What Can Multiple Price Lists Really Tell Us about Risk Preferences?

Drichoutis, Andreas and Lusk, Jayson

Agricultural University of Athens, Oklahoma State University

27 March 2012

What Can Multiple Price Lists Really Tell Us about Risk Preferences? *

Andreas C. Drichoutis[†]
Agricultural University of Athens

Jayson L. Lusk[‡]
Oklahoma State University

Abstract

Multiple price lists have emerged as a simple and popular method for eliciting risk preferences. Despite their popularity, a key downside of multiple price lists has not been widely recognized - namely that the approach is unlikely to generate sufficient information to accurately identify different dimensions of risk preferences. The most popular theories of decision making under risk posit that preference for risk are driven by a combination of two factors: the curvature of the utility function and the extent to which probabilities are weighted non-linearly. In this paper, we show that the widely used multiple price list introduced by Holt and Laury (2002) is likely more accurate at eliciting the latter, and we construct a different multiple price list that is likely more accurate at eliciting the former. We show that by combining information from different multiple price lists, greater predictive performance can be achieved.

Keywords: expected utility theory; multiple price list; probability weighting; rank dependent utility; risk.

JEL Classification Numbers: C91, D81.

*The authors would like to thank Peter Wakker for helpful comments on the manuscript and John Hey for helping shorting out estimation related questions.

[†]Department of Agricultural Economics, Agricultural University of Athens, Iera Odos 75, 11855, Athens, Greece, e-mail: adrihout@aua.gr.

[‡]Department of Agricultural Economics, Oklahoma State University, Stillwater, OK 74078, e-mail: jayson.lusk@okstate.edu.

1 Introduction

The abundance of uncertainty in life has prompted a great many investigations into humans' response to risk. The interest in understanding risk preferences has created a latent demand for effective, easy-to-use risk preference elicitation devices. Following a long line of previous research by Becker et al. (1964), Binswanger (1980, 1981), and many others, in 2002 Holt and Laury (H&L) introduced a risk preference elicitation method that has subsequently become a mainstay. In a testament to the general interest in risk preference elicitation and to the specific appeal of the approach introduced by H&L, their work has been cited more than 2,200 times according to Google Scholar and is the second most highly cited paper published by the American Economic Review since 2002 according to ISIs Web of Knowledge. The approach used by H&L has subsequently come to be referred to as a type of multiple price list (MPL) (Andersen et al., 2006; Harrison and Rutström, 2008), an approach thought to have been first used by Miller et al. (1969).[1] The key advantage of the MPL is its ease of use. Respondents make a series of consecutive choices between two outcomes, where the expected value of one outcome increases at a higher rate than the other. The point at which an individual switches from choosing one outcome over the other is often used as a measure of risk aversion.

Despite the fact that MPLs are easy to use and easy for participants to understand, the approach has some weaknesses. Harrison et al. (2005) pointed out that inferences from MPLs can be influenced by order effects (see also Holt and Laury, 2005), and Andersen et al. (2006) discussed the potential for choices in MPLs to be influenced by the ranges of values used. Here, we point to a more fundamental problem with MPLs that seems to have been overlooked by practitioners. In particular, the H&L approach is subject to Wakker and Deneffe's (1996) critique that many risk preference elicitation methods confound estimates of the curvature of the utility function (i.e., the traditional notion of risk preference) with an

[1]The word "multiple" in multiple price list is redundant since the word "list" already implies repetitive choices. Nevertheless, we adopt the phrasing MPL in this paper as it is more commonly used in the literature than other variants such as "choice list."

estimate of the extent to which an individual weights probabilities non-linearly. These are two conceptually different constructs that have different implications for individuals' behavior under risk, and without controlling for one, biased estimates of the other are obtained.

This observation about MPLs is well known to experts in the field of risk preference elicitation, and yet in our experience, it is not well known to newcomers or those outside the field. The purpose of this paper is to further elucidate some of these issues and more widely disseminate this knowledge among the (apparently large) audience of individuals interested in risk preference elicitation. Moreover, while we agree that the use of a single "choice list" or MPL, may not perform well in fully capturing the multidimensional aspects of risk preferences, it must be acknowledged that their popularity results from ease of use. Accordingly, in this paper, we show that different types of MPLs are better able to capture some risk dimensions than others and that by using two (or more) easy to use MPLs, a researcher might achieve a more balanced picture of risk preferences, and thus might attain improved predictive validity.[2]

In what follows, we show that H&L's original MPL is, perhaps ironically, not particularly well suited to measuring the traditional notion of risk preferences - the curvature of the utility function. Rather, it is likely to provide a better approximation to the curvature of the probability weighting function. We then introduce an alternative MPL that has exactly the opposite property. By combining the information gained from both types of MPLs, we show that greater prediction performance can be attained.

2 Effect of Probability Weighting in MPLs

In the base-line MPL used by H&L, individuals were asked to make a series of 10 decisions between two options (see Table 1). In option A, the high payoff amount is fixed at \$2 and the low payoff amount is fixed at \$1.60 across all 10 decision tasks. In option B, the high

[2]If interest rests solely in creating a single index of risk preference without committing to a single theory, there are some relatively simple methods available such as the one shown in exercise 3.6.3 in Wakker (2010).

payoff amount is fixed at \$3.85 and the low payoff amount is fixed at \$0.10. The only thing changing across the 10 decisions are the probabilities assigned to the high and low payoffs. Initially the probability of receiving the high payoff is 0.10 but by the tenth decision task, the probability is 1.0.

Table 1: The H&L Multiple Price List

Lottery A				Lottery B				EVA (€)	EVB (€)	Difference (€)	Open CRRA interval if subject switches to Lottery B (assumes EUT)	
p	€	p	€	p	€	p	€					
0.1	2.00	0.9	1.60	0.1	3.85	0.9	0.10	1.640	0.475	1.17	$-\infty$	-1.71
0.2	2.00	0.8	1.60	0.2	3.85	0.8	0.10	1.680	0.850	0.83	-1.71	-0.95
0.3	2.00	0.7	1.60	0.3	3.85	0.7	0.10	1.720	1.225	0.50	-0.95	-0.49
0.4	2.00	0.6	1.60	0.4	3.85	0.6	0.10	1.760	1.600	0.16	-0.49	-0.15
0.5	2.00	0.5	1.60	0.5	3.85	0.5	0.10	1.800	1.975	-0.18	-0.15	0.14
0.6	2.00	0.4	1.60	0.6	3.85	0.4	0.10	1.840	2.350	-0.51	0.14	0.41
0.7	2.00	0.3	1.60	0.7	3.85	0.3	0.10	1.880	2.725	-0.85	0.41	0.68
0.8	2.00	0.2	1.60	0.8	3.85	0.2	0.10	1.920	3.100	-1.18	0.68	0.97
0.9	2.00	0.1	1.60	0.9	3.85	0.1	0.10	1.960	3.475	-1.52	0.97	1.37
1	2.00	0	1.60	1	3.85	0	0.10	2.000	3.850	-1.85	1.37	$+\infty$

Note: Last four columns showing expected values and implied CRRA intervals were not shown to subjects.

The expected value of lottery A exceeds the expected value of lottery B for the first four decision tasks. Thus, someone who prefers lottery A for the first four decision tasks and then switches and prefers lottery B for the remainder is often said to have near-risk neutral preferences. Analysts often use the number of "safe choices" (the number of times option A was chosen) or the A-B switching point to describe risk preferences and to infer the shape of an assumed utility function (Bellemare and Shearer, 2010; Bruner et al., 2008; Eckel and Wilson, 2004; Glockner and Hochman, 2011; Lusk and Coble, 2005).

Perhaps the first thing that should be noted about the original H&L MPL is that it entails choices made over only four dollar amounts (0.10, 1.60, 2.00 and 3.85). Because a utility function is unique only up to an affine transformation, one must fix two of these points and can only identify the relative difference implied by the other two. Stated differently, the original H&L MPL reveals little information about the curvature of the utility function.[3] By contrast, the H&L MPLs entails choices over 11 different probability amounts (from 0 to 1 in increments of 0.1). Thus, the approach contains much more information about the potential shape of the probability weighting function over the entire probability domain.

