PRINCE: AN IMPROVED METHOD FOR MEASURING INCENTIVIZED PREFERENCES

Cathleen Johnson, Aurélien Baillon, Han Bleichrodt, Zhihua Li, Dennie van Dolder, Peter P. Wakker[1]

April 2015

This paper introduces the Prince incentive system for measuring preferences. Prince clarifies consequences of decisions and incentive compatibility of experimental choice questions to subjects. It combines the efficiency and precision of matching with the improved clarity and validity of choice questions. It helps distinguish between (a) genuine deviations from classical economic theories (such as the endowment effect) and (b) preference anomalies due to fallible measurements (such as preference reversals). Prince avoids the opaqueness in Becker-DeGroot-Marschak and strategic behavior in adaptive experiments. It helps reducing violations of isolation in the random incentive system. Using Prince we shed new light on willingness to accept and the major components of decision making under uncertainty: utilities, subjective beliefs, and ambiguity attitudes.

JEl-Classification: C91; D81
Keywords: incentive compatibility, random incentive system, BDM, choice list, matching

Cathleen Johnson: Philosophy, Politics, Economics and Law program, Univ of Arizona, PO Box 210108, Tucson, AZ 85721-0108, USA; cathleen@email.arizona.edu
Aurélien Baillon: Erasmus Sch. Econ., Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands; baillon@ese.eur.nl;
Han Bleichrodt: Erasmus Sch. Econ., Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands; bleichrodt@ese.eur.nl
Zhihua Li: Warwick Business Sch.; Univ. of Warwick; Gibbet Hill Road, Coventry CV4 7AL UK; Zhihua.Li@wbs.ac.uk
Dennie van Dolder: Nottingham Sch. of Econ., Univ. of Nottingham, Sir Clive Granger Building, University Park, Nottingham, NG7 2GP, UK; dennie.vandolder@nottingham.ac.uk
Peter P. Wakker: Erasmus Sch. Econ., Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands; wakker@ese.eur.nl

---

[1] Helpful discussions with Chen Li and participants at: Decisions—Theory, Experiments and Applications workshop (HEC Paris), seminar at Institute for Empirical Research in Economics of the University of Zürich, Decision Research at Warwick Forum, School of Economics of University of Nottingham, and Behavioral Economics seminar of London School of Economics are gratefully acknowledged.

1 Introduction and background

Behavioral economics challenges the classical revealed preference paradigm in economics. Many challenges were handled by incorporating irrationalities in decision models, as for instance in Tversky & Kahneman's (1992) prospect theory. Preference reversals (Lichtenstein & Slovic 1971), however, were a more fundamental challenge — the first signs in the literature of a discrepancy between choice and matching[2], calling into question the very existence of preferences. Experimental economists pointed to flaws in the original experiments conducted by behavioral economists: lack of real incentives, insufficient learning opportunities, and sometimes deception, to defend the revealed preference paradigm. However, more carefully conducted experiments supported these challenges (Grether & Plott 1979).

Although some authors blame choice-based procedures (Fischer et al. 1999), preference reversals are more often blamed on matching techniques. As a result, binary choices are now the most common way to measure preferences, recommended by Arrow et al. (1993) and others. Indifferences can then be derived indirectly from switching values in choice lists. But binary choices have drawbacks. They are more cumbersome to administer. They give interval rather than point estimates. They have their own biases.[3]

A pervasive difficulty in economic experiments is that real incentives as implemented in the laboratory are decontextualized and accordingly hard to understand for subjects. The problem is greatest for matching, where the Becker-DeGroot-Marschak (1964) mechanism (BDM) is often criticized for this reason.[4] Both choice and matching experiments commonly involve more than one decision. Paying on every decision leads to income effects. For this reason the random incentive system (RIS; proposed by Savage 1954 p. 29) is now commonly employed.

---

[2] In matching questions, subjects directly indicate indifference values. Attema & Brouwer (2013) provide a review.

[3] These biases have an older history in psychophysics (Gescheider 1997 Ch. 3). From the beginning (Fechner 1860), psychophysicists used binary comparisons besides matching to measure subjective values. The Nobel laureate von Békésy (1947) introduced bisection (the staircase method), to avoid the biases in choice lists (limiting methods).

[4] See, for instance, Bardsley et al. (2010 p. 271 ff.), Camerer (1995 p. 695 ff.), and Seidl (2002 p. 630 ff.). Theoretical problems of the BDM method have been known long ago (Karni & Safra 1987).

In this system, only one of the experimental decisions, randomly selected at the end, is implemented for real. If subjects treat each experimental decision as the only real one (isolation), then incentive compatibility follows. However, subjects may conceive of the set of decisions as a meta-lottery (Holt 1986) where, for instance, some decisions can be used to hedge others, and spillover effects can result.

The Prince incentive method, introduced in this paper, reduces and avoids the aforementioned problems by combining and improving a number of features from existing incentive systems, particularly the random incentive system (RIS), the BDM, and Bardsley's (2000) conditional information system. In brief, where capitalized letters explain the acronym Prince: (1) the choice question implemented for real is randomly selected PRior to the experiment; (2) subjects' answers are framed as INstructions to the experimenter about the real choice to be implemented at the end; (3) the real choice question is provided in a Concrete form, e.g., in a sealed envelope; (4) the Entire choice situation, rather than only one choice option, is described in that envelope. Incentive compatibility can now be crystal clear, not only to homo economicus but also to homo sapiens, and isolation is maximally salient.

Prince combines the tractability and precision of matching with improvements to binary choice's clarity and validity. Binary choice is the current standard for preference measurement (Arrow et al. 1993). Prince aims to reinvent matching. Further, for adaptive experiments (where the sequence of questions is path dependent) subjects cannot answer strategically, and subjects know this. We thus resolve the incentive compatibility problem for adaptive experiments. Wakker and Deneffe's (1986) tradeoff method (TO) for measuring utility under ambiguity now becomes available to experimental economists in properly incentivized form.[5] Finally, not only does our experiment avoid deception, but nondeception is verifiably transparent to subjects.

A difficulty for preference theory is that, unlike with many other empirical domains, there is no gold standard (Tversky & Kahneman 1981, 2nd to last para; Thaler & Sunstein 2008). Hence, there is no current consensus about best methods for measuring true preferences. Therefore, there is no benchmark for a new method to

[5] An advantage of the tradeoff method is that utility is not affected by probability weighting, so that the measurements are valid for virtually all (non)expected utility theories. In particular, collinearities between utility and probabilities are avoided.

beat.[6] Defenses of new theories and measurement methods are primarily based on internal validity arguments, using coherence criteria and general insights from psychology. We intentionally implement stimuli from classic economic decisions throughout, to illustrate the novelty nd validity of Prince. Further details are in §8.

For non-experimentalists, our improved measurements of preferences shed new light on general economic concepts. First, we confirm that utility is closer to linear than traditionally thought (Abdellaoui 2000). Second, we confirm that ambiguity attitudes display likelihood insensitivity besides the well-known aversion (Trautmann and van de Kuilen 2014). Third, Prince helps reveal which known choice anomalies reflect genuine deviations from homo economicus and which are artifacts of measurement problems. Our tests suggest that preference reversals were due to measurement problems. They do not reflect genuine intransitivities. Prince does not, however, remove the endowment effect, which may reflect genuine preference and, hence, a robust discrepancy between homo sapiens and homo economicus that is no mere artifact caused by measurement problems. Here we disagree with Plott & Zeiler (2006). We emphasize that Prince aims to improve measurements of preferences but not preferences themselves.

All stimuli and material of our experiments not presented in the main text or appendix is in the Web Appendix, in particular in part WE there.

2 Prince explained

This section introduces the Prince system. We explain its principles in the first two subsections, and define them formally in §2.3. Discussion is in §§6-7.

2.1 Prince defined

The experiment begins with a *real choice situation* (*RCS*) randomly selected from a set of possible choice situations for each subject. In our experiments the RCS is written on a slip of paper and put in a sealed envelope (following Bardsley, 2000 p. 224). The RCS describes a number of choice options (two in our experiments; for

[6] Holt & Laury (2002) and other introductions of new measurement methods share this aspect with us.

instance, a mug versus a money amount in our first experiment). The subject will receive one of these options and her goal in the experiment is to get the most preferred one. Although the subject does not know her particular RCS, she does receive some information on the potential choice situations (such as average or range of outcomes employed). The partial description about the RCS is constructed so that each choice situation considered during the experiment can possibly be the RCS. The subject need not know the exact probabilities of the latter possibility, and such probabilities need not be uniform, but they should be salient enough to motivate subjects to truthfully answer the experimental questions (Bardsley et al. 2010 p. 220). It is important that the slip of paper in the selected prior envelope describes the entire RCS, with *all* choice options (§7), not just one choice option.

During the experiment, various possible real choice situations are presented to the subject. We explicitly ask subjects to give “instructions” about the real choice to be implemented at the conclusion of the experiment. This real choice is concrete with the envelope in hand. At the end of the experiment, the experimenter opens the prior envelope and uses the instructions provided by the subject to select the desired option. We never ask “what would you do if,” referring to nonconcrete choice situations. A script with statements such as “If you say what you want then you get what you want,” or “If you give wrong instructions, then you don’t get what you want” further emphasize the connection between decision and outcome. Incentive compatibility is crystal clear to the subjects.

2.2 Prince for adaptive experiments: Problems and solutions

In adaptive experiments, stimuli depend on subject responses to previous stimuli. If traditional RISs are used, subjects may benefit (or think they benefit) from answering a question untruthfully so as to improve future stimuli. Such gaming is impossible with Prince, and this is obvious to the subjects, because the RCS has been determined prior to the experiment.

For adaptive experiments, experimenters will not know exactly which choice situations will occur during the experiment. This raises two *overlap problems*.

(1) The *indeterminacy overlap problem* entails the possibility that none of the instructions from the subject pertain to the RCS, leaving the choice from the RCS

unspecified. This solution is simple: subjects may choose between the options in the prior envelope on the spot.

(2) The *exclusion overlap problem* arises if the partial information about the RCS excludes some choice situations generated during the experiment, thereby reducing salience and motivation for truthfulness in these excluded choice situations. To combat the exclusion overlap problem, experimenters must frame the partial information concerning the RCS by anticipating the range of possible choice situations generated in the experiment. They do this by using descriptive theory and pilots. For example, in our adaptive experiment (§5) we informed subjects about a large possible outcome (> €3000). Choice situations with very large monetary amounts could arise in our experiment, depending on subjects' answers.

