

Ortmann, Andreas; Prokosheva, Alexandra; Rydval, Ondřej; Hertwig, Ralph

Working Paper

Valuing a risky prospect less than its worst outcome: uncertainty effect or task ambiguity?

Jena Economic Research Papers, No. 2007,038

Provided in Cooperation with:

Max Planck Institute of Economics

Suggested Citation: Ortmann, Andreas; Prokosheva, Alexandra; Rydval, Ondřej; Hertwig, Ralph (2007) : Valuing a risky prospect less than its worst outcome: uncertainty effect or task ambiguity?, Jena Economic Research Papers, No. 2007,038, Friedrich Schiller University Jena and Max Planck Institute of Economics, Jena

This Version is available at:

<https://hdl.handle.net/10419/25606>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



JENA ECONOMIC RESEARCH PAPERS



2007 – 038

Valuing A Risky Prospect Less Than Its Worst Outcome: Uncertainty Effect or Task Ambiguity?

by

**Andreas Ortmann
Alexandra Prokosheva
Ondřej Rydval
Ralph Hertwig**

www.jenecon.de

ISSN 1864-7057

The JENA ECONOMIC RESEARCH PAPERS is a joint publication of the Friedrich-Schiller-University and the Max Planck Institute of Economics, Jena, Germany. For editorial correspondence please contact m.pasche@wiwi.uni-jena.de.

Impressum:

Friedrich-Schiller-University Jena
Carl-Zeiß-Str. 3
D-07743 Jena
www.uni-jena.de

Max-Planck-Institute of Economics
Kahlaische Str. 10
D-07745 Jena
www.econ.mpg.de

© by the author.

Valuing A Risky Prospect Less Than Its Worst Outcome: Uncertainty Effect or Task Ambiguity?[†]

Andreas Ortmann, Alexandra Prokosheva, Ondřej Rydval and Ralph Hertwig*

July 12, 2007

Abstract

Gneezy, List and Wu [Q. J. Econ. 121 (2006) 1283-1309] document that lotteries are often valued less than the lotteries' worst outcomes. We show how to undo this result.

Keywords: Risky choice, framing, experiments, task ambiguity, subject confusion

JEL classification: C81; C91; C93; D83

[†] This research was supported by a grant from Bank Austria. We thank Pavlo Blavatskyy, John Duffy, Uri Gneezy, Glenn Harrison, John List, Tim Salmon, Lise Vesterlund and participants of the ESA World Meeting 2007 at LUISS in Rome for comments. Any remaining errors are our own.

* Ortmann and Prokosheva – Center for Economic Research and Graduate Education, Charles University, and Economics Institute, Academy of Sciences of the Czech Republic (CERGE-EI). Rydval – Max Planck Institute of Economics (Strategic Interaction Group), Jena, Germany, and CERGE-EI. Hertwig – University of Basel, Switzerland. Please send correspondence to A. Ortmann, CERGE-EI, Politických veznu 7, 11121 Prague 1, Czech Republic, Tel: +420-224-005-117, Fax: +420-224-227-143, E-mail: andreas.ortmann@cerge-ei.cz or aortmann@yahoo.com.

1. Introduction

Gneezy, List and Wu (2006, henceforth GLW) document systematic violations of what they call the *internality axiom* (henceforth IA). Underlying most theories of risky choice, the IA requires that the value of any risky prospect lie between the value of that prospect's best and worst outcomes.¹ Across various valuation "goods" (gift certificates and sports cards, time preferences and work effort), elicitation modes (pricing and choice), and implementation variants (hypothetical and real-stakes, laboratory and field experiments), GLW demonstrate that people value lotteries with intermediate probability mixes *less* than other people value the lotteries' worst outcomes. Our results suggest that the IA violations documented by GLW are vulnerable to subject pool differences and that they can be completely undone by rewording GLW's lottery instructions.

2. Design and implementation

GLW's most prominent demonstration of IA violations is the valuation of gift certificates. We focus on GLW's pricing tasks (see pp. 1286–1289) rather than their binary choice tasks (see pp. 1292–1293) because the former yield a direct valuation of the gift certificates and, *ceteris paribus*, supply higher statistical power for our analysis.

