

## Eliciting decision weights by adapting de Finetti's betting-odds method to prospect theory

Enrico Diecidue · Peter P. Wakker ·  
Marcel Zeelenberg

Published online: 4 May 2007  
© Springer Science + Business Media, LLC 2007

**Abstract** This paper extends de Finetti's betting-odds method for assessing subjective beliefs to ambiguous events. Thus, a tractable manner for measuring decision weights under ambiguity is obtained. De Finetti's method is so transparent that decision makers can evaluate the relevant tradeoffs in complex situations. The resulting data can easily be analyzed, using nonparametric techniques. Our extension is implemented in an experiment on predicting next-day's performance of the Dow Jones and Nikkei stock indexes, where we test the existence and nature of rank dependence, finding usual patterns. We also find violations of rank dependence.

**Keywords** Ambiguity · Prospect theory · Rank-dependent utility · Inverse-S · Pessimism · Optimism

**JEL Classification** D81 · C60

The importance of decision making with unknown probabilities has been emphasized by Keynes (1921), Knight (1921), and many after. In most of our decisions we face uncertainties about relevant future events. The probabilities of those events are rarely known, and we usually have to act on our best beliefs and

---

Han Bleichrodt made helpful comments.

E. Diecidue  
INSEAD, Decision Sciences, Fontainebleau, France

P. P. Wakker (✉)  
Econometric Institute, Erasmus University, P.O. Box 1738, Rotterdam 3000 DR, The Netherlands  
e-mail: Wakker@few.eur.nl  
url: <http://www.few.eur.nl/few/people/wakker/>

M. Zeelenberg  
Department of Economic and Social Psychology, Tilburg University, Tilburg, The Netherlands

subjective assessments. Keynes and Knight stressed that it is important to distinguish between subjective uncertainties and objective probabilities.

Although Knight called subjective uncertainties unmeasurable, Borel (1924), Ramsey (1931), and de Finetti (1931) demonstrated soon after that subjective uncertainties can be measured after all, at least in principle. De Finetti proposed a betting-odds system for quantifying subjective uncertainties that has been widely used ever since (Winkler 1972). In its simplest form, it implies that the subjective probability of an event is  $p$  if any betting odds more favorable than  $p:1-p$  are accepted, and any betting odds less favorable are declined.

De Finetti's system, and its applications up to now, have been based on the Bayesian model of expected utility. They are distorted by the many violations of expected utility that have been found empirically (Camerer and Weber 1992; Starmer 2000), and that have hampered wider applications. This paper adapts de Finetti's system to rank-dependent utility (Schmeidler 1989) and prospect theory (Luce and Fishburn 1991; Tversky and Kahneman 1992), which can account for many violations of expected utility.

Another restriction of de Finetti's betting-odds system is that it assumes linear utility. This assumption is commonly adopted in studies of belief elicitation (Nyarko and Schotter 2002). Under expected utility, also assumed by most modern studies on belief elicitation, linear utility implies risk neutrality. Risk neutrality is very unconvincing for large stakes, where there is pronounced risk aversion and where utility must be concave. It is more plausible, but still problematic, for moderate stakes as considered in our experiment. Empirical studies then still find considerable risk aversion. This may explain why de Finetti's method has not been used more widely in the economics literature. Our study will maintain de Finetti's assumption of linear utility, but will relax his assumption of expected utility. Violations of risk neutrality for moderate stakes, also found in our data, can then be explained by factors other than nonlinear utility. These alternative explanations are more convincing (Rabin 2000). Thus, we disentangle linear utility and risk neutrality, and resolve the major restriction of de Finetti's betting-odds system. For further comments, see the discussion section.

The general estimation of nonadditive decision weights under uncertainty from data, when no probabilities need to be given so that weights are not transforms thereof, is very complex, even if utility is known. It involves many unknown parameters that quickly become intractable for large state spaces. De Finetti's method, however, can still give tractable measurements for nonadditive decision weights. In his clever design, the resulting equalities are analytically tractable because unknowns conveniently drop from equations, allowing for nonparametric analyses. This will be demonstrated in Section 3.

