The Binary Lottery Procedure does not induce risk neutrality in the Holt & Laury and Eckel & Grossman tasks*

Oliver Kirchkamp† Joerg Oechssler‡ Andis Sofianos§

23 December 2020 14:35:26 `6cbb7f8`

We test whether the binary lottery procedure makes participants behave as if they are risk neutral in the Holt and Laury (2002) and Eckel and Grossman (2002) tasks. Depending on the task, we find that less than half of the participants behave as if risk neutral. In fact, when we compare the distribution of choices, we find no significant difference to standard experiments that did not use the binary lottery procedure. Using a structural model we find modest evidence that the binary lottery procedure might move participants at least slightly towards risk neutrality.

Keywords: risk elicitation, binary lottery procedure; experimental economics.
JEL-Classification:C91; C81

*We would like to thank three anonymous referees, Glenn Harrison and Karim Sadrieh for valuable comments. We use R version 4.0.3 (2020-10-10) for the empirical analysis. We use `runjags` 2.0.4-6 to interface with `JAGS 4.3.0`. Funding for this study was provided by the University of Heidelberg.

†University of Jena: email: oliver@kirchkamp.de
‡University of Heidelberg: email: oechssler@uni-hd.de
§University of Heidelberg: email: andis.sofianos@uni-heidelberg.de

1. Introduction

The *Binary Lottery Procedure* (BLP henceforth) is, in theory, an ingenious method of inducing risk neutrality of participants in experiments. If an experiment requires that participants behave as if they are risk neutral,[1] one can pay them by lottery tickets rather than directly with money. Each lottery ticket then gives participants an objective probability of winning a high prize in a binary lottery. If participants satisfy two axioms (monotonicity and the reduction of objective compound lotteries (ROCL), axioms which are satisfied, in particular, for expected utility maximisers),[2] they should simply maximise the probability of winning the high prize, in other words, they should maximise the expected number of lottery tickets.

Since Smith (1961) proposed and Roth and Malouf (1979) implemented (and independently proposed) the BLP, many experimenters have employed it in order to induce risk neutrality.[3] However, already Kahneman and Tversky (1979) show that human decision makers do not always follow ROCL. After Selten et al. (1999) found in an experiment that the BLP did not work at all as intended, the use of the BLP by experimenters came to an effective halt.[4] Recently, there seems to be some kind of resurrection of the method.Harrison et al. (2013) have re-examined the procedure and found evidence that, at least, it moves risk preferences towards risk neutrality. This has in turn inspired the use of its premise for incentive compatible belief elicitation through the binarized scoring rule (see Hossain and Okui, 2013; Schlag and van der Weele, 2013; Harrison et al., 2014), also discussed in Schotter and Trevino (2014). Given the conflicting evidence, it seems appropriate to give the method another test.

We study the BLP in combination with two of the most popular experimental methods to measure risk attitudes, the Holt and Laury (2002) method and the Eckel and Grossman (2002) method, which was originally developed by Binswanger (1980). These are both standard risk preference elicitation tasks that are used in numerous experimental papers.[5] If the BLP actually succeeds in making participants behave as if they are risk neutral, we should ideally observe only one type of choice: the choice that maximises the expected number of lottery tickets.[6] The short answer is: this is not what we observe. In our main online treatment, not even half of the participants (depending on the method between 29.70% and 43.55%) behave as if they were risk neutral. The share of participants who behave as if they were risk neutral across *both* tasks is only 25.33%. Furthermore, comparing our main online treatment to a control treatment, where the 'standard' procedure is implemented for both the Holt and Laury (2002) and the Eckel and Grossman (2002) methods, we find no significant difference in risk neutral choices.

We also compare the proportion of risk neutral choices in a lab-based treatment with the BLP to other studies that have been implemented in recent years utilizing the same subject

[1]Often, theories are to be tested under the auxiliary assumption that participants are risk neutral.
[2]See Selten et al. (1999) for formal statements.
[3]It was also discussed in various experimental textbooks (e.g. Kagel and Roth, 1995), while Berg et al. (2008) specifically argue strongly for its merits.
[4]See also Loomes (1998) where it is also found that the BLP did not have the predicted effect.
[5]For recent surveys of risk preference elicitation see Charness et al. (2013) and Holt and Laury (2014)
[6]Of course, this presumes that participants are able to understand the tasks and choose the options that are best for them.

pool in our laboratory. These studies implemented similar risk elicitation tasks as part of their various experimental designs but with prizes rather than tokens as is standard in the literature. We can contrast the proportion of risk neutral choices while implementing BLP (in our current data) and while not (comparison studies). The results suggest that implementing the BLP does not result in a significant difference in the proportion of risk neutral choices.

Using a structural model we find that the BLP moves participants' behaviour, if at all, only slightly towards risk neutral choices.

2. Experimental Design

In this paper, we compare risky choices with and without implementing the BLP. We use data from an online experiment with 248 participants, from a lab experiment with 119 participants, and from earlier comparison studies consisting of a total of 1348 participants (see Table 3 for details). All participants were recruited from the same subject pool at the University of Heidelberg.

The two risk elicitation tasks administered are a multiple price list type task (Holt and Laury, 2002) and an Eckel and Grossman (2002) type task. Participants in the online and the lab experiments are asked to respond to both risk elicitation tasks, with one of the tasks chosen randomly for payment.[7]

In the online experiment, we have two treatments: one with the BLP (*onlineBLP*) and one with standard monetary payments (*onlineStd*). The lab experiment is run with pen and paper using the BLP for all participants (*labBLP*). Participants are recruited through SONA and hroot (Bock et al., 2014) for the online treatments and the lab experiment, respectively. Participants receive a show-up fee of 2.50 Euro (*labBLP*: 5 Euro) and can earn an additional 5 Euro depending on their decisions and the lottery realization.[8] Translated instructions are included in Appendix C.

The binary lottery procedure is implemented as follows. Instead of listing lotteries for monetary prizes, we let participants choose among lotteries that pay out "tokens". Each token corresponds to a probability of 0.01 for winning the monetary prize of 5 Euro. The instructions explicitly mention that the greater number of tokens they earn, the higher the chance for winning a prize of 5 Euro. For example, if a participant earns 72 tokens, the participant would have a probability of 0.72 to earn 5 Euro.[9]

[7]This is done by a virtual coin toss in the online experiment and by an actual coin toss conducted by one of the participants in the lab experiment. Participants in the comparison studies responded to only one of the risk elicitation tasks. This procedure is incentive-compatible under fairly mild conditions. In particular, participants are assumed to respect first-order stochastic dominance, see Azrieli et al. (2019) and the literature cited there.

[8]We ran a total of 6 sessions in the lab. For the first two sessions of the data reported, participants had just completed an unrelated experiment and were asked if they were willing to spend few more minutes in the lab and participate in a new short experiment. These participants did not receive an additional show-up fee.

[9]In the lab treatment this chance move is implemented by letting each participant throw two 10-sided dice, one determining the tens and one determining the ones. If the participant rolls any number below or equal to 72, they would earn 5 Euro. In the online experiment the random number between 0 and 99 is drawn by the computer.

Lottery	Tokens if coin shows Tails	Tokens if coin shows Heads	Exp. value
1	38	38	38
2	28	52	40
3	16	72	44
4	0	84	42

Table 1: Eckel and Grossman (2002) Lottery Task: Participants choose one of the lotteries 1-4
Each lottery has equal chance of either outcome, determined by a coin flip. In *onlineStd*, the word "token" is replaced by "Taler" which is our experimental currency unit with an exchange rate of 20 Taler = 1 Euro.

For our Eckel and Grossman (2002) task, Table 1 lists the 4 lotteries the participants have to choose from. Participants are asked to choose their preferred one out of the four lotteries listed. Each lottery has a 50% chance to reward the participants with either the low or high prize and this is determined by a coin flip at the end (virtual coin flip for the online treatments, while real coin flip in the lab). In the standard procedure, where tokens are monetary prizes, choosing lotteries 1 or 2 is indicative of risk averse preferences as they entail lower (or even zero) variance, choosing lottery 3 implies risk neutrality as it is the lottery that maximises expected value and finally, choosing lottery 4 is indicative of risk seeking behaviour as it is the lottery with the highest variance and a lower expected value than lottery 3.

For the Holt and Laury (2002) task, Table 2 displays the choice list participants see. Participants have to choose option A or B for each of the 10 situations listed. The last column lists the difference in expected values of lotteries A and B. This is not shown to participants. Any participant who wants to maximise expected value would choose Lottery A for rows 1 through 4 and then switch to Lottery B for rows 5-10. For consistent decision makers we call the first row for which a participant chose Lottery B the switch point.[10] In the standard procedure, where tokens are monetary prizes, participants with switch point of less than 5 would be classified as risk seekers, those with a switch point of 5 would be classified as risk neutral, and those with a switch point higher than 5 as risk averse.[11]

For the comparison studies without the BLP, we collect results and data from other recent studies that took place in the same lab, which we refer to as *otherStd*. Given the anonymity of participants we cannot exclude that some participated in more than one study. However, given that the experiments were stretched over many years and given the substantial turnover in the subject pool of more than 1000 participants, the fraction of multiple participation is likely to be small. These studies implemented similar risk elicitation tasks but with prizes rather than tokens as is standard in the literature. Table 3 lists the studies surveyed, with information on which risk elicitation task each implemented and the number of observations. Table 14 in Appendix B shows the payoffs from the comparison studies. The Holt

[10]Participants who switch more than once are coded as "inconsistent". Among 367 participants in both our online and lab treatments, we find 32 that are inconsistent. If we instead code multiple switchers by the sum of the safe choices plus 1 (similar to Holt and Laury (2002, p. 1648)), the relative share of risk neutral choices slightly increases (see Table 4).

[11]Participants who always choose Lottery B have a switch point of 0. A switch point of 10 (always choose A) would imply a dominated choice for the last row.

	Lottery A	Your choice	Lottery B	EV(A) − EV(B)
1	40 tokens of the die shows 1 32 tokens if the die shows 2-10	A or B	77 tokens of the die shows 1 2 tokens if the die shows 2-10	23.3
2	40 tokens of the die shows 1-2 32 tokens if the die shows 3-10	A or B	77 tokens of the die shows 1-2 2 tokens if the die shows 3-10	16.6
3	40 tokens of the die shows 1-3 32 tokens if the die shows 4-10	A or B	77 tokens of the die shows 1-3 2 tokens if the die shows 4-10	9.9
4	40 tokens of the die shows 1-4 32 tokens if the die shows 5-10	A or B	77 tokens of the die shows 1-4 2 tokens if the die shows 5-10	3.2
5	40 tokens of the die shows 1-5 32 tokens if the die shows 6-10	A or B	77 tokens of the die shows 1-5 2 tokens if the die shows 6-10	−3.5
6	40 tokens of the die shows 1-6 32 tokens if the die shows 7-10	A or B	77 tokens of the die shows 1-6 2 tokens if the die shows 7-10	−10.2
7	40 tokens of the die shows 1-7 32 tokens if the die shows 8-10	A or B	77 tokens of the die shows 1-7 2 tokens if the die shows 8-10	−16.9
8	40 tokens of the die shows 1-8 32 tokens if the die shows 9-10	A or B	77 tokens of the die shows 1-8 2 tokens if the die shows 9-10	−23.6
9	40 tokens of the die shows 1-9 32 tokens if the die shows 10	A or B	77 tokens of the die shows 1-9 2 tokens if the die shows 10	−30.3
10	40 tokens of the die shows 1-10	A or B	77 tokens of the die shows 1-10	−37.0

Table 2: Holt and Laury (2002) task: Participants choose one of the lotteries A or B for each of the 10 possible situations

Note: One row is chosen for payment by use of a 10-sided die. Suppose choice 3 is chosen and the participant chooses A. Then another 10-sided die determines whether they would get 40 tokens or 32 tokens. In *onlineStd*, the word "token" is replaced by "Taler" which is our experimental currency unit with an exchange rate of 20 Taler = 1 Euro.