To more formally address these issues, assume peoples preferences are represented by rank-dependent utility theory introduced by Quiggin (1982) and incorporated into cumulative prospect theory by Tversky and Kahneman (1992). Applying the theory to the H&L MPL, the rank-dependent utility of option A is $RDU_A = w(p)U(2) + (1 - w(p))U(1.6)$ and the rank-dependent utility of option B is $RDU_B = w(p)U(3.85) + (1 - w(p))U(0.1)$, where p is the probability of receiving the higher payoff amount in each option. A person chooses option A over B when $RDU_A > RDU_B$ or when $w(p)U(2) + (1 - w(p))U(1.6) > w(p)U(3.85) + (1 - w(p))U(0.1)$. Re-arranging, one can see that option A is chosen when:

$$\frac{w(p)}{1 - w(p)} < \frac{U(1.6) - U(0.1)}{U(3.85) - U(2)} \tag{1}$$

Equation (1) reveals two important facts. First, the choice between option A and B in the H&L task is driven both by the shape of $w(p)$ and the shape of $U(x)$ - i.e., it does not separately identify only the curvature of the utility function or the coefficient of relative risk aversion as is often presumed. Second, equation (1) shows that, at most, one can identify only two utility differences $U(1.6) - U(0.1)$ and $U(3.85) - U(2)$, which is clearly a small amount of information to be gleaned about the shape of $U(x)$.

To illustrate the first point, note that many experimental studies have estimated the shape of $w(p)$ using functional forms such as $w(p) = p^\gamma/[p^\gamma + (1 - p)^\gamma]^{1/\gamma}$. Estimates of γ typically fall in the range of 0.56 to 0.71 (e.g., see Camerer and Ho, 1994; Tversky and Kahneman, 1992; Wu and Gonzalez, 1996), which implies an S-shaped probability weighting

[3] One can of course utilize several MPLs and scale up the payoffs as H&L did to allow for a wider range of dollar amounts (thus providing more information on the shape of the utility function). However, those researchers interested in adding a quick and simple risk preference elicitation devise to their studies are unlikely to want to add numerous MPLs simply to get an informed shape of the utility function.

function that over-weights low probability events and under-weights high probability events.

Now, consider a simple example where individuals have a linear utility function (i.e., they are risk neutral in the traditional sense), $U(x) = x$. With the traditional H&L task, a risk neutral person with $U(x) = x$ and $\gamma = 1$ would switch from option A to B at the fifth decision task. However, if the person weights probabilities non-linearly, say with a value of $\gamma = 0.6$, then they would instead switch from option A to B at the sixth decision task. Thus, in the original H&L decision task, an individual with $\gamma = 0.6$, will *appear* to have a concave utility function (if one ignores probability weighting) *even* though they have a linear utility function, $U(x) = x$. The problem is further exasperated as γ diverges from one. Of course in reality, people may weight probabilities non-linearly and exhibit diminishing marginal utility of earnings, but the point remains: simply observing the A-B switching point in the H&L decision task is insufficient to identify the shape of $U(x)$ and the shape of $w(p)$. The two are confounded. While it is possible to use data from the H&L technique to estimate these two constructs, $U(x)$ and $w(p)$, *ex post*, we argue that more information is contained about $w(p)$ than $U(x)$ in the original H&L MPL.

In addition to the above arguments that choices in the H&L MPL are likely to provide more information on the shape of $w(p)$ than $U(x)$ relates to the moderate level of payoffs used in many experiments using MPLs. Several authors have argued that the utility function should be linear over relatively low payoff amounts (Selten et al., 1999; Wakker, 2010).[4] If true, this would suggest that the risk averse behavior previously observed in H&L tasks may well relate to $w(p)$ than to $U(x)$. A final piece of evidence suggesting that the original H&L task is more likely to elicit probability weights than utility curvature are the findings that in repeated choice tasks people are more likely to pay attention to the factors changing across the tasks (which in the case of H&L are the probabilities). Because probabilities are

[4] Rabin (2000) argues that, assuming EUT, anything but risk-neutrality over modest stakes implies absurd levels of risk aversion over larger stakes. Cox and Sadiraj (2006) show that the same implications do not follow for the EUT model of income but only for the terminal wealth model. Of course, there have been several quibbles about this issue (Palacios-Huerta and Serrano, 2006; Rubinstein, 2006; Watt, 2002; Wakker, 2010, pp. 242-245).

changing in the original H&L task, people are more likely to pay attention to this dimension of choice (Bleichrodt, 2002).

2.1 A payoff-varying MPL

Given the preceding discussion, one might ask if there is a simple way to use a MPL that yields more information about $U(x)$ and, at least in some special cases, avoids the confound between $w(p)$ and $U(x)$. One can indeed achieve such an outcome by following an approach like the one used by Wakker and Deneffe (1996) in which probabilities are held constant. Using this insight, we modify the H&L task such that probabilities remain constant across the ten decision tasks and instead change the dollar payoffs down the ten tasks. Our approach is similar to that used in prior research such as that by where certainty equivalents are elicited from subjects by using repeated choices with varying payoff amounts (Cohen et al., 1987). [5]

Table 2 shows a payoff-varying MPL. In this MPL, the probabilities of all payouts are held constant at 0.5. We constructed the payoff-varying MPL shown in table 2 so that it matched the original H&L MPL in terms of the coefficient of relative risk aversion (CRRA) implied by a switch between choosing option A and option B under the assumption of expected utility (EU) preferences.[6]

What are the advantages and disadvantages of the payoff-varying MPL compared to the H&L MPL? At the onset, one can see that because the payoff-varying MPL only utilizes one probability level, 0.5, it cannot reveal much about the shape of the probability weighting function. However, the payoff-varying MPL entails choices over 22 dif-

[5]There a few other papers that have constructed tasks that vary the payoff amounts and hold probabilities constant albeit their aim was different than this paper. For example Bruner (2009) asks whether equivalent changes in the expected value of a lottery achieved by either changing the probability of a reward or by changing the reward itself, will be preferred by risk averse agents as predicted by EUT (he finds that they do). More recently, Bosch-Domènech and Silvestre (2013) compare a standard H&L task with a task they adopt from Abdellaoui et al. (2011) (which in turn is similar to the certainty equivalents method of Cohen et al. (1987)) for embedding bias. They find that the H&L task is susceptible to embedding bias while the Abdellaoui et al. (2011) task is not.

[6]For example, if an individual (with EU preferences) switched from choosing option A to option B on the sixth row of the original H&L task, it would imply a CRRA between 0.14 and 0.41. Likewise, in the payoff-varying MPL with constant probabilities, a switch from choosing option A to option B on the sixth row would also imply (assuming EU preferences) a CRRA between 0.14 and 0.41.

Table 2: Payoff-varying MPL with Constant Probabilities

Lottery A				Lottery B				EVA (€)	EVB (€)	Difference (€)	Open CRRA interval if subject switches to Lottery B (assumes EUT)	
p	€	p	€	p	€	p	€					
0.5	1.68	0.5	1.60	0.5	2.01	0.5	1.00	1.640	1.506	0.13	$-\infty$	-1.71
0.5	1.76	0.5	1.60	0.5	2.17	0.5	1.00	1.680	1.583	0.10	-1.71	-0.95
0.5	1.84	0.5	1.60	0.5	2.32	0.5	1.00	1.720	1.658	0.06	-0.95	-0.49
0.5	1.92	0.5	1.60	0.5	2.48	0.5	1.00	1.760	1.738	0.02	-0.49	-0.15
0.5	2.00	0.5	1.60	0.5	2.65	0.5	1.00	1.800	1.827	-0.03	-0.15	0.14
0.5	2.08	0.5	1.60	0.5	2.86	0.5	1.00	1.840	1.932	-0.09	0.14	0.41
0.5	2.16	0.5	1.60	0.5	3.14	0.5	1.00	1.880	2.068	-0.19	0.41	0.68
0.5	2.24	0.5	1.60	0.5	3.54	0.5	1.00	1.920	2.272	-0.35	0.68	0.97
0.5	2.32	0.5	1.60	0.5	4.50	0.5	1.00	1.960	2.748	-0.79	0.97	1.37
0.5	2.40	0.5	1.60	0.5	4.70	0.5	1.00	2.000	2.852	-0.85	1.37	$+\infty$

Note: Last four columns showing expected values and implied CRRA intervals were not shown to subjects.

ferent dollar payouts. To consider these ideas more formally, again assume individuals have rank-dependent preferences and note that option A will be chosen over option B if $w(0.5)U(A^H) + (1 - w(0.5))U(1.6) > w(0.5)U(B^H) + (1 - w(0.5))U(1)$, where A^H and B^H are the higher payoffs for options A and B, respectively (values which changes over the 10 decision tasks), and where $A^H > 1.6$, $B^H > 1$, and $B^H > A^H$. Re-arranging terms, one can see that option A is chosen if:

$$\frac{w(0.5)}{1 - w(0.5)} < \frac{U(1.6) - U(1)}{U(B^H) - U(A^H)} \tag{2}$$

Comparing equation (2) with equation (1), one can see that the original H&L task can utilize 10 points to estimate the function for $w(p)$ but by contrast, the payoff-varying task can only estimate a single point, $w(0.5)$. In contrast, whereas the original H&L task can only estimate two utility differences, the payoff-varying task can estimate 11. Thus, the payoff-varying MPL reveals more information about the shape of $U(x)$ than the original H&L MPL, but the original H&L MPL reveals more information about the shape of $w(p)$ than does the payoff-varying task.[7]

[7]There is one additional feature of the payoff-varying MPL shown in Table 2 that bears mention. Al-

3 Experiment

To investigate some of the issues discussed above, a laboratory experiment was conducted to compare behavior in the original H&L MPL and our payoff-varying MPL. Moreover, the experiment was designed to see which MPL (or whether a combination of the two) could better predict a hold-out sample of choices. The next sub-section describes the subjects, recruiting, and experimental environment. Then, we describe the different treatments used in the study.

