Payoff and Presentation Modulation of Elicited Risk Preferences in MPLs

Sameh Habib[1], Daniel Friedman[2], Sean Crockett[2] and Duncan James[3]

[1]Economics Department, University of California, Santa Cruz, sfhabib@ucsc.edu, 401 Engineering 2 Building, 1156 High Street, Santa Cruz, CA 95064
[2]Economics Department, University of California, Santa Cruz, dan@ucsc.edu
[2]Zicklin School of Business, Baruch College sean.crockett@baruch.cuny.edu
[3]Economics Department, Fordham University, dujames@fordham.edu

October 5, 2016

Abstract

Since Holt & Laury (2002), the multiple price list (MPL) procedure has been widely used to elicit individual risk preferences. We assess the impact of varying list order and spacing, and of presentation via text or graphs. Relative to the original MPL baseline, some nonlinear transformations of lottery prices systematically increase elicited risk aversion, while some graphical displays tend to reduce it.

Keywords Multiple Price List, Elicitation, Risk Aversion, Experiment
JEL Classification C91, D81, D89

Acknowledgement We are grateful for support from the National Science Foundation under grant SES-1357867. As usual, the NSF played no role in the design of our study, nor in the collection, analysis and interpretation of data, nor in the writing of the report, nor in the decision to submit the article for publication. We also thank Tobias Schmidt for a helpful pointer to previous literature, and two reviewers and an editor of this journal for very helpful comments and suggestions.

1 Introduction

To judge by citation counts (over 3600 on Google Scholar as of July 2016), the Holt & Laury (2002) Multiple Price List (MPL) procedure may now be the most widely used tool to elicit risk preferences. The standard MPL procedure confronts a subject with ten rows, in each of which the subject is to choose either Lottery A or B. Given the parameters in the first six columns of Table 1, the vast majority of subjects will choose lottery A in the first row and lottery B in the last row. An expected utility maximizing subject would switch from column A to column B at most once at some point in between. A subject who switches between rows 3 and 4 is revealed to be slightly risk seeking, for example, while one who switches between rows 8 and 9 is highly risk averse.

		Lottery A		Lottery B		Calculated	
Row	prob1	A-prize1	A-prize2	B-prize1	B-prize2	EV[A]-EV[B]	$\hat{r}$
1	0.1	2.00	1.60	3.85	0.10	1.16	-1.71
2	0.2	2.00	1.60	3.85	0.10	0.83	-0.95
3	0.3	2.00	1.60	3.85	0.10	0.49	-0.49
4	0.4	2.00	1.60	3.85	0.10	0.16	-0.15
5	0.5	2.00	1.60	3.85	0.10	-0.17	0.15
6	0.6	2.00	1.60	3.85	0.10	-0.51	0.41
7	0.7	2.00	1.60	3.85	0.10	-0.84	0.68
8	0.8	2.00	1.60	3.85	0.10	-1.18	0.97
9	0.9	2.00	1.60	3.85	0.10	-1.51	1.37
10	1	2.00	1.60	3.85	0.10	-1.85	-

Table 1: The original Holt & Laury (2002) Multiple Price List parameters appear in the first six columns. The probability in each row of receiving the lesser prize (1.60 in Lottery A and 0.10 in Lottery B) is prob2 = (1-prob1). The remaining two columns show the difference in expected values of the two lotteries, and the approximate solution $\hat{r}$ to the equation EU[A] = EU[B] at that row, where $U(x) = \frac{x^{1-r}}{1-r}$.

Given the widespread use of MPLs, it is important to explore its sensitivity to perturbations. Bosch-Domènech & Silvestre (2013), like Andersen *et al.* (2007) , finds that cropping the original Holt-Laury list of lottery choices leads to systematically different switching points than in the original, and thus to different inferred preferences. Lévy-Garboua *et al.* (2012) finds that reversing the order in which the choices between lotteries are presented changes the inferred preferences, as does changing how much of the list the subjects can see at any one time. (The original Holt-Laury procedure presents all 10 choices at once, in a single list, as in Table 1.)

The present paper continues this exploration. We too modulate list construction, but not by cropping the the list. We retain ten rows and consider five different parameter schedules. Some of the schedules involve the nonlinear payoff transformations seen previously in Cox *et al.* (1988) treatments of first price auction and in James (2007) treatments of the Becker-DeGroot-Marschak procedure. One of our schedules, like Lévy-Garboua *et al.* (2012), reverses the order of the MPL.

Equally important, we consider five different display options. Besides the usual text (or tabular) display as in Figure 1, we also offer a variety of graphical displays as in Figure 2. These display options, and the different MPL schedules, are detailed in the next section, along with testable hypotheses.

None of our modulations or display options would affect the elicited risk preferences of expected utility maximizers, but in Section 3 we will see that several of them have economically and statistically significant effects on our human subjects. The last section summarizes, and offers some interpretations and suggestions for future work.

2 Experiment

Our parametric treatments include the original Holt & Laury (2002) MPL schedule (MPLa) and four alternative schedules that are detailed in Appendix C. Treatment MPLa is shown in the first six columns of Table 1. Treatment MPLb subtracts 0.05 from all prizes and switches the placement of A prizes (originally columns 3-4 of the Table) and B prizes (originally columns 5-6). Treatment MPLc adds 0.15 to all prizes in MPLa and reverses the row order. Treatment MPLd replaces all prize amounts in MPLa by their square roots (rounded to the nearest 0.01), and treatment MPLe replaces all prize amounts in MPLa by their (rounded) squares.

Our display treatments include the original text-only presentation of each row in a schedule (Text) as in Figure 1. The four additional display treatments use rotatable, 3-d pie charts to represent each gamble. Treatment TT includes text labels as in Figure 2, while treatment TF suppresses the text label for prize amounts, FT suppresses the text label for the probabilities, and FF suppresses both labels in the graphical display.

Each *trial* consists a single price list, i.e., ten rows of pairwise lottery choices, using one of the five schedules and one of the five display options. A *sequence* consists of five trials, using all five schedules and all five displays once each, e.g., Text-MPLa, TF-MPLc, FT-MPLb, TT-MPLe, FF-MPLd. There are $5! \times 5! = 14400$ different possible sequences, so a full factorial design is impractical.

We employ a fractional factorial design using just 25 of the sequences, shown in on-line Appendix A. Each MPL schedule and each display occurs in a given trial position (first, second, third, fourth, fifth) in 5 of these 25 sequences, and schedules and displays vary independently. The idea is to eliminate sequence effects by balancing the order in which the schedules and displays occur.

Each subject completes two of the chosen 25 sequences, selected randomly subject to the constraint that no subject sees any MPL twice in the same display. For example, if she sees Text-MPLa somewhere in the first sequence, then in the second sequence (trials 6-10) MPLa appears once in one of the graphical displays (TT, TF, FT or FF). Thus each subject makes a total of 100 binary choices, comprising two sequences of five trials each, and each trial elicits the subject's risk preferences via 10 choices between some lottery A and another lottery B.

