

Volume 35, Issue 3

Measuring risk preferences in field experiments: Proposition of a simplified task

Timm Teubner

Karlsruhe Institute of Technology (KIT), Germany

Marc T. P. Adam

The University of Newcastle, Australia

Claudia Niemeyer

Karlsruhe Institute of Technology (KIT), Germany

Abstract

Individual risk preferences can serve as an effective control variable in order to describe human decisions and behavior. Due to limited participants' attention and time, using standard procedures may be difficult. This paper hence proposes a risk preference elicitation task, aiming to assess individual risk preferences in experiments conducted outside the lab. The test is evaluated against a well-established task by means of two online experiments comprising a total of 490 participants.

We thank Simon Schwendner and Mathias Janke for their help with conducting the experiment. Financial support by the Institute of Information Systems and Marketing (IISM) at the Karlsruhe Institute of Technology (KIT) is gratefully acknowledged.

Citation: Timm Teubner and Marc T. P. Adam and Claudia Niemeyer, (2015) "Measuring risk preferences in field experiments: Proposition of a simplified task", *Economics Bulletin*, Volume 35, Issue 3, pages 1510-1517

Contact: Timm Teubner - tim.teubner@kit.edu, Marc T. P. Adam - marc.adam@newcastle.edu.au, Claudia Niemeyer - claudia.niemeyer@kit.edu.

Submitted: March 18, 2015. **Published:** July 11, 2015.

1. Introduction

Risk preferences are ubiquitous in economic and every-day decision making. In experimental economics, assessing individuals' risk preferences can be useful (i) for delineating the effects of the investigated focus variables from individual risk preferences, and (ii) for investigating possible interaction effects between the focus variables and risk preferences. Typically, human risk preferences are neither consistent across different decision situations (e.g., recreational, health, or financial risks, cf. Weber et al., 2002) nor necessarily stable over time (Andersen et al., 2008; Berg et al., 2005). Moreover, dealing with risk has shown to be dependent on social influences and other context-related factors (Trautmann & Vieider, 2011; Krawczyk & Le Lec, 2010; Gandelman et al., 2012; Adam et al. 2014). Hence, measuring individuals' risk preferences is important to maintain high standards of experimental control.

A number of recent studies demonstrated the importance of context-related factors in economic decision making using field experiments (Cohn et al., 2014; Bellemare & Shearer, 2010; Harrison et al., 2007b). In particular, it has been argued that lab experiments sometimes may create an artificial environment in which participants do not always act naturally but rather feel to be “playing some strange game” (Friedman & Cassar, 2009, p. 26) and—being under observation—also taking into account which type of behavior might be desired by their co-participants or the experimenter (Fisher & Katz, 2000). Field experiments, in contrast, enable the experimenter to investigate behavior for larger sample sizes and in more realistic, on-the-spot situations as they address participants in their natural habitat, i.e., by approaching or observing them in everyday situations such as shopping malls, pedestrian areas, and internet platforms such as Mechanical Turk (Paolacci et al., 2010).

Harrison and List (2004) proposed a taxonomy for field experiments, classifying them as *artefactual* (comparable to lab experiments but with non-standard subject pool), *framed* (including a field context of commodity at stake, rules, information set, etc.), or *natural* (actual task, pursued by subjects, no knowledge about experiment). These different forms each have advantages and disadvantages among each other and in comparison to traditional laboratory experiments. Following Harrison and List (2004, p. 1010), we thus consider field experiments as “methodologically complementary” to standard lab experiments. We argue that the more an experiment represents a natural field experiment, the more important it becomes to “not disturb” subjects or even not to reveal the experimental context at all. Measuring risk preferences in such a situation unobtrusively thus represents a challenge. To address this challenge, we identify the following criteria.