Reference	Time of exp.	Acronym	Risk task	Cons. obs.	Incons. obs
Current paper	Nov, Dec 2018	labBLP	HL	110	9
Current paper	Sep 2020	onlineBLP	HL	101	23
Current paper	Sep 2020	onlineStd	HL	102	22
Brunner et al. (2014)	2011	BHO	HL	318	39
Dürsch et al. (2012)	2011	DOV	HL	199	10
Dürsch et al. (2017)	Sep 2012, Jan 2013	DRR	HL	133	11
Proto et al. (2020)	Apr, Nov 2018	PRS	HL	215	34
Roth et al. (2016)	Feb, Mar 2012	RTV	HL	104	0
Current paper	Nov, Dec 2018	labBLP	EG	119	
Current paper	Sep 2020	onlineBLP	EG	124	
Current paper	Sep 2020	onlineStd	EG	124	
Apesteguia et al. (2019)	Jan, May, Jun, Jul, Aug 2017	AOW	EG	176	
Kersting-Koenig et al. (2019)	Oct, Nov 2017	LMK	EG	199	
Schmidt (2019)	Jan, Feb 2018	S	EG	158	

Table 3: Summary of the risk elicitation studies used.
All participants were recruited from the same subject pool at the University of Heidelberg.
Note: EG stands for Eckel and Grossman (2002) task and HL stands for Holt and Laury (2002) task.

and Laury (2002) tasks used in all studies are exactly the same as the one in the current study up to a scaling factor. The Eckel and Grossman (2002) tasks are less comparable since each study used a different number and kind of lotteries which is unfortunately typical for Eckel and Grossman tasks in the literature.[12]

3. Results

Risk neutral choices When employing the BLP, participants in our experiment should have a switch point of 5 in the Holt and Laury (2002) task (HL henceforth) task and should have chosen lottery 3 in the Eckel and Grossman (2002) (EG henceforth) task.[13] In Table 4 we report the proportions of choices that are compatible with this payoff maximizing prediction as "Risk neutral".

In the EG task, in the onlineBLP treatment only 43.55% (confidence interval CI_{95}=[34.67, 52.74]) of choices can be classified as risk neutral. In the labBLP treatment this percentage is even lower at 34.45% (confidence interval CI_{95}=[25.98,43.72]). In the HL task the proportion

[12]We chose the payoffs for our experiment to be somewhere in the middle of the comparison studies (see Figure 5).
[13]Under the assumptions of monotonicity and the reduction of objective compound lotteries.

Risk Task	Treatment	Sum	Risk Av.	Risk Neut.	Risk Seek.	Risk Av. [%]	Risk Neut.[%]	Risk Seek.[%]
EG	onlineBLP	124	58	54	12	46.77	43.55	9.68
EG	onlineStd	124	66	41	17	53.23	33.06	13.71
EG	labBLP	119	69	41	9	57.98	34.45	7.56
EG	otherStd	533	391	139	3	73.36	26.08	0.56
HL (all)	onlineBLP	124	76	33	15	61.29	26.61	12.10
HL (all)	onlineStd	124	82	24	18	66.13	19.35	14.52
HL (all)	labBLP	119	86	23	10	72.27	19.33	8.40
HL (all)	otherStd	815	553	171	91	67.85	20.98	11.17
HL (cons.)	onlineBLP	101	59	30	12	58.42	29.70	11.88
HL (cons.)	onlineStd	102	64	23	15	62.75	22.55	14.71
HL (cons.)	labBLP	110	81	19	10	73.64	17.27	9.09
HL (cons.)	otherStd	732	486	161	85	66.39	21.99	11.61

Table 4: Elicited Risk Preference Classifications.

Note: HL (cons.) are only the “consistent” choices in the HL task. HL (all) are all choices. We code multiple switchers in HL (all) by the number of times they choose Lottery A.

of risk neutral choices is even smaller, at 29.70% (confidence interval CI_{95}=[21.02,39.61]) for the onlineBLP and 17.27% (confidence interval CI_{95}=[10.73,25.65]) for the labBLP; certainly far away from the ideal 100%. Furthermore, strictly speaking the BLP should make participants risk neutral in both risk elicitation tasks simultaneously. However, this works only for 25.33% (confidence interval CI_{95}=[15.99,36.70]) of the participants in the onlineBLP and for 10.00% (confidence interval CI_{95}=[5.10,17.19]) of the participants in the labBLP.[14] Thus, it seems that the BLP clearly fails in letting all – or even a majority of – participants behave as if they were risk neutral.

Given the different findings in the literature whether the BLP at least shifts the distribution of preferences towards risk neutrality, it is of course interesting to compare the share of risk-neutral choices in studies with and without the BLP. For this purpose, we report in Figure 1 the proportion of risk neutral participants in our three treatments contrasted with the studies outlined in Table 3 (otherStd). Contrasting the two online treatments, we find that for both tasks, the HL and the EG task, the number of risk neutral choices is higher with the BLP, however only by a small amount. Comparing the labBLP treatment with otherStd, we still find an increase in the number of risk neutral choices for the EG task. However, we find a small decrease for the HL task.[15] The top part of Figure 1 shows results from the online

[14] The switch point in the HL task and the chosen lottery in the EG task are however correlated with a coefficient of -0.326 (confidence interval $CI_{95} = [-0.55, -0.115]$) in onlineBLP and -0.198 (confidence interval $CI_{95} = [-0.393, -0.005]$) in labBLP. Notice that a negative correlation is to be expected. A higher switch point in HL implies higher risk aversion, while a lottery choice of higher value in EG implies lower risk aversion

[15] The LMK study has a much lower proportion of risk neutral participants than our data as well as the other two comparison studies (AOW and S). This can be because in LMK they implement a longer list of lotteries to choose from. In LMK participants were choosing one out of 11 lottery options. Any noise in choices

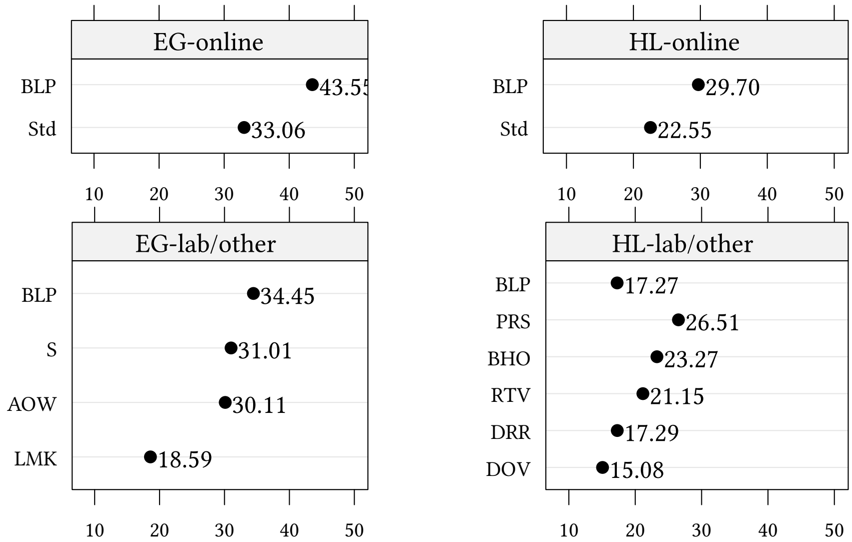

Figure 1: Proportion of risk neutral choices for consistent decision makers.

experiment.

Distribution of choices in the HL task Figure 2 shows the frequency distribution of switch points in the HL task. The top panels show the BLP treatments. The bottom panels shows the standard treatments. On the left we have the online treatments, while on the right we report the labBLP treatment which we compare with otherStd.

Figure 3 shows the distribution for the individual studies. Comparing the two panels of Figure 2 it seems pretty obvious that there is no substantial shift towards the risk-neutral choice (indicated by the vertical line). A Kolmogorov-Smirnov test shows that we cannot reject the hypothesis that the two distributions are the same ($p = 0.982$ for onlineBLP vs. onlineStd, $p = 0.342$ for labBLPvs. otherStd).[16]

Distribution of choices in the EG task The left part of Figure 4 shows choices for our online implementation of the EG task. As with the HL task, we use a Kolmogorov-Smirnov test to compare the BLP with the standard treatment. We cannot reject the hypothesis that the two distributions are the same ($p = 0.959$). The right part of Figure 4 compares the EG task in the labBLP treatment with otherStd. Since payoffs for the otherStd implementation of the EG task are all different (see Table 14), we cannot compare the distribution of choices

would then explain the lower percentage of risk neutral choices.

[16] This confirms the results of Dickhaut et al. (2013). Although their study had a different intention, they also implement a BLP procedure in an HL task and find that the distribution of choices in their Low treatment is similar to standard procedure HL tasks.

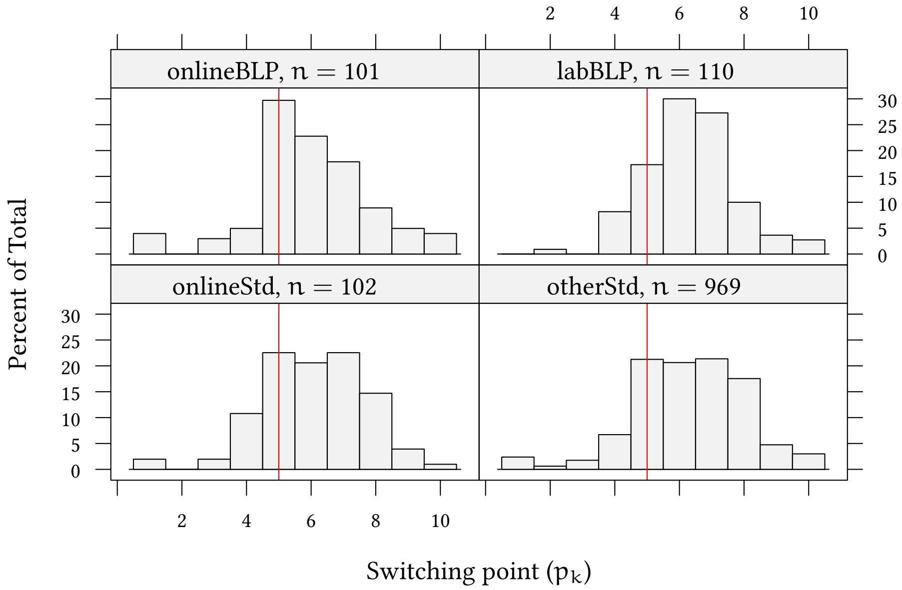

Figure 2: Switching points in the HL task for consistent players.

The left part of the figure shows only the online treatments. The right part shows the remaining treatments Figure 3 shows the individual experiments. The panel "otherStd" shows the pooled data from the HL task in otherStd (studies listed in Table 3).

The risk neutral switching point is denoted by a vertical line.

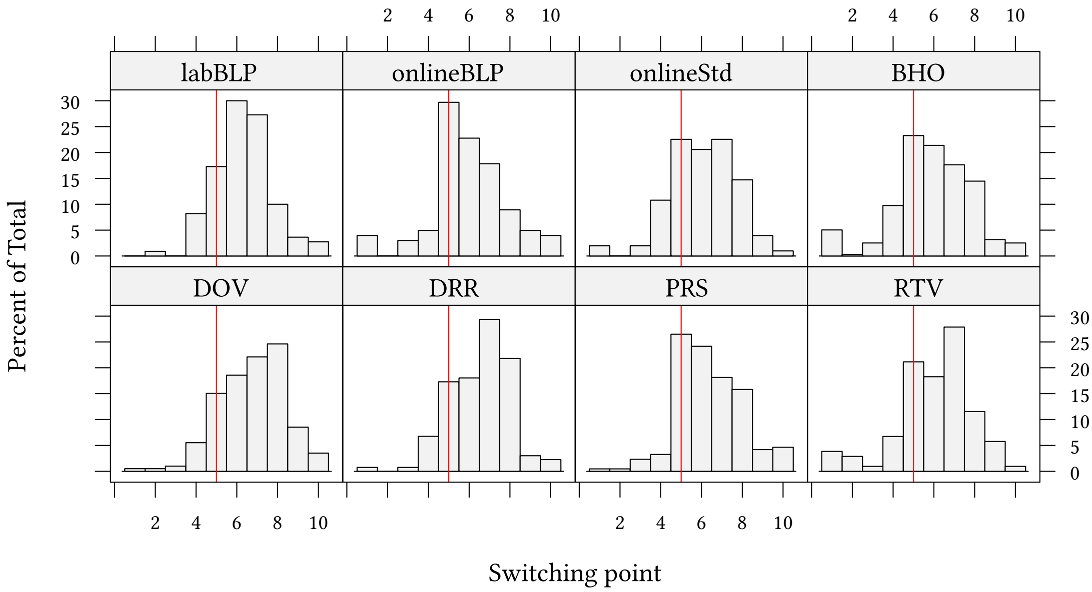

Figure 3: Switching points in the HL task for consistent players.