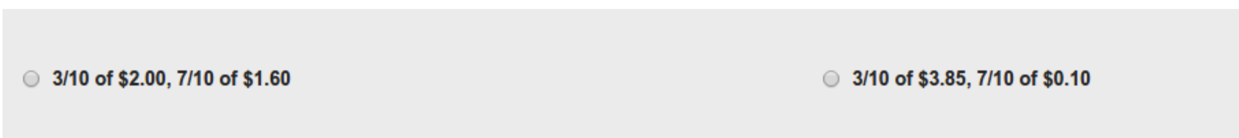

Figure 1: Text display for MPLa, row 3 only. Actual display stacks all 10 rows on the same page. Subject clicks a radio button to choose between the two lotteries.

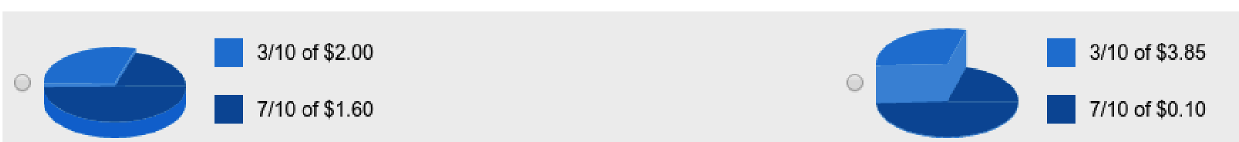

Figure 2: Graphical display TT for MPLa, row 3 only. Actual display stacks all 10 rows on the same page. Subjects can rotate each pie by mousing over it. Wedge angles encode probability and wedge height encodes prize amount.

Subjects are drawn from the LEEPS lab pool at UCSC. We report results from four sessions with a total of 47 subjects. Sessions begin with instructions describing the MPL choices and all display types. Subjects then proceed to make practice (unpaid) decisions in a sequence of five displays of the same practice MPL; the (hypothetical) prize amounts can be seen in the Instructions reproduced in on-line Appendix B. After questions have been answered, subjects complete live five trials in sequence I, and then five more trials in sequence II. When all subjects have completed both sequences, ten-sided dice are rolled to determine which trial counts for payment, which row in that trial counts, and which prize (high or low) will be paid for the chosen lottery (A or B). Subjects receive the prizes so determined in cash, plus a $7 show-up fee. Sessions lasted about 45-60 minutes and average payment was around $10 with a maximum payment of $22.

2.1 Testable Hypotheses

The natural null hypothesis is that none of our treatments has any systematic effect on subjects' elicited preferences. However, it is possible that different displays will engage subjects differently.

Recall that a risk-neutral subject would always choose the lottery with larger expected value. Our graphical displays should make it easier for the subject to apprehend the expected value of each gamble. To the extent that our subjects find the expected value of a lottery to be salient, that ease of apprehension would encourage them to reveal preferences closer to risk neutrality.[1] Thus an alternative (research) hypothesis is

H1: On average, elicited preferences in the TT, TF, FT and FF treatments will be closer to risk neutral than in the Text treatment.

The prize amounts are explicit in the Text, FT and TT treatments but not in the TF and FF treatments. If subjects are averse to payoff ambiguity (i.e., to observing only approximate relative prize amounts rather than exact amounts), then we would see lower certain equivalents in the non-explicit treatments, especially for the riskier gambles. This leads to another research hypothesis,

H2: On average, elicited preferences in the TF and FF treatments will indicate greater risk aversion than in the other treatments.

Of course, the premise here is not that subjects actually are more risk averse in those two treatments, but rather that a particular sort of ambiguity aversion would produce choices "as if" more risk averse.

Does similar reasoning apply to lack of explicit probabilities in the FT and FF treatments? We don't think so, because the probabilities of the two possible prizes in any gamble always sum to 1.0, corresponding to pie slice angles that always sum to 360^o. Rotating the pies allows subjects to gauge the probabilities quite accurately in all treatments.

It is also possible that different MPL schedules will engage subjects differently. For constant relative risk averse preferences (CRRA, defined by the last expression in Table 1's

[1] Camerer (1989) offers a similar conjecture, but does not elaborate or test it. In his experiment expected values are proportional to displayed areas, while in our experiment they are proportional to the volumes of the 3-d pies.

caption), positive (or negative) parallel shifts in all prize amounts create income effects that in each row will reduce (or increase) the parameter level $\hat{r}$ that makes an expected utility maximizer indifferent between the two gambles. On-line Appendix C shows that these effects are minor for the shifts (plus 15 cents and minus 5 cents) used in MPLb and MPLc. However, an empirical finding in Lévy-Garboua *et al.* (2012) suggests that the reverse lottery ordering in MPLc might reduce elicited risk aversion.

H3: On average, elicited preferences in the MPLc treatment will be closer to risk neutral than in the MPLa treatment.

Likewise, based on the empirical results reported in Cox *et al.* (1988) and in James (2007), we conjecture that the nonlinear transformation of prizes will affect observed behavior as follows.

H4: On average, elicited preferences in the MPLd treatment will, as in Cox *et al.* (1988) and James (2007) but counter to EUT, be closer to risk neutral than in the MPLa treatment, and those in the MPLe treatment will indicate greater risk aversion than in MPLa.

It may be worth emphasizing that, like the other research hypotheses, H4 is a behavioral prediction, and not a consequence of expected utility theory. Suppose, for example, that a subject always chooses so as to maximize the CRRA utility function $U(x) = x^{0.5}$. That subject would make row-by-row choices in MPLe as if maximizing $U(x) = (x^2)^{0.5} = x$ in MPLa — she would switch after row 4 in MPLe, which is the same row at which a risk-neutral subject would switch in MPLa. However, when confronted with MPLa, that subject would switch after row 5, and in both cases the elicited CRRA risk preference parameter (explained below) would be in the vicinity of 0.5.

3 Results

Each trial consists of 10 binary choices: either column A or column B in each row of an MPL display. Recall from the caption to Table 1 that, for a given row, $\hat{r}$ is the risk parameter value that makes a CRRA expected utility maximizer indifferent between the two columns. We summarize each trial outcome in two performance variables:

- r = the mean of $\hat{r}$ for the row just before and just after the unique column switch, and
- s = the number of safe choices (of the lottery with prizes closer together, e.g., 1.60 or 2.00 in MPLa) less the number of rows where the safe choice has the larger EV.

	r			s		
Treatment	mean	s.d.	Nobs.	mean	s.d.	Nobs.
FF	0.45	0.59	40	1.80	2.20	47
FT	0.50	0.55	40	1.93	1.89	47
Text	0.55	0.52	41	2.13	1.84	47
TF	0.43	0.60	38	1.64	1.98	47
TT	0.56	0.55	38	1.98	1.84	47
MPLa	0.49	0.39	40	1.62	1.39	47
MPLb	0.66	0.52	38	1.71	1.67	47
MPLc	0.41	0.38	40	0.59	1.43	47
MPLd	0.30	0.91	40	1.37	1.71	47
MPLe	0.64	0.28	39	4.20	1.46	47

Table 2: Summary statistics by treatment in display sequence 1. For the parametric (r) and non-parametric (s) performance variables defined in the text, the average (mean) and standard deviation (s.d.) across all (Nobs) subjects in each treatment are shown.