2.3 Prince summarized

We now formally list the principles that define Prince.

(1) [PRiority] The RCS is determined at the start of the experiment, *before the subject makes any decision*.

(2) [INstructions to experimenter] We explicitly request "instructions" from the subjects, asking what the experimenter should select from their envelope at the end, rather than asking vague "what would you prefer if" questions referring to unspecified situations.

(3) [Concreteness] A description of the RCS is handed out to the subject in tangible form, such as in a sealed envelope (the *prior envelope*).

(4) [Entirety] The description handed out to the subjects describes the entire RCS and not just one option in it[7].

For adaptive measurements, two criteria are added to the definition of Prince.

(5) [No indeterminacy] If subjects do not give instructions during the course of the experiment that pertain to the RCS, then they can give instructions on the spot, after the envelope has been opened.

(6) [No exclusion] The initial description about the RCS should be framed so as not to exclude potential choice situations faced during the experiment.

---

[7] Such an option is for instance a random price in BDM.

While parts of Prince have been used before (§6), their integration into Prince is new, and their combination achieves incentive compatibility in a transparent manner and proper conditioning on the RCS. For example, all BDM implementations that we are aware of violate Principle 4 [entirety], leading subjects to condition the wrong way (enhancing rather than avoiding meta-lottery perceptions; §7), which hampers BDM's internal validity and arguably accounts for its bad performance.

In our experiments, not only the prior envelopes, but all stimuli are physical. Other researchers may prefer computerized implementations of Prince. The physical availability of the RCS to every subject such as in a prior envelope is essential though, which is why it is listed as principle (3) above.

We avoid deception and, hence, the partial information about the RCS provided must be true. Although it is not a defining principle of Prince, in our implementation subjects could completely verify the absence of deception. First, before they received the real payment, they could verify the correctness of the information provided about the stimuli (Web Appendix E). Second, unlike computer randomizations, our physically generated randomizations were fully verifiable and carried out by the subjects themselves.

3 Experiment 1: Reinventing matching (WTA)

Experiment 1 implements Prince for one of the most used value concepts for nonmarket goods: willingness to accept (WTA). We measured WTA for a university mug that could be bought on campus for €5.95. Mugs are suited to test the endowment effect (tested later) because people quickly develop attachments to them. WTA measures how much money a subject would accept in lieu of the mug, which according to traditional theories should be the mug's cash equivalent.

N = 30 subjects (40% female), recruited from undergraduates in the School of Economics, Erasmus University Rotterdam, the Netherlands, participated in one classroom session. Advertisement of the study promised a €10 show-up fee plus either a mug or additional money. Participants immediately received a mug (*endowment*) along with the show-up fee.

Next, the experimenter presented 50 sealed envelopes, visibly numbered 1-50. These were separated into 5 piles of 10 each (1-10, …, 41-50). Five subjects each checked one pile to verify that each number between 1 and 50 occurred once. Appendix A4 explains how, using this verification, subjects could completely verify that there was no deception. The subjects placed the envelopes into a large opaque bag, shuffled them, and randomly redistributed them over a number of smaller bags (one for each row in the classroom). Each subject, in turn, randomly took one envelope, the prior envelope, from a bag (without replacement). Subjects were told that their envelope described two options, and that at the end of the session we would give them one of those two, based on instructions that subjects would give us.

Subjects received a questionnaire reproduced in Figure 3.1, and were given a short written explanation along with a PowerPoint presentation on the procedure. They were told that they could give up the mug for a price: "You will write instructions, for each possible content of your envelope (for each money amount), which of the two options you want. At the end, we will give you what you instructed. … If you write what you want, then you get what you want!" We call the question in Figure 3.1 Question 1 for later comparisons with Experiment 2.

At the end of the experiment, subjects handed in their questionnaire (instructions to experimenter). An experimenter opened their envelope, observes the real choice situation specified in the envelope, and follows the instructions in the questionnaire.

RESULT. The average WTA was 4.99 (SD 2.41). Further results are in §4.3.

DISCUSSION OF PROVIDING RANGE 0-10 FOR ANSWERS. Whereas specifying a range cannot be avoided for choice lists, it is optional for matching. We chose to specify it here, but for comparison will not specify it later in Questions 5 and 6 in §4.5. There are pros and cons (Birnbaum 1992). We chose the specific range to facilitate comparability with choice lists presented later.

FIGURE 3.1. Instructions for WTA with matching and endowment (Question 1)

In each of the 50 envelopes, one option is to keep your mug, and the other option is to give up your mug for a money amount. The note in each envelope is as follows.

Option 1: Keep your mug

Option 2: Give up your mug for €x

The money amount x varies between €0 and €10 in different envelopes. Five of the envelopes contain a randomly generated amount between €0 and €1, five envelopes contain a randomly generated amount between €1 and €2, five contain a randomly generated amount between €2 and €3, and so on, with finally five envelopes containing a randomly generated amount between €9 and €10. Thus the amount in your envelope can be any amount, in cents, between €0 and €10.

Please give us instructions, for each possible envelope that your envelope may be, whether we should let you keep your mug, or we should give you that money amount in exchange for your mug. Do so by specifying a threshold (in cents).

My threshold is € ……,……

If the money amount x in my envelope is equal to or above the threshold, then give me that money amount in exchange for my mug.

If the money amount x in my envelope is below the threshold then let me keep my mug.

4 Experiment 2: Prince implemented in a large experiment

In Prince, only one choice is implemented for real. Experiment 2 shows how Prince can be used in large experiments with many measurements.

4.1 General procedure

N = 80 subjects (41.2% female), recruited from undergraduates in the School of Economics, Erasmus University Rotterdam, the Netherlands, were randomly divided into two groups. Each group participated in one classroom session. They received a

€10 show-up fee and could gain an additional offering: money, mug, or chocolate. Experimental instructions including a short presentation were given by the experimenter (Web Appendix WE). For each of the two sessions there were 90 envelopes, numbered 1-90 in random order. As with Experiment 1, these envelopes were separated into piles of 10, checked by subjects, shuffled, explained, and randomly distributed without replacement.

The two groups of subjects received different versions of the first question, *1-match* or *1-choice* (see Figures 4.1 and 4.2). These questions were part of a between-subject test in this large experiment. Question 1 appears in Experiment 1. The remaining eight questions, 2-9, were asked in randomized orders to all subjects in the two groups. Each of the nine questions corresponded to a type (the term used with subjects) of envelope, and there were 10 envelopes of each type. The numbering (1-match/choice, 2,3,…,9) of types/questions used in this paper was not communicated to subjects. Thus each subject randomly drew an envelope, their prior envelope containing their RCS, from 90 envelopes and then gave 9 instructions in response to 9 types/questions.

At the end of the experiment, each subject handed in their questionnaire bundle and prior envelope. An experimenter opened the envelope, searched for the instruction in the questionnaire made operational by the RCS, and carried it out.

4.2 The endowment effect

Question 1-match measured subjects' WTA for a mug, now without endowment. It was asked of 41 of the 80 subjects. Figure 4.1 presents Question 1-match. Results and discussion are at the end of §4.3.

FIGURE 4.1. Instructions for cash equivalent with matching and no endowment (Question 1-match)

Instructions for envelopes of type γ

In each of the 10 envelopes of type γ, one option is the mug you just saw, and the other option is a money amount. The note in each envelope of type γ is as follows.

Type γ

Option 1: The mug

Option 2: €x

The money amount x varies between €0 and €10 in different envelopes. One of the envelopes contains a randomly generated amount between €0 and €1, one envelope contains a randomly generated amount between €1 and €2, one contains a randomly generated amount between €2 and €3, and so on, with finally one envelope containing a randomly generated amount between €9 and €10. Thus the amount in your envelope can be any amount, in cents, between €0 and €10.

Please give us instructions, for each possible envelope of type γ that your envelope may be, whether we should give you the money amount or the mug. Do so by specifying a threshold (in cents).

My threshold is € ……,……

If the money amount x in my envelope is equal to or above the threshold, then give me that money amount.
If the money amount x in my envelope is below the threshold then give me the mug.

4.3 Matching versus choice lists between subjects

Question 1-choice repeats Question 1-match, again with no endowment, but now using choice lists instead of matching, for the remaining 39 subjects. Figure 4.2 presents Question 1-match. The sure amount of money (the alternative to the mug) increases with each option presented. At first, nearly all subjects preferred the mug, but by the end nearly all subjects preferred the money. Somewhere, they switched, and the midpoint between the two money amounts where they switched was taken as their indifference point.

FIGURE 4.2. Instructions for cash equivalent with choice list and no endowment (Question 1-choice)

Instructions for envelopes of type δ

In each of the 10 envelopes of type δ, one option is the mug you just saw, and the other option is a money amount. The money amount x varies between €0.50 and €9.50 in different envelopes (see below).

The note in each envelope of type δ is as follows.

Type δ

Option 1: the mug

Option 2: €x

In the following list, each line describes the content of one envelope of type δ. On each line, cross out the square before the option that we should give you if that line describes the two options in your envelope.

1. □MUG	□€0.50
2. □MUG	□€1.50
3. □MUG	□€2.50
4. □MUG	□€3.50
5. □MUG	□€4.50
6. □MUG	□€5.50
7. □MUG	□€6.50
8. □MUG	□€7.50
9. □MUG	□€8.50
10. □MUG	□€9.50

An inconsistency results if a subject takes the money when the money offer is small but then switches to the mug when more money is offered. We allowed such inconsistencies so as to be able to detect subjects' misunderstandings. The number of misunderstandings provides information about the transparency of Prince.

RESULTS OF QUESTIONS 1, 1-match, and 1-choice. In the 119 choice lists presented in this experiment (39 subjects here and all 80 subjects in §4.4), there was only one inconsistency—that is, only one switch in the wrong direction (by subject 59). In otherwise comparable studies, typically 10% of subjects have inconsistent switches (Holt and Laury 2002). Because this one subject exhibited other anomalies as well (violating stochastic dominance in a later question), we removed her from our analyses. Leaving her in would not alter our results. Table 4.1 reports summary statistics, and Table 4.2 reports tests.