In the seven between-subject treatments of the hypothetical pricing task, GLW's subjects stated their willingness-to-pay for a lottery ticket granting a Barnes & Noble gift certificate worth either \$50 or \$100 with the following probability mixes: (1,0), (0.99,0.01), (0.6,0.4), (0.5,0.5), (0.4,0.6), (0.01,0.99) and (0,1). GLW found that the three lotteries with intermediate probability mixes were valued significantly *less* than the \$50 gift certificate. While the mean (median) willingness-to-pay for the \$50 gift certificate was \$26.1 (\$25), the mean (median) willingness-to-pay for the lotteries with the probability mixes (0.6, 0.4), (0.5,0.5) and (0.4,0.6) was only \$16 (\$6), \$16.1 (\$5) and \$20.8 (\$10), respectively. A similarly striking IA violation was reported for the real-stakes implementation.²

GLW hypothesize that the IA violations are caused by what they call the *uncertainty effect*, attributable to two lottery design characteristics that obstruct the application of the IA.

¹ The IA seemingly derives from deterministic risky choice theories but the empirical results reported below have implications for stochastic theories as well.

² While the mean (median) willingness-to-pay for the \$50 gift certificate was \$38 (\$40), the mean (median) willingness-to-pay for the (0.5,0.5) lottery was only \$28 (\$25).

First, the lotteries feature non-monetary outcomes – e.g., gift certificates – and having to translate them into monetary outcomes may induce higher cognitive demands or perception of uncertainty. Second, the between-subject design does not explicitly prompt subjects to value the lotteries based on valuing their outcomes. GLW argue that the two characteristics jointly trigger a “risk and return” lottery valuation process incompatible with most theories of risky choice:³ Rather than valuing the lottery outcomes, subjects are hypothesized to value the expectation of the outcomes’ face values and subsequently discount it for the risk involved in the lottery. This valuation process can indeed explain the observed IA violations if a high risk premium is imposed on the lotteries with intermediate probability mixes.

Without disputing the potential relevance of GLW’s uncertainty effect argument, we agree with Levitt and List (2007) that striking empirical findings, such as those reported by GLW, deserve careful scrutiny as to whether they generalize across subject pools and whether they could be driven by inadvertent ambiguities of subject instructions. In fact, we conjecture that GLW’s results may have been triggered by their lottery instructions. Keeping other design characteristics intact, we therefore examine whether simplifying GLW’s lottery instructions alleviates the IA violation.

Consider GLW’s instructions for subjects in the hypothetical pricing task:

Imagine that we offer you a lottery ticket that gives you a 50 percent chance at a \$50 gift certificate for Barnes and Noble, and a 50 percent chance at a \$100 gift certificate for Barnes and Noble. Whichever gift certificate you win is good for use within the next two weeks. What is the highest amount of money you would be willing to pay for this lottery ticket?

We conjecture that by making the conceptual divide between the lottery ticket and the gift certificates, GLW’s instructions themselves induced higher cognitive demands or perception of uncertainty or both. Hence our rewording of the instructions, though formally equivalent to GLW’s instructions, avoids the lottery ticket formulation and simplifies the lottery by assigning the probabilities directly to the gift certificates’ face values:

³ GLW document evidence suggesting that neither of the two characteristics can separately induce the IA violation. Indeed, GLW undo the IA violation when replicating the hypothetical pricing task with equal probabilities *within* subjects (see pp. 1290-91).

Imagine that we offer you a gift certificate for Barnes and Noble. With a chance of 50 percent it is a certificate worth \$50, and with a chance of 50 percent it is a certificate worth \$100. Whether the gift certificate is worth \$50 or \$100 is determined by flipping a fair coin. Whichever gift certificate you receive will be good for use within the next two weeks. What is the highest amount of money you would be willing to pay for this gift certificate?