The variable r is often used in the literature, but it depends on a particular parametric form (CRRA) and is sometimes not well defined. Following a practice frequently seen in previous work, our r analysis excludes trials with more than one column switch, or none: about 16% of observations. The non-parametric variable s is always well defined, even in the cases of multiple switches or no switches. For example, in MPLa, the safe choice is Lottery A, and it has the larger expected value only in the first 4 rows. Suppose that the subject chose lottery A in rows 1-5 and 7 of an MPLa trial. Then $s = 6 - 4 = 2$ for that trial, while r is not defined due to multiple switches. Zero values of s or r indicate risk neutrality, while higher values indicate a greater degree of risk aversion.

Table 2 collects summary statistics. One sees preliminary evidence favoring Hypothesis 1 (both r and s are a bit higher in Text than in most of the graphical treatments), and disfavoring Hypothesis 2 (both r and s are lowest in the "ambiguous" treatments TF and FF). Consistent with Hypothesis 3, both performance values are smallest for schedule MPLc. Hypothesis 4 also seems consistent with the summary data in that MPLe has a high r value and a huge s value.[2]

[2]Why is the gap in mean s between MPLe and, say, MPLa so much larger than the corresponding gap in r? We see two reasons. First, the 8 MPLe trials excluded for r exhibit higher mean values of s than the other trials. Second, the convex transform behind MPLe compresses the range of the $\hat{r}$'s, artificially lowering the mean (and standard deviation) of r in Table 2.

3.1 Treatment Effects

We now test the research hypotheses H1 – H4 via regressions of the form

$$y_{it} = \alpha + \beta \cdot x_t + \epsilon_{it}. \tag{1}$$

Here y_{it} is measured risk aversion, either s or r, obtained for subject i in trial t. The explanatory variable x_t is the 8-vector of dummy variables for the five MPL schedule treatments and the five display treatments in trial t; the two omitted dummies are for Text and MPLa. Thus the β estimates capture the first-order treatment effects relative to Text and MPLa. Equation (1) is estimated via OLS with standard errors clustered at the subject level. The focus in this subsection is on the nonparametric measure s; on-line Appendix A reports qualitatively similar results using the parametric measure r.

Table 3 reports the estimated β coefficients for s. Although few of them differ from zero at conventional significance levels, the signs of all estimated display coefficients are consistent with H1, that graphical displays encourage more nearly risk-neutral choices. The estimated impact is larger — about half a row on average in the first 5-trial sequence — in the graphical treatments (FF and TF) that suppress the text for prize size. This is contrary to Hypothesis H2, which predicts that behavior will be further away from risk neutrality in those treatments. Display treatment effects are consistently smaller in the second sequence, as one might expect in a within-subject design. Even with our relatively small data sample,[3] the TF coefficient is significant at the 10% confidence level in the first sequence and overall.

MPL treatments a and b are indistinguishable in Table 3; evidently switching the left-right positions of the A and B columns does not change revealed risk preference estimates. However, consistent with H3 and prior work (Lévy-Garboua *et al.* (2012)), reversing the top-to-bottom ordering of the lotteries in treatment MPLc increases the mean number of risky choices by a whole row, a highly significant shift towards risk neutrality. Consistent with H4, Treatment MLPe has a huge (>2 rows!) and highly significant impact in the opposite direction (contrary to EUT prediction), implying far more risk averse choices than in MPLa. Support for the other part of H4 is more equivocal: MPLd indeed shifts choices towards risk neutrality, but the shift in s is not significant in either sequence.[4]

[3] Power tests collected in on-line Appendix A suggest that, given the observed means and standard deviations in sequence I, a sample twice as large as ours would suffice to reject the null hypothesis at the 5 percent level in 90 percent of samples for the TF and FF treatments, but it would take a sample about ten times as large to do so for the FT and TT treatments.

[4] Power calculations collected in Appendix A suggest that the current sample size suffices to reject the null hypothesis for MPLd in sequence II, but that the effect size in sequence I is so small that no reasonable

	Dependent variable: s		
	seq. I	seq. II	full sample
MPLb	0.199 (0.302)	−0.008 (0.336)	0.097 (0.225)
MPLc	−1.004*** (0.302)	−1.056*** (0.336)	−1.028*** (0.225)
MPLd	−0.047 (0.302)	−0.432 (0.336)	−0.238 (0.225)
MPLe	2.826*** (0.302)	2.323*** (0.336)	2.575*** (0.225)
FF	−0.470 (0.302)	−0.103 (0.336)	−0.288 (0.225)
FT	−0.189 (0.302)	−0.147 (0.336)	−0.170 (0.225)
TF	−0.534* (0.302)	−0.336 (0.336)	−0.436* (0.225)
TT	−0.225 (0.302)	−0.076 (0.336)	−0.151 (0.225)
Constant	1.876*** (0.288)	1.784*** (0.316)	1.830*** (0.213)
Observations	235	235	470
R^2	0.451	0.343	0.392
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 3: Slope (β) and intercept (α or Constant) coefficient estimates (and standard errors) for performance variable $y = s$ in equation (1). The eight slope coefficients are estimated treatment effects relative to the omitted treatments Text and MPLa for trials $t = 1, .., 5$ in the "seq. I" column, $t = 6, ..., 10$ in the "Seq. II" column, and $t = 1, ..., 10$ in the last column.

As a robustness check, taking a cue from Hey & Orme (1994), we run a fully disaggregated specification that breaks each trial into its 10 component binary choices. From this perspective, our full sample contains 47 Subjects × 10 trials per subject × 10 observations/trial = 4700 observations. The regression equation is

$$RN_{igp\tau} = \alpha_i + \beta \cdot x_\tau + \theta \text{ AEVD}_{p\tau} + \gamma\tau + \epsilon_{igp\tau}, \tag{2}$$

where RN is a dummy variable indicating whether (=1) or not (=0) subject i's choice is for the gamble (A or B) with weakly larger expected value, given graphical display g, MPL p, and choice number τ out of the subject i's 100 binary choices. As before, x is an 8-vector

sample would suffice.

of treatment dummies (with MPLa and Text omitted), and β is the coefficient vector to be estimated. To account for expected payoff differences across rows and schedules, and for time trend, here we also include AEVD (= absolute value of [EV of gamble A minus EV of gamble B]), and τ, as explanatory variables.