TABLE 4.1. Statistics for Questions 1 (matching with endowment), 1a (matching without endowment), and 1b (choice list without endowment)

	Groups	N	Mean	SD
Experiment 1	Question 1	30	4.99	2.41
Experiment 2	Question 1-match	41	3.19	1.96
	Question 1-choice	39	3.61	2.51

TABLE 4.2. Tests of equality of means

Questions	Treatment	mean difference	t	df	p
1 vs 1-match	endowment or not	1.81	3.48	69	0.001
1-match vs 1-choice	matching versus choice	–0.43	–0.86	78	0.40

DISCUSSION. Prince confirms the endowment effect.[8] Rational or not, it reflects a genuine property of preference (Brosnan et al. 2012; Korobkin 2003 p. 1244), not a bias in measurement.

Prince reduces errors of measurement and resolves the discrepancy between choice and matching. Our matching questions are very similar to the choice questions, directly referring to the choice in the prior envelope held in hand. Accordingly, their equality is no surprise. Our contribution here is of a

[8] References are in Camerer (1995 p. 665 ff.) and Schmidt & Traub (2009).

methodological nature: we made matching look like choice, combining the virtues of both.

The test of choice versus matching presented here was between subjects. For its result, not rejecting the null, to be convincing, statistical power should be sufficient. It is sufficient because, first, the endowment effect is very significant. Second, in §4.4 we confirm our finding in a within-subject test for 80 subjects, increasing power. Several tests reported later confirm non-rejection of the null.

4.4 Matching versus choice lists within subjects

Questions 2 and 3 replicate Questions 1-match and 1-choice with chocolate (price €6.25) instead of a mug. Chocolates and mugs were used by Kahneman, Knetsch, & Thaler (1990), and many follow-up studies. Here we follow suit. Questions 2 and 3 were asked to each subject, allowing within-subject comparisons. The stimuli are in Web Appendix WB. The average cash equivalent was 3.31 for matching and 3.26 for the choice list (t_{79} = 0.28, p = 0.78), unable to reject the null hypothesis of equality.

4.5 Testing preference reversals

We used Prince to test the classical preference reversal of Lichtenstein & Slovic (1971). Details are in Web Appendix WA. For Question 4, the choice question, we used an analog of Figure 4.1 without the description of x. Option 1 was $4_{0.97}0$ (receiving €4 with probability 0.97 and €0 otherwise), called the *P-bet* in the literature because the gain probability is high. Option 2 was $16_{0.3}0$, called the *$-bet* because it has a high minimum possible gain (in dollars when receiving its name; Lichtenstein & Slovic 1971). We also measured their cash equivalents in Questions 5 and 6, again using analogs of Figure 4.1, but without ranges for amount x, writing only “The amount x varies between the envelopes.” Although in consequence almost nothing is known about x’s randomness, that affects neither the compatibility nor the transparency of incentives.

Normal preference reversals (higher CE of the $ bet but, paradoxically, choosing the P bet) occurred for 11% of the subjects, and the opposite preference reversals (higher CE of the P bet but choosing the $ bet) happened for 7% of the subjects. These percentages are not significantly different (p = 0.55) and are infrequent enough to be explained as random choice inconsistencies (Schmidt & Hey 2004). We find no evidence of genuine preference reversals. As regards not having specified a range for matching, we found more choice anomalies here than for Question 1 where we had specified a range. Our finding illustrates that providing context can reduce distortions (Birnbaum 1992).

Our finding deviates from other studies of preference reversals, where normal preference reversals are found in large majorities (surveyed by Seidl 2002). Preference reversals reflect errors in the measurement of preferences (procedural variance) rather than genuine properties of preferences such as intransitivities (Tversky, Slovic, & Kahneman 1990).[9] Prince restores consistency between choice and matching, thus resolving preference reversals.

4.6 Measuring subjective probabilities and ambiguity attitudes

Using questions 7, 8, and 9 we replicate the measurements of subjective probabilities and ambiguity attitudes by Baillon & Bleichrodt (2014 Study 1). They used classical choice lists, whereas we use Prince and matching. Karni (2009) recommended the BDM method. Details are in Web Appendix WC. We measured the probability p such that

$10_E0 \sim 10_p0.$

E denotes an event explained as an observation from the Dutch AEX stock index, and 10_E0 means that the subject receives €10 if E happens, and nothing otherwise. 10_p0 means that the subject receives €10 with objective probability p. The probability p giving the preceding indifference is called the *matching probability* of event E, denoted m(E). We measured it for three events:

E = A (Question 7): The Dutch AEX stock index increases or decreases by no more than 0.5% during the experiment.

E = B (Question 8): The Dutch AEX stock index increases by more than 0.5% during the experiment.

E = $A \cup B$ (Question 9): the AEX stock index decreases by no more than 0.5% during the experiment.

---

[9] Intransitivity is a genuine property of preference in regret theory (Loomes & Sugden 1982).

Our presentation of questions was similar to Figure 4.1, with option 1 being 10_E0 and option 2 being 10_p0, requesting that a threshold for p (instead of x) be specified. Baillon and Bleichrodt (2014) showed how we can use these observations to analyze ambiguity attitudes, using a nonadditivity index $m(A) + m(B) – m(A \cup B)$. We replicated all their findings. In particular, the nonadditivity index was mostly positive, rejecting expected utility, and confirming Tversky and Fox's (1995) subadditivity and Abdellaoui et al.'s (2011) a(mbiguity-generated likelihood)-insensitivity. These properties are genuine properties of preferences and not artifacts of measurement. Hence Prince did not remove them. Validity is confirmed because we found the same phenomena on subjective probabilities as other experimental studies did. Here, as throughout, the advantage of Prince is that we obtained our results more quickly (using matching instead of choice) and more precisely than preceding papers did.

5 Experiment 3: Prince implemented in an adaptive experiment; measuring utility

We use an adaptive method to measure utility and show how Prince can resolve incentive compatibility problems by ruling out strategic answering. Exact stimuli, instructions, and details are in Web Appendix WE. We first piloted the following procedures in two sessions, each with approximately 10 graduate students from ERIM research institute of Erasmus University. These students had considerable exposure to decision theory. After the pilot, as an assignment, they were tasked with criticizing the procedures, especially in ways the experimenter could deceive or manipulate. They were unsuccessful in their attempts to find weaknesses in the procedures.[10] These students, as well as colleagues in informal pilots, confirmed procedural transparency and absence of biases.

We use Wakker & Deneffe's (1996) adaptive tradeoff (TO) method to measure utility. This method is robust to violations of expected utility and provides a correct

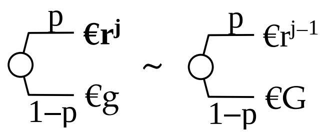
[10] There were humorous suggestions such as "pull the fire alarm just when you have to pay €3000."

utility function irrespective of whether a subject maximizes expected utility, prospect theory, or most other nonexpected utility theories. Implementations were as yet not incentive-compatible, and adding incentive compatibility is our contribution here. Integrating incentive compatibility makes the method suited for economics.

5.1 The preferences to be elicited for the tradeoff method

We measure indifferences $\mathbf{r^j}_p g \sim r^{j-1}{}_p G$, j = 1, …,4 (see Figure 5.1, with the conventional notation for bets using circles as chance nodes). Consistent with the notation used in the stimuli of the experiment, superscripts indicate the sequence of outcomes r^j.

FIGURE 5.1. Tradeoff method indifferences



The experimenter chooses some pre-set values $0 < p < 1$, $0 < g < G$ (*gauge outcomes*), and $r^0 > G$. Then the bold-printed outcomes $\mathbf{r^1}$, $\mathbf{r^2}$, $\mathbf{r^3}$, $\mathbf{r^4}$ are elicited sequentially from each subject over four stages. The experiment is adaptive because r^1, r^2, and r^3, after having been elicited, serve as input to the next question in stages 2, 3, and 4. We assume a *weighted utility* model:

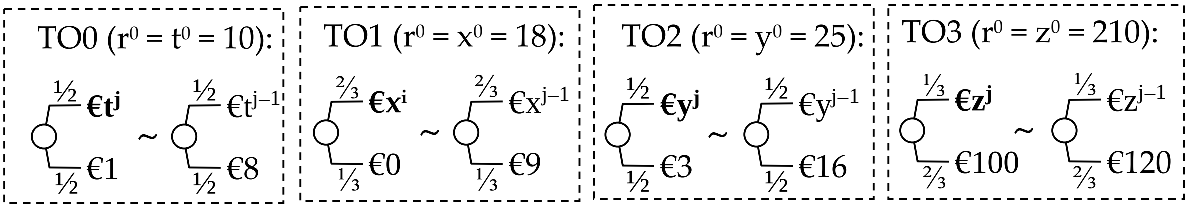
$$\text{for } x \geq y,\ x_p y \text{ is evaluated by } \pi U(x) + \rho U(y)) \ (\pi > 0,\ \rho > 0). \qquad (5.1)$$

This model includes expected utility, prospect theory for gains (Tversky & Kahneman 1992), and most other generalizations of expected utility (Wakker 2010 §7.11). Algebraic manipulations show that the r^js are equally spaced in utility units under Eq. 5.1 (Wakker 2010 §4.3, §7.11, §10.6):

$$U(r^4) - U(r^3) = U(r^3) - U(r^2) = U(r^2) - U(r^1) = U(r^1) - U(r^0). \qquad (5.2)$$

A nonparametric measurement of utility results (§5.5, §5.7) that is valid for virtually all risky choice theories. The observations can also be used for parametric fitting (§5.6, §5.8). The TO method avoids collinearity between utility U and probability weighting (π and ρ in Eq. 5.1): Eq. 5.2 is not affected by the probability weights π and ρ, and we need not even estimate them. For other measurements of prospect theory in the literature, collinearity is a serious problem (demonstrated by Zeisberger, Vrecko, & Langer 2012; p. 366 ff.). For a sophisticated recent measurement, see Bruhin, Fehr-Duda, & Epper (2010).

FIGURE 5.2. The values used for TO0-TO3; j = 1, …, 4

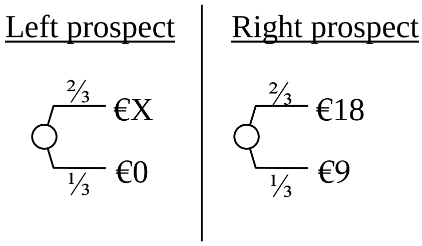


We carried out the TO measurement with four sets of pre-determined values, one training set and three observational sets: TO0 (with t^j for r^j, t means training), TO1 (with x^j for r^j), TO2 (with y^j for r^j), and TO3 (with z^j for r^j) depicted in Figure 5.2. Wakker & Deneffe (1996) used the same stimuli but scaled up and choices were hypothetical.