3.1 Description of the experiment set-up

A lab experiment was conducted using the z-Tree software (Fischbacher, 2007). Subjects consisted of undergraduate students at the University of Ioannina, Greece and were recruited using the ORSEE recruiting system (Greiner, 2004). During the recruitment, subjects were told that they would be given the chance to make more money during the experiment.[8] Harrison et al. (2009) have shown that stochastic and non-stochastic fees can significantly affect self-selection of subjects with respect to risk attitudes.

Subjects participated in sessions of group sizes that varied from 9 to 11 subjects per session (all but two sessions involved groups of 10 subjects). In total, 100 subjects participated in 10 sessions that were conducted between December 2011 and January 2012. Each session

though it does not totally do away with the aforementioned confound between $w(p)$ and $U(x)$ when we assume rank-dependent preferences, the confound completely disappears if people weight probabilities as in original prospect theory. In this case, in the payoff-varying MPL people will choose option A when $w(0.5)U(A^H) + w(0.5)U(1.6) > w(0.5)U(B^H) + w(0.5)U(1)$. One can divide both sides of this inequality by $w(0.5)$ to see that option A will be chosen when $U(A^H) + U(1.6) > U(B^H) + U(1)$, or rewriting: $1 < \frac{U(1.6)-U(1)}{U(B^H)-U(A^H)}$. This last inequality does not contain the term w(0.5), thus, the choice of option A over B cannot be explained by probability weighting. Stated differently, *even* if an individual weights probabilities non-linearly in the fashion given by original prospect theory, only the shape of $U(x)$ will dictate their choices in the payoff-varying MPL shown in Table 2. This condition could only be obtained because of our choice of the probability value 0.5. For any other probability value the weighting function does not drop out and the confound remains. Thus, in the original H&L task (which uses probabilities from 0 to 1), the confound between $w(p)$ and $U(x)$ remains even if preferences are given by original prospect theory. Note that since most empirical estimates suggest that $w(0.3) \approx 0.3$, it is possible to also use this empirical relation to create a MPL that avoids probability weighting.

[8]Subjects were told that "In addition to a fixed fee of 10, you will have a chance of receiving additional money up to 25. This will depend on the decisions you make during the experiment."

lasted about 45 minutes and subjects were paid a 10€participation fee. Subjects were given a power point presentation explaining the risk preferences tasks as well as printed copies of instructions. They were also initially given a five-choice training task to familiarize them with the choice screens that would appear in the real task. Subjects were told that choices in the training phase would not count toward their earnings and that this phase was purely hypothetical.

Full anonymity was ensured by asking subjects to choose a unique three-digit code from a jar. The code was then entered at an input stage once the computerized experiment started. The experimenter only knew correspondence between digit codes and profits. Profits and participation fees were put in sealed envelopes (the digit code was written on the outside) and were exchanged with printed digit codes at the end of the experiment. No names were asked at any point of the experiment. Subjects were told that their decisions were independent from other subjects, and that they could finish the experiment at their own convenience. Average total payouts including lottery earnings were 15.2€ (S.D.=4.56).

3.2 Risk preference elicitation

Our experiment entailed a 2x2 within-subject design, where each subject completed two different multiple price lists (MPL) at two payout (low vs. high) amounts. As shown in Table 3, the baseline (or control) involved the original H&L task at their low payoff amounts (a task we refer to as H&L1).

Table 3: Treatments in experiment

	Payout	
Multiple Price List	**low(x1)**	**high (x5)**
H&L	H&L1	H&L5
payoff-varying MPL with constant probabilities	pvMPL1	pvMPL5
Hold-out task	H1	H5

The baseline H&L MPL presented subjects with a choice between two lotteries, A or B. For each lottery choice shown in Table 1, a subject chose A, B or could state indifference between A and B. The last choice shown in Table 1 is a simple test of whether subjects understood the instructions correctly.[9] The second treatment (H&L5) is identical to the first (H&L1) except that all payouts are scaled up by a magnitude of five.

In addition to the choices in treatments H&L1 and H&L5, subjects also completed the payoff-varying MPL shown in Table 2 (pvMPL1) and another set of choices identical to the ones shown in Table 2 except that all payoffs were scaled up by a magnitude of five (pvMPL5).

Instead of providing a table of choices arrayed in an ordered manner all appearing on the same page as in H&L, each choice was presented separately showing probabilities and prizes (as in Andersen et al., 2011b). Subjects could move back and forth between screens in a given table but not between tables. Once all ten choices in a table were made, the table was effectively inaccessible. The order of appearance of the treatments for each subject was completely randomized to avoid order effects (Harrison et al., 2005). An example of one of the decision tasks is shown in Figure 1.

One of the implicit arguments made thus far is that the original H&L task can better estimate the probability weighting function and the payoff-varying MPL can better estimate the utility function. As such, a combination of the insights attained by the two approaches might result in a better overall model. To determine whether this combination is indeed "better" than either used alone, we used out-of-sample prediction as our measure of performance. Thus, as shown in Table 3, the study also included two hold-out tasks which we use as the basis of measuring prediction performance. We constructed these hold-out tasks by creating yet another MPL that modified the original H&L design such that the probability of receiving the higher payout option increased nonlinearly down the list (see Table 4). The MPL is constructed so that it matched the original H&L task in terms of the coefficient of relative risk aversion (CRRA) implied by a switch between choosing option A and option B under the assumption that subjects have prospect-theory preferences where they weigh

[9] 16 out of 100 subjects failed to pass this test concerning comprehension of lotteries and were omitted from our sample.

Figure 1: Example Decision Task

probabilities nonlinearly with $w(p) = p^{0.6}/[p^{0.6} + (1-p)^{0.6}]^{1/0.6}$.

Because each subject completed three MPLs (with 10 choices each) at two payouts, they each made 60 binary choices. For each subject, one of the 60 choices was randomly chosen and paid out.

4 Data analysis and results

4.1 Descriptive analysis

Figure 2 illustrates the proportion of subjects choosing option A for the original H&L task and the payoff-varying task for small and large payoff amounts. Note that all four tasks were designed to elicit the same switching point for a given risk aversion coefficient under the assumption of expected utility preferences (an assumption we will show later to be descriptively invalid).

The two H&L tasks and the pvMPL5 tasks, imply significant risk averse behavior as

Table 4: Holdout Task

Lottery A				Lottery B				EVA (€)	EVB (€)	Difference (€)	Open CRRA interval if subject switches to Lottery B (assumes EUT)	
p	€	p	€	p	€	p	€					
0.03	2.00	0.97	1.60	0.03	3.85	0.97	0.10	1.610	0.194	1.42	$-\infty$	-1.71
0.09	2.00	0.91	1.60	0.09	3.85	0.91	0.10	1.636	0.439	1.20	-1.71	-0.95
0.20	2.00	0.80	1.60	0.20	3.85	0.80	0.10	1.678	0.835	0.84	-0.95	-0.49
0.34	2.00	0.66	1.60	0.34	3.85	0.66	0.10	1.735	1.365	0.37	-0.49	-0.15
0.50	2.00	0.50	1.60	0.50	3.85	0.50	0.10	1.800	1.975	-0.17	-0.15	0.14
0.66	2.00	0.34	1.60	0.66	3.85	0.34	0.10	1.865	2.585	-0.72	0.14	0.41
0.80	2.00	0.20	1.60	0.80	3.85	0.20	0.10	1.922	3.116	-1.19	0.41	0.68
0.91	2.00	0.09	1.60	0.91	3.85	0.09	0.10	1.964	3.512	-1.55	0.68	0.97
0.97	2.00	0.03	1.60	0.97	3.85	0.03	0.10	1.990	3.756	-1.77	0.97	1.37
1	2.00	0	1.60	1	3.85	0	0.10	2.000	3.850	-1.85	1.37	$+\infty$

Note: Last four columns showing expected values and implied CRRA intervals were not shown to subjects.

subjects switch, on average, far after task four. However, for the low-payoff payoff-varying task, pvMPL1, where probabilities are held constant, a different picture emerges. Subjects exhibit what appears to be risk loving behavior in the constant-probability task than the risk averse behavior in the conventional H&L task.