First, the risk elicitation task has to be *conductible in short time* (i.e., within few minutes) as the available time is typically much shorter compared to lab experiments. Second, the task has to be *easy to explain and understand*, because the participants' attention span may be very limited due to various distractions. Third, the task should *discriminate* well between different types and lead participants to state their risk preference in a *valid* way (i.e. not contradictory from a theoretical perspective). This relates to the visual presentation of the task: items should be placed in a structured and logical way. Fourth, the relation to an established procedure, e.g., the Holt and Laury (2002) risk elicitation task (henceforth referred to as HL task) is desirable. Hence, the task should exhibit a *high correlation* with its benchmark and, in order to meet the complexity and visual presentation requirements, possibly be based on a multiple price list (MPL) design.

In this paper, we propose and evaluate a risk elicitation task that meets these criteria in order to assess individual risk preferences in experiments conducted outside the lab. The task builds on the well-established HL task, but differs in two systematic ways. First, we reduced the number of rows in the MPL from 10 to 5. Second, we reduced lottery complexity. The results of an evaluation study with 490 participants indicate that the test can be conducted in short time and exhibits a high correlation with the HL task.

The remainder of this article is organized as follows. In Section 2, we review established methods for assessing individual risk preferences. Based on the HL risk preference elicitation task, we propose a condensed version in Section 3. In Section 4, we outline an experimental design to benchmark this task against its template—and present its results in Section 5. Section 6 concludes.

2. Risk Preference Elicitation

Over the last two decades, several procedures have been established for measuring risk preferences in the context of economic decision making—see Charness et al. (2013) for a comprehensive overview. In the task proposed by Holt & Laury (2002), participants face a series of 10 choices between two lotteries—A or B—each. It works as follows: While lottery A represents a gamble with payoffs of either US\$2.00 or US\$1.60, lottery B entails payoffs of US\$3.85 or US\$0.10. The associated probabilities for the high and low payoff in each row are p and $1-p$, respectively, where p increases in steps of 10% per row. Lottery B thus becomes increasingly attractive. A risk neutral decision maker would choose A in the first 4 rows and then switch to B. The HL task can arguably be regarded as the standard procedure for risk preference elicitation in experimental economics, as it “has been widely implemented in recent laboratory experiments and involves a relatively transparent task” (Harrison et al., 2007a, p. 437). It has found wide adoption in behavioral economics and game theoretic laboratory experiments (e.g., Shafran, 2010; Adam et al., 2014). Its applicability to field experiments is limited, however, since the HL task requires considerable attention and the participants must be willing to carefully read and comprehend rather extensive and complex instructions.

In another class of risk elicitation tasks, participants step-by-step decide whether to claim additional profits at the risk of losing all what has already been accumulated, or to cash out the amount currently at stake. Examples are the Balloon Analogous Risk Task (BART, Lejuez et al., 2002) and the Bomb Risk Elicitation Task (BRET, Crosetto & Filippin, 2013). This task concept appears more intuitive, however, without identifying a participant’s risk preferences as systematically as MPL formats and thus not fully meeting the requirements above. Other formats were proposed, e.g., by Figner et al. (2009, Columbia Card Task), Becker et al. (1964, using willingness to pay), Gneezy and Potters (1997, investing a flexible part of an endowment into a lottery), or Binswanger (1980) and Eckel and Grossman (2008), both using a design in which subjects pick one out of a set of coin toss lotteries).

3. Proposition of a Short Version of the HL Task

The proposed short version of the HL task is derived from its original in two steps. First, the number of rows is reduced from 10 to 5. The condensed version of the task will thus be referred to as HL5, whereas the original version is referred to as HL10. Since most participants

in the HL10 task switch from option A to B in the middle rows, it appears plausible to cut off the rows at the ends of the list. In the original study by Holt and Laury (2002), for instance, 94% of the participants exhibited between 3 and 7 safe choices. Naturally, the task will lose discriminatory power and precision with regard to risk attitude, particularly for extreme preferences.