Figure 5.3 displays the first two questions, TO1.1 and TO1.2, of the TO1 quadruple, as presented to the subjects. Question TO1.2 immediately followed TO1.1 on a separate page. Not only is the experiment adaptive, but also it is obviously so to subjects. Each subjects had to impute the answer they gave to the first question ($x^1 = r^1$) before answering the next question (determining r^2). The third and fourth questions were like the second, requesting information of the previous answer.

T01.1

Determining first number x^1



Your envelope may contain two prospects of the above form.

For each nr. X, instruct which prospect you want to be taken from the envelope if its content is as above.

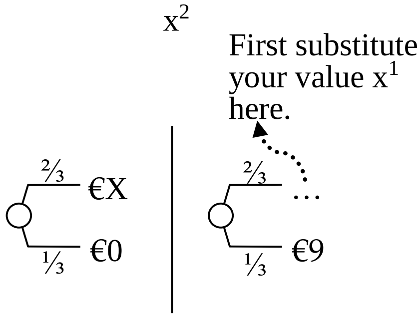
For small values of X you prefer the right prospect. For large values of X you prefer the left prospect. For some value of X, which we call x^1, your preference switches from left to right. Fill this switching value in below, and then on page TO1.0.

x^1 = ...

FIG. a

TO1.2

Determining second number x^2



- You have substituted your value x^1.
- To indicate your instructions, determine your switching value of X again.
- We call it x^2 (obviously, $x^2 > x^1$).
- Fill it in below and on page TO1.0.

x^2 = ...

FIG. b

FIGURE 5.3. Figures used in the tradeoff method

5.2 Procedure and real incentives

We used Prince in two one-hour, pen and paper sessions (25 and 55 subjects) so as to implement performance-contingent incentives. Subjects were undergraduate students of Erasmus University in Rotterdam who were enrolled in an economics class. They received a €5 show-up fee in addition to their performance-based payoff. They first chose a sealed envelope with their RCS. Then they received written instructions, accompanied by an explanatory PowerPoint presentation.

Subjects filled out the training questions of TO0, jointly and simultaneously, exactly as in Wakker & Deneffe (1996), guided by the PowerPoint presentation. Subjects wrote their answers on pp. TO0.1-TO0.3, which they kept, but also on the front page TO0.0, which they tore off and gave to the experimenter at the end of the experiment. We explained how the performance payment procedure worked, and how subjects' answers to the questionnaire would determine the selection from the RCS in their envelope. Only then did subjects receive the three sets of questions TO1, TO2, TO3 (ordered randomly, subject-dependent), which they completed at their own pace. Three subjects in the first group, and six in the second, were randomly selected for real play at the end.

5.3 Construction and use of envelopes for real incentives, and avoiding the two overlap problems

In preparation for each session, we constructed 100 envelopes, from which each subject would randomly choose one (without replacement). Each envelope contained a slip with two bets written on it (the RCS). We used popular theories of risky choice, mostly expected value and prospect theory, and pilot studies to determine the contents of the envelopes that minimize both overlap problems. Because the details depend on particularities of the experiment, we present them in Appendix A.

5.4 Experiment with hypothetical choice

We did two sessions (10 and 44 subjects) with hypothetical choice. Subjects were unaware that other subjects played for real incentives. There was no role for Prince techniques here, as no incentives needed explaining or implementing. We only describe the differences with the incentivized experiment. Subjects received €10 for participation. They made less on average than the real incentive condition but the session took less time. The results that follow concern the incentivized sessions, unless stated otherwise.

5.5 Non-parametric analysis

An advantage of the TO method is that we can infer the utility function nonparametrically, i.e., without a commitment to any family of, or any shape of, utility functions (Wakker 2010 §9.4.2), using Eq. 5.2. Figure 5.7.1 presents utility graphs generated by average answers.

To develop a nonparametric test of concavity, note that for strictly concave utility we have (with r = x, y, or z, respectively)

$$r^{i+2} - r^{i+1} > r^{i+1} - r^{i} \qquad (5.3)$$

for all i, and for strictly convex utility we have

$$r^{i+2} - r^{i+1} < r^{i+1} - r^{i} \qquad (5.4)$$

for all i. We classified a subject's utility as concave if Eq. 5.3 was satisfied more often than Eq. 5.4, and as convex if the opposite held, with Eq. 5.4 satisfied more often than Eq. 5.3. The remaining subjects were irregular or linear.

5.6 Parametric analysis

Whereas the TO method allows nonparametric utility analyses, it can also be used for parametric analyses. We used Eq. 5.1 with the two most common parametric utility families, CARA (constant absolute risk aversion, or linear-exponential) and CRRA (constant relative risk aversion, or log-power) utility. They are defined as follows.

CARA utility:

for $\alpha > 0$, $U(r) = 1 - e^{-\alpha r}$; (5.5)

for $\alpha = 0$, $U(r) = r$; (5.6)

for $\alpha < 0$, $U(r) = e^{-\alpha r} - 1$. (5.7)

CRRA utility:

for $\rho > 0$, $U(r) = r^{\rho}$; (5.8)

for $\rho = 0$, $U(r) = ln(r)$; (5.9)

for $\rho < 0$, $U(r) = -r^{\rho}$. (5.10)

We used Eq. 5.2 as the basis of our parametric analysis, rewriting it as

$$r^j = U^{-1}(2U(r^{j-1}) - U(r^{j-2})) \text{ for } j = 2, 3, 4. \tag{5.11}$$

We assumed a Fechnerian error model with the error term directly imposed on answers r^j that subjects produced:

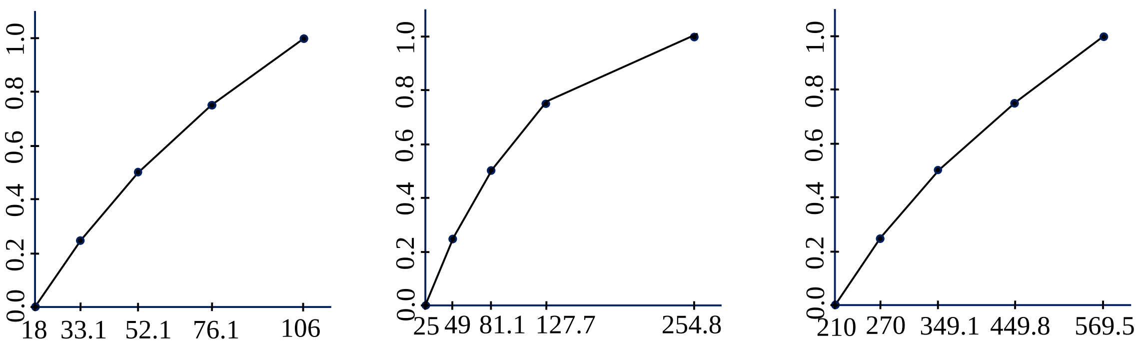
$$r^j = U^{-1}(2U(r^{j-1}) - U(r^{j-2})) + \varepsilon_R \text{ for } j = 2, 3, 4. \tag{5.12}$$

We chose to impose error terms upon the direct measured values, being r^j (Eq. 5.11), rather than upon utilities as often done in the literature. Our error model is similar to Wilcox's (2011) contextual approach. The error term ε_R has expectation 0 and standard deviaton $\sigma > 0$ per €. That is, σ is to be interpreted as the standard deviation when differences $r^j - r^{j-1}$ are in the order of magnitude of €1. Our error model is thus close to the thinking process of subjects, because the TO method enhances the direct comparison of r^j to r^{j-1}during the experiment. Given that $x^j - x^{j-1} = 4.5$ under expected value maximization, we chose the standard deviation ε_x for $r^j = x^j$ equal to 4.5σ. We similarly used expected value approximations to set $\varepsilon_y = 13\sigma$ and $\varepsilon_z = 40\sigma$. We estimate σ and the utility parameter so as to maximize likelihood.

5.7 Results of the non-parametric analysis

As regards the indeterminacy overlap problem of §2.2, for eight out of the nine envelopes opened during our experiment, the questionnaire answers determined the choice from the envelope, which was implemented. For the indeterminate case, the subject chose on the spot.

FIGURE 5.7.1. Nonparametric utility curves for the TO method



a. Utility curve for average x^j answers

b. Utility curve for average y^j answers

c. Utility curve for average z^j answers

Figure 5.7.1 depicts the utility graphs resulting from average answers to the x, y, and z questions, based on Eq. 5.2, with utility normalized to be 0 at outcomes x^0, y^0, and z^0, and to be 1 at x^4, y^4, and z^4. Note that these graphs do not involve parametric assumptions. They can also be produced for every individual. We can use overlaps of the x, y, and z regions to combine such curves into one overall curve. As one would expect from the overall concavity of curves in Figure 5.7.1, most participants exhibit concave utility (Eq. 5.3) versus convexity over outcomes (Eq. 5.4): 37 versus 13 for the x's, 29 versus 12 for the y's, and 21 versus 14 for the z's. This is significant for both x and y ($p \leq 0.01$). Thus our findings confirm concavity of utility, although the concavity is not pronounced. For the x's, about 20% of our subjects had equality for all i in Eqs. 5.3 and 5.4, giving perfectly linear utility, and the same held for the y's and z's. The subjects not classified exhibited irregular (or linear with noise) utilities. The hypothetical choice groups' results were in line, but with more concavity for x and y stimuli, and not for z stimuli, than for incentivized groups.

5.8 Results of the parametric analysis

Our main analysis is at the individual level, finding the best utility parameter and error variance for all observations x^j, y^j, and z^j simultaneously. Normal distributions of the parameters are rejected and, hence, we report medians, Wilcoxon signed rank, Mann-Whitney, and Kruskal-Wallis tests.

The median CARA parameter α is 0.0001 (more precisely, $\alpha = 0.000099$), yielding risk tolerance $1/\alpha = €10000$. It suggests concave utility, but the value is not significantly different from linearity ($\alpha = 0$). The median power ρ is 0.96, again suggesting weak concavity but not deviating significantly from linearity ($\rho = 1$).