One striking difference in pvMPL1 task is the fact that the percent choosing option A remains at about 50% for the first five decision task, and, in fact, slightly increases over this range. One explanation for this trend is that the payoff-varying task generated more multiple switching points than the standard H&L task.[10] If we calculate the number of choices that violate monotonicity, we find that the average subject made 0.21 and 0.11 such violations in the original H&L task at low and high payouts, respectively. By contrast, in our payoff-varying MPL tasks with constant probabilities, the average subject made 0.85 and 0.69 such violations in the low and high payout tasks, respectively. Over the first few choices in the payoff-varying decision task at low payoffs (pvMPL1), the difference in the expected values between lottery options A and B were relatively small, and this might partially explain why

[10]In our experiment, we did not impose monotonicity on choices or provide warnings when monotonicity was violated. Although such a procedure could be implemented, it is unclear if it is superior to simply observing how people behave when unconstrained.

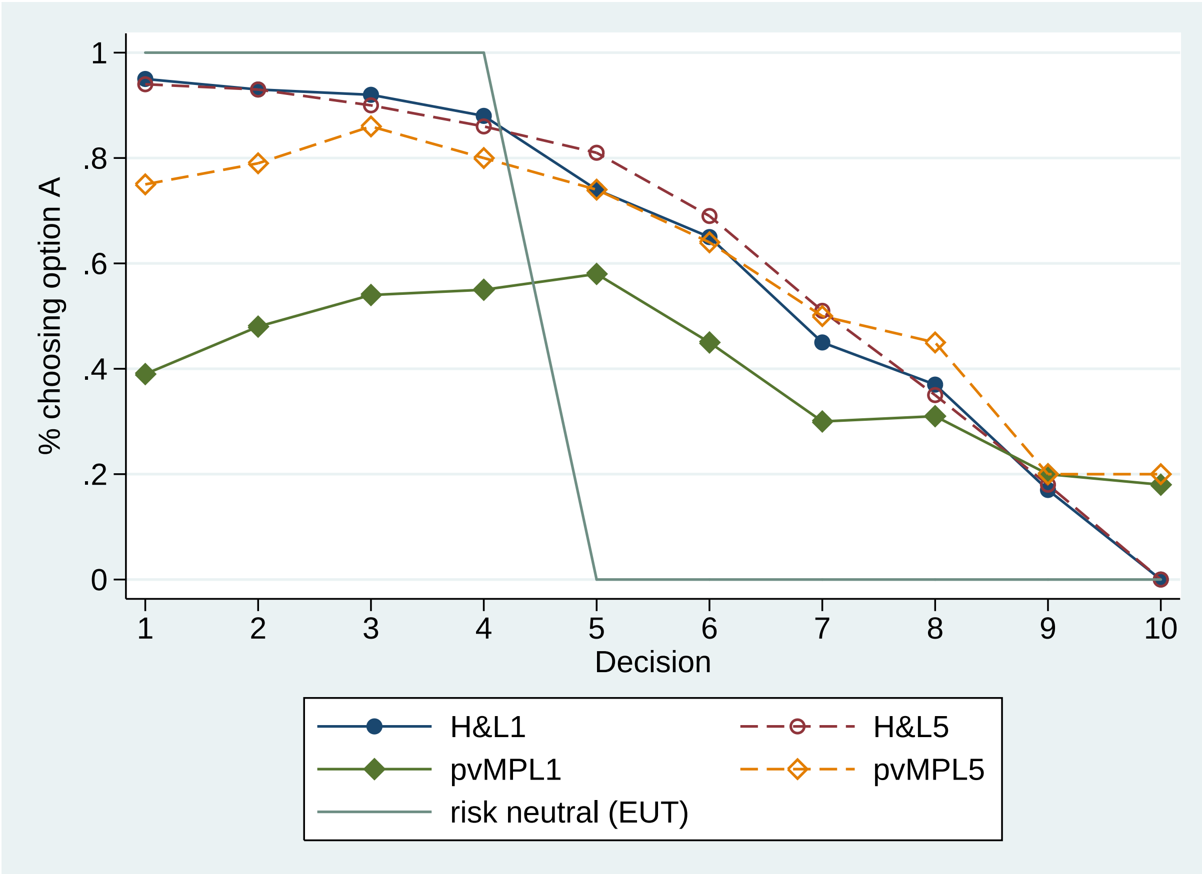

Figure 2: Percentage of respondents choosing option A for each decision task

the task generated more switching behavior. However, it should be noted that such small differences in expected values were *required* to generate the same implied CRRA intervals as the original H&L task given the overall payout magnitudes. Thus, this is not a feature of the payoff-varying task *per se* but rather a feature of constant relative risk aversion and expected utility theory applied to lotteries with payouts of the magnitude considered in the original H&L task but with constant probabilities. Importantly, we have analyzed our data removing individuals that significantly violated monotonicity (i.e., made three or more inconsistent choices), and our econometric estimates (discussed momentarily) are virtually unchanged.

Figure 2 also illustrates the effects of scaling off payoffs. For the traditional H&L task, increasing payoffs had very little effect on the percentage of times option A was chosen.

However, increasing payoffs had a much larger effect on our payoff-varying MPL. The issue of monotonicity does not appear as problematic in the payoff-varying MPL when payoffs are scaled up by a factor of five. This might be because the expected value differences between options A and B (shown in table 2) are also scaled up by a factor of five in this task.

4.2 Econometric modeling approach

To explore the results in terms of the curvature of the utility and probability weighting functions, we utilize the random utility approach also used by Andersen et al. (2008) and use the rank-dependent utility model as the base-line model of analysis. We let the random rank-dependent utility of option k experienced by individual i in choice j be: $V_{ij}^k = Z_{ij}^k + \varepsilon_{ij}^k$ for $k = A, B$ where ε_{ij} is a stochastic error term assumed to be known to the individual but unobservable to the analyst. Z_{ij}^A and Z_{ij}^B are the systematic portions of the utility functions assumed to follow rank-dependent preferences, i.e., $Z_{ij}^A = w(p_j)U(A_j^H) + (1 - w(p_j))U(A_j^L)$, where A_j^H is the high payoff and A_j^L is the low payoff for option A in choice j (and likewise for Z_{ij}^B).

The probability of option A being chosen over option B can be given by $P_{ij}^A = \Phi((Z_{ij}^A - Z_{ij}^B)/\sigma)$, where the difference in the error terms is distributed i.i.d. normal with standard deviation equal to σ. Thus, a log likelihood function can be defined for estimation: $LF = \sum_{i=1}^N \sum_{j=1}^J [y_{ij} ln(P_{ij}^A) + (1 - y_{ij}) ln(1 - P_{ij}^A)]$ where $y_{ij} = 1$ if option A is chosen, $y_{ij} = 0$ if option B is chosen and $y_{ij} = 0.5$ if an individual indicates indifference to A and B.

In the analyses that follow, we consider several specifications for $w(p)$ and $U(x)$. The base-line specifications for the utility function is the constant relative risk aversion specification: $U(x) = \frac{x^{1-r}}{1-r}$, where r is the coefficient of relative risk aversion.[11] In the payoff-varying MPLs we have many more points on the utility function and can also estimate an expo-power

[11] In the original H&L task, we can also estimate a "non-parametric" utility function and instead estimate the two utility differences shown in equation (1): $[U(1.6) - U(0.1)]$ and $[U(3.85) - U(2)]$. In this latter case, however, the standard deviation, σ, is no longer separately identified and must be normalized to one. In the H&L MPL, this formulation is actually observationally equivalent to the CRRA specification with σ freely estimated; both utility specifications give identical maximum likelihood function values and probability weighting estimates. These results are shown in the appendix Table A.1.

utility function (Saha, 1993): $U(x) = (1 - exp(-\alpha x^{1-r}))/\alpha$.

For the probability weighting function in the H&L MPLs, we consider the function used by Tversky and Kahneman (1992) and others: $w(p) = p^{\gamma}/[p^{\gamma}+(1-p)^{\gamma}]^{1/\gamma}$. We also estimated the probability weighting function proposed by Prelec (1998): $w(p) = exp(-\beta(-ln(p)^{\tau}))$. For the payoff-varying MPL, there is only a single probability point and thus we need only estimate a single parameter, θ, representing the weight placed on the 0.5 probability, i.e., $w(0.5) = \theta$.