Second, the lotteries' complexity is reduced. For the riskier lottery B, the low payoff (\$0.10) is replaced by a zero payoff. This enables the lottery to be introduced in a more concise way, i.e., "You receive \$3.85 with a probability of 30%, and nothing, otherwise." Moreover, the less risky lottery A, is replaced by an entirely certain payoff in such a way that its value *retains the expected value differences between alternatives A and B from the original design*. Figure 1 summarizes all payoffs and the associated probabilities for both the HL10 and HL5 task.

Option A	Option B	Option A	Option B
1/10 of \$2.00, 9/10 of \$1.60	1/10 of \$3.85, 9/10 of \$0.10		
2/10 of \$2.00, 9/10 of \$1.60	2/10 of \$3.85, 9/10 of \$0.10		
3/10 of \$2.00, 9/10 of \$1.60	3/10 of \$3.85, 9/10 of \$0.10	\$ 1.65	30% \$ 3.85
4/10 of \$2.00, 9/10 of \$1.60	4/10 of \$3.85, 9/10 of \$0.10	\$ 1.70	40% \$ 3.85
5/10 of \$2.00, 9/10 of \$1.60	5/10 of \$3.85, 9/10 of \$0.10	\$ 1.75	50% \$ 3.85
6/10 of \$2.00, 9/10 of \$1.60	6/10 of \$3.85, 9/10 of \$0.10	\$ 1.80	60% \$ 3.85
7/10 of \$2.00, 9/10 of \$1.60	7/10 of \$3.85, 9/10 of \$0.10	\$ 1.85	70% \$ 3.85
8/10 of \$2.00, 9/10 of \$1.60	8/10 of \$3.85, 9/10 of \$0.10		
9/10 of \$2.00, 9/10 of \$1.60	9/10 of \$3.85, 9/10 of \$0.10		
10/10 of \$2.00, 9/10 of \$1.60	10/10 of \$3.85, 9/10 of \$0.10		

Figure 1: HL10 (cf. Holt and Laury, 2002) and HL5 tasks in comparison.

In order to assess risk taking in the different tasks, the risk parameter a may be used, representing the exponent in utility functions with constant relative risk aversion (CRRA), e.g., $u(x) = x^a$. Figure 2 illustrates the ranges of this parameter for both tasks cover and how they are partitioned.

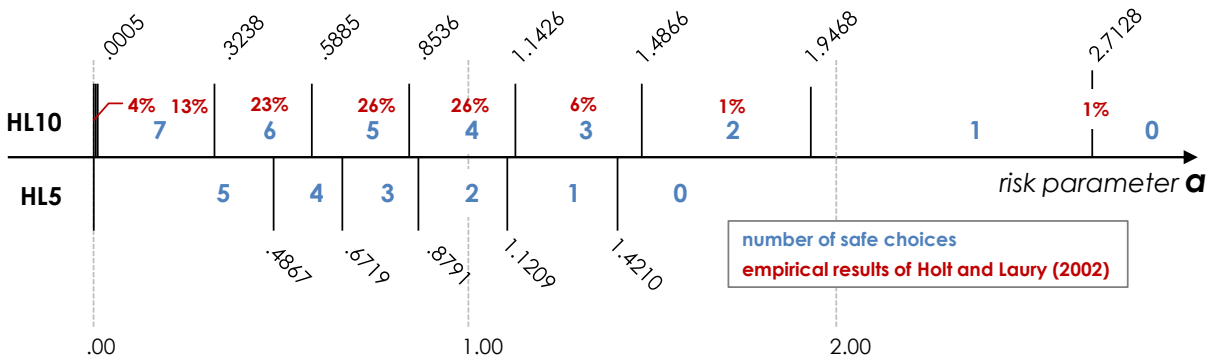


Figure 2: Risk parameter ranges and partitioning for the HL5 and HL10 tasks.