The standard deviation σ is larger for CARA utility than for CRRA utility ($p = 0.001$)[11]. Thus CRRA utility fits the data better than CARA utility. We also fitted parameters to the x stimuli, y stimuli, and z stimuli separately, and compared them. Kruskal-Wallis tests reveal that CARA's α marginally depends on x, y, and z ($p = 0.09$), being higher for x than for y stimuli ($p = 0.03$), marginally higher for x than for z stimuli ($p = 0.09$), and not different between y and z stimuli. These results confirm

[11] Their medians happen to be the same (0.88), due to different skewnesses. Note that the standard deviations have the same unit (1 per €) and, hence, can be compared.

the commonly assumed decreasing absolute risk aversion (references in Wakker 2010 p. 83). A reason that we find no differences for the z stimuli may be because the z stimuli, while at higher levels of wealth, also involve bigger differences, which in itself will enhance risk aversion. CRRA's ρ does depend on x, y, and z ($p = 0.01$), being smaller for the x than for the y and z stimuli ($p \leq 0.01$), and not different between the y and z stimuli. This finding implies decreasing relative risk aversion. Although most authors have conjectured increasing, rather than decreasing relative risk aversion, several studies find the opposite (Gollier 2001 §4; Ogaki & Zhang 2001) and no consensus has emerged on this point.

Our findings comparing incentivized and hypothetical choices agree with common findings. The hypothetical data were noisier and contained more outliers. Further: (1) There was some more risk seeking for hypothetical choice than for real incentives in the z stimuli, reaching marginal significance ($0.05 < p < 0.10$) both for CARA and CRRA utility. (2) No significant differences were found in risk attitudes or standard deviations of parameters for further comparisons between real and hypothetical choice.

5.9 Discussion of adaptive utility measurements

As regards the problem of strategic answering in adaptive experiments, Toubias et al. (2013 p. 629) and Wang, Filiba, & Camerer (2010) provide suggestions alternative to ours for mitigating this problem. One of these suggestions, deriving a preference functional from the experimental answers and implementing this functional in the RCS, was implemented by Ding (2007). Then subjects cannot directly understand the effects of their answers on the RCS during the experiment, and have to trust the relevance of the derived functional.

In experiments where subjects cannot really influence stimuli, they may mistakenly think they can, e.g. due to magical thinking (Rothbart & Snyder 1970) or illusions of control (Stefan & David 2013). Such distortions are more likely with future than with past uncertainties. Thus Prince helps to avoid such distortions by determining the RCS before to the subject makes actual decisions.

By classical economic standards it may be surprising that we find near-linear utility, whereas classical estimates, based on expected utility, usually find more concavity. Recent studies find that risk aversion is mostly generated by factors other than utility for the moderate stakes considered in our experiment. With these factors filtered out, as in Eq. 5.2, utility turns out to be almost linear. Epper, Fehr-Duda, & Bruhin (2011) who, like us, correct for deviations from expected utility, argue for the reasonableness of this finding.

Unlike most measurements of utility in the literature, our analysis does not need to correct for deviations from expected utility. The TO stimuli were carefully devised such that those deviations have no bearing on our analysis, giving the same Eq. 5.2 under expected utility and nonexpected utility. The deviations are avoided rather than corrected for. This point further supports our claim of high internal and external validity for our utility findings.

6 Parts of Prince used in preceding studies

Virtually all choice experiments using the RIS randomly select the RCS at the conclusion of the experiment, thus violating our Principle 1 (priority), and then also 3 (concreteness). All (to our knowledge) violate Principle 2 (instructions). Many satisfy Principle 4 (entirety), randomly selecting, for instance, a row in a choice list, which constitutes the whole choice situation. Virtually all matching experiments (mostly using BDM) violate Principles 1, 2, and 3, and none that we know of satisfy Principle 4. The remainder of this section focuses on studies (partly) satisfying Principle 1 by providing envelopes to subjects at the start of the experiment.

Loomes, Starmer, & Sugden's (1989) procedure in their experiments 1 and 2 comes close to Prince, also aiming to enhance isolation. The envelope selected a priori by each subject contained a number indicating the RCS, which concerned a choice between two lotteries. Thus Principles 1, 3, and 4 were partly satisfied. A problem for Principle 1 is that the RCS assigned to the number in the envelope may be determined only during the experiment, such as in adaptive designs. Hence subjects cannot know if it is determined a priori, and Principles 3 and 4 also are not fully satisfied. Bardsley (2000) partly satisfied Principle 1 too.[12] His purpose was to

[12] The first experiment with a prior envelope may have occurred earlier, by Johann Wolfgang von Goethe (January 16, 1797, letter cited by Mandelkow (1968, p. 254). Goether wrote: "I am inclined to offer Mr. Vieweg from Berlin an epic poem, Hermann and Dorothea … Concerning the royalty we will

implement lying without lying, so to say. Bardsley could not determine the choice options in the RCS for a given subject beforehand because the latter depended on choices made by other subjects during the experiment. Thus Bardsley could neither satisfy our Principles 2-4. He recommended Principle 3 (concreteness) for future studies (last paragraph of his §7).

In Schade, Kunreuther, & Koellinger (2012; first version 2001), options were determined a priori in an envelope (lying on a desk in the front of the room), but not whole choice situations. What was real (sculpture/painting) was determined only at the end of the experiment, and with a small probability. Hence Principle 1 was partly satisfied, and Principle 3 was approximately, but Principles 2 and 4 were not. Wang, Venkatesh, & Chatterjee (2007) also used Schade, Kunreuther, & Koellinger's (2012) design, referring to Schade et al.'s 2001 working paper.

Bohnet et al. (2008) determined the RCS a priori. One choice option was inserted in an envelope that was visibly posted on a blackboard while subjects answered the experimental questions. Thus Principle 1 is satisfied, and Principle 3 is approximately so. Principle 4 is not satisfied. The authors first asked subjects what subjects would "pick," but later formulated these as instructions to the experimenters, thus partly satisfying Principle 2. Hao & Houser (2012) satisfy our Principles 1 and 3. They also used a formulation in the spirit of Principle 2. However, to optimize other goals in their research, they deviate from Principle 4. They present a meta-lottery B before explaining choices, then present a single strategy-choice between meta-lotteries, explicitly deviating from isolated binary choices.

Bleichrodt et al. (2013) considered Gilboa & Schmeidler's (2001) case-based decision theory. Their questions concern varying memories for the same two choice options, rather than varying choice options as in classical revealed-preference experiments. They used Bardsley's (2000) method, satisfying Principles 1 and 3. They did not satisfy Principles 2 and 4. Their experiment was adaptive, but such that

---

proceed as follows: I will hand over to Mr. Counsel Böttiger a sealed note which contains my demand, and I wait for what Mr. Vieweg will suggest to offer for my work. If his offer is lower than my demand, then I take my note back, unopened, and the negotiation is broken. If, however, his offer is higher, then I will not ask for more than what is written in the note to be opened by Mr. Böttiger." We thank Uyanga Turmunkh for this citation.

the two overlap problems (indeterminacy and exclusion) did not arise, thus avoiding strategic behavior.

In Camerer (1989), subjects, when selected for real play, could choose their preferred option only then, on the spot. We did the same but only in the special cases where the experimental choices did not specify the choice in the RCS, which can happen in adaptive experiments. Camerer's experiment contained no adaptive questions, and he offered the ex-post choice to all subjects, thus testing isolation. His test involved revising choices, which entailed a slight deviation from the prior announcement that experimental choices would be outcome relevant, involving a mild form of deception.

7 General discussion

The principles of Prince listed in §2.3, and in general each detail of Prince, serve to enhance isolation by enhancing psychological conditioning upon the RCS, increasing internal validity. Although Starmer & Sugden (1991) found isolation satisfied in the RIS, violations have been found.[13] In fact, any finding of learning, order effect, or spillover effect (Cox, Sadiraj, & Schmidt 2014b) entails a violation of isolation, and such effects have been widely documented. Hence improvements of isolation are desirable.

Regarding Principle 1 (priority), many studies have shown that conditioning works better for events determined in the past, even if yet uncertain, than for events to be determined in the future.[14] In the case of future determination, a meta-lottery is

---

[13] They assumed one single choice per subject as gold standard as regards the implementation of real incentives, but Birnbaum (1992) criticized this assumption. This gold standard is also impractical for collecting rich data.

[14] See Bardsley et al. (2010 p. 277), Cubitt, Starmer, & Sugden (1998), and Shafir & Tversky (1992 p. 463). In Bardsley et al.'s (2010) terminology, Prince uses the direct decision approach and avoids the strategy method. That the timing of the resolution of uncertainty, even if of no strategic or informational relevance, still affects subjects, has been demonstrated in many studies (Bosman & van Winden 2010; Dillenberg 2010; Grant, Kajii, & Polak 2000; Kreps & Porteus 1979; Strzalecki 2013), and plays a role in time inconsistencies. In particular, prediction and postdiction are perceived differently (Brun & Teigen 1990; Heath & Tversky 1991 p. 9; Rothbart & Snyder 1970). Importantly

realistically perceived because the situation is still unresolved. More generally, we want the RCS to be felt as realistically as possible. Planning prior to the determination of the RCS (rather than after as with Prince's Principle 1), which amounts to the strategy method, can matter (Bellemare, Kroger, & Van Soest 2008; footnote 4). It generates a psychological distance (Bardsley et al. 2010 §6.4.3). Strategy choice further obstructs isolation by referring to random options (as with the random prizes of BDM) rather than to random choice situations. Principles 3 (concreteness) and 4 (entirety) reduce such obstructions.

There have been several implementations of real incentives using prior envelopes (§6) after Bardsley (2000), but all describe only one choice option in the envelope. If the randomization concerns the entire choice situation as with Prince (Principle 4), then subjects immediately condition on it, serving isolation. BDM randomizes a choice option (the price) rather than the choice situation, leading subjects to condition the wrong way. It obfuscates the choice situation, with the random price draw enhancing the undesirable perception of meta-lotteries. Principle 4 (entirety) is crucial for Prince.

Researchers in decision theory will immediately see that Prince is strategically equivalent to RIS, soliciting real preferences. Homo economicus will behave the same in both procedures, and for her Prince need not be developed. However, as Bardsley et al. (2010 p. 270-271) wrote: "the effects of incentive mechanisms can depend on features of their implementation which are irrelevant from a conventional choice-theoretical point of view." Prince minimizes the biases generated by those features. It targets homo sapiens.