By phrasing the lottery structure directly in terms of the gift certificates, our rewording arguably prompts subjects to value the lottery based on valuing the certificates, thus rendering the application of the IA more transparent. Accordingly, we hypothesize that, *ceteris paribus*, our rewording produces less violation of the IA compared to GLW's original instructions.⁴

To test our hypothesis, we conduct two between-subject treatments with GLW's and our reworded lottery instructions, calling the treatments *Replication* and *Rewording*, respectively. In either treatment, we substitute "Luxor Book Palace at Wenceslas Square" for "Barnes & Noble," and we also use local-currency gift certificates worth 500CZK and 1000CZK with local purchasing power matching GLW's.

To match GLW's sample size, we recruited thirty-two subjects for each treatment – students of the Institute of Economic Studies at the Faculty of Social Sciences, Charles University, Prague.⁵ Following GLW, we used a pen-and-paper design. The instructions were presented to subjects in Czech by the third author. Subjects first completed an anonymous demographic questionnaire and then received the instructions prompting them to write down their willingness-to-pay. Each subject received a show-up fee of 100 CZK (approximately \$10 in purchasing power terms).⁶

⁴ Of course, an even more transparent way to represent the lotteries would be through Venn diagrams, as is common in contemporary lottery choice research. However, our goal was to preserve the basic features of GLW's implementation and examine purely the effect of our rewording.

⁵ We conducted four sessions, two each for *Replication* and *Rewording*. Since another experiment following this one also featured two treatments, we guarded ourselves against potential (though unlikely) spillover effects by conducting the two experiments in a 2x2 factorial design.

⁶ The show-up fee is about ten times higher than the \$1 fee used by GLW but our fee also applied to the follow-up experiment where the subjects earned additional \$15 on average. The average total earnings were more than three times the average hourly wage in the Czech Republic. The sessions lasted about 40 minutes.

While we are well aware of potential behavioral differences between hypothetical and real-stakes implementations (e.g., Hertwig and Ortmann, 2001, Rydval and Ortmann, 2004), we focus on the hypothetical pricing task for which we can provide an alternative book store that is as well-known and has similar product variety as Barnes & Noble.⁷ We also conveniently circumvent the potential confounds introduced in GLW's real pricing task through their use of the Becker-DeGroot-Marschak (1964) elicitation mechanism. We further note that only one out of twenty subjects in GLW's real-stakes implementation was randomly selected and paid based on his/her answer, making it a rather low-stakes implementation (in expected-value terms) resembling a hypothetical implementation.

3. Results

Table 1 displays our willingness-to-pay (WTP) data and below it our statistical tests. We get strikingly higher WTP in *Replication* compared to GLW's results. Relative to the face value of the worst gift certificate – 500CZK for our case and \$50 for GLW – the 95% confidence interval for WTP is (78.46%, 102.17%) in *Replication*, compared to merely (18.83%, 45.66%) in GLW's hypothetical pricing task.

While our implementation is tightly in line with GLW's, including the purchasing power of our gift certificates, we naturally cannot rule out implementation differences stemming from experimenter effects, language effects, or subject pool differences. Among a multitude of potential reasons for the differences in WTP, our subjects might have a greater affection for reading, or fewer of them might have had trouble understanding the lottery implementation in *Replication* compared to GLW's Chicago students.

Indeed, we find that WTP in both *Replication* and *Rewording* is significantly positively affected by subjects' year of study, while WTP is unrelated to their age and gender. Nevertheless, we demonstrate in Table 1 that our results stand regardless of the influence of the demographic characteristics on WTP.