4. Experimental Design

We benchmark the HL5 against the original version of the test. For that matter, an online experiment with a total of 490 participants was conducted (370 male, 120 female). Participants were recruited from a pool of students at (blinded for review), using ORSEE (Greiner, 2004). All participants conducted both the HL10 and the HL5 tasks (within-subject design), whereas the

sequence of these tasks was selected randomly with equal probabilities (cf. Holt and Laury, 2005). It was made clear in the instructions that in either task, participants could earn real money, depending on their decisions and lottery outcomes. At registration, participants provided real names and student IDs. The payoffs could be picked up one week after the experiment, or by individual appointment any time after this date. Since the study was conducted in the euro zone, we use EUR as payoff currency.

There were two experimental conditions which differed in the height of the stakes and the payoff mode: high stakes (396 participants) and low stakes (94 participants). In the low stakes condition, payoffs ranged between €0.00 and €3.85 and every participant was paid accordingly. In the high stakes condition, payoffs were multiplied by a factor of 10. Here, however, the probability of being selected for payoff was 10%. The expected payoff was thus equal in both conditions. Whether a specific participant was selected for payoff or not, was determined and communicated after task completion. Subsequent to the risk preference tasks, participants provided demographic information. The process is depicted schematically in Figure 3.

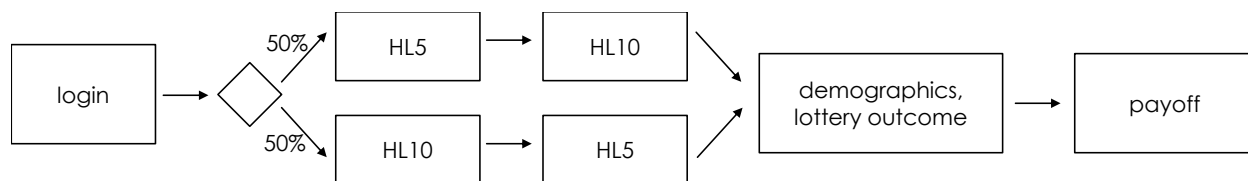


Figure 3: Schematic representation of the experiment process.

5. Results

Our experiment uses a within-subject design, that is, each of the 490 subjects has completed both the HL5 and HL10 task. Using (non-parametric) Spearman correlation, we analyze the association of the number of safe choices from the HL5 and the HL10 tasks, yielding a coefficient of $\rho=.631$ ($p<.001$, $n=490$), indicating a *substantial* or *large* correlation (Kotrlík & Williams, 2003). This main effect is robust against using different subsets of the data: high stakes ($\rho=.626$), low stakes ($\rho=.584$), male participants ($\rho=.620$), female participants ($\rho=.661$), HL10 task first ($\rho=.648$), HL5 task first ($\rho=.618$); all p -values $<.001$. Broadly speaking, the safe choices correlation between 5- and 10-item version of the task may be estimated as above 60%.

We now consider the identified criteria execution time and consistency. The HL5 task was conducted in 48 seconds on average, whereas the HL10 task took 106 seconds (paired-samples t -test: $t(489)=-9.559$, $p<.001$). The percentage of consistent responses, i.e., no switching back from option B to A, was high and comparable for both tasks (HL5: 99.2%, HL10: 99.8%). Regarding the distributions of (0, 1, ..., 5/10) safe choices, we find (.01, .01, .30, .25, .26, .17) for the HL5 and (.00, .00, .01, .04, .17, .13, .24, .24, .12, .03, .02) for the HL10 task.

We now take a slightly different perspective on the data set, transforming it into long-format, yielding 2 observations per subject. To better understand the association of lottery choices among the tasks and the time needed to complete, we conduct two regression analyses (labelled i) and ii) in Table I), accounting for subject random effects and cluster robust standard errors. As our experiment yields risk *intervals* for each subject, see Figure 2, we apply an interval regression (Harrison and Rutström, 2008).