Throughout the history of preference measurement, there have been discussions of the pros and cons of matching versus choice.[15] Choice is less precise. It takes

---

for Prince, people more readily condition on uncertainties determined in the past than in the future, and take future uncertainty more as a meta-lottery (Keren 1991). This phenomenon underlies several findings in game theory (Weber, Camerer, & Knez 2004: virtual observability). Techniques for enhancing conditional isolated thinking are commonly used in insurance and marketing. For instance, pseudo-certainty is generated by making clients ignore background risks and exceptions to reimbursements (Tversky & Kahneman 1981 p. 456).

[15] These discussions include Bostic, Herrnstein, & Luce (1990) , Noussair, Robin, & Ruffieux (2004), and Poulton (1989). There is also extensive literature in the health domain (Stevens, McCabe, & Brazier 2007; Gold et al. 1996) and psychophysics (Gescheider 1997 Ch. 3).

more time to elicit preferences, requires a specification of range and initial values which generates biases, and it enhances the use of qualitative noncompensatory heuristics (lexicographic choice and misperception of dominance). Matching is harder for subjects to understand, as are its incentive compatible implementations. Further, the matching environment can lead subjects to ignore qualitative information and to resort to inappropriate arithmetical operations.

Prince avoids an important misperception of matching: subjects may misperceive matching as bargaining.[16] In Prince, with the choice situation (the price therein being one option) specified beforehand in an envelope held in hand, it is perfectly obvious that this price is not subject to bargaining or any other influence.

Several experimental economists have implemented more than one choice situation for real, which is acceptable if the distortions due to the income effect are smaller than other distortions.[17] A systematic study is Cox, Sadiraj, & Schmidt (2014a), which is close in spirit to our study in seeking to reduce distortions in the RIS. It considers alternative incentive systems that imply particular income effects, and investigates circumstances in which these income effects generate smaller distortions than the regular RIS does. Our study seeks to improve the RIS while avoiding any income effect, thus preserving incentive compatibility for homo economicus, rather than replacing RIS by another system with some income effect.

This paper focused on individual choice. Li, Turmunkh, & Wakker (2015) prepare an application in game theory. The envelopes for different players in a game are then chosen jointly.

---

[16] Because the link to the RCS is not clear in classical implementations, subjects think of what is closest to their everyday life, and this is probably bargaining. See Engelmann, & Hollard's (2010; trade uncertainty), Korobkin (2013 p. 1243), and Sayman & Öncüler (2005 §2.2); also see Bardsley et al. (2010 p. 273).

[17] Repeated payment is common in game and market experiments. In individual choice it is not very common, but still has been used in several studies, including Epper, Fehr-Duda, & Bruhin (2011) and Mosteller & Nogee (1951).

8 Pros and cons of Prince summarized

We have provided theoretical arguments for Prince showing (1) its internal validity (§2 and §7); (2) that it combines the pros of choice and matching, resolving a long-standing debate; (3) that it avoids the problem of strategic answering in adaptive experiments. We have also provided empirical arguments for Prince: (1) It induces highly consistent reporting (only one of 119 choice lists was inconsistent); (2) debriefings and discussions in pilots confirmed its transparencies; (3) it confirms well-established preference findings; (4) it reconciles choice and matching in 4 tests. The latter reconciliation implies that preferences are invariant across contexts, and are nonsuspect in the terminology of Bernheim & Rangel (2009). Although such a reconciliation need not be an improvement (Ariely, Loewenstein, and Prelec 2001), in view of the preceding arguments, we claim it is.

Direct comparisons with gold standards of true preference, or with best practice of measurement, would be desirable. Unfortunately, unlike many empirical domains, preference theory does not have a gold standard, and accordingly there is no consensus on which measurement method would be best practiced today.[18] Debates on prescriptive desirability of decision theories have therefore always been based on internal coherence criteria, such as (weakenings of) separability for risk and ambiguity, and dynamic consistency for dynamic decisions. In the terminology of Hammond (2006), we have to use coherence criteria rather than correspondence criteria. Thus, if a new measurement method's results are unlike a previous one's, then it is not easy to show that the difference is an improvement. Based on theoretical coherence criteria we conjecture that Prince comes closer to true preference values than other methods, but there is no way to prove this.[19] Given the importance of

---

[18] Introductions of new measurement methods in decision theory therefore can never refer to a gold standard (e.g., Holt & Laury 2002; Andreoni & Sprenger 2012).

[19] We know only one introduction of a new measurement method in decision theory that immediately included a comparison with another method (Attema et al. 2015), because the authors thought that the other method (Epper, Fehr-Duda, & Bruhin 2011) was clearly best practice at that moment. Holt & Laury (2002) did not provide such a comparative test. Andreoni, Kuhn, & Sprenger (2013) dedicated a separate paper to compare Andreoni & Sprenger's (2012) new method for measuring intertemporal preference with one alternative existing method. Their comparison is of limited interest to the readers who would have preferred another existing method for comparison.

preference measurement, for instance for consultancy and for providing inputs for policy decisions, there nevertheless exists an extensive literature on developing new methods, and our paper contributes to this literature.

We have argued that Prince improves internal validity. Investigations of external validity are highly desirable. Useful insights into the descriptive performance of Prince can be obtained by investigating out-of-sample predictive power (especially regarding real-life decisions), extensive consistency checks to assess noise, and manipulations of Prince with separate principles turned on and off, where Prince is compared with existing methods in these regards. Further, documented spillover effects and violations of isolation can be reduced using Prince. Given the size of this paper, showing that Prince can be implemented for virtually every preference measurement, we prefer to leave such desirable tests to future studies, where contributions by objective outsiders will be especially useful.

The main drawback of Prince is that it requires preparation: Envelopes with different choice situations have to be prepared for every session.

9 Conclusion

The Prince incentive system improves on the random incentive system, the Becker-DeGroot-Marschak system, and Bardsley's (2000) conditional information system. Our subjects understand that there is only one real choice situation: the one they hold in hand. Prince resolves or reduces: (a) violations of isolation; (b) misperceptions of bargaining; (c) strategic answering in adaptive experiments. Incentive compatibility is completely transparent to subjects. Hence there were virtually no irrational preference switches in choice lists. Prince reconciles choice and matching, combining the efficiency and precision of matching with an improved clarity and validity of choice. This resolves the classical preference reversals and reinvents matching. Prince avoids the major weakness of BDM: BDM randomizes choice options not choice situations, whereas Prince, by randomizing choice situations, leads subjects to condition properly.

Despite the absence of a gold standard for true preference, theoretical coherence arguments suggest the following conjectures: Prince provides more valid and transparent measurements of preferences without affecting those preferences themselves. The endowment effect and nonadditivity of subjective probabilities are genuine properties of preferences, entailing genuine deviations from classical principles. As with aversion to ambiguity, insensitivity is also real. Apparent preference reversals, to the contrary, are measurements errors. Utility is closer to linear than commonly thought, and decreasing absolute risk aversion is confirmed.

Many incentivized experimental measurements of preference or value can be improved using Prince. We used it for WTA, subjective probabilities (§4.6), utilities (§5), and ambiguity attitudes (§4.6). Prince sheds new light on which phenomena are to be incorporated in behavioral models.

Appendix A Procedures for the TO measurement

The x stimuli contain an outcome 0, at which utility is $-\infty$ for $\rho \le 0$. Because this outcome cancels from the equations, we still allow for power $\rho \le 0$ in our data fitting. Only five subjects had $\rho \le 0$ (minimum -0.75), and they do not affect any result.

A.1. Preceding studies and real incentives

Wakker & Deneffe (1996) introduced the TO method. If real incentives are implemented using the traditional BDM method and RIS, then subjects can answer strategically and may rationally overstate the values of r^1, r^2, and r^3. Hence most experiments using the method were hypothetical. Four exceptions are Abdellaoui (2000), Bleichrodt, Cillo, & Diecidue (2010), Schunk & Betsch (2006), and van de Kuilen & Wakker (2011) where the RIS was used with real incentives but subjects apparently did not notice the chance to answer strategically. The latter study used a strategy-check question to verify so (end of their §4). Toubias et al. (2013) also used the RIS in an adaptive experiment, but here it was, unlike with the TO method, impossible for subjects to notice the possibility to answer strategically.

We use Prince to implement real incentives. Then the RCS has been randomly selected for each subject before decision making. Subjects possess the RCS but with details remaining unknown. Subjects then obviously do not have any opportunity to influence the RCS and, furthermore, this is perfectly clear to them. Wang, Venkatesh, & Chatterjee (2007 p. 203) pointed out that providing the RCS prior to the experiment also rules out any illusion of strategic answering in a nonadaptive experiment.

A.2 Construction and use of envelopes for real incentives

For the 100 envelopes constructed before the experiment, we chose 12 types of envelopes, one for each question from TO1.1 – TO3.4. The envelope of type TOi.j (i = 1-3; j = 1-4) contains the two bets of Question TOi,j with values substituted as indicated in Table A2.1. The values $\overline{x}^0 = x^0 = 18$, $\overline{y}^0 = y^0 = 25$, and $\overline{z}^0 = z^0 = 210$ are as in Figure 5.2. For example, the envelopes of type TO1.2 results from Figure 5.3b and contains the bets $32_{\frac{2}{3}}0$ and $27_{\frac{2}{3}}9$. The only exception is type TO3.4, which contains $\bar{z}^{3\prime} = 3050$ instead of $z^3 = 342$ for r^3. The # numbers in Table A2.1 indicate how many of the 100 envelopes were of the particular type. For example, there were three envelopes of type TO1.2.

TABLE A2.1. The pre-set TO values

	$j=1$	$j=2$	$j=3$	$j=4$
TO1	$\bar{x}^1 = 23$ for x^1 (#33)	$\bar{x}^2 = 27$ for x^2 (#3)	$\bar{x}^3 = 32$ for x^3 (#3)	$\bar{x}^4 = 36$ for x^4 (#3)
TO2	$\bar{y}^1 = 46$ for y^1 (#33)	$\bar{y}^2 = 69$ for y^2 (#3)	$\bar{y}^3 = 92$ for y^3 (#3)	$\bar{y}^4 = 116$ for y^4 (#3)
TO3	$\bar{z}^1 = 253$ for z^1 (#9)	$\bar{z}^2 = 297$ for z^2 (#3)	$\bar{z}^3 = 342$ for z^3 (#3)	$\bar{z}^4 = 3100$ for z^4* (#1)

* In this choice question (Figure 5.2, TO3, $j=4$), we used $\bar{z}^{3\prime} = 3050$ (instead of $\bar{z}^3 = 342$) for z^3.