We will use the new version of prospect theory (Tversky and Kahneman 1992). It corrects some theoretical problems of original prospect theory (Kahneman and Tversky 1979), using Quiggin's (1981) rank-dependent probability weighting. More importantly, the new version of prospect theory, unlike the original version, can deal with *uncertainty* (unknown probabilities or ambiguity), using Schmeidler's (1989) rank-dependent weighting of events. Schmeidler introduced rank dependence for the context of uncertainty, a context which is more important but also more difficult to analyze than risk (known probabilities). The approach of Schmeidler, Kahneman,

and Tversky, developed 60 or 70 years after Keynes (1921) and Knight (1921), resulted in the first full-blown and empirically testable theory for uncertainty that reckons with ambiguity attitudes.<sup>1</sup> Ambiguity attitudes concern situations of uncertainty where no objective probabilities are known but where it is, in deviation from the Bayesian approach, also problematic to assign subjective (additive) probabilities to events (Camerer and Weber 1992). This paper will report on an empirical test of rank dependence for subjective decision weights. It considers only positive outcomes, where Schmeidler's rank-dependent utility coincides with prospect theory, the term used throughout this paper. Our findings, therefore, apply to both theories.

When restricted to two outcomes, the rank-dependent model had been known long before (Allais 1953, Eq. 19.1; Pfanzagl 1959, p. 288). The novelty of rank dependence only shows up for prospects with three or more outcomes (Gonzalez and Wu 2003). With such prospects, direct measurements can be obtained of decision weights in various middle "ranking positions," and not just with the best or worst ranks, the only possible ranks for two-outcome prospects, as will be explained in Section 1.

General prospects with three or more outcomes, while prevailing in practice, are hard to implement in experiments. They are usually investigated in special designs that make their characteristics transparent. Studies for decision under risk using such prospects include Chew and Waller (1986), Gonzalez and Wu (2003), Lopes and Oden (1999), and several others. For decision under uncertainty, the domain of our study, there have only been a few studies with prospects yielding more than two outcomes (MacCrimmon and Larsson 1979, p. 364–365; Tversky and Kahneman 1992, Section 1.3; Wu and Gonzalez 1999).

We developed a design for three-outcome prospects that incorporates rank dependence but at the same time makes de Finetti's betting odds transparent. Thus, subjects can still relate to the choices in a meaningful manner without major cognitive effort, and direct elicitation of nonlinear decision weights are obtained. Such elicitation is desirable for tractable empirical applications of prospect theory when probabilities are unknown. No direct quantitative elicitation of decision weights when in middle ranking positions have been provided in the literature before. Such elicitation gives insights into the novelty of rank dependence relative to preceding theories for uncertainty. In this way, our study is to some extent a counterpart to Gonzalez and Wu (2003). These authors considered decision under risk, and used three-outcome prospects to test the novelty of new prospect theory against the 1979 version of this theory. They found mixed results. Our study concerns unknown probabilities, and gives new insights into the distortions of the widely used elicitation of subjective beliefs through de Finetti's betting-odds method. Clemen and Lichtendahl (2005) emphasized the importance of quantitative measurements of biases in belief elicitation, so as to develop quantitative corrections.

We implement our method in an experiment on subjective probability estimations for the performance of stocks. Shiller et al. (1996, p. 163) argued for the importance

---

<sup>1</sup> The multiple-priors model existed before (Wald 1950), and became popular after Gilboa and Schmeidler (1989) established its theoretical soundness. While this model has proved to be useful in many theoretical studies, we are not aware of empirical measurements thereof, or a tractable way to obtain those.

of measuring subjective probability estimates for stock performances. We test the presence of rank dependence and, thereby, the desirability to extend the classical Bayesian methods for eliciting beliefs. We also investigate what the deviations from Bayesianism were, regarding some widely discussed properties of rank dependence. In particular, we investigate whether decision weights are convex (“pessimistic” or “uncertainty averse”), a condition mostly assumed in theoretical studies, and whether decision weights are likelihood-insensitive (“inverse-S” or “boundedly subadditive”), a condition suggested by most empirical studies. The latter condition entails a bias of beliefs and decision weights in the direction of fifty-fifty.

Many studies of rank dependence have considered situations where rank dependence is most prone to appear. Some even deliberately and explicitly manipulated stimuli so as to maximally enhance rank dependence (Weber and Kirsner 1997). Our strategy is opposite. We consider regular stimuli that are not particularly targeted towards enhancing rank-dependence effects. Our layout and presentation always maximize transparency and cognitive ease for the subjects. Thus, we move subjects in the direction of Bayesianism (which we consider rational), at the cost of enhancing rank dependence. Our estimations and significance levels for the existence and nature of rank dependence will, therefore, be conservative. We also include a number of situations that are especially prone to generate violations of prospect theory, so that we can critically test this theory. In these ways, we make it hard for prospect theory to perform well.