See Table 15 for payoffs. The risk neutral switching point is denoted by a vertical line.

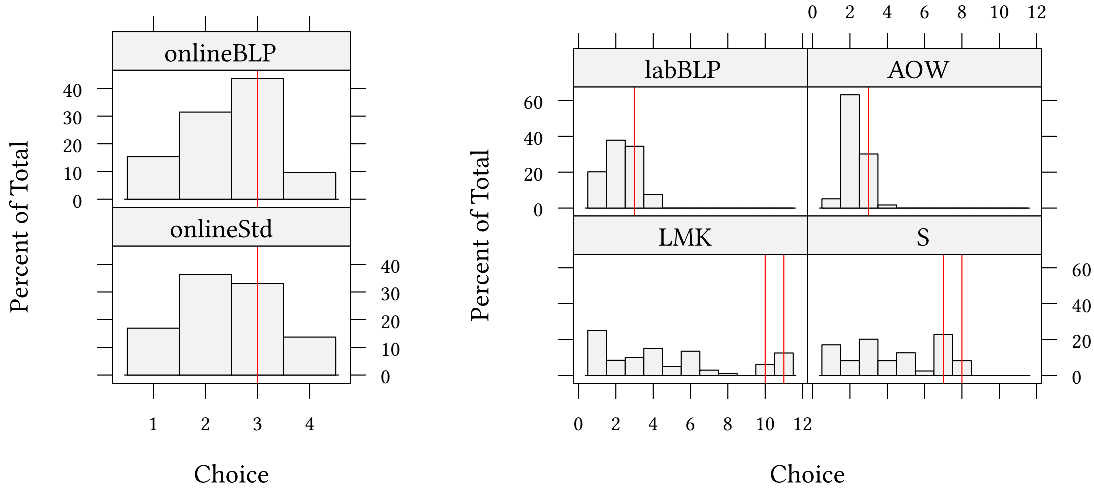

Figure 4: Choices in the EG task.
See Table 14 for payoffs. The risk neutral choices are denoted by a vertical line.

between labBLP and otherStd with the help of a Kolmogorov-Smirnov test.

4. Structural model

4.1. Model

Utility functions The data we use here elicits preferences for risk in different ways. To make results comparable we translate the choices for each decision maker i into a parameter of a CRRA expected utility function. This does not imply that decision makers would always follow this utility function. The utility function is just a device to facilitate comparison across different lotteries with different payoffs and the different estimation methods that we implement.

Bayesian inference We have two major reasons to use Bayesian inference to present our results: unbiasedness and transparency. We would like to have an unbiased estimator for the non-linear structural model we estimate below. We also want to transparently communicate our estimation strategy and to allow other researchers to easily reproduce our results.

Arminger and Muthén (1998) provide an insightful discussion on the shortcomings of ML, Pseudo ML, and weighted least squares in the context of non-linear structural models. The Bayesian estimator provides unbiased results even for small and medium sized samples, even for estimates of parameters at the boundary of the parameter space. Any ML estimator could deliver only biased results.

Transparency is another important reason to use the Bayesian framework: We are not aware of any standard tool to estimate this problem. We fear that any ML estimator could only be communicated as a purpose-built, hard to comprehend and hard to replicate optimisation problem. Within the Bayesian framework we can transparently communicate our models using the well known BUGS language. We provide the data for our experiment and our methods on-line. Any reader who wants to compare the Bayesian estimate of our non-linear structural model with an alternative approach of his or her liking can easily perform this comparison. Finally, the communication of results of the Bayesian estimation might lead to less confusion than Frequentist Null Hypothesis Testing. In this paper Bayesian methods help us to present results in a way that is transparent and easy to understand.[17]

Bayesian inference has been used by experimental economists for a number of tasks, e.g. to study markets (Smith, 1964; Cipriani et al., 2012; Farjam and Kirchkamp, 2018), risk (Harrison, 1990; Engel and Kirchkamp, 2019), learning (El-Gamal et al., 1994), auctions (Kirchkamp and Reiß, 2019) and to help with the design of experiments (El-Gamal and Palfrey, 1996).[18]

Comparing lotteries In the HL task, each choice reflects one comparison of two lotteries, $\mathfrak{L}_A$ and $\mathfrak{L}_B$. For each HL task we include here, payoffs are the same up to a scaling factor. Even if our estimation approach introduced a bias, we should expect that estimations from the different studies are affected by such a bias in the same way. Below we will compare four different estimation approaches. For the HL task, all approaches come to very similar conclusions.

Our two online treatments of the EG task differ only in the payment mechanism: BLP or standard. The EG task in our labBLP treatment is, however, less comparable with the other studies from the same lab (otherStd). Each of the studies implementing the EG task is based on a different set of lotteries (see Figure 5 and Table 14). Ideally, if the model for our estimation represented the true data generating process, this difference among lotteries should not matter. The estimator should still yield an unbiased estimate. In practice, each model is only an approximation of the truth and, hence, each estimation is only a possibly biased approximation of the truth. Different lotteries might be affected by a bias in different ways. Nevertheless, we carried out all the estimations we did for the online treatments to also compare labBLP with otherStd choices in the EG task. Estimation results of this last comparison are shown in Appendix A.2. These results are, more or less, in line with estimation results for the HL task. In this section we will focus on results of our online treatments for both tasks. Additionally, we compare the labBLP with the otherStd choices for the HL task.

Each lottery $\mathfrak{L}_k$ has two possible outcomes $j \in \{1, 2\}$. If, in any of these comparisons of two lotteries, $\mathfrak{L}_k$ and $\mathfrak{L}_{k'}$, lottery $\mathfrak{L}_k$ is chosen, outcome j realises with probability p_{jk}, yielding a payment x_{jk}. The utility from choosing lottery $\mathfrak{L}_k$ for decision maker i with CRRA utility function is then

$$u_i(\mathfrak{L}_k|r_i) = \sum_j p_{jk} \frac{x_{jk}^{1-r_i} - 1}{1 - r_i} \tag{1}$$

where r_i is the coefficient of relative risk aversion. A risk neutral decision maker would be

[17] On `https://www.kirchkamp.de/research/blp.html` we provide the Bayesian model in BUGS notation, the R commands, and the data.

[18] Vallois and Jullien (2018) present an interesting discussion of the development of statistical methods in experimental economics.

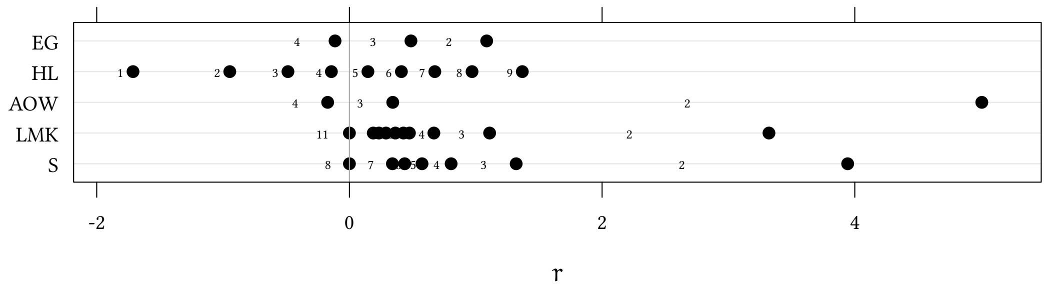

Figure 5: Choices for decision makers with CRRA preferences.
The figure shows the critical values of the coefficient of relative risk aversion r as given by (1) where a decision maker is indifferent between two lotteries. For the HL task: Each of the ten choices corresponds to one dot and is denoted with the index of that choice. A decision maker with this value of r is, for that choice, indifferent between A and B. For the EG tasks (including AOW, LMK and S): Numbers denote the range of r where this choice is the preferred choice.

characterised by $r_i = 0$. A more risk averse decision maker would have an $r_i > 0$. Figure 5 shows the critical coefficients of relative risk aversion r where a decision maker with CRRA preferences is indifferent between two lotteries.

Random utility Faced with the choice between two lotteries, $\mathfrak{L}_k$, and $\mathfrak{L}_{k'}$, we assume the probability to choose $\mathfrak{L}_k$ follows a logistic model:

$$P(\mathfrak{L}_k \succ_i \mathfrak{L}_{k'}|r_i) = \mathcal{L}\left((u_i(\mathfrak{L}_k|r_i) - u_i(\mathfrak{L}_{k'}|r_i)) \cdot \sqrt{\tau_{\mathfrak{L},i}}\right) \tag{2}$$

Here $\mathcal{L}$ is the logistic function. Each decision maker is characterised by two parameters: r_i describes i's preference towards risk. $\tau_{\mathfrak{L},i}$ describes the precision of i's preferences. Equation (2) is what Becker et al. (1963) call a Fechner model. A decision maker with a very large precision $\tau_{\mathfrak{L},i}$ will appear consistent and will almost always choose the lottery with the higher utility. A decision maker with a small precision $\tau_{\mathfrak{L},i}$ will appear less consistent. A decision maker with a $\tau_{\mathfrak{L},i} = 0$ will choose both lotteries, $\mathfrak{L}_k$ and $\mathfrak{L}_{k'}$, with the same probability.

We are mainly interested in the distribution of r_i under different elicitation procedures. Since we use Bayesian inference, we have to think about a prior. We use vague priors, i.e. we assume that we have almost no prior information at all. Nevertheless, to demonstrate the robustness of the model we compare four different approaches, M1, M2, M3, M4:

Baseline (M1) As a standard case we assume that the individual coefficient of relative risk aversion r_i is drawn from a normal distribution $r_i \sim N(\mu_r, \tau_r)$ where μ_r is the mean and $\tau_r = 1/\sigma_r$ is the precision of the Normal distribution. We are interested in the population mean of this distribution μ_r. As a prior for μ_r we assume $\mu_r \sim N(0, .1)$, i.e. our prior expectation for the population mean is, on average, risk neutral behaviour. The precision of this prior is only 0.1, i.e. we allow that the prior population mean μ_r is with probability 50% between -2.13 and 2.13.

Changing the prior (M2) Since we use vague priors, it does not matter much which mean we assume for the prior of μ_r. Still, M2 is an attempt to convince the sceptical reader that the prior for μ_r has (as long as it is a vague prior) practically no influence. If all we knew was Table 3 in Holt and Laury (2002), we might expect an average coefficient of relative risk aversion μ_r between 0.25 and 0.55, depending on the size of the stakes and whether payoffs are hypothetical or real. While M1 assumes (on average) a risk neutral prior, M2 assumes that $\mu_r \sim N(.5, .1)$, i.e. M2 assumes a priori a quite risk averse population mean with an average coefficient of relative risk aversion of .5. As in M1, the precision of that prior distribution is very small. We still assume $r_i \sim N(\mu_r, \tau_r)$. Now the prior population average μ_r is with probability 50% between -1.63 and 2.63.

Robust estimation (M3) Figures 2 and 3 suggest that, at least for the HL task, risk aversion might include some outliers. An adaptive robust version of the model takes outliers into account, assuming that r_i is not necessarily normal but drawn from a t distribution: $r_i \sim t(\mu_r, \tau_r, \nu)$ (Lange et al., 1989) with endogenous degrees of freedom ν. For $\nu \to \infty$ this model includes the normal model. For $\nu < \infty$ outliers are downweighted. As in M1, we assume $\mu_r \sim N(0, .1)$.