Our key result is the effect of our rewording on WTP. Particularly, WTP is significantly higher in *Rewording* than in *Replication*. The upward shift in WTP is so strong that the 95%

⁷ Our book store, Luxor Book Palace, is one of the largest bookstores in the Czech Republic, located within two-minute walking distance of the experiment's site and the subjects' University (see below for details) and also allowing online purchases. Note that instead of Barnes & Noble, GLW's real-stakes implementation used a smaller local bookstore of unknown quality.

confidence interval for WTP in *Rewording* lies completely above the face value of the 500CZK gift certificate. To the extent that transfers in cash are preferred to those in kind, our subjects would value the 500CZK gift certificate less than 500CZK. Hence our *Rewording* essentially eliminates the possibility of the IA violation documented by GLW.⁸

4. Discussion and conclusion

GLW show that when lotteries involve non-monetary outcomes and people are not explicitly prompted to value them based on valuing the outcomes, lotteries with equal or similar probabilities are often valued less than their worst outcomes. Our results for *Replication* suggest that this IA violation is vulnerable to subject pool differences. Our results for *Rewording* suggest that the IA violation can be completely undone by rewording GLW's lottery instructions in a way that arguably makes the application of the IA more transparent.

We do not dispute that GLW's lottery design may involve high cognitive demands or perception of uncertainty or both, which may trigger a valuation process incompatible with risky choice theories. After all, like GLW, we did not trace the valuation process that our subjects actually used. However, our results suggest that at least part of what GLW call the uncertainty effect might be triggered by their lottery instructions obstructing the application of the IA.

We conjecture that GLW's instructions might have been misinterpreted as describing a compound lottery featuring zero outcome(s) and uncertainty. For instance, some subjects might have misinterpreted GLW's task as featuring two lotteries with unknown probabilities: either a (0.5, 0.5) lottery with outcomes being the \$50 gift certificate and zero, or a (0.5, 0.5) lottery with outcomes being the \$100 gift certificate and zero. Alternatively, some subjects might have misinterpreted GLW's task as featuring a 50% probability of winning one of the two gift certificates with unknown probabilities, and a 50% probability of receiving zero. Hastie and Dawes (2001) refer to such misinterpretations as triggering "ambiguous uncertainty." The anomalously low lottery valuation documented by GLW might therefore

⁸ In fact, the *Replication* confidence interval also marginally contains the face value of the 500CZK certificate, so if subjects value the 500CZK gift certificate less than 500CZK, even our *Replication* would be unlikely to produce the IA violation documented by GLW (at reasonable significance levels).

stem from ambiguity aversion, similarly to a common interpretation of the Ellsberg's paradox (Ellsberg, 1961).

Our results contribute to a bigger picture that has emerged in psychology where verbally described tasks (word problems) similar to GLW's are ubiquitously used in research on probabilistic or logical reasoning. Many psychologists have emphasized that a primary source of ambiguity in word problems stems from the use of ambiguous natural language terms and experimenters' violations of conversational norms (e.g., Adler, 1991; Hilton, 1995; Evans, 2002; Schwarz, Strack, Hilton and Naderer, 1991; but see also Mellers, Hertwig and Kahneman, 2001). To interpret people's behavior as violating rules of logic, probability theory or axioms of rational choice theory, the experimenter needs to assume that the word problem represents nothing more than instantiations of normative rules. However, word problems come with the *semantic* and *pragmatic* ambiguity of natural language. Consequently, unlike experimenters who know the meaning of the key terms in logical and probabilistic word problems such as *and*, *or*, *if-then* and *probability*, experimental subjects have to infer the meaning. Crucially, natural language frequently offers multiple meanings, and to determine the intended one, people appear to use conversational maxims (i.e., Grice's 1975, 1989, theory of conversational reasoning, based on the premise that the communicator attempts to meet certain general standards of communication). In inferring the intended meaning of words or utterances, experimental subjects may arrive at interpretations that diverge from those of experimenters (e.g., Evans, 2002; Hertwig and Gigerenzer, 1999; Politzer and Noveck, 1991). Thus experimenters, in simply equating their and subjects' understanding of the task, may *erroneously* interpret the subjects' behavior as irrational (see Harrison 2005).

