PRINCE: AN IMPROVED METHOD FOR MEASURING INCENTIVIZED PREFERENCES

CATHLEEN JOHNSON, Philosophy, Politics, Economics and Law program, Univ of Arizona, Tucson, AZ, USA

AURÉLIEN BAILLON, Erasmus Sch. Econ., Erasmus Univ. Rotterdam, the Netherlands

HAN BLEICHRODT, Erasmus Sch. Econ., Erasmus Univ. Rotterdam, the Netherlands

ZHIHUA LI, Birmingham Business Sch.; Univ. of Birmingham; JG Smith Building, Birmingham B15 2TT, UK

DENNIE VAN DOLDER, Sch. of Business and Econ., VU Amsterdam, Amsterdam, the Netherlands

PETER P. WAKKER (corresp. author), Erasmus Sch. Econ., Erasmus Univ. Rotterdam, 3000 DR, the Netherlands, 31.10.408.12.65 (O), 31.10.408.91.62 (F), wakker@ese.eur.nl

February 2019

This paper introduces the Prince incentive system for measuring preferences. Prince is a variation of the random incentive system that enhances isolation and makes incentive compatibility more transparent to subjects. It allows for the precise and direct elicitation of indifference values as with matching while having the clarity and validity of choice lists. Prince avoids the opaqueness of Becker-DeGroot-Marschak's mechanism and precludes strategic behavior in adaptive experiments. Using Prince, we shed new light on willingness to accept and the major components of decision under uncertainty: utilities, subjective beliefs, and ambiguity attitudes. Prince outperforms a classical implementation of the random incentive system.

JEl-Classification: C91; D81

Keywords: incentive compatibility, random incentive system, BDM, choice list, matching

1. INTRODUCTION AND BACKGROUND

Behavioral economics challenges the classical revealed preference paradigm in economics. Many of the challenges were handled by incorporating irrationalities in decision models, as for instance in Tversky and Kahneman's (1992) prospect theory. Preference reversals, revealing systematic differences between choice and maching,[1] entailed a more fundamental challenge. They casted doubt on the basic concept of preference. Although some researchers blamed choice-based procedures for preference reversals (Fischer et al. 1999), most researchers nowadays prefer choice to matching, following the recommendations by Arrow et al. (1993) and others. Binary choices are not without drawbacks, though: they take more time to administer, give interval rather than point estimates, and have their own biases.[2] For this reason, some recent studies revealed choices from linear budget sets, an intermediate between binary choice and matching (Choi et al. 2007; Epper and Fehr-Duda 2015; Miao and Zhong 2015). The present paper introduces a new incentive system to measure preferences that combines the greater validity of choice with the greater efficiency and precision of matching. It reconciles the two.

A pervasive difficulty in economic experiments is that real incentives as implemented in the laboratory are often hard to understand for subjects. This problem is greatest for matching, where the Becker-DeGroot-Marschak (1964) mechanism (BDM) has often been criticized for this reason.[3] 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. 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

---

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

[2] 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").

[3] See, for instance, Bardsley et al. (2010 p. 271 ff.), Cason and Plott (2014), Noussair, Robin, and Ruffieux (2004), and Seidl (2002 p. 630 ff.).

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 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, with the one envelope with the one real choice situation in hand, isolation is maximally enhanced.

Prince makes it possible to combine the tractability and precision of matching with the (improved) clarity and validity of binary choice. 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—strategic answering is impossible, and this is patently clear for any subject who might think of it. We thus resolve the incentive compatibility problem for adaptive experiments.

We apply Prince to incentivize Wakker and Deneffe's (1996) adaptive tradeoff method (TO) for measuring utility. This method provides parameter-free measurements of utility for expected utility that remain valid if expected utility is violated. It, thus, also provides the right utility function under prospect theory without requiring knowledge of the probability weighting function. However, as yet the method could not be implemented in an incentive compatible manner. Using Prince this becomes possible, so that the method becomes available to experimental economics. We can now measure risky utility in a parameter-free manner that, unlike other existing methods (Holt and Laury 2002), is not distorted by the extensive violations of expected utility that have been found empirically (Starmer 2000). Finally, not only does our experiment avoid deception, but nondeception is verifiably transparent to subjects.

For non-experimentalists, our improved measurements of preferences shed new light on general economic concepts. In particular, all our findings support the modern behavioral deviations from classical models. 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 2015). Third, Prince helps to clarify which known choice anomalies reflect genuine deviations from homo economicus and which are artifacts of measurement problems. Our tests suggest that preference reversals are due to measurement problems, and that they do not reflect genuine intransitivities. Prince does not, however, remove the endowment effect. This suggests that the endowment effect reflects genuine preference (Anagol, Balasubramaniam, and Ramadorai 2018; Rees-Jones 2018) and, hence, a robust discrepancy between homo sapiens and homo economicus, and that it is no mere artifact caused by measurement problems. On this point we deviate from Plott and Zeiler (2005). We emphasize that Prince primarily aims to improve measurements of preferences, rather than the quality (rationality) of preferences themselves, although improvement of preference may occur to some extent as a secondary effect.