Random preferences (M4) Models M1, M2 and M3 are based on random utility (Equation (2)). Random preferences are an alternative to random utility. An empirical comparison of models with random preferences and random utility can be found, e.g., in Loomes et al. (2002) and Butler et al. (2012). Theoretical properties of models with random utility and random preferences have recently been discussed by Wilcox (2011), Apesteguia and Ballester (2018), and Conte and Hey (2018). Although econometrically not always as convenient as models with random utility, models with random preferences seem to fit the data well and can be attractive from an axiomatic viewpoint. Therefore, we consider M4 as an alternative to M1, M2 and M3. M4 is based on random preferences as follows:

$$P(\mathfrak{L}_k \succ_i \mathfrak{L}_{k'} | r_i) = \int_{u_i(\mathfrak{L}_k|\tilde{r}_i) > u_i(\mathfrak{L}_{k'}|\tilde{r}_i)} \phi(r_i, \tau_{\mathfrak{L},i}) \, d\tilde{r}_i \tag{3}$$

As in M1, M2 and M3, the parameter r_i denotes the mean of the risk preference of decision maker i. The random preference used for a specific comparison, $\tilde{r}_i$, follows a normal distribution with density ϕ, mean r_i and precision $\tau_{\mathfrak{L},i}$.

For specifications M1, M2, M3 and M4 we are primarily interested in the population parameter μ_r (the population average coefficient of relative risk aversion r_i). We also estimate the population precision of this preference τ_r, i.e. how much preferences in the population differ from each other.[19]

Figure 6 shows estimation results for M1. [20] From this figure it is clear that the BLP does

[19]The remaining priors are $\tau_{\mathfrak{L},i} \sim \Gamma(.01, .01), \tau_r \sim \Gamma(.01, .01)$, $\nu \sim \Gamma(1/30, 1/30)$. $\Gamma(\alpha, \beta)$ denotes the Gamma distribution with shape α and rate β. $N(\mu, \tau)$ is the normal distribution with mean μ and precision τ. $t(\mu, \tau, \nu)$ is the t-distribution with mean μ, precision τ and ν degrees of freedom.

[20]Results for priors M2, M3 and for random preferences M4 are very similar and are shown in Figures 7 and 8 in Appendix A.1. Figure 9 in Appendix A.2 shows estimation results for the EG task. Table 9 in Appendix A.1

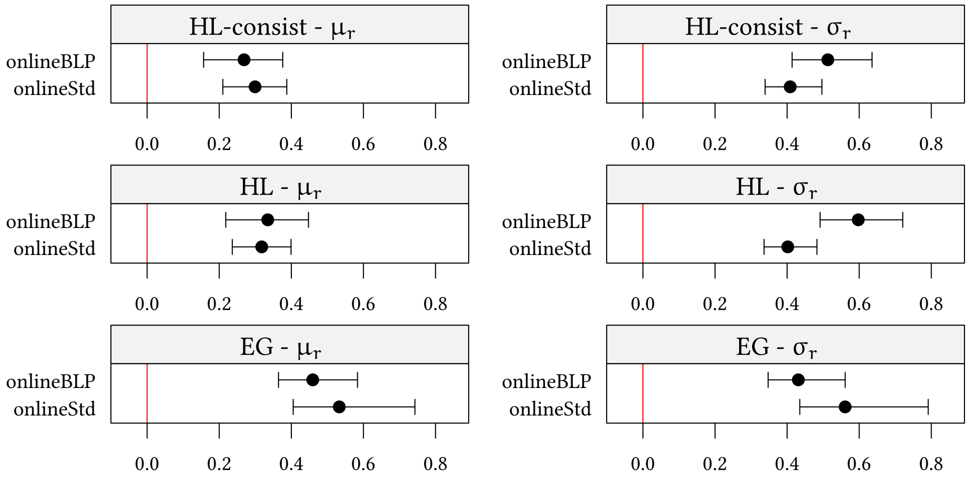

Figure 6: Population parameters for onlineBLP and onlineStd treatments (M1)
The figure shows for the two treatments the median value of population parameters μ_r and $\sigma_r = 1/\sqrt{\tau_r}$ together with a 95% credible interval. Risk neutrality corresponds to $\mu_r = 0$. Tables 9 and 10 in Appendix A.1 provide convergence diagnostics. Figures 7 and 8 in Appendix A.1 show results for the other treatments and the other priors for the HL task. Figure 9 in Appendix A.2 shows corresponding results for the EG task.

not achieve complete risk neutrality. For no procedure does the 95% credible interval contain $\mu_r = 0$. Instead, with the BLP the median value of μ_r is about 0.357 for the HL task.

4.2. The effect of the binary lottery procedure

As we have seen in Figure 6, the binary lottery procedure does not clearly differ from the standard procedure. Let us, nevertheless, try to assess the size of the effect of the binary lottery procedure. To do this, we assume that the individual parameter r_i is drawn from $r_i \sim N(\mu_r^{BLP}, \tau_r^{BLP})$ if preferences are elicited with the binary lottery procedure and r_i is drawn from $r_i \sim N(\mu_r^{std}, \tau_r^{std})$ otherwise.

We will compare μ_r^{BLP} with μ_r^{std} in Section 4.2.1. We will also compare the distribution of the different r_i for BLP and standard in Section 4.2.2.

4.2.1. The average effect on μ_r

Absolute effect size If the BLP procedure does what it is supposed to do, then we should expect μ_r^{BLP} to be closer to 0 than μ_r^{std}. In our sample the posterior for the average relative risk aversion μ_r is positive almost always. Nevertheless, to be on the safe side, we compare absolute values of μ_r^{BLP} and μ_r^{std} as follows:

$$\Delta_r = |\mu_r^{std}| - |\mu_r^{BLP}| \tag{4}$$

shows diagnostics for the MCMC sampler (effective sample size, Gelman et al., 2013, p. 287, and potential scale reduction factor, Gelman and Rubin, 1992) for M1, M2, M3, and M4.

mean $\Delta_r = |\mu_r^{std}| - |\mu_r^{BLP}|$		mean Δ_r/σ_r^{std}	CI_{95} Δ_r	odds $\Delta_r > 0$	effss Δ_r	psrf Δ_r	median ν
Eckel and Grossman (2002) task, online data							
M1	0.08507	0.13768	$[-0.10034, 0.30294]$	4.13 : 1	2407	1.00481	∞
M2	0.08082	0.13127	$[-0.10275, 0.29567]$	3.83 : 1	2512	1.00179	∞
M3	0.08519	0.17746	$[-0.08548, 0.28808]$	4.87 : 1	2411	1.00130	11.4
M4	0.02593	0.04398	$[-0.12784, 0.17904]$	1.71 : 1	9447	1.00049	∞
Holt and Laury (2002) task, online data							
M1	-0.01619	-0.04027	$[-0.15850, 0.12411]$	1 : 1.43	23714	1.00013	∞
M2	-0.01535	-0.03823	$[-0.15541, 0.12600]$	1 : 1.42	25144	1.00010	∞
M3	0.04995	0.22621	$[-0.06445, 0.16197]$	4.18 : 1	9643	1.00007	1.46
M4	-0.01286	-0.02760	$[-0.16727, 0.14405]$	1 : 1.31	12198	1.00057	∞
Holt and Laury (2002) task, online data, consistent only							
M1	0.03096	0.07666	$[-0.10999, 0.17224]$	2.01 : 1	22502	1.00020	∞
M2	0.03104	0.07699	$[-0.10946, 0.17355]$	1.99 : 1	24284	1.00015	∞
M3	0.04990	0.19038	$[-0.07209, 0.17114]$	3.79 : 1	12857	1.00009	1.89
M4	0.04491	0.10444	$[-0.10793, 0.19905]$	2.51 : 1	11936	1.00061	∞

Table 5: Average effect of the BLP procedure on the average coefficient of relative risk aversion μ_r.

CI_{95} is the 95%-credible interval (equal-tailed). ν are the degrees of freedom of the t-distribution for the robust model. To assess convergence of the MCMC sampler for Δ_r, the table provides effective sample size (effss) (Gelman et al., 2013, p. 287) and potential scale reduction factor (psrf) (Gelman and Rubin, 1992)). Table 6 shows results for the HL task when comparing labBLP and otherStd. Table 11 in Appendix A.2 shows results for the EG task when comparing labBLP and otherStd.

If the binary lottery procedure has no effect, $E[\Delta_r]$ should be zero. If in the binary lottery procedure decision makers choose in a more risk neutral way, then $E[\Delta_r]$ should be positive.

Table 5 summarises estimation results for Δ_r for the onlineBLP and onlineStd treatments. The effect of the BLP procedure is already small for the EG task (for M1 a reduction of 0.085 for μ_r). It is even smaller, depending on the model even negative, for the HL task. Table 6 summarises estimation results for the HL task when comparing labBLP and otherStd. We find that effect sizes are even smaller than effect sizes for the online treatments. Table 11 in Appendix A.2 shows results for the EG task.

Relative effect size: To relate the average effect shown in Table 5 to the different risk preferences in the population, the third column in Table 5 shows the ratio Δ_r/σ_r^{std}. We see that the effect of the BLP procedure on risk aversion is in any case not larger than 22.6% of the standard deviation of risk aversion in our sample.

Strength of evidence To interpret posterior odds whether $\Delta_r > 0$, i.e. whether BLP leads to a reduction of the average population risk aversion at all, we follow the terminology of

mean $\Delta_r = \lvert\mu_r^{std}\rvert - \lvert\mu_r^{BLP}\rvert$		mean Δ_r/σ_r^{std}	CI_{95} Δ_r	odds $\Delta_r > 0$	effss Δ_r	psrf Δ_r	median ν
Holt and Laury (2002) task, lab/other							
M1	0.01490	0.03166	$[-0.06221, 0.09109]$	1.85 : 1	21027	1.00016	∞
M2	0.01510	0.03210	$[-0.06221, 0.09284]$	1.85 : 1	21249	1.00005	∞
M3	0.03286	0.10716	$[-0.03880, 0.10298]$	4.57 : 1	10164	1.00008	2.13
M4	0.00010	0.00025	$[-0.08245, 0.08177]$	1.01 : 1	9928	1.00008	∞
Holt and Laury (2002) task, lab/other, consistent only							
M1	−0.00332	−0.00707	$[-0.08403, 0.07780]$	1 : 1.13	23374	1.00013	∞
M2	−0.00370	−0.00783	$[-0.08303, 0.07655]$	1 : 1.16	22699	1.00006	∞
M3	0.02266	0.07207	$[-0.05084, 0.09612]$	2.71 : 1	11733	1.00004	2.29
M4	−0.01627	−0.03089	$[-0.10033, 0.06863]$	1 : 1.82	10607	1.00029	∞

Table 6: Average effect of the BLP procedure on the average coefficient of relative risk aversion μ_r comparing labBLP with otherStd.

CI_{95} is the 95%-credible interval (equal-tailed). ν are the degrees of freedom of the t-distribution for the robust model. To assess convergence of the MCMC sampler for Δ_r, the table provides effective sample size (effss) (Gelman et al., 2013, p. 287) and potential scale reduction factor (psrf) (Gelman and Rubin, 1992)). Table 11 in Appendix A.2 shows results for the EG task.

Kass and Raftery (1995).[21]

For the EG task, the odds with prior M3 that $\Delta_r > 0$ are 4.87:1, i.e. we have "positive" evidence that the average attitude towards risk with the BLP procedure is closer to risk neutrality. The odds are even smaller for the other priors (see Table 5). For the HL task, the odds with prior M3 that $\Delta_r > 0$ are 4.18:1, i.e. we have "positive" evidence that the average attitude towards risk with the BLP procedure is closer to risk neutrality. The odds are even smaller for the other priors (see Table 5).

4.2.2. The individual effect on r_i

We might not only care about effects on the average but also about the effect on the individual. After all, we cannot rule out that a few participants find the BLP procedure hard to understand but that these participants affect the average behaviour substantially. The experimenter might prefer that a larger share of participants gains from a procedure, even if a few participants lose. Here we try to estimate the size of the population that gains from the BLP procedure. To better understand the effect on the individual, we call $\text{rank}(r_i^2)$ the rank of the individual distance of the actual risk preference r_i to risk neutrality ($r = 0$). Each of our n participants has a rank, denoted $\text{rank}(r_i^2)$. Decision makers closer to risk neutrality $r = 0$ will have a small rank, those further away will have a larger rank.