We estimated each of these competing specifications separately for high and low payoffs (note: likelihood ratio tests reject the hypothesis of equality of parameters across high and low payoffs for each specification). Moreover, we used the AIC and BIC model selection criteria to determine the best fitting model for each dataset.

4.3 Econometric Results

For the traditional H&L MPLs, each of the aforementioned model variations was estimated (see the appendix Table A.1). For both high and low payoffs, the AIC and BIC model selection criteria indicate a preference for the rank-dependent models over the prospect-theory models. Within the rank-dependent models, the Tversky and Kahneman (1992) probability weighting function is preferred to the Prelec (1998) weighting function according to the AIC and BIC.

The preferred model for both the H&L1 and H&L5 treatments is the CRRA model with the Tversky and Kahneman (1992) probability weighting function. For low-payoffs, the estimate of the coefficient of relative risk aversion ($r = 0.004$) was not statistically different from zero, but the estimated parameter on the probability weighting function, $\gamma = 0.501$, was statistically different from one, indicating a rejection of the expected utility model in favor of the rank-dependent model. Similarly, for high-payoffs, the coefficient of relative risk aversion ($r = -0.135$) was not statistically different from zero, but the estimated parameter on the probability weighting function,$\gamma = 0.45$, was statistically different from one. Thus, at least for our subjects, the apparent risk averse behavior shown in Figure 2 is solely a result of probability weighting rather than utility function curvature for the conventional H&L tasks. The implication is that practitioners using the H&L task to infer curvature of the utility function would have arrived at erroneous conclusions had they not also jointly estimated the extent to which people weight probabilities non-linearly.

Figure 3 plots the estimated probability weighting functions for the low-payoff H&L1 MPL. In addition to the two aforementioned functional forms, we also show the results of a "non-parametric" estimation in which a single parameter is estimated for each of the 11 probability points available in the H&L task (with the lowest normalized to zero and the highest normalized to one). Although the "non parametric" form is not preferred to the parametric forms according to AIC and BIC, the results reveal the level of information about probability weights obtainable from the H&L task. In all specifications, the results reveal significant over-weighting of low probability events and under-weighting of high probability events.

Turning to the payoff-varying MPLs, the AIC and BIC indicate that the most preferred models (see full results in appendix Table A.2) are the models assuming constant relative risk aversion with the error variance normalized to one for both the low and high payoff tasks. The results reveal that at low payoffs, the coefficient of relative risk aversion ($r = 0.116$) was not statistically different from zero and the estimated weight applied to probability 0.50 was $w(0.5) = 0.581$, a difference $(0.581 - 0.50 = 0.081)$ which is not statistically different from zero, implying linear probability weighting in the vicinity of $p = 0.5$. Taken together, for low payoffs, the estimates imply near risk-neutral behavior for the payoff-varying MPL.

At high payoffs, however, a different picture emerges. For our payoff-varying MPL which utilizes much more variation over payoff amounts than the H&L task, we find that for high payoffs, a statistically significant estimate for the coefficient of relative risk aversion ($r = 0.233$) emerges. Moreover, we find that the estimated weight applied to probability 0.50 was $w(0.5) = 0.366$. This estimate of probability weighing is very similar to that implied

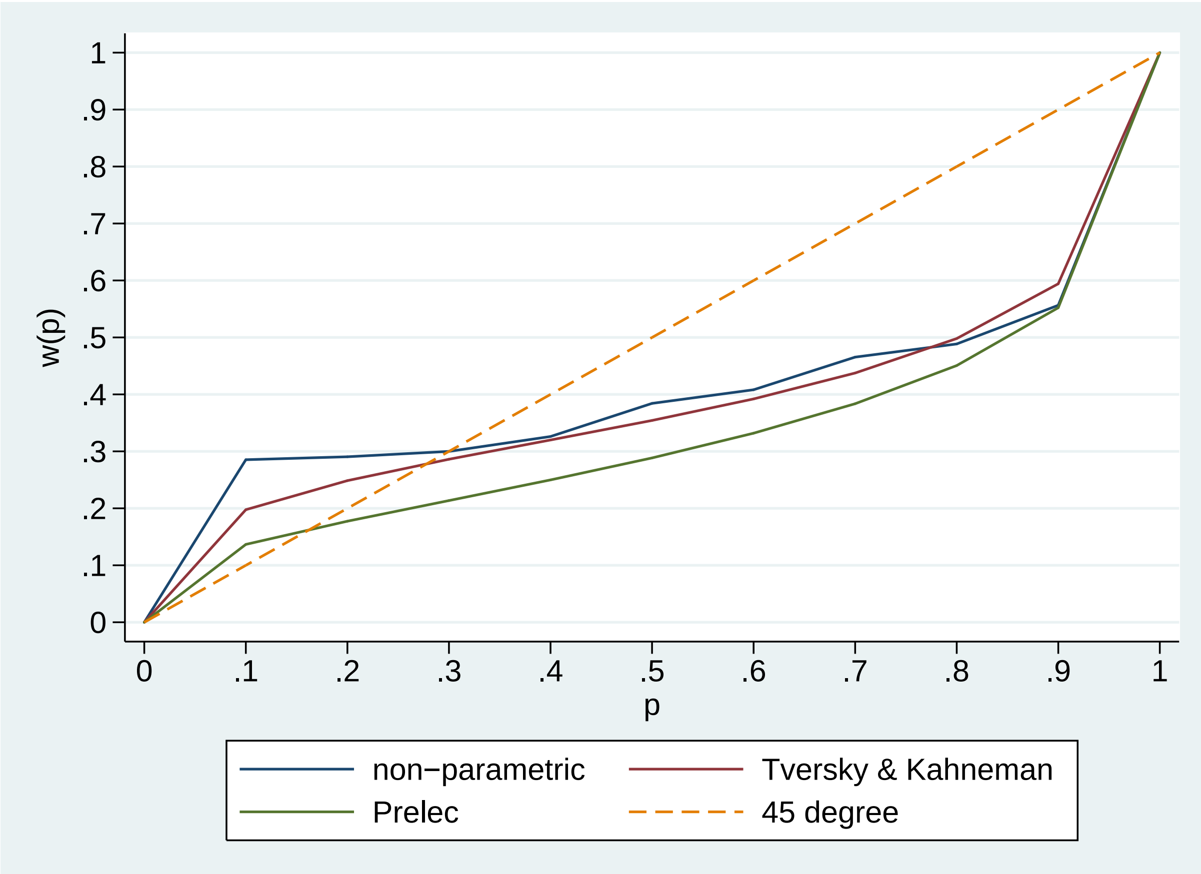

Figure 3: Non-linear probability weighting implied by three different models for the H&L1 MPL

by the high-payoff H&L task (with $\gamma = 0.45$, the H&L5 task implies $w(0.5) = 0.313$).

4.4 Prediction performance

Because of the larger variation in probabilities in the H&L task, we have argued that this task should yield better estimates of the curvature of the probability weighting function. By contrast, because of the larger variation in the dollar amounts in our payoff-varying MPL, we have argued that this task should yield better estimates of the curvature of the utility function. To put these conjectures to the test, we now see how well the aforementioned estimates are able to predict the holdout tasks at low (H1) and high (H2) payoff amounts.

In particular, we compare the predictive performance of three models: i) a model based on the estimate of r and γ from the H&L MPL, ii) a model based on the estimate of r (and for lack of a better choice assuming $\gamma = 1$) from the pvMPL, and finally iii) a composite model in which we use the estimate of r from our payoff-varying MPL and the estimate of γ from the H&L MPL.[12] To judge predictive fit, we use two criteria: 1) the percent of correct predictions and 2) the value of the likelihood function observed at out-of-sample values — the out-of-sample log-likelihood function (OSLLF). The out-of-sample log-likelihood function approach has long been used in the marketing literature for model selection (Erdem, 1996; Roy et al., 1996) and further elucidated in the economics literature (Norwood, Lusk, and Brorsen, 2004; Norwood, Roberts, and Lusk, 2004). The OSLLF has desirable properties in judging the predictive fit of discrete choice models and it is our preferred selection criteria.

Table 5 shows the performance of the three models in predicting the out-of-sample hold-out choices. For low payoffs, the composite model generates the same % correct predictions but has a lower OSLLF than the H&L MPL. Although a paired-test indicates no significant difference in the composite-model and H&L1 OSLLF values, the non-parametric sign-rank test indicates the two are significantly different (p-value< 0.01). The composite model outperforms the pvMPL both in terms of percent of correct predictions and in terms of OSLLF (both the t-test and signed-rank test indicate OSLLFs are significantly different at $p < 0.01$ level).