This paper is organized as follows. Section 2 defines Prince. Then, to show its general applicability, the following sections implement it in various experiments, involving 251 subjects in total. For brevity reasons, each experiment is described concisely in the main text, with details in the online appendix (110 pages). Section 3 implements Prince in a small experiment, showing how it combines the pros of matching and choice. A large experiment, with many stimuli and measurements, is in §4. Section 5 shows how Prince solves the problem of manipulation in an adaptive utility measurement. A comparison of Prince with a traditional RIS is in Section 6. Section 7 is on preceding incentive systems that share components with Prince, such as using envelopes to specify choice situations prior to the experiment. A general discussion is in §§8-9, followed by a conclusion. We end with a link to an Online Appendix, referred to as Appendix O in the main text, which gives all stimuli and provides additional analyses.

2. PRINCE EXPLAINED

This section introduces the Prince system. We explain its principles in the first two subsections, and define them formally in §2.C. Further discussion is in §§8-9.

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). This envelope is given to subjects at the beginning of the experiment, and remains sealed until the end. The RCS describes a number of choice options (two in our experiments; for 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 during the experiment, she does receive some information on the potential choice situations (such as average or range of outcomes employed) beforehand. 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 unambiguously describes the entire RCS, with *all* choice options specified. For instance, in implementations of BDM using prior envelopes, only one random prize is commonly specified in the envelope and, thus, not the entire choice situation. Other studies specified a number of the choice situation in the envelope, which leaves ambiguity to subjects about which RCS corresponds with their number. Further discussion and references are in §8.

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 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. This way, 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, held in their hand, 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. The solution is simple: subjects may then choose 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 this overlap problem, experimenters must frame the partial information about the RCS by anticipating the range of possible choice situations generated in the experiment. They do this by using descriptive theories and pilots. For example, choice situations with very large monetary amounts could arise in our adaptive experiment (§5), depending on subjects' answers. We hence informed subjects about the existence of a large possible outcome (> €3000).

2.3. Prince Summarized

We 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 completely and unambiguously describes the entire RCS

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

(5) [No indeterminacy] If none of the instructions that subjects give during the experiment that pertain to the RCS, then they can choose 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.

While parts of Prince have been used before (§7), their integration into Prince is new. Their combination achieves incentive compatibility in a transparent manner, involving proper conditioning on the RCS. For example, all BDM implementations that we are aware of violate Principle 4 [entirety], which can lead subjects to condition in the wrong way, enhancing rather than avoiding meta-lottery perceptions (see §8). Such mistakes hamper BDM’s internal validity and arguably account for its bad performance.

In the experiments in this paper, not only the prior envelopes, but all stimuli are physical, including the implementations of lotteries, but this is not essential for Prince. Other researchers may prefer computerized implementations. The physical availability of the RCS to every subject such as in a prior envelope is essential though for Prince, 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. They could always verify the correctness of the information provided about the stimuli. Second, unlike computer randomizations, our physically generated randomizations were also fully verifiable and were carried out by the subjects themselves.

3. EXPERIMENT 1: REINVENTING MATCHING (WTA)

Experiment 1 illustrates how to implement Prince in a small experiment that uses a matching question to elicit 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. 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 Erasmus School of Economics, 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 OD.4 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 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!" The question in Figure 1 is called 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, observed the real choice situation specified in the envelope, and followed 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 either way (Birnbaum 1992). We chose the specific range to facilitate comparability with choice lists presented later.

FIGURE 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 (Appendix OH). 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, and randomly distributed without replacement.

The two groups of subjects received different versions of the first question, *1-match* or *1-choice* (see Figures 2 and 3), which were variants of the question used in Experiment 1. These questions were part of a between-subject test in this large experiment. To facilitate comparison with Experiment 1, these questions were always asked at the start of Experiment 2 (prior to the other questions in this larger experiment). 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 the 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 (Figure 2) measured subjects' WTA for a mug, however, now without endowment. It was asked of 41 of the 80 subjects. Results and discussion are at the end of §4.3.

FIGURE 2. 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 (Figure 3) repeats Question 1-match, again with no endowment, but now using choice lists instead of matching, for the remaining 39 subjects. 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 3. 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 this analysis. Leaving her in would not alter our results. Table 1 reports summary statistics, and Table 2 reports tests.

TABLE 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	38	3.61	2.54

TABLE 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	–0.42	–0.82	77	0.41

DISCUSSION. Prince confirms the endowment effect because WTA with endowment exceeds WTA without endowment.[4] This suggests that the endowment effect, rational

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

or not, reflects a genuine property of preference (Brosnan et al. 2012; Korobkin 2003 p. 1244), and not a bias in measurement.

Prince shows no difference 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 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. The fact that we obtain a highly significant results for the endowment effect suggests that power is sufficient. Furthermore, in §4.4 we confirm our finding in a higher-powered within-subject test for all 80 subjects.

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, and 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 Appendix OB. 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 and Slovic (1971). Details are in Appendix OA. For Question 4, the choice question, we used an analog of Figure 2. The only difference was that we now omitted the complete description of how many envelopes contained particular values 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 and Slovic 1971). We also measured their cash equivalents in Questions 5 and 6, again using analogs of Figure 2, 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, this 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 ($p = 0.37$) and are infrequent enough to be explained as random choice inconsistencies (Schmidt and 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 thus 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, and Kahneman 1990). Prince restores consistency between choice and matching, thus resolving preference reversals.