## 1 A reformulation of prospect theory through decision weights and rank dependence

This section presents rank-dependent utility and the new version of prospect theory in an elementary manner so as to highlight the central role of rank dependence. Tversky and Kahneman’s (1992) explanation is more complex. Given the importance of prospect theory, a simple explanation, accessible to a wide audience, is desirable.

The three uncertain events in our experiment (U, D, and R) are related to the performance of the Dow Jones industrial average and the Nikkei 225. U denotes the “Up” event that both stock indexes will go up tomorrow, D the “Down” event that both will go down, and R the “Rest” event that either one will go up and the other one will go down or at least one will remain constant. A *prospect*  $(u,d,r)$  yields  $\$u$  if U obtains,  $\$d$  if D obtains, and  $\$r$  if R obtains. The *outcomes*  $u,d,r$ , are always positive in this paper. In applications, outcomes usually depend on stock-index changes in more complex manners. For the sake of exposition, and to be consistent with the experiment described later, we confine our attention to the three-outcome prospects as just described. Generalizations to more outcomes are straightforward. Outcomes  $x$  are sometimes equated with constant (riskless) prospects  $(x,x,x)$ .

*Subjective expected utility* holds if there exists a *utility function*  $v$  and *subjective probabilities*  $\pi_U$ ,  $\pi_D$ , and  $\pi_R$  that are nonnegative and sum to 1, such that a prospect  $(u,d,r)$  is evaluated by  $\pi_U v(u) + \pi_D v(d) + \pi_R v(r)$ . Prospect theory generalizes subjective expected utility by allowing the  $\pi$ s to depend not only on the subjective beliefs about the occurrence of the event, but also on the “rank” of the events. Formally, the *rank* of an event is defined through the event that is ranked better in

the sense of yielding better outcomes. The term rank dependence refers to this dependence. We use the term *decision weight* instead of subjective probability to reflect this dependence. To what extent decision-based quantities such as subjective probabilities and decision weights reflect beliefs or other factors has been a topic of many debates and speculations (Fox and Tversky 1998; Karni 1996; Nau 1995). At any rate, decision weights are relevant to decisions, and are the focus of this paper.

To illustrate the above evaluation, consider the prospect (5,9,7). Event D yields the best outcome and has the best rank. Event R has the middle rank, and event U has the worst. The *prospect theory* value of the prospect is

$$(5, 9, 7) \rightarrow \pi_U^w v(5) + \pi_D^b v(9) + \pi_R^{m,D} v(7)$$

where the superscript w reflects the worst rank where all other events are better, b the best one where no other event is better, and m the middle one where in this case event D yields a better outcome. The middle decision weight can depend on which of the other events is best, indicated by the superscript D in this case. In this manner, there are four decision weights for event U,  $\pi_U^b, \pi_U^w, \pi_U^{m,D}$ , and  $\pi_U^{m,R}$ , and, similarly, there are four decision weights for the events D and R.

The general formula for prospect theory is

$$(u, d, r) \rightarrow \pi_U v(u) + \pi_D v(d) + \pi_R v(r) \tag{1.1}$$

where superscripts are to be added to the  $\pi$ 's according to the ranks of events, described in Table 1. For events that yield the same outcomes, such as events D and R that both yield outcome 1 in prospect (0,1,1), the ranking can be chosen arbitrarily. Observation 1.1 below will ensure that each possible ranking leads to the same evaluation; we do not elaborate on this point. An example of a prospect-theory evaluation is

$$(5, 3, 6) \rightarrow \pi_U^{m,R} v(5) + \pi_D^w v(3) + \pi_R^b v(6). \tag{1.2}$$

Schmeidler (1989) and Tversky and Kahneman (1992) stated their theories in terms of a weighting function. This function assigns to each event E the decision weight  $\pi_E^b$  (when E has the best rank). Our presentation in terms of decision weights is equivalent. Decision weights for a single prospect should sum to 1.