A similar result is obtained for the high-payoff values. Although all three models generate similar performance in terms of the percentage of correct predictions, the composite model far outperforms the H&L5 and pvMPL5 tasks in isolation according to the OSLLF values (the composite model yields significantly different OSLLF values as compared to the H&L5 and pvMPL5 tasks according to t-tests and signed-rank tests at the $p < 0.01$ level).

Taken together, the results in Table 5 largely confirm our intuition that better predictions

[12] Our composite model takes the estimate of the curvature of the probability weighting from the H&L task and the estimate of the curvature of the utility function from the payoff-varying MPL. An alternative approach is to pool the two data sets and estimate a combined model. When we do this for the low payoff task, we find an estimate of $r = 0.3249$ and $\gamma = 0.6904$, and $\sigma = 0.5883$, all of which are significantly different from zero. However, this model exhibits significantly poorer out of sample predictions with the OSLLF$= -0.5222$ and % correct predictions=76.42% than the composite model discussed in the main text. A similar result holds for the high payoff task.

Table 5: Out-of-sample prediction performance of three competing models

	H&L predictions	pvMPL predictions	Composite model predictions
low payoffs			
OSLLF	-0.4068	-0.4677	-0.4025
% correct	83.57%	76.43%	83.57%
high payoffs			
OSLLF	-0.4696	-1.0565	-0.3876
% correct	83.09%	83.09%	82.60%

can be made by using the H&L task to infer the curvature of the probability weighting function and the payoff-varying MPL to infer the curvature of the utility function.

5 The Random Lottery Incentive Mechanism

In our experiment subjects were paid for one out of 60 tasks. There is a caveat with this payoff mechanism that we haven't discussed so far. One line of criticism with the Random Lottery Incentive Mechanism (RLIM) as first put forward by Holt (1986), is that given the reduction axiom, RLIM is incentive compatible if and only if the Independence Axiom holds. Given that RDU does not include the independence axiom, then RLIM is inappropriate on theoretical grounds. Starmer and Sugden (1991) investigated the empirical validity of Holt's (1986) critique. They designed an experiment to test whether subjects' behavior under the RLIM is consistent with the reduction axiom. They rejected the reduction axiom i.e., that subjects make every decision as if the whole experiment is one big lottery over which they are optimizing.

The results in Starmer and Sugden (1991) are not to say that the RLIM is immune to cross-task contamination since reduction is just the extreme case of such contamination. Cubitt et al. (1998) extended the work of Starmer and Sugden (1991) by generating data from an experiment with more than two choice tasks (contamination may be more of an issue with many choice tasks). They found "...no evidence of cross-task contamination in the random lottery design." Further evidence in favor of the separation hypothesis as well as against cross-task contamination are provided in Hey and Lee (2005b) and Hey and Lee (2005a), respectively.

The RLIM has since been used without second thought by many experimental economists with only few exceptions (see for example the discussion in the Appendix in Conlisk (1989)). Admittedly, either the Independence Axiom holds or it does not. This concern is not specific to our study but to the hundreds of studies that have used the RLIM (and introduced a theoretical confound) when they set out to test non-EUT alternatives. Researchers have different views about this issue. For example, Wakker (2007) argued that this issue has unduly hindered many papers in the review process and that it is counter-productive to re-hash the issue each and every time.

The random selection mechanism has been favored because it is known to avoid wealth and portfolio effects that arise with alternative payoff mechanisms, such as if one decides to pay all decisions sequentially during the experiment or all decisions at the end of the experiment. No incentive-payment scheme is likely to be a "holy grail" and the RLIM has emerged as the mechanism that best compromises all the competing issues. However, criticism of the RLIM seems to have been continually debated, with referees invoking arguments along the lines of Holt (1986) which obviously prompted Hey and Lee (2005b) to explicitly conclude that: "...experimenters can continue to use the random lottery incentive mechanism and that this paper can be used as a defense against referees who argue that the procedure is unsafe".

In essence, researchers that use the RLIM under non-EUT specifications invoke the assumption of the isolation effect i.e., that a subject views each choice in an experiment as independent of other choices in the experiment. If the isolation assumption holds, then RLIM is incentive compatible even under non-EUT specifications. However, if the isolation effect does not hold, then there is no alternative for multi-decision experiments where behavior departs from EUT. The debate then boils down to an empirical question of whether the isolation effect holds or not.

As mentioned above, several studies have defended the validity of RLIM. Nevertheless, the issue has been re-opened recently by one group of researchers (Cox et al., 2014; Harrison and Swarthout, 2012). Cox et al. (2014) tested a one-task (OT) decision environment which is the only alternative payoff mechanism that is valid under any behavioral decision theory. In OT, a subject makes one (and only one) choice and is paid for that one choice. However, this task is practically infeasible with our design. Cox et al. (2014) use five lottery pairs and in five out of their six treatments, subjects make choices between all five lottery pairs. Each of these treatments employs a sample size of roughly 40 subjects. To make things comparable, their OT treatment employs a sample size roughly five times of that employed in the other treatments (5 lottery pairs$\times$40 subjects$\approx$231 subjects). Given that our experiment entails each subject making choices between 60 lottery pairs, the sample size implied by the rule of thumb in Cox et al. (2014) is 6000 subjects! That's far beyond many researchers' resources combined. In addition, our research is implicitly about multiple price lists (MPL), and by definition, MPLs entail several repeated choices between ordered lotteries. Thus, even from a practical standpoint, it is unclear to us how a OT design could successfully critique a MPL, which (rightly or wrongly) has become the most popular method for providing an index of risk preferences.

Furthermore, a OT design precludes within-subjects hypothesis tests, since in such experiments a subject only makes one decision. Another assumption is therefore required pertaining to homogeneous preferences across subjects (or at least large samples with adequate randomization to treatment). Harrison and Swarthout (2012) point out that "...plausible estimates of the degree of heterogeneity in the typical population imply massive sample sizes for reasonable power, well beyond those of most experiments." The implications for experimental practice, in terms of the cost of having much larger sample sizes coupled with the need to conduct high stakes experiments, are profound.

So how can one ease concerns about the use of RLIM under non-EUT? First, in the context of our experiment, it seems implausible to assume *a priori* that subjects are sophisticated enough to consider the whole experiment as being one single question. We can reasonably assume that such overall optimization over 60 tasks could easily exceed subjects' information processing capacities. In the words of Hey and Lee (2005b), if the subjects were choosing the best strategy for the experiment as a whole then they would have to choose between 2^{60} different strategies (ignoring the fact that a third option of indifference was also available), which seems unlikely.

A more plausible pattern of choices would be cross-task contamination in the form of serially-correlated responses that are inconsistent with isolation. To tackle this issue we used the procedures set forth in Hey and Lee (2005a). Hey and Lee (2005a) advance three hypothesis which they call Mark 0, Mark 1 and Mark 2, respectively. Mark 0 advances the isolation hypothesis i.e., subjects view each task independently of other choice tasks. Under Mark 1 the subject considers all past choice tasks but treats the current task as if it is the last one. Under Mark 2 the subject considers the past tasks but treats all forthcoming tasks as if these are going to be similar to the current task. Under EUT the subject should behave the same under the three marks but not under RDU. As in Hey and Lee (2005a) we check to see which of the three hypotheses appears to be most consistent with the data.

We used a maximum score (MS) method of estimation for two different preference functionals: the parametric EUT with a CRRA utility function and the corresponding RDU functional. These are the two functionals that were favored by our data (see Tables A.1 and A.2 in the paper). The MS method involves choosing the parameters of the preference functionals (r for EUT and r, γ for RDU) in such a way that the number of correct predictions (of the subject's behavior) is maximized. Our algorithm iteratively tried all possible values of r in the interval [-7, 7] with a step of 0.001 and all possible values of γ in the interval [0.2, 3] with a step of 0.01 to obtain a maximum score. This maximization was performed for each subject (77 subjects),[13] per mark (Mark 0, Mark 1, Mark 2) and per preference

[13] This number is 7 subjects less than the 84 subjects that were used in our original estimations reported in Tables A.1 and A.2 in the paper. Subjects that expressed at least one indifference when choosing between

functional (EUT and RDU).

To compare between preference functionals we then converted the maximum scores into (corrected) log-likelihoods as in Hey and Lee (2005a). Table 6 below displays average and total corrected log-likelihoods (corrected to account for the number of different parameters between the EUT and RDU functionals).