We might say that the binary lottery procedure performs well if participants under this

[21] Kass and Raftery (1995) suggest the following terminology: odds $\in [1 : 1, 2.72 : 1]$: only anecdotal evidence, odds $\in [2.72 : 1, 20.1 : 1]$: positive evidence, odds $\in [20.1 : 1, 148 : 1]$: strong evidence, odds $\in [148 : 1, \infty : 1]$: very strong evidence.

	mean ρ	$CI_{95}(\rho)$	odds $(\rho < 1/2)$	effss(ρ)	psrf(ρ)	median(ν)
EG task						
M1	0.42166	[0.33234, 0.50741]	27 : 1	1775	1.00592	∞
M2	0.42371	[0.33448, 0.50878]	23.8 : 1	1834	1.00283	∞
M3	0.42172	[0.33090, 0.50520]	29.6 : 1	1717	1.00074	11.4
M4	0.45647	[0.41506, 0.49740]	52.7 : 1	2369	1.00079	∞
HL task						
M1	0.52844	[0.48400, 0.57271]	1 : 8.61	12346	1.00001	∞
M2	0.52828	[0.48407, 0.57265]	1 : 8.49	13512	1.00029	∞
M3	0.50002	[0.44836, 0.55144]	1 : 1	9909	1.00024	1.46
M4	0.52853	[0.48810, 0.56959]	1 : 11.1	3589	1.00023	∞
HL task, consistent only						
M1	0.49737	[0.45108, 0.54319]	1.19 : 1	10631	1.00008	∞
M2	0.49758	[0.45117, 0.54368]	1.17 : 1	12005	1.00046	∞
M3	0.48397	[0.43584, 0.53184]	2.91 : 1	13074	1.00002	1.89
M4	0.49710	[0.45797, 0.53562]	1.25 : 1	2362	1.00077	∞

Table 7: Relative rank ρ of BLP procedure – online experiments.
Estimations for the lab version of the Holt and Laury (2002) task are shown in Table 8. Estimations for the comparison of labBLP and otherStd implementation of the EG task are shown in Appendix A.2 in Table 12.

procedure have mainly small ranks (i.e. have a small distance to risk neutrality) while participants under the standard procedure have larger ranks. We use the following measure:

$$\rho = \frac{\left(\sum_{i\in \mathrm{BLP}} \mathrm{rank}\,(r_i^2)\right) - \underline{R}}{\overline{R} - \underline{R}} \tag{5}$$

where $\underline{R}$ is the smallest possible sum of ranks BLP could obtain (i.e. all individual decision makers with BLP have a smaller r_i^2 than the other decision makers) and $\overline{R}$ is the highest possible sum of ranks (i.e. all individual decision makers with BLP have a larger r_i^2 than the other decision makers).

If the binary lottery procedure has no effect at all, then ρ should be $1/2$. If the binary lottery works perfectly, i.e. all individuals with BLP are closer to $r = 0$ than all the others, then ρ should be 0. Estimation results are shown in Table 7 for the online treatments and in Table 8 we contrast labBLP and otherStd.

For prior M1 and the EG task in the online treatments we find a change from 0.5 to 0.423, i.e. an effect of 0.077. The effect itself might be small, but the odds that $\rho < 1/2$ are 27:1, i.e. we have "strong" evidence that under the BLP procedure indeed more participants behave in a more risk neutral way.

For the HL task in the online treatments, if we include consistent and non-consistent observations, the effect is, actually, negative. If we consider consistent observations only, then the BLP procedure changes ρ from 0.5 to 0.497, i.e. an effect of 0.003. The odds that $\rho < 1/2$ are 1.19:1, i.e. we have "only anecdotal" evidence that under the BLP procedure indeed more participants behave in a more risk neutral way.

	mean ρ	$CI_{95}(\rho)$	odds $(\rho < 1/2)$	effss(ρ)	psrf(ρ)	median(ν)
HL task						
M1	0.44568	[0.41353, 0.47759]	2100 : 1	12850	1.00017	∞
M2	0.44606	[0.41391, 0.47817]	2100 : 1	13719	1.00013	∞
M3	0.44431	[0.40862, 0.48021]	929 : 1	9936	1.00007	2.13
M4	0.44669	[0.41906, 0.47469]	3640 : 1	3009	1.00038	∞
HL task, consistent only						
M1	0.45624	[0.42433, 0.48787]	281 : 1	13843	1.00043	∞
M2	0.45603	[0.42394, 0.48749]	363 : 1	14036	1.00044	∞
M3	0.45391	[0.41977, 0.48762]	252 : 1	11893	1.00005	2.29
M4	0.45537	[0.42844, 0.48190]	1290 : 1	2275	1.00028	∞

Table 8: Relative rank ρ of BLP procedure – lab/other experiments.
Estimations for the online version are shown in Table 7. Estimations for the comparison of labBLP and otherStd implementation of the EG task are shown in Appendix A.2 in Table 12.

Things look a bit better for the HL task if we consider the comparison of labBLP and otherStd. There the BLP reduces ρ from 0.5 down to 0.446, i.e. an effect of 0.054. This is a small effect, but we can be rather certain that there is at least an effect in the right direction. For the HL task the odds that $\rho < 1/2$ are 2100:1, i.e., we have "very strong" evidence that under the BLP procedure indeed more participants behave in a more risk neutral way.

Results are very similar for alternative priors M2 and M3 and for the model with random preferences, M4.

To summarise, our structural estimation suggests that for both the EG task and the HL task there is a rather small effect.

5. Conclusion

We design and report a simple experiment where we test whether the Binary Lottery Procedure induces participants to behave as if they are risk neutral in the Holt and Laury (2002) and Eckel and Grossman (2002) tasks. Namely, we ask participants to respond to an Eckel and Grossman (2002) type task as well as a Holt and Laury (2002) type task. The only tweak to these procedures that we make is to have participants choose across lotteries of tokens rather than prize money, exactly as the Binary Lottery Procedure proposes. If the procedure works as intended, we should observe all (or at least most) participants choosing lotteries that maximise expected earnings and thus all appear to be risk neutral agents. We compare choices in this tweaked procedure of the two tasks with a standard procedure where participants choose while faced with monetary prizes rather than tokens. Furthermore, we use a structural model to formalize our analysis on the efficacy of the Binary Lottery Procedure.

Putting together the results of the proportion of risk neutral agents, both in terms of the modal response within our data and the comparison with other studies should put one in serious doubt on whether BLP does indeed work in practice. Far from the majority of participants are found to be risk neutral and in fact we do not find large differences in the occurrence of risk neutral preferences between our own data and other comparison studies where similar risk elicitation procedures were implemented but with – as is standard – prize money. Given how the Binary Lottery Procedure can complicate tasks for participants and can be tedious in implementation and as it importantly appears to not work as hoped, it seems redundant for experimenters to be utilizing it. However, an experimenter who is already satisfied if participants behave in an only slightly more risk neutral way (rather than fully risk neutrally) might still prefer the BLP. For this case we have only mildly encouraging evidence that shows that the BLP moves participants in the right direction.

There may be, however, alternative methods for making participants behave as if there were risk neutral and there is some evidence that they work as intended. Kirchkamp et al. (2006) tell participants that their earnings in an auction are determined by the average of their payoffs from playing the auction multiple times (e.g. 50 times) with the same bidding function. They then find that bidding functions are much closer to the risk neutral equilibrium bids. Similarly, Niemeyer et al. (2019) let participants choose lotteries in a Holt and Laury (2002) task where the payoff is determined by the average result of many drawings from the chosen lottery. They find that most participants behave as if they were risk neutral.

References

Apesteguia, J. and Ballester, M. A. (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy*, 126(1):74–106.

Apesteguia, J., Oechssler, J., and Weidenholzer, S. (2019). Copy trading. *Management Science*. Forthcoming.

Arminger, G. and Muthén, B. (1998). A bayesian approach to non-linear, latent variable models using the gibbs sampler and the metropolis-hastings algorithm. *Psychometrika*, 63:271–300.

Azrieli, Y., Chambers, C. P., and Healy, P. J. (2019). Incentives in experiments with objective lotteries. *Experimental Economics*.

Becker, G. M., Degroot, M. H., and Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral Science*, 8(1):41–55.

Berg, J. E., Rietz, T. A., and Dickhaut, J. W. (2008). On the performance of the lottery procedure for controlling risk preferences. In Plott, C. R. and Smith, V. L., editors, *Handbook of experimental economics results*, volume 1, chapter 115, pages 1087–1097. Elsevier.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural india. *American Journal of Agricultural Economics*, 62(3):395–407.

Bock, O., Baetge, I., and Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, 71:117–120.

Brunner, C., Hu, A., and Oechssler, J. (2014). Premium auctions and risk preferences: an experimental study. *Games and Economic Behavior*, 87:467–484.

Butler, D., Isoni, A., and Loomes, G. (2012). Testing the 'standard' model of stochastic choice under risk. *Journal of Risk and Uncertainty*, 45(3):191–213.

Charness, G., Gneezy, U., and Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization*, 87:43–51.

Cipriani, M., Costantini, R., and Guarino, A. (2012). A Bayesian approach to experimental analysis: trading in a laboratory financial market. *Review of Economic Design*, 16(2):175–191.

Conte, A. and Hey, J. (2018). Rehabilitating the random utility model. A comment on Apesteguia and Ballester (2018). Discussion Papers in Economics 18/12, Department of Economics and Related Studies, University of York.

Dickhaut, J., Houser, D., Aimone, J. A., Tila, D., and Johnson, C. (2013). High stakes behavior with low payoffs: Inducing preferences with holt–laury gambles. *Journal of Economic Behavior & Organization*, 94:183–189.

Dürsch, P., Oechssler, J., and Vadovic, R. (2012). Sick pay provision in experimental labor markets. *European Economic Review*, 56(1):1–19.

Dürsch, P., Römer, D., and Roth, B. (2017). Intertemporal stability of uncertainty preferences. *Journal of Economic Psychology*, 60:7–20.

Eckel, C. C. and Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and human behavior*, 23(4):281–295.

El-Gamal, M. A., McKelvey, R. D., and Palfrey, T. R. (1994). Learning in experimental games. *Economic Theory*, 4(6):901–922.

El-Gamal, M. A. and Palfrey, T. R. (1996). Economical experiments: Bayesian efficient experimental design. *International Journal of Game Theory*, 25(4):495–517.

Engel, C. and Kirchkamp, O. (2019). How to deal with inconsistent choices on multiple price lists. *Journal of Economic Behavior & Organization*, 160:138 – 157.

Farjam, M. and Kirchkamp, O. (2018). Bubbles in hybrid markets: How expectations about algorithmic trading affect human trading. *Journal of Economic Behavior & Organization*, 146:248 – 269.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis*. CRC Texts in Statistical Science. Chapman & Hall, 3 edition.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.

Harrison, G. W. (1990). Risk Attitudes in First-Price Auction Experiments: A Bayesian Analysis. *The Review of Economics and Statistics*, 72(3):541–546.

Harrison, G. W., Martínez-Correa, J., and Swarthout, J. T. (2013). Inducing risk neutral preferences with binary lotteries: A reconsideration. *Journal of Economic Behavior & Organization*, 94:145–159.

Harrison, G. W., Martínez-Correa, J., and Swarthout, J. T. (2014). Eliciting subjective probabilities with binary lotteries. *Journal of Economic Behavior & Organization*, 101:128–140.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Holt, C. A. and Laury, S. K. (2014). Assessment and estimation of risk preferences. In *Handbook of the economics of risk and uncertainty*, volume 1, pages 135–201. Elsevier.

Hossain, T. and Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, 80(3):984–1001.

Kagel, J. H. and Roth, A. E., editors (1995). *The Handbook of Experimental Economics*. Princeton University Press.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–91.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kersting-Koenig, C., Merkel, A. L., and Lohse, J. (2019). Active and passive risk taking. mimeo.

Kirchkamp, O. and Reiß, J. P. (2019). Heterogeneous bids in auctions with rational and boundedly rational bidders: theory and experiment. *International Journal of Game Theory*.

Kirchkamp, O., Reiss, J. P., and Sadrieh, A. (2006). A pure variation of risk in first-price auctions. Technical report, Jena economic research papers.

Lange, K. L., Little, R. J. A., and Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84(408):881–896.

Loomes, G. (1998). Probabilities vs money: A test of some fundamental assumptions about rational decision making. *The Economic Journal*, 108(447):477–489.