4.6. Measuring Subjective Probabilities and Ambiguity Attitudes

The RIS has been especially criticized in the study of ambiguity attitudes (Bade 2015; Oechssler and Roomets 2014), the topic of this section. Using questions 7, 8, and 9 we replicate the measurements of subjective probabilities and ambiguity attitudes by Baillon and Bleichrodt (2015 Study 1). They used classical choice lists, whereas we use Prince and matching. Details are in Appendix OC. 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 does not decrease by more than 0.5% during the experiment.

Our presentation of questions was similar to Figure 2, with option 1 being 10_E0 and option 2 being 10_p0, requesting that a threshold for p (instead of x) be specified. The desirability to reckon with violations of classical Bayesianism in such belief measurements had been widely felt (e.g., Costa-Gomes and Weizsäcker 2008 p. 731). Baillon and Bleichrodt (2015) showed how to do so, using a nonadditivity index $m(A) + m(B) - m(A \cup B)$ to capture ambiguity attitudes. We replicated all their findings. In particular, the nonadditivity index was mostly positive, rejecting expected utility, and confirming a(mbiguity-generated likelihood)-insensitivity (Trautmann and van de Kuilen 2015). 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 (point estimate instead of interval estimate) than preceding papers did.

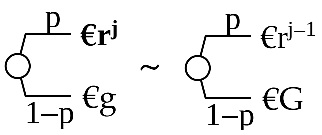
5. EXPERIMENT 3: PRINCE IMPLEMENTED IN AN ADAPTIVE EXPERIMENT; MEASURING UTILITY

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 Appendix OH. We first piloted the following procedures in two sessions, each with about 10 graduate students who had had considerable exposure to decision theory. After the pilot, as an assignment, they were tasked with criticizing the procedures, especially concerning possible deception by the experimenter or the subjects. They did not find weaknesses.[5] These students, as well as colleagues in informal pilots, confirmed procedural transparency and absence of biases.

We use Wakker and Deneffe's (1996) adaptive tradeoff (TO) method to measure utility. This method is robust to violations of expected utility and provides a correct utility function for most nonexpected utility theories. Dimmock et al. (2018) used it in a nationally-representative sample of several thousand respondents in the American Life Panel (ALP). Implementations so far were not incentive-compatible. Prince makes this method available for economists by allowing for proper incentivization.

5.1. The Preferences to be Elicited for the Tradeoff Method

FIGURE 4. Tradeoff method indifferences



We measure indifferences $\mathbf{r^j}_p g \sim r^{j-1}{}_p G$, j = 1, …,4 (Figure 4, with the usual notation for lotteries). Superscripts indicate the sequence of outcomes r^j. 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 values r^j, after having been elicited, serve as input to the next question. We assume a *weighted utility* model:

$$\text{for } x \geq y,\ x_p y \text{ is evaluated by } \pi U(x) + \rho U(y)\ (\pi > 0,\ \rho > 0). \tag{1}$$

This model includes expected utility, prospect theory for gains (Tversky and Kahneman 1992), and most other generalizations of expected utility (Miyamoto 1988; Wakker 2010 §7.11). Eq. 1 implies that the r^js are equally spaced in utility units (Wakker 2010 §4.3, §7.11, §10.6):

$$U(r^4) - U(r^3) = U(r^3) - U(r^2) = U(r^2) - U(r^1) = U(r^1) - U(r^0). \tag{2}$$

[5] A humorous suggestions was: "pull the fire alarm just when you have to pay €3000."

A nonparametric measurement of utility can be derived (§5.5, §5.6) that is valid for most risky choice theories, using Basu's (1982) derivation of uniqueness. The observations can, of course, also be used for parametric fitting (Appendix OD.5 and OD.6). The TO method avoids collinearity between utility U and probability weighting (π and ρ in Eq. 1): Eq. 2 is not affected by the probability weights π and ρ, and these do not even need to be estimated. For other measurements of prospect theory in the literature, collinearity is a serious problem (demonstrated by Zeisberger, Vrecko, and Langer 2012; p. 366 ff.). For a sophisticated recent measurement, see Bruhin, Fehr-Duda, and Epper (2010).

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. Wakker and Deneffe (1996) used the same stimuli but scaled up, and their choices were hypothetical.

Figure 6 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 subject 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.

5.2. Procedure and Real Incentives

We used Prince in a one-hour pen and paper session in a classroom. We conducted two sessions, one with 25 and one with 55 subjects. Subjects were undergraduate students of Erasmus University 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 explanations, accompanied by an explanatory PowerPoint presentation.