Table 6: Average and total corrected log-likelihoods

	EUT			RDU		
	Mark 0	Mark 1	Mark 2	Mark 0	Mark 1	Mark 2
Average	-26.2684	-26.2684	-26.2684	-26.1162	-25.6111	-26.0292
Total	-2022.67	-2022.67	-2022.67	-2010.95	-1972.05	-2004.25

First, note that the RDU functional is to be preferred, which is consistent to what we have already found in Tables A.1 and A.2. Although it appears that both Mark 1 and Mark 2 (which are consistent with some form of cross task contamination) are favored when compared with Mark 0 (which is consistent with the isolation hypothesis), a closer look at our results reveals that this average behavior is being driven by 19 subjects for which we can clearly reject the isolation hypothesis. For the rest of the subjects (58 subjects), Mark 0 does equally well as Mark 1 and Mark 2.

We then re-estimated our models without the 19 subjects who showed evidence of violation of the isolation effect and we also removed the 7 subjects that expressed at least one indifference and were excluded from the maximum score estimation. As shown in Table A.3, including or excluding the subjects that do not isolate and/or the subjects with indifferent choices, makes little difference to our results and conclusions. Although the main point of our paper is not specific to the RLIM and incentive compatibility, our conclusions, methods and main point of the paper remain intact.

An additional issue with the RLIM is that presenting the payment scheme as a random lottery may dilute incentives if expected earnings become negligible. Harrison (1994) uses

lotteries were dropped from the maximum score estimation due to the inability of the method of ever predicting an indifference correctly.

this doctrine as the basis of a critique of the random lottery design and points out that in a random lottery experiment the incentives offered by the face values of the options are diluted by the fact that each task has only a small probability of being binding. While this practice is common in the literature, it is rarely evaluated. We are aware of only one study that varied the probability of payment for a discounting task from 10% to 100% and found no significant difference in the responses between these treatments (Andersen et al., 2011b). Several other studies have used a RLIM without second thought about diluted incentives. For example, in Andersen et al. (2013) there were 40 discounting choices and 40 risk attitude choices, and each subject had a 10% chance of being paid for one of each set. Similarly, in Andersen et al. (2011a) there were 40 discounting choices, 40 atemporal risk attitude choices and 40 intertemporal risk attitude choices, and each subject had a 10% chance of being paid for one choice in each set of 40 choices. In Harrison et al. (2009) there was a 10% chance of having 1 out of 40 choices realized. In Cheung (2013), 1 out of 40 risk preference tasks was paid and 1 out of 80 time preference tasks was paid. For half of the time preference tasks, the prizes were given with a 50% probability. Andreoni and Sprenger (2012) administered 87 choice tasks. Each of the first 45 choice tasks had 1/47 chance of being binding, each of 22 choices had 1/(47*22) chance of being binding and the next 20 choices had 1/(47*20) chance of being binding. Thus, our study does not differ significantly with other studies in terms of monetary incentives. Whether paying one task randomly results in a problem of diluted incentives, remains a problem that requires proper evaluation in future studies.

6 Conclusion

Although H&L introduced a useful tool for characterizing risk taking behavior, their approach is limited in being able to identify why a particular behavior under risk was observed. Risk averse behavior could result from curvature of the utility function, curvature of the probability weighting function, or both. The obvious implication is that caution should be taken in directly using a single number like "number of safe choices" from H&Ls risk preference elicitation method to infer curvature of the utility function, the theoretical concept that is often of interest, because risk averse behavior may be driven by probability weighting. In fact, we show that, if anything, the H&L task is probably best suited to measuring the curvature of the probability weighting function.

We constructed a modified version of the H&L task which held probabilities constant at 0.50 and provided much more variation in the payoff amounts. By providing more variation in payoff amounts, we hoped to obtain better estimates of the curvature of the utility function. By and large, that's what our experimental results imply. At both low and high payoff amounts, econometric estimates suggest that behavior is almost totally driven by the curvature of the probability weighting function (the estimated CRRA is not significantly different from zero in either case). Only with our payoff-varying MPL under high payoffs did we observe significant curvature in the utility function.

To test our intuition about the relative merits of the two elicitation approaches, we sought to determine whether a composite model that combined the estimate of the curvature of the utility function from our payoff-varying MPL with the estimate of the curvature of the probability weighting function from the H&L task would exhibit better out-of-sample prediction performance with a hold-out task than either model used in isolation. Our results implied that the composite model did indeed generate lower OSLLF values than the estimates from the conventional H&L task or the MPL used alone.

References

Abdellaoui, M., A. Driouchi, and O. L'Haridon (2011). Risk aversion elicitation: reconciling tractability and bias minimization. *Theory and Decision 71*(1), 63–80.

Andersen, S., G. W. Harrison, M. Lau, and E. E. Rutström (2011a). Multiattribute utility theory, intertemporal utility and correlation aversion. *Center for the Economic Analysis of Risk, Working Paper 2011-04, Robinson College of Business, Georgia State University*.

Andersen, S., G. W. Harrison, M. Lau, and E. E. Rutström (2013). Discounting behavior and the magnitude effect. *Economica 80*(320), 670–697.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2006). Elicitation using multiple price list formats. *Experimental Economics 9*(4), 383–405.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2008). Eliciting risk and time preferences. *Econometrica 76*(3), 583–618.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2011b). Discounting behavior: A reconsideration. *Center for the Economic Analysis of Risk, Working Paper 2011-03, Robinson College of Business, Georgia State University*.

Andreoni, J. and C. Sprenger (2012). Estimating time preferences from convex budgets. *The American Economic Review 102*(7), 3333–3356.

Becker, G. M., M. H. Degroot, and J. Marschak (1964). Measuring utility by a single-response sequential method. *Behavioral Science 9*(3), 226–232.

Bellemare, C. and B. Shearer (2010). Sorting, incentives and risk preferences: Evidence from a field experiment. *Economics Letters 108*(3), 345–348.

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

Binswanger, H. P. (1981). Attitudes toward risk: Theoretical implications of an experiment in rural india. *Economic Journal 91*(364), 867–890.

Bleichrodt, H. (2002). A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Economics 11*(5), 447–456.

Bosch-Domènech, A. and J. Silvestre (2013). Measuring risk aversion with lists: a new bias. *Theory and Decision 75*(4), 465–496.

Bruner, D., M. McKee, and R. Santore (2008). Hand in the cookie jar: An experimental investigation of equity-based compensation and managerial fraud. *Southern Economic Journal 75*(1), 261–278.

Bruner, D. M. (2009). Changing the probability versus changing the reward. *Experimental Economics 12*(4), 367–385.

Camerer, C. F. and T.-H. Ho (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty 8*(2), 167–196.

Cheung, S. L. (2013). On the elicitation of time preference under conditions of risk. *American Economic Review (forthcoming)*.

Cohen, M., J.-Y. Jaffray, and T. Said (1987). Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses. *Organizational Behavior and Human Decision Processes 39*(1), 1–22.

Conlisk, J. (1989). Three variants on the allais example. *The American Economic Review 79*(3), 392–407.

Cox, J., V. Sadiraj, and U. Schmidt (2014). Paradoxes and mechanisms for choice under risk. *Experimental Economics*, 1–36.

Cox, J. C. and V. Sadiraj (2006). Small- and large-stakes risk aversion: Implications of concavity calibration for decision theory. *Games and Economic Behavior 56*(1), 45–60.

Cubitt, R. P., C. Starmer, and R. Sugden (1998). On the validity of the random lottery incentive system. *Experimental Economics 1*(2), 115–131.

Eckel, C. C. and R. K. Wilson (2004). Is trust a risky decision? *Journal of Economic Behavior & Organization 55*(4), 447–465.

Erdem, T. (1996). A dynamic analysis of market structure based on panel data. *Marketing Science 15*(4), 359–378.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Glockner, A. and G. Hochman (2011). The interplay of experience-based affective and probabilistic cues in decision making. *Experimental Psychology 58*(2), 132–141.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer and V. Macho (Eds.), *Forschung Und Wissenschaftliches Rechnen. Gwdg Bericht 63. Ges. Fr Wiss*, pp. 79–93. Datenverarbeitung, Gttingen.

Harrison, G. W. (1994). Expected utility theory and the experimentalists. *Empirical Economics 19*(2), 223–253.

Harrison, G. W., E. Johnson, M. M. McInnes, and E. E. Rutström (2005). Risk aversion and incentive effects: Comment. *The American Economic Review 95*(3), 897–901.

Harrison, G. W., M. I. Lau, and E. E. Rutström (2009). Risk attitudes, randomization to treatment, and self-selection into experiments. *Journal of Economic Behavior & Organization 70*(3), 498–507.

Harrison, G. W. and E. E. Rutström (2008). Risk aversion in the laboratory. In J. C. Cox and G. W. Harrison (Eds.), *Research in Experimental Economics Vol 12: Risk Aversion in Experiments*, Volume 12, pp. 41–196. Bingley, UK: Emerald Group Publishing Limited.