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

Niemeyer, C., Reiss, P., and Sadrieh, K. (2019). Reducing risk in experimental games and individual choice. Technical report, KIT.

Proto, E., Rustichini, A., and Sofianos, A. (2020). Intelligence, errors and strategic choices in the repeated prisoners' dilemma. CESifo Working Paper Series 8068, CESifo Group Munich.

Roth, A. E. and Malouf, M. W. (1979). Game-theoretic models and the role of information in bargaining. *Psychological review*, 86(6):574.

Roth, B., Trautmann, S. T., and Voskort, A. (2016). The role of personal interaction in the assessment of risk attitudes. *Journal of Behavioral and Experimental Economics*, 63:106–113.

Schlag, K. H. and van der Weele, J. (2013). Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality. *Theoretical Economics Letters*, 3:38–42.

Schmidt, R. J. (2019). Identifying the ranking of focal points in coordination games on the individual level. University of Heidelberg: AWI Discussion Paper Series No. 660.

Schotter, A. and Trevino, I. (2014). Belief elicitation in the laboratory. *Annu. Rev. Econ.*, 6(1):103–128.

Selten, R., Sadrieh, A., and Abbink, K. (1999). Money does not induce risk neutral behavior, but binary lotteries do even worse. *Theory and Decision*, 46(3):213–252.

Smith, C. A. (1961). Consistency in statistical inference and decision. *Journal of the Royal Statistical Society: Series B (Methodological)*, 23(1):1–25.

Smith, V. L. (1964). Effect of market organization on competitive equilibrium. *The Quarterly Journal of Economics*, 78(2):182–201.

Vallois, N. and Jullien, D. (2018). A history of statistical methods in experimental economics. *The European Journal of the History of Economic Thought*, 25(6):1455–1492.

Wilcox, N. T. (2011). 'Stochastically more risk averse:' A contextual theory of stochastic discrete choice under risk. *Journal of Econometrics*, 162(1):89–104.

A. Alternative priors and convergence

A.1. Results for the Holt and Laury (2002) task

Figure 6 in Section 4.2 shows population parameters for the online treatments if priors follow M1. Figure 7 in Section A.1 extends Figure 6, adding more treatments and more priors (M1, M2, M3, and M4). Figure 8 does the same, taking into account only consistent decision makers. Table 9 shows diagnostics (effective sample size, Gelman et al., 2013, p. 287, and potential scale reduction factor, Gelman and Rubin, 1992) for the results shown in Figures 6, 7. Table 10 does the same for Figure 6 and 8.

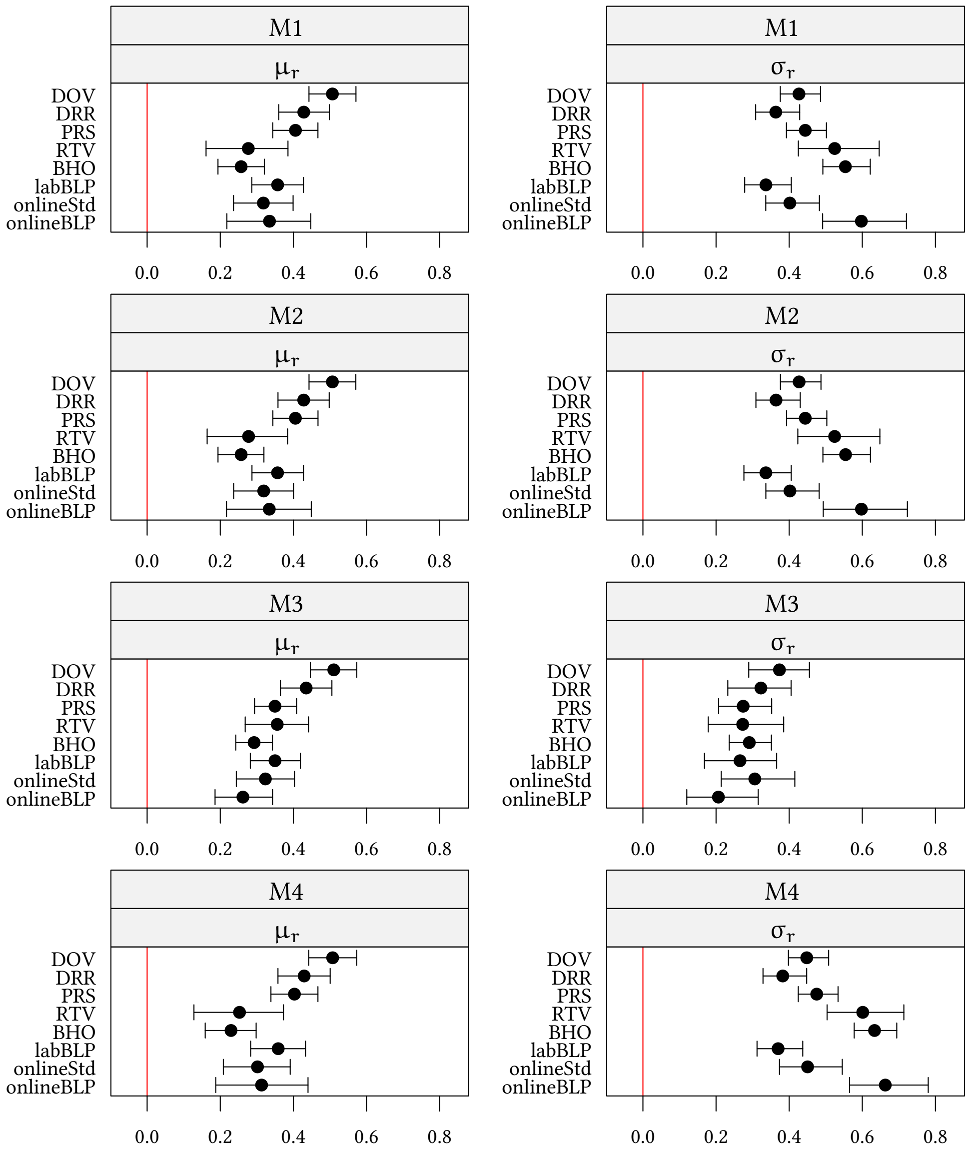

Figure 7: Population parameters for the Holt and Laury (2002) task.
This figure extends Figure 6. This figure shows, based on all decision makers, for each approach and for each procedure in the Holt and Laury (2002) task the median value of parameters μ_r and $\sigma_r = 1/\sqrt{\tau_r}$ together with a 95% credible interval. Table 9 shows convergence diagnostics.

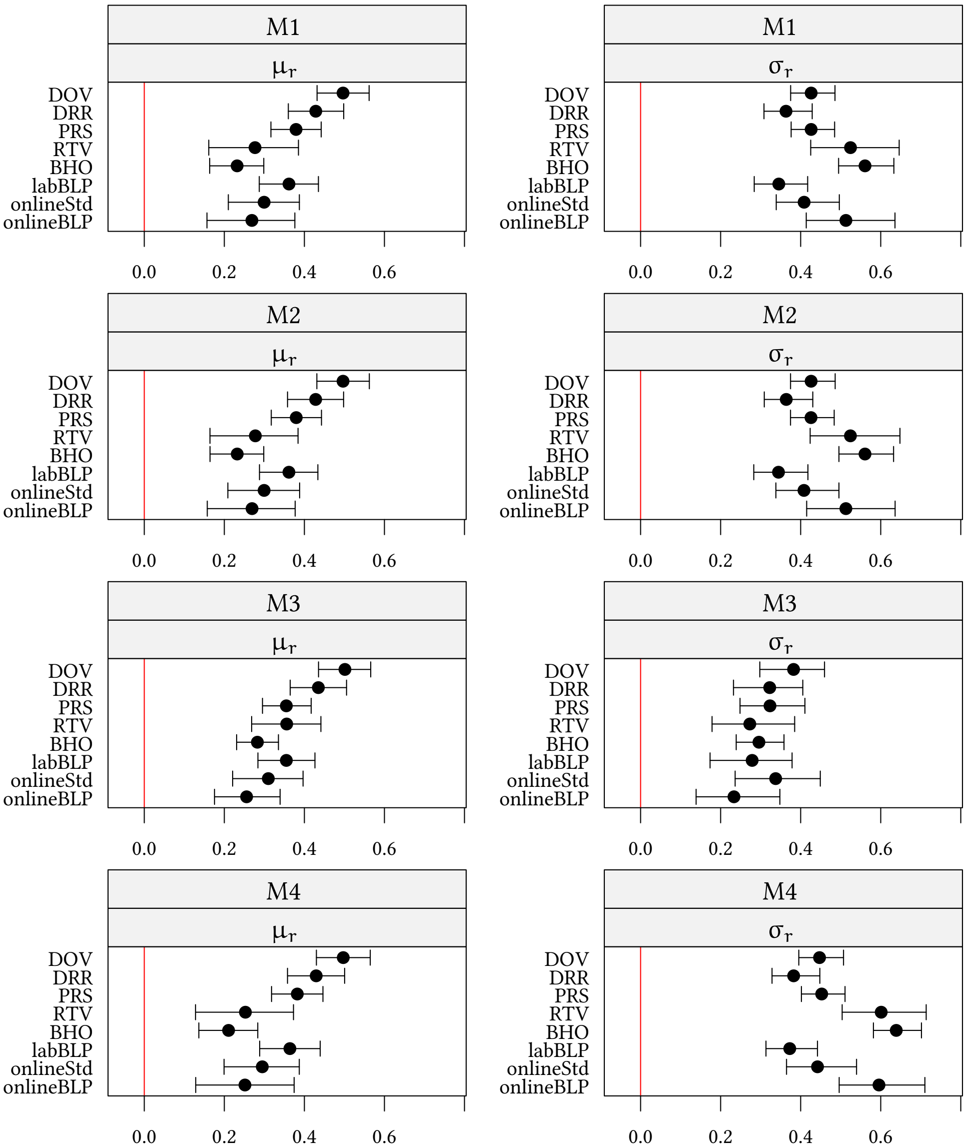

Figure 8: Population parameters for the Holt and Laury (2002) task, consistent only.
This figure extends Figure 6. This figure shows, based on only consistent decision makers, for each approach and for each procedure in the Holt and Laury (2002) task the median value of parameters μ_r and $\sigma_r = 1/\sqrt{\tau_r}$ together with a 95% credible interval. Table 10 shows convergence diagnostics.

	μ_r		σ_r	
	effss	psrf	effss	psrf
DOV M1	24152	1.00005	10609	1.00002
BHO M1	22607	1.00011	3536	1.00011
RTV M1	21796	1.00024	4635	1.00044
DRR M1	22974	1.00030	10070	1.00014
PRS M1	23843	1.00006	9330	1.00004
labBLPM1	19942	1.00012	7419	1.00059
onlineStd M1	21505	0.99999	8053	1.00088
onlineBLP M1	25064	1.00023	5860	1.00038
DOV M2	24222	1.00000	10634	1.00005
BHO M2	23680	1.00024	4046	1.00059
RTV M2	22829	1.00010	4556	1.00016
DRR M2	22689	1.00002	9968	1.00047
PRS M2	23977	1.00008	9229	0.99999
labBLPM2	20913	1.00016	7148	1.00002
onlineStd M2	20456	1.00012	7734	1.00039
onlineBLP M2	26567	1.00001	5366	1.00074
DOV M3	17079	1.00022	1986	1.00445
BHO M3	12583	1.00013	4315	1.00190
RTV M3	12129	1.00049	3987	1.00314
DRR M3	14607	1.00008	2507	1.00093
PRS M3	7664	1.00083	2460	1.00380
labBLPM3	12702	1.00024	1455	1.00497
onlineStd M3	13097	1.00021	2472	1.00632
onlineBLP M3	6206	1.00014	2492	1.00182
DOV M4	12275	1.00007	2170	1.00035
BHO M4	14821	1.00009	2372	1.00034
RTV M4	12043	1.00019	1791	1.00270
DRR M4	11981	1.00039	3052	1.00059
PRS M4	12853	1.00107	2304	1.00152
labBLPM4	9265	1.00014	2660	1.00170
onlineStd M4	10245	1.00036	1217	1.00223
onlineBLP M4	14581	1.00019	1715	1.00100

Table 9: Effective sample size and convergence for the Holt and Laury (2002) task.
The table shows effective sample size (effss) and potential scale reduction factor (psrf) for the estimation results shown in Figures 6 and 7.