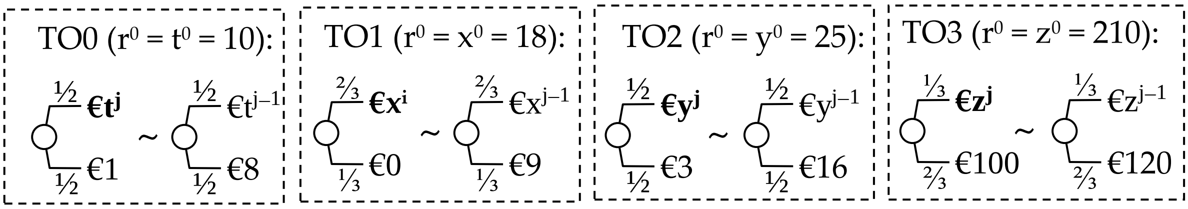
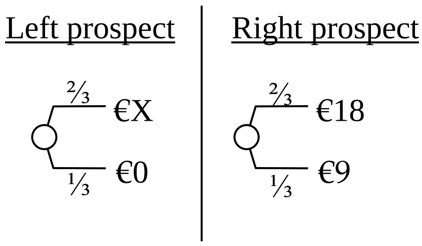


FIGURE 5. The values used for TO0-TO3; $j = 1, \ldots, 4$

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.

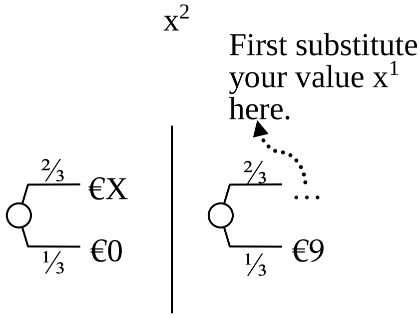
For small values of X you prefer the right prospect.
For large values of X you prefer the left prospect.
For some value of X, which we call x^1, your preference switches from left to right.
Fill this switching value in below, and then on page TO1.0.

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 6. Figures used in the tradeoff method

Subjects filled out the training questions of TO0, jointly and simultaneously, exactly as in Wakker and 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. Their expected payoff if playing randomly (but subjects could of course do better), was €58.27. Under random play, the expected payoff over all 80 subjects was then in total €10.99, in agreement with common policies of sufficient saliency of real incentives.

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 lotteries 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. The details depend on particularities of the experiment, and are in Appendixes OD.2 and OD.3.

5.4. Experiment with Hypothetical Choice

Besides the aforementioned sessions, we also conducted two sessions with hypothetical choices, one with 10 and one with 44 subjects. Subjects were unaware that other subjects played for real incentives. There was of course no role for Prince here. 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. Analysis

As all methods, the TO method can be used for parametric analyses. An advantage is that it can also be used for nonparametric analyses, i.e., without a commitment to any family of, or any shape of, utility functions (Wakker 2010 §9.4.2). Because this is a novelty of the method, we report it here. Parametric analyses are in Appendix OD.5 and OD.6.

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

$$r^{i+2} - r^{i+1} > r^{i+1} - r^{i} \quad (3)$$

for all i, and for strictly convex utility we have

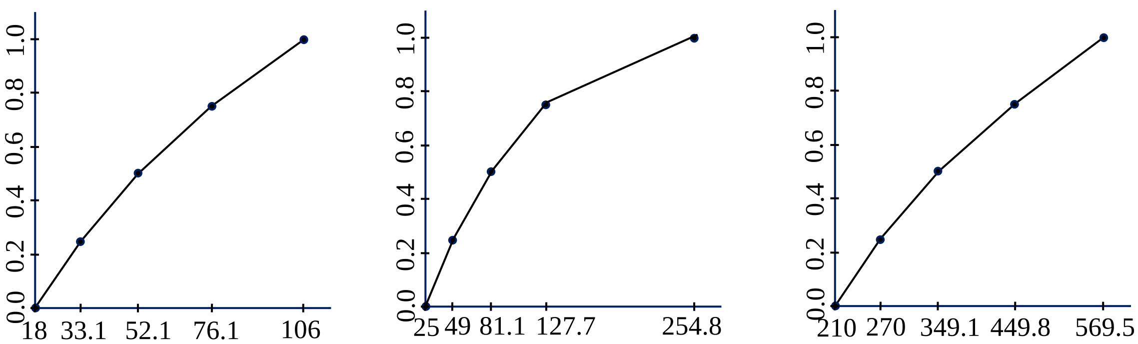
$$r^{i+2} - r^{i+1} < r^{i+1} - r^{i} \quad (4)$$

for all i. We classified a subject's utility as concave if Eq. 3 was satisfied more often than Eq. 4, and as convex if the opposite held, with Eq. 4 satisfied more often than Eq. 3. The remaining subjects were irregular or linear.

5.6. Results

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 7. 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 7 depicts the utility graphs resulting from average answers to the x, y, and z questions, based on Eq. 2, with normalizations $U(x^0) = 0$ and $U(x^4) = 1$, $U(y^0) = 0$ and $U(y^4) = 1$, and $U(z^0) = 0$ and $U(z^4) = 1$, respectively. 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 on the union of domains.

As one would expect from the overall concavity of curves in Figure 7, most participants exhibit concave (Eq. 3) rather than convex (Eq. 4) utility: 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 ($p \leq 0.01$), but not for z ($p = 0.31$). Our findings thus confirm moderate concavity of utility. For the x's, the y's, and also the z's, about 20% of our subjects had equality for all i in Eqs. 3 and 4, giving perfectly linear utility. The unclassified subjects 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.