The stimuli of TO1 were used by Abdellaoui (2000) and those of TO2 were used by Booij, van Praag, & van de Kuilen (2010), both scaled up. Those of TO3 were not used before. They serve to study high outcomes.

Subjects were informed that: (a) at least one of every ten of them would play for real; (b) if playing for real, the average gain under random (“blind”) choosing is 53.27; (c) at least one prize exceeded 3000. They were told that the types of envelope did not occur equally often, which should be obvious with 100 envelopes of 12 types.

In other respects the organization was as with the Prince experiments reported before. For example, the questionnaire asked subjects to give us instructions about which option from their own envelope to give them at the end of the experiment. One more difference was as follows. In the previous experiments we demonstrated how Prince can be used with a regular performance-contingent real payment for every subject. We now implement Prince with performance-contingent real payments only for some subjects, but then having these payments large. Abdellaoui et al. (2011, Web Appendix) suggested that such payment schemes work best to motivate subjects. For this purpose, at the end we collected the three front pages numbered TOj.0 of each subject that contained their answers, whereas the subjects kept the rest of their questionnaires that also contained their answers.[20] The front pages of each subject were folded together so as to be unrecognizable, and were put in an opaque case and shuffled. Then we let one subject draw some (three in the first session and six in the second) triples of questionnaires from this case, after which the corresponding subjects played for real.

A.3 Maximizing overlap with experimental questions

The numbers x^j for TO1 in Table A2.1 result from expected value maximization, and those for TO2 and TO3 (except TO3.4) result from Tversky & Kahneman's (1992) prospect theory and their parameters. We rounded the x^js. We constructed the 100 envelopes, in particular the frequency of each type of envelope among the 100 envelopes, so as to generate an appropriate overall expected value of the game for the subjects.

We chose most envelopes of type TOj.1 because here three rather than two outcomes were known beforehand. Hence, each answer x^1 by the subject implied a choice from the corresponding envelope, depending on whether x^1 exceeded the corresponding value from the envelope or not.[21] The same observations hold for y^1 and z^1. We thus obtain a high overlap between experimental answers and RCSs.

In the explanations given at the beginning of the experiment an example was used that occurred in a pilot: In TO1.1 a subject in a pilot had answered $x^1 = 50$, implying an indifference $10_{\frac{1}{2}}8 \sim 50_{\frac{1}{2}}1$. Then, if the choice in the envelope had been between $10_{\frac{1}{2}}8$ and $32_{\frac{1}{2}}1$, the former, $10_{\frac{1}{2}}8$, would be given to the subject. We explained that the subject's preference would only be reinforced if outcome 10 in his preferred bet had been increased to 24, and that he would want the former even more from the pair $\{24_{\frac{1}{2}}8, 32_{\frac{1}{2}}1\}$. This pair was actually contained in the envelope of this subject, who indeed received.$24_{\frac{1}{2}}8$. In general, if the value $\overline{r}^j$ in the envelope exceeded the answer r^j given by the subject *and the value* $\overline{j^{i-1}}$ in the envelope *was below* rj^{i-1} (reinforcing

---

[20] The subjects filled out all answers twice: once on the page containing the question, and once on the front page numbered TOj.0.

[21] If indifference, subjects could choose on the spot, as always if the experimental answers did not specify a choice.

the preference), then we would give the bet with outcome r^j to the subject.

Conversely, if the value $\overline{r^j}$ in the envelope was below the answer r^j given by the subject *and the value* $\overline{j^{i-1}}$ *in the envelope exceeded* j^{i-1} (reinforcing the preference), then we would give the lottery with outcome r^{j-1} to the subject. Thus the answers given during the experiment pertained to many possible envelopes, and this was explained to the subjects.

A.4 Verifiable absence of deception

As in the previous experiments, subjects could verify that there was no deception. Again, at the beginning they verified, through sets of 10 numbered envelopes, that all numbered envelopes were present. Subjects collected, shuffled, and selected envelopes from bags themselves. At the end, when subjects who had been selected for payment came to the front of the room, a list describing the content of all 100 envelopes was handed out to the other subjects, with calculations showing that our information about expected value under random play and maximal amounts was correct. These subjects were asked to open their envelope and verify that the list correctly described its content. For the subjects who were in front of the class, everyone in the class could see that the description of their envelopes was correct. Similar procedures were followed in all experiments described in this paper.

One of the experimenters carried over €3000 in cash with him, showing to the subjects that this amount could and would be paid on the spot if the lucky case arose. The random selection of at least one per 10 subjects to play for real (three out of 25 in the first session and 6 out of 55 in the second) was done by letting the subjects put their questionnaires in an opaque bag, after which the questionnaires to be implemented for real were randomly selected by a subject. Lotteries were generated using a six-sided die, thrown by a subject selected for this role by the subject who was playing for real.

Web appendix

The web appendix can be downloaded here:
http://people.few.eur.nl/wakker/pdf/prince_web_appendix.pdf

References

Abdellaoui, Mohammed (2000) "Parameter-Free Elicitation of Utility and Probability Weighting Functions," *Management Science* 46, 1497–1512.

Abdellaoui, Mohammed, Aurélien Baillon, Laetitia Placido, & Peter P. Wakker (2011) "The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation," *American Economic Review* 101, 695–723.

Andreoni, James & Charles Sprenger (2012) "Estimating Time Preference from Convex Budgets," *American Economic Review* 3333–3356.

Andreoni, James, Michael A. Kuhn, & Charles Sprenger (2013) "On Measuring Time Preferences," Working Paper 19392, http://www.nber.org/papers/w19392, National Bureau of Economic Research, Cambridge, MA 02138, August 2013.

Ariely, Dan, George F. Loewenstein, & Drazen Prelec (2001) " 'Coherent Arbitrariness': Stable Demand Curves without Stable Preferences," *Quarterly Journal of Economics* 118, 73–106.

Arrow, Kenneth J., Robert M. Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, & Howard Schuman (1993) "Report of the NOAA Panel on Contingent Valuation," *Federal Register* 58, 4602–4614.

Attema, Arthur E. & Werner B.F. Brouwer (2013) "In Search of a Preferred Preference Elicitation Method: A Test of the Internal Consistency of Choice and Matching Tasks," *Journal of Economic Psychology* 39, 126–140.

Attema, Arthur E., Han Bleichrodt, Yu Gao, Zhenxing Huang, & Peter P. Wakker (2015) "Measuring Discounting without Measuring Utility," mimeo, Erasmus School of Economics, Erasmus University, Rotterdam, the Netherlands.

Baillon, Aurélien & Han Bleichrodt (2014) "Testing Ambiguity Models through the Measurement of Probabilities for Gains and Losses," *American Economic Journal: Microeconomics*, forthcoming.

Bardsley, Nick (2000) “Control without Deception: Individual Behavior in Free-Riding Experiments Revisited,” *Experimental Economics* 3, 215–240.

Bardsley, Nick, Robin P., Cubitt, Graham Loomes, Peter Moffat, Chris Starmer, & Robert Sugden (2010) “*Experimental Economics; Rethinking the Rules*.” Princeton University Press, Princeton, NJ.

Becker, Gordon M., Morris H. de Groot, & Jacob Marschak (1964) “Measuring Utility by a Single-Response Sequential Method,” *Behavioral Science* 9, 226–232.

Bellemare, Charles, Sabine Kroger, & Arthur van Soest (2008) “Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities,” *Econometrica* 76, 815–839.

Bernheim, B. Douglas & Antonio Rangel (2009) “Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics,” *Quarterly Journal of Economics* 124, 51–104.

Birnbaum, Michael H. (1992) “Should Contextual Effects in Human Judgment Be Avoided?,” Book Review of: E. Christopher Poulton (1989) “Bias in Quantifying Judgments,” Erlbaum, Hillsdale NJ; *Contemporary Psychology* 37, 21–23.

Bleichrodt, Han, Alessandra Cillo, & Enrico Diecidue (2010) “A Quantitative Measurement of Regret Theory,” *Management Science* 56, 161–175.

Bleichrodt, Han, Martin Filko, Amit Kothiyal, & Peter P. Wakker (2013) “Making Case-Based Decision Theory Directly Observable,” mimeo.

Bohnet, Iris, Fiona Greig, Benedikt Herrmann, & Richard Zeckhauser (2008) “Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States,” *American Economic Review* 98, 294–310.

Booij, Adam S., Bernard M.S. Van Praag, & Gijs van de Kuilen (2010) “A Parametric Analysis of Prospect Theory's Functionals,” *Theory and Decision* 68, 115–148.

Bosman, Ronald & Frans van Winden (2010) “Global Risk, Investment and Emotions,” *Economica* 77, 451–471.

Bostic, Raphael, Richard J. Herrnstein, & R. Duncan Luce (1990) “The Effect on the Preference-Reversal Phenomenon of Using Choice Indifferences,” *Journal of Economic Behavior and Organization* 13, 193–212.

Brosnan, Sarah F., Owen D. Jones, Molly Gardner, Susan P. Lambeth, & Steven J. Schapiro (2012) “Evolution and the Expression of Biases: Situational Value

Changes the Endowment Effect in Chimpanzees," *Evolution and Human Behavior* 33, 378 –386.

Bruhin, Adrian, Helga Fehr-Duda, & Thomas Epper (2010) "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion," *Econometrica* 78, 1375–1412.

Brun, Wibecke & Karl H. Teigen (1990) "Prediction and Postdiction Preferences in Guessing," *Journal of Behavioral Decision Making* 3, 17–28.

Camerer, Colin F. (1989) "An Experimental Test of Several Generalized Utility Theories," *Journal of Risk and Uncertainty* 2, 61–104.

Camerer, Colin F. (1995) "Individual Decision Making." *In* John H. Kagel & Alvin E. Roth (eds.) *Handbook of Experimental Economics*, 587–703, Princeton University Press, Princeton, NJ.

Cox, James C., Vjollca Sadiraj, & Ulrich Schmidt (2014a) "Paradoxes and Mechanisms for Choice under Risk," *Experimental Economics*, forthcoming.

Cox, James C., Vjollca Sadiraj, & Ulrich Schmidt (2014b) "Asymmetrically Dominated Choice Problems, the Isolation Hypothesis and Random Incentive Mechanisms," *PLoS ONE* 9, e90742..