The regression results reported in Table 4 are fairly consistent across the four specifications. The coefficient signs again suggest that most of the graphical display treatments encourage more risk-neutral (RN=1) choices as in hypothesis H1, but again the estimated magnitudes are insignificant. Again the evidence fails to support H2; here one "ambiguous" display treatment has the smallest (indeed negative) coefficient and the other has the largest, but both are insignificant. Once again, schedule MPLe is an outlier, with significant larger deviations from risk neutral choices; more modest but significant treatment effects are detected here for MPLb (less risk neutral behavior) and MPLd (more risk neutral behavior). Again, these results run counter to EUT predictions, but are in line with prior empirical work using the same transforms.

3.2 Within subject tests

Since our design exposes each subject to each treatment twice, we can make within subject comparisons. These are potentially sharper in that they eliminate individual idiosyncrasies not captured in fixed effects. On the other hand, the within subject comparisons would be blunted by learning dynamics if subjects transfer insights gained from one treatment to subsequent treatments.

Table 5 summarizes the within subject effects of the display and MPL treatments. All estimated coefficients for displays are positive, consistent with H1, that graphical displays push subjects towards risk neutrality. The first column estimates consistently have larger magnitude than the corresponding regression coefficients in Table 3, suggesting that the within subject comparisons tend to be a bit sharper. In the second column the largest effect again is for the TF treatment; once comparisons involving the parameter set MPLe (with its extreme elicited values) are removed, the size of the effect is almost one full row, significant at the 1% level. A nonparametric Wilcox signed rank test gives qualitatively similar results.

Do subjects reveal less risk aversion over time? The means suggest so: the mean over the 47 subjects of average s in Sequence I minus that in Sequence II is 0.17, significant at the 10% level. More specifically, do subjects reveal less risk aversion with the standard display, Text, after they are exposed to graphical displays? The mean difference for Text over the

	Linear		*Logit*	
	FE	*Pooled*	*FE*	*Pooled*
MPLb	−0.040**	−0.040	−0.250**	−0.232**
	(0.018)	(0.093)	(0.116)	(0.093)
MPLc	−0.001	−0.001	−0.013	−0.006
	(0.018)	(0.081)	(0.119)	(0.081)
MPLd	0.088***	0.088	0.573***	0.533***
	(0.018)	(0.119)	(0.126)	(0.119)
MPLe	−0.602***	−0.602***	−3.336***	−3.024***
	(0.023)	(0.188)	(0.165)	(0.188)
TT	0.014	0.014	0.079	0.086
	(0.018)	(0.114)	(0.121)	(0.114)
TF	0.016	0.016	0.108	0.101
	(0.018)	(0.126)	(0.121)	(0.126)
FT	0.014	0.014	0.090	0.090
	(0.018)	(0.131)	(0.122)	(0.131)
FF	−0.010	−0.010	−0.057	−0.060
	(0.018)	(0.121)	(0.121)	(0.121)
trend	0.0001	0.0001	0.001	0.001
	(0.0002)	(0.001)	(0.001)	(0.001)
AEVD	0.088***	0.088*	0.558***	0.509***
	(0.003)	(0.046)	(0.031)	(0.046)
Constant		0.705***		0.851***
		(0.121)		(0.121)
Observations	4,700	4,700	4,700	4,700
R^2	0.171	0.16		
Log Likelihood			−2,137.102	−2,302.467
Akaike Inf. Crit.			4,388.204	4,626.934
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 4: Estimates of equation (2). The FE columns include (but do not report) subject-specific intercepts ("fixed effects"), while the Pooled columns force α_i = Constant. The first two columns ("Linear") estimate (2) directly, while the last two columns replace the dependent variable by its log odds ("Logit"). Errors are clustered at the individual subject level.

Diff	Seq.I	Seq.Ine	Seq.II	p-values
Text vs FF	0.510	0.214	0.148	0.15, 0.48, 0.63
Text vs FT	0.234	0.357	0.170	0.52, 0.28, 0.64
Text vs TF	0.574	0.964***	0.404	0.11, 0.00, 0.20
Text vs TT	0.234	0.074	0.085	0.52, 0.79, 0.77
Text vs all Graphical	0.388	0.405*	0.202	0.18, 0.07 , 0.42
MPLa vs MPLb	-0.19	-	0.02	0.41, - , 0.90
MPLa vs MPLc	1.00***	-	1.06***	0.00, - , 0.00
MPLa vs MPLd	-0.06	-	0.44**	0.73, - , 0.03
MPLa vs MPLe	-2.82***	-	-2.31***	0.00, - , 0.00

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Within subject difference (Text-Graphical and MPLa-MPLi) in performance variable s. Column Seq.Ine, with Nobs = 27, excludes differences involving MPLe. Columns Seq.I and Seq.II each have Nobs = 47. The last column shows p-values associated with the Diff=0 null hypothesis for the preceding three columns.

47 subjects is 0.32, an economically significant decline but not statistically significant due to its large standard deviation of 2.16. A power test indicates that it would take $n = 791$ observations to reject the null hypothesis at significance level 5% in 90% of samples with that mean and standard deviation. Similar conclusions arise in such comparisons when using the parametric performance variable r or using non-parametric signs tests on s.

4 Discussion

We affirm and extend earlier work on MPL-elicited risk preferences, and demonstrate sensitivity to the substance and form of the choice list. Our data support two hypotheses suggested by previous empirical findings in other contexts: H3, that reversing the ordering of MPL rows tends to push elicited preferences towards risk neutrality; and H4, that a convex (concave) transformation of prize amounts will, in violation of EUT, push elicited preferences towards greater (lesser) risk aversion. We also have an intriguing new finding: consistent with hypothesis H1, more graphical presentations of the lotteries tend to make our subjects appear more risk-neutral. That tendency is stronger when explicit statements of prize amounts are excluded, contrary to hypothesis H2. The data are fairly noisy, so (although the signs are quite consistent across specifications) these effects are statistically significant only in some specifications.

Why might graphical presentation provoke a shift towards risk neutrality? One possible explanation is that most subjects really are nearly risk neutral towards small stakes gambles, but some of them are not sufficiently numerate to calculate and compare expected values in text based presentations. Graphical presentation might enable such subjects' visual cortex to process approximate expected values, bypassing difficult calculation, and thus aid expected value maximizing choice. See the on-line Appendix for an elaboration of this conjecture, and cites of relevant literature.

Some readers might be surprised that, if anything, our data suggest ambiguity seeking, not ambiguity aversion as in H2, in the prize dimension. We are more familiar with work on ambiguity in the probability dimension, but suspect that a behavioral theory already exists or could be constructed to accommodate our results. Behavioral theorists may also wish to consider possible explanations of our H3 and H4 findings regarding ordering and nonlinear transformation of prize amounts, especially since they corroborate earlier work in other contexts.