Harrison, G. W. and J. T. Swarthout (2012). The independence axiom and the bipolar behaviorist. *Center for the Economic Analysis of Risk, Working Paper 2012-01, Robinson College of Business, Georgia State University*.

Hey, J. D. and J. Lee (2005a). Do subjects remember the past? *Applied Economics 37*(1), 9–18.

Hey, J. D. and J. Lee (2005b). Do subjects separate (or are they sophisticated)? *Experimental Economics 8*(3), 233–265.

Holt, C. A. (1986). Preference reversals and the independence axiom. *The American Economic Review 76*(3), 508–515.

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

Holt, C. A. and S. K. Laury (2005). Risk aversion and incentive effects: New data without order effects. *The American Economic Review 95*(3), 902–904.

Lusk, J. L. and K. H. Coble (2005). Risk perceptions, risk preference, and acceptance of risky food. *American Journal of Agricultural Economics 87*(2), 393–405.

Miller, L., D. E. Meyer, and J. T. Lanzetta (1969). Choice among equal expected value alternatives: Sequential effects of winning probability level on risk preferences. *Journal of Experimental Psychology 79*(3), 419–423.

Norwood, B. F., J. L. Lusk, and B. W. Brorsen (2004). Model selection for discrete dependent variables: Better statistics for better steaks. *Journal of Agricultural and Resource Economics 29*(3), 404–419.

Norwood, B. F., M. C. Roberts, and J. L. Lusk (2004). Ranking crop yield models using out-of-sample likelihood functions. *American Journal of Agricultural Economics 86*(4), 1032–1043.

Palacios-Huerta, I. and R. Serrano (2006). Rejecting small gambles under expected utility. *Economics Letters 91*(2), 250–259.

Prelec, D. (1998). The probability weighting function. *Econometrica 66*(3), 497–528.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization 3*(4), 323–343.

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica 68*(5), 1281–1292.

Roy, R., P. K. Chintagunta, and S. Haldar (1996). A framework for investigating habits, "the hand of the past," and heterogeneity in dynamic brand choice. *Marketing Science 15*(3), 280–299.

Rubinstein, A. (2006). Dilemmas of an economic theorist. *Econometrica 74*(4), 865–883.

Saha, A. (1993). Expo-power utility: A flexible form for absolute and relative risk aversion. *American Journal of Agricultural Economics 75*(4), 905–913.

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

Starmer, C. and R. Sugden (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation. *The American Economic Review 81*(4), 971–978.

Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty 5*(4), 297–323.

Wakker, P. and D. Deneffe (1996). Eliciting von neumann-morgenstern utilities when probabilities are distorted or unknown. *Management Science 42*(8), 1131–1150.

Wakker, P. P. (2007). Message to referees who want to embark on yet another discussion of the random-lottery incentive system for individual choice. Available at http://people.few.eur.nl/wakker/miscella/debates/randomlinc.htm.

Wakker, P. P. (2010). *Prospect theory for risk and ambiguity*. Cambridge, UK: Cambridge University Press.

Watt, R. (2002). Defending expected utility theory. *The Journal of Economic Perspectives 16*(2), 227–229.

Wu, G. and R. Gonzalez (1996). Curvature of the probability weighting function. *Management Science 42*(12), 1676–1690.

A Appendix

Table A.1: Competing estimates for the H&L1 and H&L5 MPLs

	Rank Dependent Theory				Prospect Theory	
	T&K weight function		Prelec weight function		T&K weight function	Prelec weight function
Parameters	Non-Parametric	CRRA	Non-Parametric	CRRA	Non-Parametric	Non-Parametric
low payoff						
r	—	0.004 (0.183)	—	-0.162 (0.320)	—	—
$U(1.6) - U(1)$	3.312** (0.27)	—	2.696 (1.995)	—	5.159** (1.729)	4.388 (2.705)
$U(3.85) - U(2)$	4.063** (1.286)	—	4.150 (3.444)	—	3.419** (1.359)	3.681 (2.764)
γ	0.501** (0.066)	0.501** (0.066)	—	—	0.479** (0.099)	—
σ	1	0.454** (0.06)	1	0.529 (0.391)	1	—
α	—	—	0.392 (0.337)	0.392 (0.337)	—	1.31** (0.391)
τ	—	—	1.435** (0.583)	1.435** (0.583)	—	0.164 (0.11)
LLF	-349.9	-349.9	-349.15	-349.15	-354	-351.65
AIC	705.8	705.8	706.3	706.3	714	711.3
BIC	720	720	725.2	725.2	728.2	730.2
high payoff						
r	—	-0.135 (0.218)	—	-0.314 (0.324)	—	—
$U(8) - U(5)$	3.447** (0.300)	—	2.077** (0.826)	—	5.792** (2.131)	3.864** (1.582)
$U(19.25) - U(10)$	5.118** (1.846)	—	3.901 (2.507)	—	3.809** (1.653)	3.119* (1.622)
γ	0.450** (0.062)	0.450** (0.062)	—	—	0.445** (0.093)	—
σ	1	2.592** (0.608)	1	5.483* (2.869)	1	—
α	—	—	0.461* (0.279)	0.461* (0.279)	—	1.548** (0.385)
τ	—	—	1.818** (0.487)	1.818** (0.487)	—	0.157** (0.073)
LLF	-347.85	-347.85	-347.1	-347.1	-354.1	-350.85
AIC	701.7	701.7	702.2	702.2	714.2	709.7
BIC	715.9	715.9	721.2	721.2	728.4	728.6

Note: **(*) Statistically significant at the 5%(10%) level. Numbers in parenthesis denote standard errors.

Table A.2: Competing estimates for the pvMPL1 and pvMPL5 tasks

	CRRA	normalized CRRA	expo power
low payoff			
r	-0.610 (1.518)	0.116 (0.180)	0.368 (1.341)
θ	—	—	-0.340 (2.275)
$w(0.5)$	0.517** (0.153)	0.581** (0.072)	0.581** (0.074)
σ	2.376 (4.339)	1	1
LLF	-539.6	-539.7	-539.65
AIC	1085.2	1083.4	1085.3
BIC	1099.4	1092.9	1099.5
high payoff			
r	1.022 (0.949)	0.233** (0.030)	-0.083 (0.171)
θ	—	—	0.048** (0.009)
$w(0.5)$	0.526** (0.192)	0.366** (0.011)	0.434** (0.062)
σ	0.134 (0.337)	1	1
LLF	-477.1	-477.45	-476.9
AIC	960.2	958.9	959.8
BIC	974.4	968.4	974

Note: **(*) Statistically significant at the 5%(10%) level.
Numbers in parenthesis denote standard errors.

Table A.3: Results comparing the full model with the subsamples

	H&L			pvMPL		
low payoff						
	All data	Isolating subset + indifferences	Isolating subset	All data	Isolating subset + indifferences	Isolating subset
r	0.004 (0.183)	0.035 (0.156)	0.054 (0.191)	-0.610 (1.518)	-0.899 (1.480)	-0.797 (1.547)
γ	0.501** (0.066)	0.536** (0.081)	0.534** (0.081)	-	-	-
$w(0.5)$	-	-	-	0.517** (0.153)	0.483** (0.152)	0.476** (0.168)
σ	0.454** (0.060)	0.500** (0.076)	0.504** (0.076)	2.376 (4.339)	2.650 (4.704)	2.289 (4.124)
N	840	650	580	840	650	580
LLF	-349.9	-284.3	-253.4	-539.6	-405	-362.3
AIC	705.8	574.6	512.8	1085.2	816	730.6
BIC	720	588	525.9	1099.4	829.4	743.7
high payoff						
r	-0.135 (0.218)	-0.247 (0.259)	-0.174 (0.267)	1.022 (0.949)	1.691 (0.881)	1.670* (0.737)
γ	0.450** (0.062)	0.443** (0.075)	0.460** (0.082)	-	-	-
$w(0.5)$	-	-	-	0.526** (0.192)	0.673** (0.145)	0.670** (0.121)
σ	2.592** (0.608)	2.957** (0.791)	2.686** (0.727)	0.134 (0.337)	0.019 (0.050)	0.019 (0.042)
N	840	650	580	840	650	580
LLF	-347.85	-276.4	-242.75	-477.1	-365.9	-318.5
AIC	701.7	558.8	491.5	960.2	737.8	643.9
BIC	715.9	572.3	504.6	974.4	751.2	657

Note: **(*) Statistically significant at the 5%(10%) level.
Numbers in parenthesis denote standard errors.
Columns labeled “All data” correspond to results reported in Tables A.1 and A.2 in the paper.