Cubitt, Robin P., Chris Starmer, & Robert Sugden (1998) "Dynamic Choice and the Common Ratio Effect: An Experimental Investigation," *Economic Journal* 108, 1362–1380.

Dillenberger, David (2010) "Preferences for One-Shot Resolution of Uncertainty and Allais-Type Behavior," *Econometrica* 78, 1973–2004.

Ding, Min (2007), "An Incentive-Aligned Mechanism for Conjoint Analysis," *Journal of Marketing Research* 44, 214-223.

Engelmann, Dirk & Guillaume Hollard (2010) "Reconsidering the Effect of Market Experience on the 'Endowment Effect'," *Econometrica* 78, 2005–2019.

Epper, Thomas, Helga Fehr-Duda, & Adrian Bruhin (2011) "Viewing the Future through a Warped Lens: Why uncertainty Generates Hyperbolic Discounting," *Journal of Risk and Uncertainty* 43, 163–203.

Fechner, Gustav Th. (1860) "*Elemente der Psychophysik.*" Von Breitkopf und Härtel, Leipzig.

Fischer, Gregory W., Ziv Carmon, Dan Ariely, & Gal Zauberman (1999) "Goal-Based Construction of Preferences: Task Goals and the Prominence Effect," *Management Science* 45, 1057–1075.

Gescheider, George A. (1997) "*Psychophysics: the Fundamentals"* (3rd edn.). Lawrence Erlbaum Associates.

Gilboa, Itzhak & David Schmeidler (2001) "*A Theory of Case-Based Decisions."* Cambridge University Press, Cambridge.

Gold, Marthe R., Joanna E. Siegel, Louise B. Russell, & Milton C. Weinstein (1996) "*Cost-Effectiveness in Health and Medicine."* Oxford University Press, New York.

Gollier, Christian (2001) "Wealth Inequality and Asset Pricing," *Review of Economic Studies* 68, 181–203.

Grant, Simon, Atsushi Kajii, & Ben Polak (2000) "Temporal Resolution of Uncertainty and Recursive Non-Expected Utility Models," *Econometrica* 68, 425–434.

Grether, David M. & Charles R. Plott (1979) "Economic Theory of Choice and the Preference Reversal Phenomenon," *American Economic Review* 69, 623–638.

Hammond, Kenneth R. (2006) "*Beyond Rationality*." Oxford University Press, New York.

Hao, Li & Daniel Houser (2012) "Belief Elicitation in the Presence of Naïve Respondents: An Experimental Study," *Journal of Risk and Uncertainty* 44, 149–160.

Heath, Chip & Amos Tversky (1991) "Preference and Belief: Ambiguity and Competence in Choice under Uncertainty," *Journal of Risk and Uncertainty* 4, 5–28.

Holt, Charles A. (1986) "Preference Reversals and the Independence Axiom," *American Economic Review* 76, 508–513.

Holt, Charles A. & Susan K. Laury (2002) "Risk Aversion and Incentive Effects," *American Economic Review* 92, 1644–1655.

Kahneman, Daniel, Jack L. Knetsch, & Richard H. Thaler (1990) "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy* 98, 1325–1348.

Karni, Edi (2009) "A Mechanism Design for Probability Elicitation," *Econometrica* 77, 603 –606.

Karni, Edi & Zvi Safra (1987) "Preference Reversal and the Observability of Preferences by Experimental Methods," *Econometrica* 55, 675–685.

Keren, Gideon B. (1991) “Calibration and Probability Judgments: Conceptual and Methodological Issues,” *Acta Psychologica* 77, 217–273.

Korobkin, Russell (2003) “The Endowment Effect and Legal Analysis,” *Northwestern University Law Review* 97, 1227–1293.

Kreps, David M. & Evan L. Porteus (1979) “Dynamic Choice Theory and Dynamic Programming,” *Econometrica* 47, 91–100.

Li, Chen, Uyanga Turmunkh, & Peter P. Wakker (2015), in preparation.

Lichtenstein, Sarah & Paul Slovic (1971) “Reversals of Preference between Bids and Choices in Gambling Decisions,” *Journal of Experimental Psychology* 89, 46–55.

Loomes, Graham, Chris Starmer, & Robert Sugden (1989) “Preference Reversal: Information-Processing Effect of Rational Non-Transitive Choice?,” *Economic Journal* 99, Supplement, 140–151.

Loomes, Graham & Robert Sugden (1982) “Regret Theory: An Alternative Theory of Rational Choice under Uncertainty,” *Economic Journal* 92, 805–824.

Mandelkow, Karl Robert (1968; ed) “*Goethes Briefe*.” Hamburg: Wegner, 1968.

Mosteller, Frederick & Philip Nogee (1951) “An Experimental Measurement of Utility,” *Journal of Political Economy* 59, 371–404.

Ogaki, Masao & Qiang Zhang (2001) “Decreasing Relative Risk Aversion and Tests of Risk Sharing,” *Econometrica* 69, 515–526.

Plott, Charles R. & Kathryn Zeiler (2005) “The Willingness to Pay-Willingness to Accept Gap, the “Endowment Effect,” Subject Misconceptions, and Experimental Procedures for Eliciting Valuations,” *American Economic Review* 95, 530–545.

Poulton, E. Christopher (1989) “*Bias in Quantifying Judgments.*” Erlbaum, Hillsdale NJ.

Rothbart, Myron & Mark Snyder (1970) “Confidence in the Prediction and Postdiction of an Uncertain Outcome,” *Canadian Journal of Behavioral Science* 2, 38–43.

Savage, Leonard J. (1954) “*The Foundations of Statistics.*” Wiley, New York. (2nd edn. 1972, Dover Publications, New York.)

Sayman, Serdar & Ayse Öncüler (2005) “Effects of Study Design Characteristics on the WTA-WTP Disparity: A Meta Analytic Framework,” *Journal of Economic Psychology* 26, 289–312.

Schade, Christian, Howard Kunreuther, & Philipp Koellinger (2012) "Protecting against Low-Probability Disasters: The Role of Worry," *Journal of Behavioral Decision Making* 25, 534–543.

Schmidt, Ulrich & John D. Hey (2004) "Are Preference Reversals Errors? An Experimental Investigation," *Journal of Risk and Uncertainty* 29, 207–218.

Schmidt, Ulrich & Stefan Traub (2009) "An Experimental Investigation of the Disparity Between WTA and WTP for Lotteries," *Theory and Decision* 66, 229–262.

Schunk, Daniel & Cornelia Betsch (2006) "Explaining Heterogeneity in Utility Functions by Individual Differences in Decision Modes," *Journal of Economic Psychology* 27, 386–401.

Seidl, Christian (2002) "Preference Reversal," *Journal of Economic Surveys* 16, 621–655.

Shafir, Eldar & Amos Tversky (1992) "Thinking through Uncertainty: Nonconsequential Reasoning and Choice," *Cognitive Psychology* 24, 449–474.

Starmer, Chris & Robert Sugden (1991) "Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation," *American Economic Review* 81, 971–978.

Stefan, Simona & Daniel David (2013) "Recent Developments in the Experimental Investigation of the Illusion of Control. A Meta-Analytic Review," *Journal of Applied Social Psychology* 43, 377–386.

Stevens, Katherine J., Christopher J. McCabe, & John E. Brazier (2007) "Multi-Attribute Utility Function or Statistical Inference Models: A Comparison of Health State Valuation Models Using the HUI2 Health State Classification System," *Journal of Health Economics* 26, 992–1002.

Strzalecki, Tomasz (2013) "Temporal Resolution of Uncertainty and Recursive Models of Ambiguity Aversion," *Econometrica* 81, 1039–1074.

Thaler, Richard H. & Cass R. Sunstein (2008) "*Nudge: Improving Decisions About Health, Wealth, and Happiness*." Yale University Press, New Haven.

Trautmann, Stefan T. & Gijs van de Kuilen (2014) "Ambiguity Attitudes." *In* Gideon Keren & George Wu (eds.), *Handbook of Judgment and Decision Making*, forthcoming, Blackwell, Oxford.

Toubias, Olivier, Eric Johnson, Theodoros Evgeniou, & Philippe Delquié (2013), "Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters," *Management Science* 59, 613–640.

Tversky, Amos & Craig R. Fox (1995) "Weighing Risk and Uncertainty," *Psychological Review* 102, 269–283.

Tversky, Amos & Daniel Kahneman (1981) "The Framing of Decisions and the Psychology of Choice," *Science* 211, 453–458.

Tversky, Amos & Daniel Kahneman (1992) "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5, 297–323.

Tversky, Amos, Paul Slovic, & Daniel Kahneman (1990) "The Causes of Preference Reversal," *American Economic Review* 80, 204–217.

van de Kuilen, Gijs & Peter P. Wakker (2011) "The Midweight Method to Measure Attitudes toward Risk and Ambiguity," *Management Science* 57, 582–598.

von Békésy, Georg (1947) "A New Audiometer," *Acta Otolaryngology* 35, 411–422.

Wakker, Peter P. (2010) "*Prospect Theory: For Risk and Ambiguity*." Cambridge University Press, Cambridge, UK.

Wakker, Peter P. & Daniel Deneffe (1996) "Eliciting von Neumann-Morgenstern Utilities when Probabilities Are Distorted or Unknown," *Management Science* 42, 1131–1150.

Wang, Stephanie W., Michelle Filiba, & Colin F. Camerer (2010), "Dynamically Optimized Sequential Experimentation (DOSE) for Estimating Economic Preference Parameters," working paper, California Institute of Technology.

Wang, Tuo, Ramaswamy Venkatesh, & Rabikar Chatterjee (2007) "Reservation Price as a Range: An Incentive-Compatible Measurement Approach," *Journal of Marketing Research* 64, 200–213.

Weber, Roberto A., Colin F. Camerer, & Marc Knez (2004) "Timing and Virtual Observability in Ultimatum Bargaining and "Weak Link" Coordination Games," *Experimental Economics* 7, 25–48.

Wilcox Nathaniel T. (2011) " 'Stochastically More Risk Averse:' A Contextual Theory of Stochastic Discrete Choice under Risk," *Journal of Econometrics* 162, 89–104.

Zeisberger, Stefan, Dennis Vrecko, & Thomas Langer (2012) "Measuring the Time Stability of Prospect Theory Preferences," *Theory and Decision* 72, 359–386.