What are the implications for applied researchers? Unfortunately we can not, in good conscience, offer confident recommendations such as "for most precise estimates of individual subjects' coefficients of relative risk aversion, apply maximum likelihood structural estimation to choices gathered in the TF-MPLc treatment." We can say that elicited preferences are more sensitive to details of the MPL procedure than many researchers realize, and that sensitivity does not take the tractable form of a mean-preserving spread or a consistent bias. Thus our advice is only that applied researchers should exercise caution in applying standard risk elicitation procedures.

Whether one suspects that the search for reliable elicitation techniques is futile because typical subjects do not have stable risk preferences (as conjectured in Friedman *et al.* (2014)), or that researchers should continue to search for context-dependent risk preferences, the present study highlights sensitivity to procedural details. In particular, shifts in lottery payoffs and variation in presentation format both modulate elicited preferences.

References

Andersen, Steffen, Harrison, Glenn W, Lau, Morten I, & Rutström, E Elisabet. 2007. Elicitation Using Multiple Price List Formats. *Experimental Economics*.

Benjamin, Daniel J, Brown, Sebastian A, & Shapiro, Jesse M. 2013. Who is "behavioral" Cognitive ability and anomalous preferences. *Journal of the European Economic Association*, **11**(6), 1231–1255.

Bosch-Domènech, Antoni, & Silvestre, Joaquim. 2013. Measuring risk aversion with lists: a new bias. *Theory and decision*, **75**(4), 465–496.

Camerer, Colin F. 1989. An experimental test of several generalized utility theories. *Journal of Risk and uncertainty*, **2**(1), 61–104.

Cokely, Edward T, & Kelley, Colleen M. 2009. Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, **4**(1), 20.

Cox, James C, Smith, Vernon L, & Walker, James M. 1988. Theory and individual behavior of first-price auctions. *Journal of Risk and uncertainty*, **1**(1), 61–99.

Frederick, Shane. 2005. Cognitive reflection and decision making. *The Journal of Economic Perspectives*, **19**(4), 25–42.

Friedman, Daniel, Isaac, R Mark, James, Duncan, & Sunder, Shyam. 2014. *Risky curves: On the empirical failure of expected utility*. Routledge.

Hey, John D, & Orme, Chris. 1994. Investigating generalizations of expected utility theory using experimental data. *Econometrica: Journal of the Econometric Society*, 1291–1326.

Holt, Charles A, & Laury, Susan K. 2002. Risk aversion and incentive effects. *American economic review*, **92**(5), 1644–1655.

James, Duncan. 2007. Stability of risk preference parameter estimates within the Becker-DeGroot-Marschak procedure. *Experimental Economics*, **10**(2), 123–141.

Lévy-Garboua, Louis, Maafi, Hela, Masclet, David, & Terracol, Antoine. 2012. Risk aversion and framing effects. *Experimental Economics*, **15**(1), 128–144.

Park, Joonkoo, & Brannon, Elizabeth M. 2013. Training the approximate number system improves math proficiency. *Psychological science*, 0956797613482944.

Schley, Dan R, & Peters, Ellen. 2014. Assessing "economic value" Symbolic-number mappings predict risky and riskless valuations. *Psychological Science*, 0956797613515485.

Weller, Joshua A, Dieckmann, Nathan F, Tusler, Martin, Mertz, CK, Burns, William J, & Peters, Ellen. 2013. Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, **26**(2), 198–212.

Appendix A Supplementary Analysis

A.1 Results for r

	Dependent variable: r		
	display seq. I	display seq. II	full sample
FF	−0.149	−0.102	−0.125
	(0.118)	(0.122)	(0.085)
FT	−0.009	−0.056	−0.033
	(0.118)	(0.123)	(0.085)
TF	−0.162	−0.083	−0.126
	(0.120)	(0.123)	(0.086)
TT	0.016	0.046	0.030
	(0.120)	(0.121)	(0.085)
MPLb	0.241**	0.152	0.196**
	(0.121)	(0.126)	(0.087)
MPLc	−0.048	−0.083	−0.067
	(0.119)	(0.121)	(0.085)
MPLd	0.035	−0.288**	−0.131
	(0.119)	(0.121)	(0.085)
MPLe	0.228*	0.125	0.174**
	(0.120)	(0.122)	(0.086)
Constant	0.524***	0.508***	0.519***
	(0.111)	(0.114)	(0.080)
Observations	197	199	396
R^2	0.070	0.092	0.069
Adjusted R^2	0.030	0.054	0.050
Residual Std. Error	0.532 (df = 188)	0.546 (df = 190)	0.539 (df = 387)
F Statistic	1.763* (df = 8; 188)	2.410** (df = 8; 190)	3.595*** (df = 8; 387)

Table A.1: Specification as in Table 3 but with $\hat{r}$ as dependent variable

Diff	Seq. I	Seq. Ine	Seq. II	p-values
Text vs FF	0.103	0.055	0.048	0.26, 0.60, 0.65
Text vs FT	-0.012	0.033	0.121	0.88, 0.78, 0.26
Text vs TF	0.168*	0.332***	0.052	0.10, 0.00, 0.60
Text vs TT	-0.047	0.069	-0.024	0.61, 0.62, 0.78
Text vs all Graphical	0.053	0.05	0.077	0.44, 0.44, 0.36
MPLa vs MPLb	-0.19***	-	-0.11*	0.01, - , 0.09
MPLa vs MPLc	0.07**	-	0.05	0.05, - , 0.15
MPLa vs MPLd	0.01	-	0.28***	0.89, - , 0.00
MPLa vs MPLe	-0.21***	-	-0.09	0.00, - , 0.17

*** p<0.01, ** p<0.05, * p<0.1

Table A.2: Analogue of Table 5 for performance variable r, showing within subject difference (Text-Graphical and MPLa - MPLi). Column Seq. Ine excludes differences involving MPLe. The p-values are associated with the Diff=0 null hypothesis for the preceding three columns.

A.2 Other supplementary results

Table A.3 shows how many observations (we have n=47 at best in the current study) it would take to detect at the 5% level in 90% of samples with NOBS = n an effect of the empirical size and variability shown in the columns labeled mean and sd.

In Figure A.1, the two 5x5 blocks on the main diagonal are most relevant. Text has remarkably low correlation with any of the visual presentations for the s variable (.08 - .25), but even the different visual presentations are not very correlated (.11-.48). The MPL schedules have s-correlations in the .41-.69 range, and surprisingly schedule e has correlations well within that range. The r-correlations are a bit higher, but probably because their calculations exclude the NAs (e.g., multiple switches), which eliminates a lot of variation.