We briefly summarize the results of our parametric analyses. For CRRA (constant relative risk aversion) utility, the median index of relative risk aversion was 0.04, and for CARA (constant absolute risk aversion) utility, the median risk tolerance was €10,000. CRRA fitted the data better than CARA. Both suggested weak concavity, but did not deviate from linearity significantly. We found decreasing absolute risk aversion (Wakker 2010 p. 83), and decreasing relative risk aversion (Wakker 2010 p. 83 footnote 7).

Hypothetical data were noisier and contained more outliers. Further: (1) Hypothetical choice tended to have more risk seeking than real incentives in the z stimuli ($0.05 < p < 0.10$) both for CARA and CRRA utility. (2) No other significant differences were found between real and hypothetical choice.

5.7. Discussion of Adaptive Utility Measurements

As regards the problem of strategic answering in adaptive experiments, Toubias et al. (2013 p. 629) and Wang, Filiba, and 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). The downside of this alternative approach is that 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 and Snyder 1970) or illusions of control (Stefan and David 2013). Such distortions are more likely with future than with past uncertainties. Prince helps to avoid such distortions by determining the RCS before 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. 2, utility turns out to be almost linear. Epper, Fehr-Duda, and 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. 2 under expected utility and nonexpected utility. The deviations are avoided rather than corrected for.

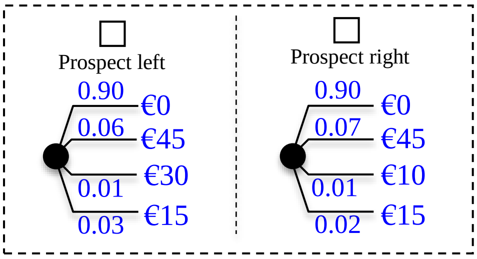
6. EXPERIMENT 4: PRINCE VS. THE TRADITIONAL RANDOM INCENTIVE SYSTEM

A difficulty for preference theory, as compared with many other empirical domains, is that there is no gold standard of true preference (Adams and Ferreira 2010 p. 885; Infante, Lecouteux, and Sugden 2016; Pedroni et al. 2017 2nd para; Thaler and Sunstein 2008; Tversky and Kahneman 1981, 2nd to last para). Hence, there is no current consensus about best methods for measuring preferences, and no clear benchmark for a new method to beat. Prediction tests, commonly used to compare different models or theories for common data, usually cannot be done for different measurement methods because those involve different stimuli. Therefore, defenses of new (theories and) measurement methods are based primarily on internal validity arguments, using coherence criteria, stylized findings that the field has converged upon, and general psychological insights.[6] Choi et al. (2007), Holt and Laury (2002), and other introductions of new measurement methods shared this aspect with us. Therefore, they typically did not compare their method with an existing method.

Even so, we carry out an experimental comparison between Prince and a standard RIS, the most popular implementation of real incentives for individual preference today. Given the absence of a gold standard, our tests are mostly exploratory. We do speculate on true preferences for two tests, which we report here in the main text. Other tests are described only briefly here. All further experimental details and descriptions are in Appendix OF.

We recruited 51 undergraduates from the Erasmus School of Economics in Rotterdam. They were randomly divided into a group for Prince (n = 26) and a group for standard RIS (n = 25). They received €5 show-up fee and could gain, if they chose randomly, an additional €11.50 on average from a lottery choice.

FIGURE 8. Violation of stochastic dominance.



For the first test, we replicated a violation of stochastic dominance (Tversky and Kahneman 1986). The choice is between two probability distributions (*lotteries*) over four outcomes depicted in Figure 8, using obvious notation. The right lottery stochastically dominates the left one and should be preferred. We presented the choice twice to each subject, with many other choices in between so that subjects could be expected to forget their previous choice and choose independently. Tversky and Kahneman (1986), and many replications (Birnbaum and Navarrete 1998), found that most subjects prefer the left lottery. Subjects appear to ignore probabilities and

[6] Hence, we intentionally used stimuli from classic economic decisions throughout this paper, to illustrate the novelty and validity of Prince.

erroneously think that there is an outcome-dominance. We claim that those majority preferences do not reflect true preferences. They result from misunderstanding due to the subtle framing of the complex lotteries. They will disappear if subjects fully understand the lotteries.

With RIS we replicated the majority preference (left). The average number of left choices per subject in the two choice situations was 1.42 (> 1 with $p = 0.02$; n = 24). However, with Prince, the average was exactly 1 ($p = 1$; n = 26), i.e., there were equally many choices for left as for right. Prince is closer to true preference (right) than RIS ($p = 0.04$) in this first test.

Our second test is a replication of Cox, Sadiraj, and Schmidt's (2015; CSS henceforth) comparative test. They considered five choices between a sure and a risky lottery and compared seven methods for implementing real incentives. CSS chose their One Task (OT) method as their gold standard for true preference. In OT, a subject carries out only one task that is implemented for real, avoiding income effects or meta-lottery perceptions. Although this gold standard is not beyond debate (Binmore, Stewart, and Voorhoeve 2012 p. 234 point 4; Birnbaum 1992), experimental economists have commonly endorsed it (Bardsley et al. 2010 p. 268; Cox, Sadiraj, and Schmidt 2014, 2015; Cubitt, Starmer, and Sugden 1998; Starmer and Sugden 1991). We, therefore, adhere to it too, and use CSS's OT results as a gold standard, to have the same criterion throughout. It gives our measurements the extra handicap of between-experiment differences.