Table I. GLS Regression models for i) time to complete (in seconds), and ii) risk parameter intervals (assuming constant relative risk aversion, $u(x) = x^a$).

	i) Time to complete (s)		ii) Interval reg. for a	
	b	SE	b	SE
Dummy: HL10	57.033 ^{***}	(5.902)	-.277 ^{***}	(.017)
Dummy: male	15.186 [*]	(6.927)	.034	(.037)
Dummy: High Stakes	-1.040	(10.498)	-.220 ^{***}	(.040)
Dummy: HL10 first	4.716	(6.826)	-.017	(.031)
Dummy: 2 nd Task	-22.367 ^{***}	(5.902)	-.041 [*]	(.017)
Constant	46.585 ^{***}	(12.987)	.893 ^{***}	(.052)
<hr/>				
σ_u			.288 ^{***}	(.015)
σ_e			.893 ^{***}	(.010)
<hr/>				
R ² overall	.192		$\rho = .573$	
N (groups)	980 (490)		980 (490)	

Significance: ^{*} $p < .05$, ^{**} $p < .01$, ^{***} $p < .001$

With respect to the risk parameter intervals, we observe more risk averse behavior in the HL10 task and in the high stakes condition. Moreover, participants tend to behave more risk averse in the second task (within-subject, regardless of which task that in fact is), but participants in the different sequence conditions (HL10 first, then HL5 vs. HL5 first, then HL10) do not significantly differ from each other (between-subject). With regard to time needed to complete the task, the regression confirms the differences between the tasks and additionally reveals that male participants tend to take more time. Furthermore, the second task (within-subject) is completed significantly faster than the first one.

6. Discussion

In this paper, we proposed and evaluated a short version of the HL risk preference elicitation task to assess individual risk preferences in experiments conducted outside the lab. Our results suggest that the short task is an appropriate means of assessing individual risk preferences in ad hoc experiment situations. It meets the requirements of being presentable concisely (cf. Figure 1, 80 vs. 15 information items), executable in short time (<1 minute), discriminates, and can be related back to the standard HL procedure reliably (60% correlation).

However, we do observe an offset between the tasks: participants showed riskier behavior in the HL5 task. This may be assigned to the HL5's limited range. In particular on the *safe end*, some participants would have preferred even less risk, whereas on the *risky end*, this does not seem to be the case (only 2% in the 0 and 1-safe-choice bins). The offset might also stem from an explicit preference for gambling per se or "attraction to chance" (Albers et al., 2000; Adam & Kroll, 2012) since the HL5 yields certain payoffs in the less risky option whereas the HL10 yields a lottery for both alternatives. It appears that risk preference elicitation may always have to consider the specific task at hand (Reynaud & Couture, 2012).

Of course, a "short and simple" approach such as ours is not capable of capturing all factors relevant to risk attitudes and preferences. As does the standard HL10 task, we measure a

risk premium, assuming constant relative risk aversion from the outset. Whether or not this assumption holds for certain subjects or not, for instance with regard to stake size or win and loss framing, may be subject to investigation too, calling for more verbose elicitation tasks. The application domain of the task presented here is thus clearly limited to gaining first insights into sign and magnitude of individual risk premiums.

For very similar tasks such as the elicitation of ambiguity preferences, Maffioletti et al. (2009) found that the method used may impact results significantly, as different methods highlight different aspects of the decision problem at hand (prices, probabilities, safe alternatives, etc.). Future research will have to address how other methods of assessing risk preferences (e.g., Lejuez et al., 2002; Crosetto & Filippin, 2013) compare and how they can be related back to standard lab procedures. Moreover, we suggest investigating systematic differences between lab and ad hoc contexts. Our task, in this regard, may help to enable risk preference elicitation outside of the lab.