The off-diagonal 5x5 blocks correlate the r and s measures for the 5 visual displays with the corresponding measures for the 5 MPL schedules. These are of less intrinsic interest, and in any case are upward biased by including overlapping observations. E.g., an observation of s = 2, say, in the treatment MPLa-TT is counted in both vectors when MPLa is correlated with TT.

A.3 A conjecture on graphics

Math problems can be solved symbolically (e.g., with numbers) or using an approximate number system (ANS) in which quantities are represented as noisy mental magnitudes.

	Seq. I			Seq. II			Seq. Ine		
Treatment	mean	sd	n	mean	sd	n	mean	sd	n
FF	0.1	0.56	620	0.04	.65	3823	0.05	0.52	1877
FT	-0.012	0.54	38336	0.12	0.66	633	0.03	0.6	6746
TF	0.16	0.6	274	0.05	0.62	2917	0.33	0.55	60
TT	-0.04	0.55	2857	-0.02	0.58	11573	-0.06	0.63	1768
Graphical	0.05	0.44	1416	0.07	0.53	1023	0.09	0.47	478
MPLb	-0.19	0.43	104	-0.11	0.4	265			
MPLc	0.07	0.24	212	0.05	0.25	380			
MPLd	0.01	0.74	46276	0.285	0.64	107			
MPLe	-0.21	0.3	45	-0.09	0.39	401			
FF	0.51	2.4	466	0.14	2.13	4325	0.21	1.59	1165
FT	0.23	2.47	2356	0.17	2.51	4583	0.35	1.72	491
TF	0.57	2.46	387	0.40	2.13	586	0.96	1.64	62
TT	0.23	2.48	2372	0.08	2.07	12467	0.07	1.49	8522
Graphical	0.38	1.95	533	0.20	1.7	1493	0.4	1.36	240
MPLb	-0.19	1.59	1462	0.02	1.27	75667			
MPLc	1	1	22	1.06	1.22	28			
MPLd	0.06	1.27	8389	0.44	1.44	219			
MPLe	-2.82	1.3	5	-2.31	1.33	8			

Table A.3: Power tests for $\alpha = .05$ and power $= 90\%$. First block reports results for $\hat{r}$ and second block reports results for s.

Park & Brannon (2013) conclude that ANS is a precursor to symbolic math, and report that training with visual representations of problems significantly improves symbolic math performance relative to symbolic training alone. In particular, subjects were trained to add or subtract large quantities of dots arranged in two arrays, without counting. When presented with follow up symbolic addition and subtraction problems, these subjects provided significantly more accurate answers than a control group with no ANS training. A subsequent treatment provided evidence that ANS training improved subsequent symbolic math performance relative to symbolic math training.

Thus we conjecture that visual representations of risky choice problems may impact preference elicitation if the expected value of a lottery is salient and visualization improves a subject's assessment of expected value. Many studies have demonstrated that subjects with lower numeracy exhibit greater small-stakes risk aversion (Benjamin *et al.* (2013); Cokely & Kelley (2009); Frederick (2005); Weller *et al.* (2013); Schley & Peters (2014)). Since these studies rely on symbolic rather than visual choice representations, it's plausible that reported small-stakes risk aversion has been, to some extent, a consequence of low-numeracy,

	FF	FT	text	TF	TT	MPLa	MPLb	MPLc	MPLd	MPLe
FF	0.821	0.613	0.438	0.425	0.653	0.663	0.694	0.765	0.688	0.634
FT	0.265	0.891	0.362	0.538	0.663	0.685	0.596	0.701	0.764	0.586
text	0.249	0.083	0.868	0.398	0.468	0.568	0.700	0.467	0.707	0.363
TF	0.154	0.229	0.198	0.912	0.428	0.579	0.579	0.727	0.552	0.765
TT	0.240	0.487	0.146	0.114	0.841	0.560	0.671	0.664	0.805	0.667
MPLa	0.576	0.610	0.312	0.467	0.438	1.000	0.540	0.740	0.604	0.569
MPLb	0.613	0.474	0.526	0.253	0.417	0.409	0.896	0.639	0.709	0.586
MPLc	0.494	0.531	0.392	0.621	0.555	0.704	0.481	1.000	0.682	0.690
MPLd	0.425	0.639	0.483	0.415	0.696	0.621	0.553	0.664	1.000	0.616
MPLe	0.394	0.465	0.301	0.582	0.515	0.512	0.428	0.665	0.609	1.000

Figure A.1: Rank Correlation between all treatments: r above the diagonal, s below diagonal, and r and s in the same treatment along the diagonal.

and that elicited responses may change with the use of visual lottery displays.

In investigating this conjecture in future work, it may be helpful to develop treatments intermediate between our TF and FF (which are ambiguous as to prize amounts) and TT and FT (which are explicit). Intermediate treatments (denoted I below) would not show prizes explicitly but would have a labelled height axes so that a subject could visually approximate the absolute prize sizes. Finding a difference between {TT, FT} and {TI, FI} would sharpen the argument for visual cortex effects, while finding a difference between {TI, FI} and {TF, FF} would support the argument for ambiguity effects.

Sequence	trial 1	trial 2	trial 3	trial 4	trial 5
1	Text-A	TF-C	FT-B	TT-E	FF-D
2	Text-B	TF-D	FT-C	TT-A	FF-E
3	Text-C	TF-A	FT-E	TT-D	FF-B
4	Text-D	TF-E	FT-A	TT-B	FF-C
5	Text-E	TF-B	FT-D	TT-C	FF-A
6	TF-A	Text-C	FF-B	FT-E	TT-D
7	TF-B	Text-D	FF-C	FT-A	TT-E
8	TF-C	Text-A	FF-E	FT-D	TT-B
9	TF-D	Text-E	FF-A	FT-B	TT-C
10	TF-E	Text-B	FF-D	FT-C	TT-A
11	FT-A	TT-C	Text-B	FF-E	TF-D
12	FT-B	TT-D	Text-C	FF-A	TF-E
13	FT-C	TT-A	Text-E	FF-D	TF-B
14	FT-D	TT-E	Text-A	FF-B	TF-C
15	FT-E	TT-B	Text-D	FF-C	TF-A
16	TT-A	FF-C	TF-B	Text-E	FT-D
17	TT-B	FF-D	TF-C	Text-A	FT-E
18	TT-C	FF-A	TF-E	Text-D	FT-B
19	TT-D	FF-E	TF-A	Text-B	FT-C
20	TT-E	FF-B	TF-D	Text-C	FT-A
21	FF-A	FT-C	TT-B	TF-E	Text-D
22	FF-B	FT-D	TT-C	TF-A	Text-E
23	FF-C	FT-A	TT-E	TF-D	Text-B
24	FF-D	FT-E	TT-A	TF-B	Text-C
25	FF-E	FT-B	TT-D	TF-C	Text-A

Figure A.2: Treatment sequences selected for experiment. Labels A-E correspond to MPL variants (MPLa-MPLe) and FT, TT, text, FF and TF denote the different graphical display treatments.