The upper left portion of Table 3 replicates CSS's Table 4. The five columns $S_1,\ldots,S_5$ correspond with their five lottery choices.[7] We use these same lotteries. Rows describe implementation methods. We add our Prince and RIS to the bottom two rows. Cells in the table give percentages of safe choices. The last column gives the Euclidean distance to the gold standard (OT). The two best-performing methods of CSS are PAS (pay all, in a maximally correlated manner) and PAC/N (PAS, but divided by the number of questions, giving an average payoff). For CSS's remaining five methods we refer to their paper.

[7] See Cox, Sadiraj, and Schmidt (2015 Table 1): $3_{0.25}0$ vs $5_{0.20}0$, $10_{0.80}0$ vs 6, $6_{0.25}0$ vs $10_{0.20}0$, $12_{0.75}6$ vs. (0.75:12, 0.20:10, 0.05:0) (using obvious notation for three-outcome lottery), and $22_{0.80}12$ vs 18, respectively.

In all five choices, Prince is closer to the gold standard than RIS. Compared with the methods considered by CSS, Prince finishes third, defeated by their PAC/N and, closely, by their PAS. Prince's result is promising, the more so as it also faced the handicap of between-experiment differences. A pro of Prince is that it is incentive compatible in revealing true preferences for homo economicus under common assumptions[8], whereas PAS and PAC/N are subject to income effects and move preferences towards linear utility (Schmidt and Hewig 2015).

TABLE 3. Distance of methods to gold standard

Mechanism	S_1	S_2	S_3	S_4	S_5	distance
OT (gold standard)	39.47	15.52	27.59	28.95	38.46	0
PORnp	37.50	45.00	47.50	32.50	60.00	41.78
PORpi	27.50	50.00	42.50	22.50	50.00	41.58
PORpas	22.50	42.50	20.00	10.00	30.00	38.78
PAS	25.64	23.08	33.33	10.26	17.95	32.42
PAC	36.84	52.63	23.68	21.05	42.11	38.41
PAC/N	37.50	35.00	35.00	22.50	45.00	22.86
PAI	36.84	52.63	36.84	34.21	52.63	41.21
RIS	*12.00*	*40.00*	*08.00*	*60.00*	*68.00*	*59.79*
Prince	*23.08*	*38.46*	*23.08*	*34.62*	*53.85*	*32.93*

Our Prince RIS comparison, Experiment 4, further reproduced several classical choice problems. We briefly summarize the results here. Prince tended to have lower risk aversion (marginally significant), and it had significantly fewer common consequence violations of expected utility. For all nonsignificant differences—regarding choice consistency, deliberate randomization, spillover effects, constant absolute risk aversion, constant relative risk aversion, and common ratio violations of expected utility—Prince was closer to consistency and expected value maximization than RIS. These findings may be interpreted as increased rationality, but their status

[8] This theoretical claim does not need to assume expected utility, but only a weak dynamic isolation assumption (Bardsley et al. 2010 p. 269; Cohen et al. 1987), which Prince seeks to maximally enhance psychologically.

viz-a-viz true preferences is unclear. We leave better calibrations of true preferences and more and better competitors for Prince to future studies.

7. 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 Principle 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 entire 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, and 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 specified later. This method violates priority because subjects cannot know if the RCS corresponding with their number is determined a priori, or is so only during the experiment, possibly depending on their answers. It violates principles 3 and 4 because the envelopes do not contain the concrete RCS in its entirety and subjects cannot know what RCS corresponds with their number. Adaptive manipulations by the experimenter, and therefore possibly by the subject, cannot be excluded.

Bardsley (2000) partly satisfied Principle 1 too.[9] Bardsley could not determine the choice options in the RCS for a given subject beforehand because the latter

---

[9] 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). Goethe wrote: "I am inclined to offer Mr. Vieweg from Berlin an epic poem, Hermann and Dorothea … Concerning the royalty we will 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.

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, and 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, and Chatterjee (2007) also used Schade, Kunreuther, and 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 and 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, and then present a single strategy-choice between meta-lotteries, explicitly deviating from isolated binary choices.

8. GENERAL DISCUSSION

The principles of Prince listed in §2.3, and in general all details of Prince, serve to enhance isolation by enhancing psychological conditioning upon the RCS, increasing internal validity. Although Starmer and Sugden (1991) found isolation satisfied in the RIS, violations have been found.[10] In fact, any finding of learning, order effect, or spillover effect (Baltussen et al. 2012; Cox, Sadiraj, and Schmidt 2014; Stewart, Reimers, and Harris 2015) in RIS implementations entails a violation of isolation, and

---

[10] 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.