References

- Adam, M. T. P., Kroll, E. B., & Teubner, T. (2014). A note on coupled lotteries. *Economics Letters*, 124(1), 96-99.
- Adam, M. T., & Kroll, E. B. (2012). Physiological evidence of attraction to chance. *Journal of Neuroscience, Psychology, and Economics*, 5(3), 152-165.
- Albers, W., Pope, R., Selten, R., & Vogt, B. (2000). Experimental evidence for attractions to chance. *German Economic Review*, 1(2), 113-130.
- Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. (2008). Lost in state Space: Are preferences stable? *International Economic Review*, 49(3), 1091-1112.
- Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3), 226-232.
- Bellemare, C., & Shearer, B. (2010). Sorting, incentives and risk preferences: Evidence from a field experiment. *Economics Letters*, 108(3), 345-348.
- Berg, J., Dickhaut, J., & McCabe, K. (2005). Risk preference instability across institutions: A dilemma. *Proceedings of the National Academy of Sciences of the United States of America*, 102(11), 4209-4214.
- Binswanger, Hans P. (1980). Attitudes toward Risk: Experimental Measurement in Rural India. *American Journal of Agricultural Economics*, 62(3), 395-407.
- Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization*, 87(2013), 43-51.
- Cohn, A., Fehr, E., & Maréchal, M. A. (2014). Business Culture and Dishonesty in the Banking Industry. *Nature* 516 (2014), 86-89.
- Crosetto, P., & Filippin, A. (2013). The “bomb” risk elicitation task. *Journal of Risk and Uncertainty*, 47(1), 31-65.
- Eckel, C. C., & Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, 68(1), 1-17.
- Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 709-730.
- Fisher, R. J., & Katz, J. E. (2000). Social-desirability bias and the validity of self-reported values. *Psychology & Marketing*, 17(2), 105-120.
- Friedman, D., & Cassar, A. (2009). *Economics Lab: An Intensive Course in Experimental Economics*. Routledge advances in experimental and computable economics. London: Routledge.
- Gandelman, N., Balsa, A., & González, N. (2012). *Peer effects in risk aversion* (No. 11/2012). Development Research Working Paper Series.
- Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2), 631-645.
- Greiner, B. (2004). An online recruitment system for economic experiments, in Forschung und wissenschaftliches Rechnen, K. Kremer, and V. Macho (eds.), 79–93.
- Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature* 42(4), 1009-1055.

- Harrison, G. W., & Rutstrom, E. E. (2008). Risk Aversion in the Laboratory. In *J. C. Cox and G. W. Harrison (eds.), Risk Aversion in Experiments (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12.*
- Harrison, G. W., Lau, M. I., & Rutström, E. E. (2007b). Estimating risk attitudes in Denmark: A field experiment. *The Scandinavian Journal of Economics*, 109(2), 341-368.
- Harrison, G. W., List, J. A., & Towe, C. (2007a). Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion. *Econometrica*, 75(2), 433-458.
- Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5), 1644-1655.
- Holt, C. A., & Laury, S. K. (2005). Risk aversion and incentive effects: New data without order effects. *American Economic Review*, 92(5), 902-904.
- Kotrlik, J. W. and H. A. Williams (2003). The incorporation of effect size in information technology, learning, and performance research and performance research. *Information Technology, Learning, and Performance Journal*, 21(1), 1-7.
- Krawczyk, M., & Le Lec, F. (2010). 'Give me a chance!' An experiment in social decision under risk. *Experimental Economics*, 13(4), 500-511.
- Lejuez, C. W., Jennifer P. Read, Christopher W. Kahler, Jerry B. Richards, Susan E. Ramsey, Gregory L. Stuart, David R. Strong, and Richard A. Brown (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, 8(2), 75-84.
- Maffioletti, A., Schmidt, U., & Schroder, C. (2009). The effect of elicitation methods on ambiguity aversion: an experimental investigation. *Economics Bulletin*, 29(2), 638-643.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5(5), 411-419.
- Shafraan, A. P. (2010). Interdependent security experiments. *Economics Bulletin* 30(3), 1950-1962.
- Smith, V. L. (1976). Experimental economics: Induced value theory. *The American Economic Review* 66(2), 274-279.
- Trautmann, S. T. and F.M. Vieider (2011). Social Influences on Risk Attitudes: Applications in Economics. In: S. Roeser (ed.), *Handbook of Risk Theory*, Springer, 1-28. (Chapter 29).
- Weber, E. U., Blais, A. R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making* 15(4), 263-290.