References

- Adler, J.E., 1991, An optimist's pessimism: conversation and conjunction, *Posnan Studies in the Philosophy of the Sciences and Humanities* 21, 251-282.
- Ellsberg, D., 1961, Risk, ambiguity, and the savage axioms, *The Quarterly Journal of Economics* 75, 643-669.
- Evans, J.St.B.T., 2002, Logic and human reasoning: an assessment of the deduction paradigm, *Psychological Bulletin* 128, 978-996.
- Gneezy, U., List, J.A. and G. Wu, 2006, The uncertainty effect: when a risky prospect is valued less than its worst outcome, *The Quarterly Journal of Economics* 121, 1283-1309.
- Grice, P.H., 1989, *Studies in the Way of Words*. (Harvard University Press, Cambridge, MA).
- Harrison, G.W., 2005, Field Experiments and Control, in: J. Carpenter, G.W. Harrison and J.A. List, eds., *Field Experiments in Economics*, Vol. 10 (JAI Press, Greenwich, CT) 17-50.
- Hastie, R. and R.D. Dawes, 2001, *Rational Choice in an Uncertain World*. (Sage, Thousand Oaks, CA).
- Hertwig, R. and G. Gigerenzer, 1999, The 'conjunction fallacy' revisited: how intelligent inferences look like reasoning errors, *Journal of Behavioral Decision Making* 12, 275-305.
- Hertwig, R. and A. Ortmann, 2001, Experimental practices in economics: a challenge for psychologists?, *Behavioral and Brain Sciences* 24, 383-403.
- Hilton, D.J., 1995, The social context of reasoning: conversational inference and rational judgment, *Psychological Bulletin* 118, 248-271.
- Levitt, S.D. and J. List, 2007, What do laboratory experiments measuring social preferences reveal about the real world?, *Journal of Economic Perspectives* 21, 153-174.
- Mellers, B., Hertwig, R. and D. Kahneman, 2001, Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration, *Psychological Science* 12, 269-275.
- Politzer, G. and I.A. Noveck, 1991, Are conjunction rule violations the result of conversational rule violations?, *Journal of Psycholinguistic Research* 20, 83-103.

Rydval, O. and A. Ortmann, 2004, How financial incentives and cognitive abilities affect task performance in laboratory settings: an illustration, *Economics Letters* 85, 315-320.

Schwarz, N., Strack, F., Hilton, D. and G. Naderer, 1991, Base rates, representativeness, and the logic of conversation: the contextual relevance of 'irrelevant' information, *Social Cognition* 9, 67-84.

Table 1: Willingness-to-pay (WTP) in *Replication* and *Rewording*

	Session	WTP (in CZK) in ascending order for each treatment (WTP ≥ 500 in bold)															
<i>Replication</i>	1	100	300	300	300	400	400	450	500	500	500	500	500	500	500	550	600
	3	250	250	250	300	300	300	300	400	500	600	600	600	650	730	750	770
<i>Rewording</i>	2	300	300	375	500	500	500	500	500	600	600	650	650	700	700	750	750
	4	250	499	500	500	500	500	500	500	550	600	600	620	650	700	700	720

	<i>Replication</i>	<i>Rewording</i>
Mean WTP	451.56	555.13
Median WTP	500.00	525.00
Standard deviation of WTP	164.44	129.35
95% C.I. for the means without demographic controls	(392.28, 510.85)	(508.49, 601.76)
95% C.I. for the means with significant demographic controls	(388.33, 507.28)	(518.41, 599.36)
95% binomial exact C.I. for the medians	(300.00, 500.00)	(500.00, 650.00)
Wilcoxon rank-sum test	Z = 2.56, p-value = 0.011	
Kolmogorov-Smirnov test	KS = 0.34, p-value = 0.045	
t-test without demographic controls	t = 2.80, p-value = 0.0068	
t-test with significant demographic controls	t = 3.07, p-value = 0.0032	
t-test with all demographic controls	t = 3.13, p-value = 0.0027	

Notes: Raw WTP data (top) are followed by summary statistics and two-sided tests (bottom). Wherever applicable, confidence intervals (C.I.) and tests are based on heteroskedasticity-robust standard errors. The second C.I. pair and the last two *t*-tests are adjusted for the influence of demographics. Session effects, higher-order moments and interactions of demographics, as well as their interactions with the treatments, are individually and jointly insignificant at the 5% significance level.