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.[11] In the case of future determination, a meta-lottery is realistically perceived because the situation is still unresolved. More generally, we want the RCS to be felt as realistically as possible. Planning beforehand generates a psychological distance (Bardsley et al. 2010 §6.4.3). Strategy choice (subjects commit to all choices at the beginning of the experiment) 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 (§7) 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 can immediately condition on it, serving isolation. BDM randomizes a choice option (the price) rather than the choice situation, leading subjects to condition in 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

---

[11] See Azevedo and Budish (2019), Bardsley et al. (2010 p. 277), Cubitt, Starmer, and Sugden (1998), and Shafir and 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 and van Winden 2010; Grant, Kajii, and Polak 2000; Kreps and Porteus 1979), and plays a role in time inconsistencies. In particular, prediction and postdiction are perceived differently (Brun and Teigen 1990; Heath and Tversky 1991 p. 9; Rothbart and Snyder 1970). Importantly 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, and Knez 2004: virtual observability).

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.[12] Choice is less precise. It takes 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.[13] In Prince, with the choice situation (the price therein being one option) specified in advance in an envelope held in hand, it is perfectly obvious that this price is not subject to bargaining or any other influence. In Bénabou and Tirole's (2001) model, it avoids crowding-out principal-agent perceptions on the part of subjects due to additional, private, information about the RCS possessed by the experimenter.

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.[14] Cox, Sadiraj, and Schmidt (2015) provide a systematic study, 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

---

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

[13] 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, and Hollard's (2010; trade uncertainty), Korobkin (2013 p. 1243), and Sayman and Öncüler (2005 §2.2); also see Bardsley et al. (2010 p. 273).

[14] 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 Cox, Sadiraj, and Schmidt (2015), Epper, Fehr-Duda, and Bruhin (2011), and Mosteller and Nogee (1951).

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 of course does not claim to once and for all settle all empirical issues about preferences that have occurred in the literature, and neither the empirical issues that we applied Prince to in this paper. Our primary aim is to show, in a broad variety of experiments, that Prince can be used in many kinds of experiments and thus has the potential to shed new light on many questions. A deeper investigation of (a) the various empirical questions used here to illustrate Prince, (b) the empirical differences between Prince and the many other existing preference measurement methods, and (c) more targeted analyses of which components of Prince are most important, are beyond the scope of what can be provided in one research papere.

This paper focuses on individual choice. The follow-up paper Li, Turmunkh, and Wakker (2018) adapts Prince to game theoretic experiments that include interactions between subjects. The envelopes for different players in a game were chosen jointly there.

9. PROS AND CONS OF PRINCE SUMMARIZED

We have provided theoretical arguments for Prince showing (1) its internal validity (§2 and §8); (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 confirm its transparencies; (3) it confirms well-established preference findings; (4) it performs better than a standard RIS in a comparative study; (5) it reconciles choice and matching in four tests. Although such a reconciliation need not always be an improvement (Ariely, Loewenstein, and Prelec 2001), in view of the preceding arguments, we claim it is.

Investigations of external validity are 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. This can reveal which component of Prince has which effect. The main purpose of this paper has been to show that Prince as a whole works well, and can be used to reduce documented violations of isolation. Given the size of this paper, showing that Prince can be implemented for virtually every preference measurement, we prefer to leave the aforementioned tests to future studies, where contributions by objective outsiders will be especially useful.

The main drawback of Prince is that it requires a nontrivial preparation by the experimenters: envelopes with different choice situations have to be prepared for every session.

10. CONCLUSION

The Prince incentive system improves on the standard 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.

An important contribution of Prince is that it revives matching. Prince makes it possible to combine the efficiency and precision of matching, with the (improved) transparency and validity of choice. Prince avoids the major weakness of BDM by not randomizing choice options but instead whole choice situations, thus leading subjects to condition properly.

Despite the usual 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, insensitivity is also real for ambiguity. Apparent preference reversals, to the contrary, are measurements errors. Decreasing absolute risk aversion is confirmed, even if utility is closer to linear than commonly thought.

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.

ONLINE APPENDIX

The Online Appendix, giving Appendixes OA-OH and RA-RG, can be downloaded here: http://personal.eur.nl/wakker/pdf/prince_online_appendix.pdf

Acknowledgment. Helpful discussions with Chen Li are gratefully acknowledged.

REFERENCES

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

Adams, Renée & Daniel Ferreira (2010) “Moderation in Groups: Evidence from Betting on Ice Break-ups in Alaska,” *Review of Economic Studies* 77, 882–913.

Anagol, Santosh, Vimal Balasubramaniam, & Tarun Ramadorai (2018) “Endowment Effects in the Field: Evidence from India’s IPO Lotteries,” *Review of Economic Studies* 85, 1971–2004.

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.

Azevedo, Eduardo M. & Eric Budish (2019) "Strategy-Proofness in the Large," *Review of Economic Studies* 86, 81–116.

Bade, Sophie (2015) "Randomization Devices and the Elicitation of Ambiguity-Averse Preferences," *Journal of Economic Theory* 159, 221–235.

Baillon, Aurélien & Han Bleichrodt (2015) "Testing Ambiguity Models through the Measurement of Probabilities for Gains and Losses," *American Economic Journal: Microeconomics* 7, 77–100.