	μ_r		σ_r	
	effss	psrf	effss	psrf
DOV M1	25932	1.00003	12417	1.00012
BHO M1	21220	1.00020	3791	1.00109
RTV M1	21796	1.00024	4635	1.00044
DRR M1	22974	1.00030	10070	1.00014
PRS M1	25500	0.99999	10913	1.00063
labBLPM1	21604	1.00016	8321	1.00065
onlineStd M1	24617	1.00035	8084	1.00112
onlineBLP M1	22956	1.00001	4546	1.00094
DOV M2	26003	1.00002	11099	0.99999
BHO M2	21948	1.00005	3896	1.00167
RTV M2	22829	1.00010	4556	1.00016
DRR M2	22689	1.00002	9968	1.00047
PRS M2	25672	0.99998	9926	1.00036
labBLPM2	22947	0.99997	7614	1.00078
onlineStd M2	24518	1.00015	7769	1.00038
onlineBLP M2	23265	1.00010	4795	1.00034
DOV M3	17129	1.00017	2686	1.00305
BHO M3	13880	1.00027	5008	1.00063
RTV M3	12129	1.00049	3987	1.00314
DRR M3	14607	1.00008	2507	1.00093
PRS M3	11066	1.00007	2501	1.00145
labBLPM3	12873	0.99998	1631	1.00157
onlineStd M3	15175	1.00023	2305	1.00097
onlineBLP M3	9382	1.00069	2814	1.00096
DOV M4	11767	1.00009	2685	1.00025
BHO M4	16927	1.00004	2274	1.00034
RTV M4	12043	1.00019	1791	1.00270
DRR M4	11981	1.00039	3052	1.00059
PRS M4	13712	1.00046	2373	1.00180
labBLPM4	12115	1.00013	2902	1.00081
onlineStd M4	10671	1.00001	1217	1.00056
onlineBLP M4	14387	1.00018	1448	1.00050

Table 10: Effective sample size and convergence for the Holt and Laury (2002) task (consistent only).

The table shows effective sample size (effss) and potential scale reduction factor (psrf) for the estimation results shown in Figures 6 and 8.

A.2. Results for the Eckel and Grossman (2002) task

In our estimations for the Holt and Laury (2002) task we assumed that each choice reflects one comparison of two lotteries, $\mathfrak{L}_A$ and $\mathfrak{L}_B$. Similarly, we model choices in the Eckel and Grossman (2002) task as a number of binary comparisons. To decide that, e.g., $\mathfrak{L}_3$ is better than $\mathfrak{L}_1$, $\mathfrak{L}_2$ and $\mathfrak{L}_4$, the decision maker makes three comparisons: $\mathfrak{L}_3 \succ \mathfrak{L}_1$, $\mathfrak{L}_3 \succ \mathfrak{L}_2$, and $\mathfrak{L}_3 \succ \mathfrak{L}_4$. Figure 6 in Section 4 shows population parameters for the online treatments. Figure 9 is an extension of Figure 6 providing credible intervals for the labBLP and otherStd versions of the Eckel and Grossman (2002) task and for the different estimation approaches M1, M2, M3, and M4. Table 5 in Section 4 shows estimates for the effect of BLP on the average coefficient of relative risk aversion μ_r for the online treatments. Table 11 extends Table 5 and shows estimation results for the labBLP and otherStd. Table 7 in Section 4 shows estimation of the relative rank ρ for the online treatments. Table 12 is an extension of Table 7 and show estimates for the labBLP and otherStd data. Equivalent results for the Holt and Laury (2002) task are shown in Tables 6 and 8. Table 13 shows convergence diagnostics for Figure 9.

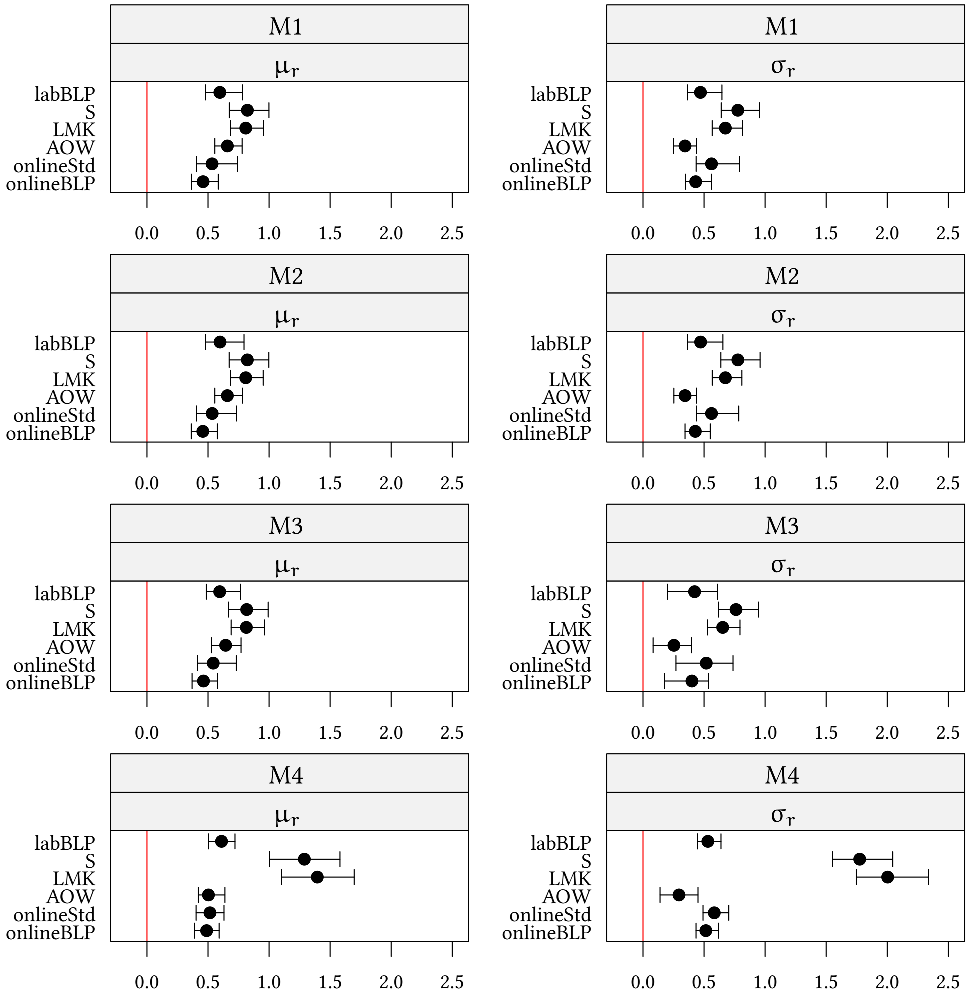

Figure 9: Population parameters for the Eckel and Grossman (2002) task.

This is an extended version of Figure 6. The figure shows for each approach and for each procedure in the Eckel and Grossman (2002) task the median value of parameters μ_r and $\sigma_r = 1/\sqrt{\tau_r}$ together with a 95% credible interval. Table 13 show convergence diagnostics.

	mean $\Delta_r = \lvert\mu_r^{std}\rvert - \lvert\mu_r^{BLP}\rvert$	mean Δ_r/σ_r^{std}	CI_{95} Δ_r	odds $\Delta_r > 0$	effss Δ_r	psrf Δ_r	median ν
Eckel and Grossman (2002) task							
M1	0.17215	0.28211	[−0.01482, 0.32344]	29.8 : 1	2607	1.00047	∞
M2	0.17185	0.28164	[−0.01745, 0.32163]	27.6 : 1	2640	1.00111	∞
M3	0.17206	0.28643	[−0.00891, 0.32324]	32.3 : 1	2371	1.00046	40.8
M4	0.83212	0.46554	[0.63449, 1.02878]	40000 : 0	2871	1.00212	∞

Table 11: Average effect of the BLP procedure on the average coefficient of relative risk aversion μ_r for Eckel and Grossman (2002) task in labBLP and otherStd.

Estimations for the online treatments are provided in Table 5.

	mean ρ	$CI_{95}(\rho)$	odds $(\rho < 1/2)$	effss(ρ)	psrf(ρ)	median(ν)
Eckel and Grossman (2002) task						
M1	0.40852	[0.34440, 0.49217]	54.1 : 1	1761	1.00036	∞
M2	0.40835	[0.34425, 0.49533]	47.7 : 1	1746	1.00156	∞
M3	0.40783	[0.34301, 0.48987]	63.3 : 1	1669	1.00070	40.8
M4	0.35985	[0.33781, 0.38257]	40000 : 0	1223	1.00747	∞

Table 12: Relative rank ρ of BLP procedure for the Eckel and Grossman (2002) task in labBLP and otherStd.

Table 8 shows results for the Holt and Laury (2002) task. Table 7 shows results for the online treatments

	μ_r		σ_r	
	effss	psrf	effss	psrf
AOW M1	2519	1.00059	1572	1.00088
S M1	6633	1.00047	3668	1.00044
labBLPM1	1844	1.00028	1551	1.00053
LMK M1	4667	1.00017	3210	1.00131
onlineStd M1	1541	1.00163	1161	1.00326
onlineBLP M1	4692	1.00060	2726	1.00110
AOW M2	2557	1.00213	1517	1.00297
S M2	6469	1.00007	3509	1.00035
labBLPM2	1711	1.00030	1495	1.00053
LMK M2	4854	1.00086	3206	1.00048
onlineStd M2	1876	1.00102	1449	1.00155
onlineBLP M2	5701	1.00011	3002	1.00011
AOW M3	1225	1.00443	268	1.02552
S M3	5313	1.00060	3007	1.00080
labBLPM3	2210	1.00097	717	1.01459
LMK M3	3494	1.00007	3145	1.00045
onlineStd M3	2178	1.00379	860	1.02530
onlineBLP M3	4853	1.00089	868	1.04741
AOW M4	226	1.00773	130	1.01871
S M4	5784	1.00063	403	1.00458
labBLPM4	9062	1.00108	2161	1.00230
LMK M4	2715	1.00142	113	1.09043
onlineStd M4	10041	1.00051	1600	1.00497
onlineBLP M4	10126	1.00018	2128	1.00124

Table 13: Effective sample size and convergence for the Eckel and Grossman (2002) task.
The table shows effective sample size (effss) and potential scale reduction factor (psrf) for the estimation results shown in Figure 9.

B. Payoffs for the experiments

Tables 1 and 2 in Section 2 show payoffs for our BLP treatments. Tables 14 and 15 in Appendix B show payoffs for the standard procedure experiments.

For labBLP treatment, the payment of participants was conducted as follows. First, one participant in each session flipped a coin to determine which task was the payoff relevant for everyone. Then each individual participant came to the front of the lab and flipped coins and/or rolled 10-sided dice to determine their payment. In addition to the show-up fee, average pay from the lottery choice was about 2.50 Euro for about 20 minutes. At the end there was a questionnaire where we asked for participants' gender.

For the online treatments the random draws were done by the computer and shown to participants immediately after they made their decisions. Participants in these treatments were paid via bank transfer.

AOW	7.2	6.4	*4	0.8							
	8	15	18.6	20.8							
BLP	38	28	*16	0							
	38	52	72	84							
LMK	6	5.4	4.8	4.2	3.6	3	2.4	1.8	1.2	*0.9	*0
	6	6.9	7.8	8.7	9.6	10.5	11.4	12.3	13.2	14.1	15
onlineBLP	38	28	*16	0							
	38	52	72	84							
onlineStd	38	28	*16	0							
	38	52	72	84							
S	4	3.5	3	2.5	2	1.5	*1	*0.5			
	4	5	6	7	8	9	10	10.5			

Table 14: Payoffs for the Eckel and Grossman (2002) task.

Choices classified as “risk neutral” are marked with a *. In labBLP and onlineBLP participants would obtain *tokens*. Each token had a probability of 1/100 to win 5€ (see Appendix C for the details). In the other experiments the prize was in €.