Appendix B Instructions

Welcome! You are about to participate in an experiment in the economics of decision-making. If you listen carefully and make good decisions, you could earn a considerable amount of money that will be paid to you in cash at the end of the experiment.

Please remain silent and do not look at other participants' screens. If you have any questions or need any assistance, please raise your hand and we will come to you. Do not attempt to use the computer for any other purpose than what is explicitly required by the experiment. This means you are not allowed to browse the Internet, check email, etc. If you interrupt the experiment by using your smart phone, talking, laughing, etc., you may be asked to leave and may not be paid. We expect and appreciate your cooperation today.

B.1 The Basic Idea

This experiment is composed of ten segments. In each of which you are asked to make a sequence of ten choices/decisions between two lotteries presented to you. Each decision is a paired choice between "Option A" and "Option B." In each segment you will make ten choices and record them by clicking the radio button next to the option you chose, A or B. To determine you payment, we will randomly pick one of the ten segments by rolling a ten-side die, we will then randomly select one of your ten choices within that segment by another roll of the ten-sided die. Final payment will be determined by the outcome of your chosen lottery, which will be determined by a third and final roll of a ten-sided die.

Before you start making your choices, please let me explain in more detail how these choices will affect your earnings. Here is a ten-sided die that will be used to determine payoffs; the faces are numbered from 1 to 10 (the "0" face of the die will serve as 10.) After you have made all of your choices, we will throw this die three times, once to select one of the ten segments to be used, a second time to determine one of the ten choices for that segment, and a third and final time to determine what your payoff is for the option you chose, A (left) or B (right), for the particular decision selected. Even though you will make ten decisions per segment, only one of these will end up affecting your earnings for that segment, but you will not know in advance which decision will be used.

A lottery is defined by probabilities and a set of payoffs associated with these probabilities. As shown in the example in figure B.1, you will be asked to choose between lottery A (on the left), which offers you a payoff of $3 with probability 0.3 (30%) and a payoff of $2.50 with probability 0.7 (70%), and lottery B (on the right), which offers you a payoff of $4.82 with probability 0.3 (30%) and a payoff of $0.10 with probability 0.7 (70%).

Again, in a given segment, you will face a list of ten pairs of lotteries. You must make one choice between the two displayed lotteries for each of the ten pairs in the list. Payment for the experiment will be based on the lotteries you choose. A random segment is picked and a random line will be chosen for payment by a roll of a die and payment will be determined for each subject by the outcome of the chosen lottery (also by a roll of a die) on the randomly selected line.

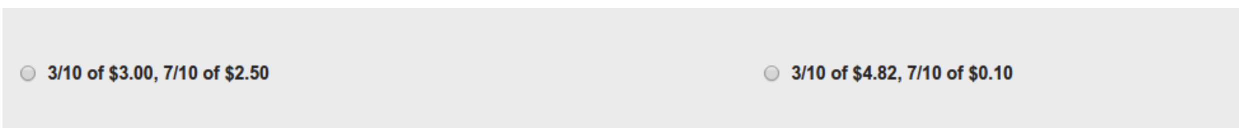

Figure B.1: Text Display

B.2 Display Types

B.2.1 text Only

The following example illustrates one of the displays you will be seeing in this experiment. In the text only display you will be facing a list of different lotteries of the same form as in B.1. As you can see the probabilities and payoffs are displayed in text. In this specific line the lottery on the left has payoffs $3 with probability 0.3 (30%) and a payoff of $2.50 with probability 0.7 (70%), and lottery B (on the right), which offers you a payoff of $4.82 with probability 0.3 (30%) and a payoff of $0.10 with probability 0.7 (70%). You are to choose one of the two lotteries on this line for payment.

B.2.2 Full graphics with payoff and probability in text

In this segment we combine a graphical and text representation of probabilities and payoffs. As demonstrated in figure B.2This display type will feature the probability of a given payoff in the size of the pie slice while the payoff is represented by the height of the pie slice. Therefore, a pie with two different (in color and volume) wedges represents each lottery; the size of a given wedge represents the probability while the height of the wedge represents the payoff. In order to get a full picture of the wedges you can rotate the pie graphs to the left and right by sliding your mouse over the figure (you don't need to click).

It's important to point out that both lotteries (A and B) are displayed on the same scale. This means that if a payoff in lottery A is \$2 and a payoff in lottery B is \$4, then the height of the \$4 payoff in lottery B is twice as high as the height of the \$2 payoff in lottery A. This is also true for payoffs in the same lottery. If lottery A pays $2 or 0.20$, then the height of the \$2 payoff will be ten times as high as the \$0.20 payoff in the same lottery.

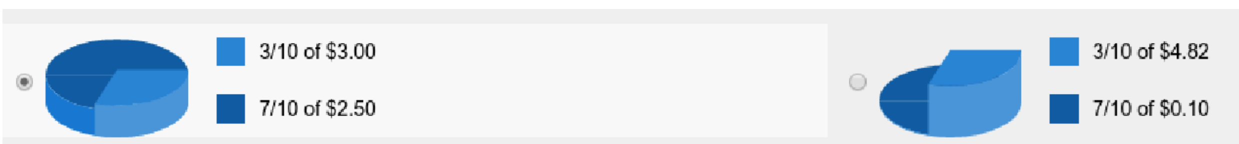

Figure B.2: TT Display: Graphical with text prize and probability

B.2.3 Full graphics with probability in text

This display type is similar to the full graphics with text. As you can see in figure B.3, the only difference is that now payoff is represented only in a graphical form (in the height of the pie slice), while the probabilities are display in text as well as graphically (in the size of the pie slice).

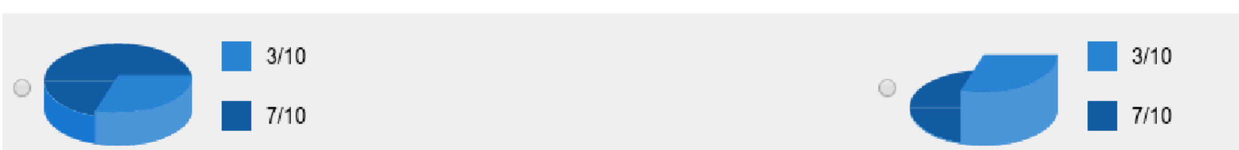

Figure B.3: TF Display: Graphical with text probability only

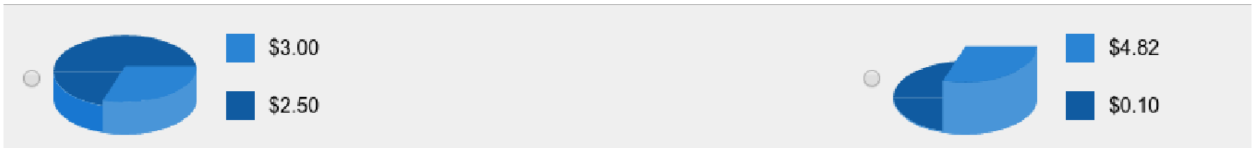

Figure B.4: FT Display: Graphical with text prize only

B.2.4 Full graphics with payoff in text

This display type is similar to the full graphics with text. The only difference is that now probabilities are represented only in a graphical form (in the size of the pie slice), while the payoffs are display in text as well as graphically (in the height of the pie slice).