Baltussen, Guido, G. Thierry Post, Martijn J. van den Assem, & Peter P. Wakker (2012) "Random Incentive Systems in a Dynamic Choice Experiment," *Experimental Economics* 15, 418–443.

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.

Basu, Kaushik (1982) "Determinateness of the Utility Function: Revisiting a Controversy of the Thirties," *Review of Economic Studies* 49, 307–311.

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

Bénabou, Roland & Jean Tirole (2001) "Intrinsic and Extrinsic Motivation," *Review of Economic Studies* 70, 489–520.

Binmore, Ken, Lisa Stewart, & Alex Voorhoeve (2012) "How Much Ambiguity Aversion? Finding Indifferences between Ellsberg's Risky and Ambiguous Bets," *Journal of Risk and Uncertainty* 45, 215–238.

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.

Birnbaum, Michael H. & Juan B. Navarrete (1998) "Testing Descriptive Utility Theories: Violations of Stochastic Dominance and Cumulative Independence," *Journal of Risk and Uncertainty* 17, 49–78.

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.

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.

Cason, Timothy N. & Charles R. Plott (2014) "Misconceptions and Game Form Recognition: Challenges to Theories of Revealed Preference and Framing," *Journal of Political Economy* 122, 1235–1270.

Choi, Syngjoo, Raymond Fisman, Douglas Gale, & Shachar Kariv (2007) "Consistency and Heterogeneity of Individual Behavior under Uncertainty," *American Economic Review* 97, 1921–1938.

Cohen, Michèle, Jean-Yves Jaffray, & Tanios Said (1987) "Experimental Comparisons of Individual Behavior under Risk and under Uncertainty for Gains and for Losses," *Organizational Behavior and Human Decision Processes* 39, 1–22.

Costa-Gomes, Miguel & Georg Weizsäcker (2008) "Stated Beliefs and Play in Normal-Form Games," *Review of Economic Studies* 75, 729–762.

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

Cox, James C., Vjollca Sadiraj, & Ulrich Schmidt (2015) "Paradoxes and Mechanisms for Choice under Risk," *Experimental Economics* 18, 215–250.

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

Dimmock, Stephen G., Roy Kouwenberg, Olivia S. Mitchell, & Kim Peijnenburg (2018) "Household Portfolio Underdiversification and Probability Weighting: Evidence from the Field," NBER Working Paper Series 24928, http://www.nber.org/papers/w24928.

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 (2015) "Risk Preferences Are not Time Preferences: Balancing on a Budget Line: comment (#12)" *American Economic Review* 105, 2261–2271.

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.

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

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.

Infante, Gerardo, Guilhem Lecouteux, & Robert Sugden (2016) "Preference Purification and the Inner Rational Agent: A Critique of the Conventional Wisdom of Behavioural Welfare Economics," *Journal of Economic Methodology* 23, 1–25.

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.

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 (2018) "Trust as a Decision under Ambiguity," *Experimental Economics*, forthcoming.

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.

Miao, Bin & Songfa Zhong (2015) "Risk Preferences Are not Time Preferences: Separating Risk and Time Preference: Comment (#13)," *American Economic Review* 105, 2272–2286.

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

Miyamoto, John M. (1988) "Generic Utility Theory: Measurement Foundations and Applications in Multiattribute Utility Theory," *Journal of Mathematical Psychology* 32, 357–404.

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

Noussair, Charles, Stephane Robin, & Bernard Ruffieux (2004) "Revealing Consumers' Willingness-to-Pay: A Comparison of the BDM Mechanism and the Vickrey Auction," *Journal of Economic Psychology* 25, 725–741.

Oechssler, Jörg & Alex Roomets (2014) "Unintended Hedging in Ambiguity Experiments," *Economic Letters* 122, 243–246.

Pedroni, Andreas, Renato Frey, Adrian Bruhin, Gilles Dutilh, Ralph Hertwig, & Jörg Rieskamp (2017) "The Risk Elicitation Puzzle," *Nature Human Behaviour* 1, 803–809.

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.

Rees-Jones, Alex (2018) "Quantifying Loss-Averse Tax Manipulation," *Review of Economic Studies* 85, 1251–1278.

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, Barbara & Johannes Hewig (2015) "Paying Out One or All Trials: A Behavioral Economic Evaluation of Payment Methods in a Prototypical Risky Decision Study," *Psychological Record* 65:245–250.

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.

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 (2000) “Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk,” *Journal of Economic Literature* 38, 332–382.

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.

Stewart, Neil, Stian Reimers, & Adam J.L. Harris (2015) “On the Origin of Utility, Weighting, and Discounting Functions: How They Get Their Shapes and how to Change Their Shapes,” *Management Science* 61, 687–705.

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 (2015) “Ambiguity Attitudes.” *In* Gideon Keren & George Wu (eds.), *The Wiley Blackwell Handbook of Judgment and Decision Making* (Ch. 3), 89–116, Blackwell, Oxford, UK.

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

Tversky, Amos & Daniel Kahneman (1986) “Rational Choice and the Framing of Decisions,” *Journal of Business* 59, S251–S278.

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.

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.

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