participants complete HL's 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 and we therefore refer to this treatment as treatment 1×. Moreover, in our treatments we only consider the choices made by students since HJMR recruited their participants from a convenience pool of students. Table 4 in Appendix D reports the interval regression results with controls for age, gender, the level of education (undergraduate or graduate), and the major field of study. We find that risk aversion in treatment 1×—mean CRRA equals 0.372—is significantly lower than risk aversion in treatments *Scale50PrUnknown* and *Scale50Pr1/15*——mean CRRA equal 0.650 and 0.715—at the 1% level of significance but it is

[13] Full-time students who study business administration are the most risk-prone students with a significantly lower estimated ratio than economic students (the difference is 0.023 with p-value < 0.05), MNE students (the difference is 0.037 with p-value < 0.01), and SSH students (the difference is 0.029 with p-value < 0.01). Similar differences are observed for part-time students except that economic students are weakly more risk-prone than business students. On the other hand, whether students are primarily responsible for the payment of tuition and living expenses or not as well as students' duration and level of education don't significantly affect risk attitudes.

significantly lower than risk aversion in treatment *Scale10Pr1/3*—mean CRRA equals 0.554—only at the 10% level of significance. On the other hand, we cannot reject the null hypothesis that risk aversion in treatment 10×—mean CRRA equals 0.574—is identical to the one estimated in any of our incentive treatments at any conventional significance level.[14] Though the difference is not statistically significant, the effects of monetary incentives are stronger in *Scale50Pr1/15* than in a 10× treatment where payment is certain though expected payoffs per lottery choice are (about) three times lower.

Second, Holt and Laury (2005) report that the average number of safe choices equals 6.7 in a 20× treatment where participants recruited from a convenience pool of students complete the HL's task. We observe that the average number of safe choices among students equals 6.5 in *Scale50Pr1/15*, 6.2 in *Scale50PrUnknown*, and 5.9 in *Scale10Pr1/3*. Hence, the effects of monetary incentives are (almost) as strong in *Scale50Pr1/15* as in a 20× treatment where payment is certain though expected payoffs per lottery choice are (about) six times lower.

These comparisons confirm once more that nominal payoffs have a bigger impact on the effects of monetary incentives than the probability of payment.

4.4 Structural Estimation Results

For the sake of parsimony, we here focus on model specifications where k is restricted to be homogeneous across individuals and treatments. Appendix E discusses the results of structural models where k varies with the experimental condition and the demographic characteristics of participants. The fit of some of the regressions improves with heterogeneous k according to likelihood-ratio tests, but estimates of the ratio of utilities are systematically (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 participants the estimated k is large.

4.4.1 Parameter estimates at the treatment level

Table 4 contains the results of the model with treatment dummies. Following von Gaudecker, van Soest, and Wengström (2011) all coefficients are on the original parameter scale, i.e. 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 $g_z\left(\beta^z_{\text{Constant}} + \beta^z_{\text{Treatment}}\right) - g_z\left(\beta^z_{\text{Constant}}\right)$, $z = r, k, w$.

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

Table 4: Estimated parameters for structural model with treatment dummies

[14]Note that the mean CRRA indices in treatments 1× and 10× that HJMR report are identical to the ones we report.

The results clearly 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 less risk averse in *Scale10Pr1/3* and more risk averse in *Scale50Pr1/15*. Therefore, similar to the interval regression results, we find evidence that the scale of nominal payoffs is more effective in motivating subjects than the level of the selection probability. 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 decision-making, estimates provide clear evidence that trembles are substantially larger on the Internet. The median subject in *Scale50PrUnknown* chooses completely at random with probability 6.3 percent compared to 1.4 percent for the median laboratory subject, and this difference is statistically significant in a strong sense. There are no significant differences between the laboratory treatments which confirms that the trembling probability does not vary with the scale of nominal payoffs. On the other hand, the estimated sensitivity to payoff differences is high in all treatments and differences are non-significant (see Appendix E).[15] More decision errors in the Internet treatment seem to result from a higher propensity of subjects to loose concentration.

4.4.2 Parameter estimates in subsamples of participants

Table 5 presents the median predicted utility ratios and trembling probabilities stratified by major demographics. For a given sub-sample of participants, the table shows the predicted sub-sample 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 4.2 respectively. The presentation conveys the total effect of varying the demographic variable, taking into account possible correlations between variables. Table 5 in Appendix E presents the marginal effects, and table 6 contains the median predicted parameters stratified by all demographic variables.

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 a 5% significance level and it is significantly larger than in *Scale50PrUnknown* (r = 0.710) at a 10% and 5% significance level 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

[15] 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 $d < 6$ ($d > 8$).

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 those of the interval regressions. 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).[16]

Note finally 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 based on observed heterogeneity alone account for less than 20 percent of the distribution of r when unobserved heterogeneity is also considered (see Appendix E).