B.2.5 Fully graphical

This segment removes all text from the list and represents the lotteries in a fully graphical manner. The lotteries are again displayed with the probabilities represented by the size of the pie slice and the payoffs represented by the height of the pie slice. There is no text displaying probabilities or payoffs. Each lottery is fully characterized by the height and size of the pie wedges.

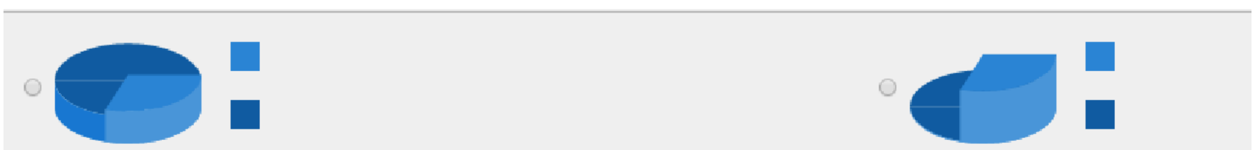

Figure B.5: FF Display: Graphical with no text

Are there any questions? Now you may begin making your choices. Please do not talk with anyone while we are doing this; raise your hand if you have a question.

Appendix C MPL variants

		Lottery A		Lottery B		Calculated		
Row	prob1	A-prize1	A-prize2	B-prize1	B-prize2	EV[A]-EV[B]	$\hat{r}$	$\hat{r}$*
1	0.1	4.00	0.25	2.15	1.75	-1.16	-1.87	-
2	0.2	4.00	0.25	2.15	1.75	-0.83	-1.05	-5.11
3	0.3	4.00	0.25	2.15	1.75	-0.49	-0.54	-2.76
4	0.4	4.00	0.25	2.15	1.75	-0.16	-0.16	-0.84
5	0.5	4.00	0.25	2.15	1.75	0.17	0.16	0.91
6	0.6	4.00	0.25	2.15	1.75	0.51	0.47	2.66
7	0.7	4.00	0.25	2.15	1.75	0.84	0.79	4.57
8	0.8	4.00	0.25	2.15	1.75	1.18	1.16	6.91
9	0.9	4.00	0.25	2.15	1.75	1.51	1.66	10.41
10	1	4.00	0.25	2.15	1.75	1.85	-	-

Table C.1: MPLb, a variation of the original Holt & Laury (2002) Multiple Price List where columns are switched with \$0.15 added to prize. The remaining columns show the difference in expected values of the two lotteries; the approximate solution $\hat{r}$ to the equation EU[A] = EU[B] at that row, where $U(x) = \frac{x^{1-r}}{1-r}$; and $\hat{r}$* is the solution when \$7 is added to all prizes.

		Lottery A		Lottery B		Calculated		
Row	prob1	A-prize1	A-prize2	B-prize1	B-prize2	EV[A]-EV[B]	$\hat{r}$	$\hat{r}$*
1	1	1.95	1.55	3.80	0.05	-1.85	-	-
2	0.9	1.95	1.55	3.80	0.05	-1.51	1.23	10.16
3	0.8	1.95	1.55	3.80	0.05	-1.18	0.86	6.74
4	0.7	1.95	1.55	3.80	0.05	-0.84	0.62	4.47
5	0.6	1.95	1.55	3.80	0.05	-0.51	0.38	2.6
6	0.5	1.95	1.55	3.80	0.05	-0.17	0.13	0.88
7	0.4	1.95	1.55	3.80	0.05	0.16	-0.13	-0.82
8	0.3	1.95	1.55	3.80	0.05	0.49	-0.46	-2.70
9	0.2	1.95	1.55	3.80	0.05	0.83	-0.91	-
10	0.1	1.95	1.55	3.80	0.05	1.16	-1.65	-

Table C.2: MPLc, a variation of the original Holt & Laury (2002) Multiple Price List where rows are reversed with $0.05 subtracted from prize.

		Lottery A		Lottery B		Calculated		
Row	prob1	A-prize1	A-prize2	B-prize1	B-prize2	EV[A]-EV[B]	$\hat{r}$	$\hat{r}$*
1	0.1	1.41	1.26	1.96	0.32	0.79	-	-
2	0.2	1.41	1.26	1.96	0.32	0.64	-2.87	-
3	0.3	1.41	1.26	1.96	0.32	0.49	-1.95	-
4	0.4	1.41	1.26	1.96	0.32	0.35	-1.26	-
5	0.5	1.41	1.26	1.96	0.32	0.20	-0.69	-5.01
6	0.6	1.41	1.26	1.96	0.32	0.05	-0.16	-1.20
7	0.7	1.41	1.26	1.96	0.32	-0.09	0.36	2.81
8	0.8	1.41	1.26	1.96	0.32	-0.24	0.95	7.55
9	0.9	1.41	1.26	1.96	0.32	-0.39	1.76	14.39
10	1	1.41	1.26	1.96	0.32	-0.54	-	-

Table C.3: MPLd, a variation of the original Holt & Laury (2002) Multiple Price List where we take the square root of all prizes.

		Lottery A		Lottery B		Calculated		
Row	prob1	A-prize1	A-prize2	B-prize1	B-prize2	EV[A]-EV[B]	$\hat{r}$	$\hat{r}$*
1	0.1	4	2.56	14.82	0.01	1.21	-0.35	-1.1
2	0.2	4	2.56	14.82	0.01	-0.12	0.02	0.08
3	0.3	4	2.56	14.82	0.01	-1.46	0.26	0.91
4	0.4	4	2.56	14.82	0.01	-2.79	0.42	1.61
5	0.5	4	2.56	14.82	0.01	-4.13	0.57	2.28
6	0.6	4	2.56	14.82	0.01	-5.47	0.70	2.96
7	0.7	4	2.56	14.82	0.01	-6.81	0.83	3.74
8	0.8	4	2.56	14.82	0.01	-8.14	0.98	4.73
9	0.9	4	2.56	14.82	0.01	-9.48	1.18	6.29
10	1	4	2.56	14.82	0.01	-10.82	-	-

Table C.4: MPLe, a variation of the original Holt & Laury (2002) Multiple Price List where we square all prizes.