Lottery A	Lottery B
In 1 out of 10 cases you earn 40 tokens / 2€, in 9 out of 10 cases you earn 32 tokens 1.6€	In 1 out of 10 cases you earn 77 tokens/ 3.85€, in 9 out of 10 cases you earn 2 tokens / 0.1€
In 2 out of 10 cases you earn 40 tokens / 2€, in 8 out of 10 cases you earn 32 tokens 1.6€	In 2 out of 10 cases you earn 77 tokens/ 3.85€, in 8 out of 10 cases you earn 2 tokens / 0.1€
In 3 out of 10 cases you earn 40 tokens / 2€, in 7 out of 10 cases you earn 32 tokens 1.6€	In 3 out of 10 cases you earn 77 tokens/ 3.85€, in 7 out of 10 cases you earn 2 tokens / 0.1€
In 4 out of 10 cases you earn 40 tokens / 2€, in 6 out of 10 cases you earn 32 tokens 1.6€	In 4 out of 10 cases you earn 77 tokens/ 3.85€, in 6 out of 10 cases you earn 2 tokens / 0.1€
In 5 out of 10 cases you earn 40 tokens / 2€, in 5 out of 10 cases you earn 32 tokens 1.6€	In 5 out of 10 cases you earn 77 tokens/ 3.85€, in 5 out of 10 cases you earn 2 tokens / 0.1€
In 6 out of 10 cases you earn 40 tokens / 2€, in 4 out of 10 cases you earn 32 tokens 1.6€	In 6 out of 10 cases you earn 77 tokens/ 3.85€, in 4 out of 10 cases you earn 2 tokens / 0.1€
In 7 out of 10 cases you earn 40 tokens / 2€, in 3 out of 10 cases you earn 32 tokens 1.6€	In 7 out of 10 cases you earn 77 tokens/ 3.85€, in 3 out of 10 cases you earn 2 tokens / 0.1€
In 8 out of 10 cases you earn 40 tokens / 2€, in 2 out of 10 cases you earn 32 tokens 1.6€	In 8 out of 10 cases you earn 77 tokens/ 3.85€, in 2 out of 10 cases you earn 2 tokens / 0.1€
In 9 out of 10 cases you earn 40 tokens / 2€, in 1 out of 10 cases you earn 32 tokens 1.6€	In 9 out of 10 cases you earn 77 tokens/ 3.85€, in 1 out of 10 cases you earn 2 tokens / 0.1€
In 10 out of 10 cases you earn 40 tokens / 2€, in 0 out of 10 cases you earn 32 tokens 1.6€	In 10 out of 10 cases you earn 77 tokens/ 3.85€, in 0 out of 10 cases you earn 2 tokens / 0.1€

Table 15: Payoffs for the Holt and Laury (2002) task.

In the binary lottery procedure participants were instructed that they would obtain *tokens*. Each token had a probability of 1/100 to win 5€. (see Appendix C for details) For RTV, DRR, PRS, and BHO the prize was 2.00€ or 1.60€ for lottery A versus 3.85€ or 0.10€ for lottery B. For DOV the prize was three times as much: 6.00€ or 4.80€ for lottery A versus 11.55€ or 0.30€ for lottery B.

C. Instructions

[This is a translation of the original German instructions. In the onlineStd treatment, [tokens] are “Taler”. In the onlineBLP and labBLP treatment, [tokens] are “tokens”.]

Welcome to the Experiment

[[*The following information is shown only in onlineBLP and onlineStd:*]]:

In this experiment we study your decisions. The experiment is carried out jointly by the Universität Heidelberg and the Friedrich-Schiller-Universität Jena. During the experiment we elicit data in Jena:

- To make sure that you participate only once in the experiment you have received an invitation to the experiment from the Universität Heidelberg.

  In the next step we will ask for the email address. You can participate only once and only with this email address.

- Name, town of residence and bank account number: You will receive your payment after the experiment from the Universität Heidelberg via SEPA transfer. To be able to transfer the money, the Universität Heidelberg needs your name, your town of residence, the zip code of your town of residence and your bank account number (IBAN). We will ask for this information.

- Decisions: We will then describe your choice situation. You will enter your decisions.

Your personal data and your choices will be transferred at least four weeks after the experiment to the Universität Heidelberg. Immediately thereafter your personal data will be deleted in Jena. For more information on processing your data in Jena contact Prof. Dr. Oliver Kirchkamp[22].

The administration of the Universität Heidelberg will transfer your payment. Your personal data will be separated from your decisions. Only your decisions will be used in our research. For more information regarding the processing of your data in Heidelberg please contact Prof. Dr. Jörg Oechssler.

If you agree with this procedure and if you want to participate in the experiment, please click on “Agree”:

Don’t Agree | Agree

To be able to pay you, the Universität Heidelberg needs your name, your town of residence and your bank account number:

[22]Here, and for Prof. Dr. Jörg Oechssler, was a link to the respective homepage.

Please enter your name here:	
Please enter your town of residence here:	
Please enter the ZIP code of your town of residence here:	
Please enter your IBAN here:	

Continue

[[*The following information is shown only in labBLP:*]]

Please read these instructions carefully. Please do not talk to other participants. Please turn off your mobile phone and leave it turned off until the end of the experiment. If you have any questions, please raise your hand, and someone will come over. All participants have received the same instructions.

[[*The following information is shown in all treatments:*]]

This experiment consists of two decision problems. One of the decision problems will be chosen for payment in the end…

onlineBLP and onlineStd:	*labBLP:*
by the computer with a probability of 50%.	by flipping a coin.

Hence, you should work through both parts carefully because both parts can be relevant for your payoff.

Additionally to the payoff from the decision problem, each participant receives…

onlineBLP and onlineStd:	*labBLP:*
2.5€ for participation.	5€ for participation.

onlineBLP and onlineStd:	*labBLP:*
	All random decisions of coin flips and dice rolls will be made at the end of the experiment by either yourself or some other volunteer with fair and genuine coins and dice.

Decision problem 1

onlineBLP and onlineStd:	*labBLP:*
	Decision problem 1 will be chosen for payoff if the flipped coin shows "Tails".

In this experiment you have to make a choice among 4 lotteries.

onlineBLP and onlineStd:	*labBLP:*
Each lottery has two possible outcomes, Heads or Tails. Both outcomes are equally probable (i.e. both have a probability of 50%. At the end of the experiment the computer chooses randomly one of the outcomes, Head or Tails.	Each lottery has two possible outcomes. The outcome will be decided by flipping a coin at the end of the experiment (of course, this is independent of the coin above).

Depending on whether the coin comes up "heads" or "tails" you can win a number of "[tokens]".

For example, in lottery 2 you win 52 [tokens] if heads comes up and 28 [tokens] if tails comes up.

onlineBLP: Tokens will later give you the opportunity to win 5 Euro by the draw of a random number between 1 and 100. Each random number between, and including, 1 and 100 is equally likely to occur.	*onlineStd:* Taler are converted into Euro at the end of the experiment. The exchange rate is 20 Taler for 1 Euro.	*labBLP:* Tokens will later give you the opportunity to win 5 Euro by the draw of a random number between 1 and 100. Each random number between, and including, 1 and 100 is equally likely to occur. In fact, you will be able to draw the two random numbers yourself by rolling two 10-sided dice. The tokens give you the chance of winning the 5 Euro.
onlineBLP: The more points you earn, the greater your chance of winning 5 Euro. In particular, if the random number is equal or less than the number of tokens you own, then you win 5 Euro. If the random number is higher, you win nothing.	*onlineStd:*	*labBLP:* The more points you earn, the greater your chance of winning 5 Euro. In particular, if the random number generated by the dice is equal or less than the number of tokens you own, then you win 5 Euro.
onlineBLP: For example, if you chose lottery 2 and tails comes up, you win 28 tokens. If the random number is 28 or less, you win 5 Euro. If the random number is higher than 28, you win nothing.	*onlineStd:* For example, if you chose lottery 2 and tails comes up, you win 28 Taler. At the end these will be converted into 1.40 Euro.	*labBLP:* For example, if you chose lottery 2 and tails comes up, you win 28 tokens. If the random number is 28 or less, you win 5 Euro. If the random number is higher than 28, you win nothing.

Now please choose one of the lotteries 1-4.

Lottery	Tokens if coin shows Heads	Tokens if coin shows Tails	Please choose exactly one lottery
1	38	38	☐
2	28	52	☐
3	16	72	☐
4	0	84	☐

Decision problem 2

onlineBLP and onlineStd:

labBLP:
Decision problem 2 will be chosen for payoff if the flipped coin shows "Heads".

In this experiment you have to make a choice across 2 lotteries, A or B, in 10 different cases. Each lottery has the same structure: with some probability, you receive a large amount of [tokens] and with the residual probability, you will receive a smaller amount of [tokens].

onlineBLP and onlineStd:
The outcome of each chosen lottery will determined randomly at the end of the experiment. Each number, 1, 2, 3,…10, will be drawn by the computer with the same probability. At the end of the experiment the computer will also determine which of the 10 choices will be the payoff relevant one.
For example, say that choice 3 was determined as payoff relevant and say that in the third row you chose lottery A. If at the end a random number between 1 and 3 was drawn, you receive 40 [tokens] and if a random number between 4 and 10 was drawn, you receive 32 [tokens].

labBLP:
Which of the 10 choices will be the payoff relevant one will be decided by throwing a 10-sided die at the end of the experiment. Further, the outcome of the chosen decision will be decided by throwing a 10-sided die.
For example, say the 10-sided die rolls a 3 and in the third row you chose lottery A. If now the 10-sided die rolls a number between 1 and 3, you receive 40 [tokens] and if it rolls a number between 4 and 10, you receive 32 [tokens].

onlineBLP:
Tokens will later give you the opportunity to win 5 Euro by the draw of a random number between 1 and 100. Each random number between, and including, 1 and 100 is equally likely to occur.

onlineStd:
If you have won 40 Taler, these will be converted into 2 Euro. If you have won 32 Taler, these will be converted into 1.60 Euro. We still use the exchange rate of 20 Taler for one Euro.

labBLP:
Tokens will later give you the opportunity to win 5 Euro by the draw of a random number between 1 and 100. Each random number between, and including, 1 and 100 is equally likely to occur.

onlineBLP:	onlineStd:	labBLP:
If the random number is equal or less than the number of tokens you own, then you win 5 Euro. If the random number is larger, you win nothing. Hence, the more points you earn, the greater your chance of winning 5 Euro.		In fact, you will be able to draw the two random numbers yourself by rolling two 10-sided dice. The tokens give you the chance of winning the 5 Euro. The more points you earn, the greater your chance of winning 5 Euro. In particular, if the random number generated by the dice is equal or less than the number of tokens you own, then you win 5 Euro.

onlineBLP and onlineStd:	labBLP:
Now please choose one of the lotteries A or B across the different cases below by clicking on your preferred lottery in all 10 cases.	Now please choose one of the lotteries A or B across the different cases below by underlining your preferred lottery in the middle, in all 10 cases.

Lottery A	Your choice	Lottery B
	(please [choose/ underline] one lottery in each row)	
40 tokens if die shows 1 32 tokens if die shows 2-10	A or B	77 tokens if die shows 1 2 tokens if die shows 2-10
40 tokens if die shows 1-2 32 tokens if die shows 3-10	A or B	77 tokens if die shows 1-2 2 tokens if die shows 3-10
40 tokens if die shows 1-3 32 tokens if die shows 4-10	A or B	77 tokens if die shows 1-3 2 tokens if die shows 4-10
40 tokens if die shows 1-4 32 tokens if die shows 5-10	A or B	77 tokens if die shows 1-4 2 tokens if die shows 5-10
40 tokens if die shows 1-5 32 tokens if die shows 6-10	A or B	77 tokens if die shows 1-5 2 tokens if die shows 6-10
40 tokens if die shows 1-6 32 tokens if die shows 7-10	A or B	77 tokens if die shows 1-6 2 tokens if die shows 7-10
40 tokens if die shows 1-7 32 tokens if die shows 8-10	A or B	77 tokens if die shows 1-7 2 tokens if die shows 8-10
40 tokens if die shows 1-8 32 tokens if die shows 9-10	A or B	77 tokens if die shows 1-8 2 tokens if die shows 9-10
40 tokens if die shows 1-9 32 tokens if die shows 10	A or B	77 tokens if die shows 1-9 2 tokens if die shows 10
40 tokens if die shows 1-10	A or B	77 tokens if die shows 1-10