Table 5: Median structurally estimated parameters stratified by major 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)

Continued on next page

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

Table 5: Continued

		Scale50PrUnknown		*Scale50Pr1/15*		*Scale10Pr1/3*	
		r	w	r	w	r	w
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)
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)
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)

Continued on next page

Table 5: Continued

		Scale50PrUnknown		*Scale50Pr1/15*		*Scale10Pr1/3*	
		r	w	r	w	r	w
Field of studies:	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)
Other	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)		

5 Conclusion

This paper presents experimental evidence that clarifies the relative impact of nominal payoffs and the probability of payment on the effects of monetary incentives implemented by BRIS. We conclude that nominal payoffs have a bigger impact on incentive effects than the probability of payment meaning that the random selection of subjects for real payment only mildly dilutes the effects of monetary incentives. Thus, a BRIS with a 50× scale of nominal payoffs and a 1/15 probability of payment induces significantly stronger incentive effects for full-time students than a BRIS with a 10× scale of nominal payoffs and a 1/3 probability of payment. Our conclusion is additionally supported by the finding that incentive effects are non-significantly stronger for students in a BRIS with a 50× scale of nominal payoffs and a 1/15 probability of payment than in a 10× payoff scale treatment where payment is certain. The clearest support for our conclusion however is the finding that 50× nominal payoffs with an unknown and (likely to be perceived as) tiny probability of payment induce non-significantly stronger incentive effects than a 10× payoff scale treatment where payment is certain.

Earlier experimental results corroborate that subjects do not fully reduce nominal payoffs to account for the probability of payment though results should be interpreted with the important caveat that there is no unequivocal way to ascertain incentive effects in many of the former studies. 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. 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 participants 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.[17]

We acknowledge that the bulk of the experimental evidence favoring the use of BRIS comes from simple choices. Still, a preliminary analysis of the choices made by our subjects after completion of HL's risk elicitation task supports the use of BRIS even in complex and dynamic environments. As a follow-up task, subjects traded in prediction markets. We observe that in incentive treatment *Scale50PrUnknown* predictions markets provide highly accurate forecasts of sporting events. In fact, the forecasting perfor-

[17] See Appendix F for more details on the literature. Some economic experiments have been conducted in countries with lower standards of living (Slonim and Roth, 1998) and some used non-monetary incentive schemes in the hope of adequately motivating subjects. The former approach raises the obvious question of potential cultural effects and evidence obtained with the latter approach indicates that behavioral patterns observed with non-monetary incentives differ significantly from those with monetary incentives (Duersch, Oechssler, and Schipper, 2009).

mance of our prediction markets compares favorably with those of earlier studies (Cyranek, 2013). In contrast with our preliminary results, 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 capacity to motivate subjects in a dynamic choice experiment. Further research is needed to assess the dilution of the effects of monetary incentives implemented by BRIS in complex and dynamic tasks.

References

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

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2008): "Eliciting Risk and Time Preferences," *Econometrica*, 76, 583–618.

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.

Bellemare, C., S. Kröger, and A. Van Soest (2008): "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities," *Econometrica*, 76, 815–39.

Bosch-Domènech, A., J. G. Montalvo, R. Nagel, and A. Satorra (2002): "One, Two, (Three), Infinity, . . . : Newspaper and Lab Beauty-Contest Experiments," *American Economic Review*, 92, 1687–1701.

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

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.

Cyranek, R. (2013): "Three Essays in Economic Internet and Field Experiments," Dissertation, LMU München: Faculty of Economics.

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.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. Wagner (2011): "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences," *Journal of the European Economic Association*, 9, 522–550.

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

Duersch, P., J. Oechssler, and B. C. Schipper (2009): "Incentives for Subjects in Internet Experiments," *Economics Letters*, 105, 120–122.

Greiner, B. (2004): "An Online Recruitment System for Economic Experiments," in *Forschung und wissenschaftliches Rechnen 2003, GWDG Bericht 63*, ed. by K. Kremer, and V. Macho, pp. 79–93. Ges. für Wiss. Datenverarbeitung, Göttingen, Germany.

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

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., and J. T. Swarthout (2014): "Experimental Payment Protocols and the Bipolar Behaviorist," *Theory and Decision*, 77, 423–38.

Hergueux, J., and N. Jacquemet (2014): "Social Preferences in the Online Laboratory: A Randomized Experiment," *Experimental Economics*, forthcoming.

Hermann, B., C. Thoni, and S. Gächter (2008): "Antisocial punishment across societies," *Science*, 319, 1362–67.

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

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.

Landry, C., A. Lange, J. List, M. Price, and N. Rupp (2006): "Towards an Understanding of the Economics of Charity: Evidence from a Field Experiment," *Quarterly Journal of Economics*, 121, 747–82.

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

Niza, C., C. Rudisill, and P. Dolan (2014): "Vouchers versus Lotteries: What Works Best in Promoting Chlamydia Screening? A Cluster Randomized Controlled Trial," *Applied Economic Perspectives and Policy*, 36, 109–24.

Onderstal, S., A. J. Schram, and A. Soetevent (2013): "Bidding to Give in the Field," *Journal of Public Economics*, 105, 72–85.

Slonim, R., and A. Roth (1998): "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic," *Econometrica*, 66, 569–96.

Volpp, K. G., L. K. John, A. B. Troxel, L. Norton, J. Fassbender, and G. Loewenstein (2008): "Financial Incentive-Based Approaches for Weight Loss: A Randomized Trial," *Journal of the American Medical Association*, 300, 2631–7.

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.

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.