*Observation 1.1* All rows in Table 1 sum to 1.  $\square$

**Table 1** Decision weights for a prospect (u,d,r), depending on the ranks of U, D, R

	$\pi_U$	$\pi_D$	$\pi_R$
$u \geq d \geq r$	$\pi_U^b$	$\pi_D^{m,U}$	$\pi_R^w$
$u \geq r \geq d$	$\pi_U^b$	$\pi_D^w$	$\pi_R^{m,U}$
$d \geq u \geq r$	$\pi_U^{m,D}$	$\pi_D^b$	$\pi_R^w$
$d \geq r \geq u$	$\pi_U^w$	$\pi_D^b$	$\pi_R^{m,D}$
$r \geq u \geq d$	$\pi_U^{m,R}$	$\pi_D^w$	$\pi_R^b$
$r \geq d \geq u$	$\pi_U^w$	$\pi_D^{m,R}$	$\pi_R^b$

Because of this observation, the new version of prospect theory avoids the violations of stochastic dominance that hampered the developments of original prospect theory.

Decision weights for middle ranks only show up for prospects with three or more outcomes. Consequently, such prospects are needed for direct elicitations of such decision weights, and for direct tests of such decision weights. Earlier elicitations of middle decision weights were indirect, deriving them from nonadditive measures elicited from two-outcome prospects (Abdellaoui 2000; Bleichrodt and Pinto 2000; Fox and Tversky 1998; Gonzalez and Wu 1999; Tversky and Kahneman 1992). We will follow the assumption of linear utility underlying de Finetti’s betting-odds system and discussed elsewhere, i.e., we set  $v(x)=x$ .

## 2 Our hypotheses

We now turn to empirical phenomena that cannot be modeled by expected utility, but can be by prospect theory. The following analysis concerns concepts discussed by Gonzalez and Wu (1999) for risk (given probabilities), and by Einhorn and Hogarth (1985) for uncertainty, the context of our paper. We discuss these concepts for event U. Similar observations apply to the events D and R. *Pessimism* holds if decision weights increase as events get ranked worse. For U this means

$$\pi_U^w \geq \{\pi_U^{m,R}, \pi_U^{m,D}\} \geq \pi_U^b, \tag{2.1}$$

where the braces  $\{\pi_U^{m,R}, \pi_U^{m,D}\}$  indicate that the inequalities hold for both  $\pi_U^{m,R}$  and  $\pi_U^{m,D}$ . Pessimism reflects the attitude of a person who (erroneously) believes that events get more likely as their outcomes are more unfavorable, or of a person who, deliberately and more rationally, decides that more attention should be given to unfavorable events in decisions, such as in worst-case scenarios. It can be seen that pessimism is equivalent to convexity of weighting functions, and extends Gonzalez and Wu’s (1999) “low elevation” from risk to uncertainty. It is modeled through the parameter  $\beta$  in Einhorn and Hogarth (1985). *Optimism* refers to the opposite phenomenon, with

$$\pi_U^w \leq \{\pi_U^{m,R}, \pi_U^{m,D}\} \leq \pi_U^b. \tag{2.2}$$

(*Likelihood*) *insensitivity*, a mix of the above two phenomena, holds if

$$\pi_U^w \geq \{\pi_U^{m,R}, \pi_U^{m,D}\} \text{ and } \pi_U^b \geq \{\pi_U^{m,R}, \pi_U^{m,D}\}. \tag{2.3}$$

It implies an overweighting of extreme outcomes, both worst (as under pessimism) and best (as under optimism), and an underweighting of intermediate outcomes. It corresponds with inverse-S shaped, or bounded subadditive, weighting functions, and extends “low discrimination” from risk to uncertainty (Chateauneuf et al. 2005; Gonzalez and Wu 1999; Tversky and Fox 1995; Tversky and Wakker 1995; Viscusi and Evans 2006). It is modeled through the parameter  $\theta$  in Einhorn and Hogarth (1985).

Theoretical studies mostly assume pessimism (Dow and Werlang 1992), and terms such as uncertainty aversion and ambiguity aversion have been used. Empirical

studies have suggested that insensitivity is prevailing (Abdellaoui et al. 2005; Einhorn and Hogarth 1985; Gonzalez and Wu 1999; Tversky and Fox 1995). On the basis of the above, a mix of pessimism and insensitivity can be expected, with strong inequalities  $\pi_U^w > \{\pi_U^{m,R}, \pi_U^{m,D}\}$  and weaker inequalities  $\pi_U^b \geq \{\pi_U^{m,R}, \pi_U^{m,D}\}$ ; the latter may be reversed if the effect of pessimism is stronger than that of insensitivity. Ellsberg (2001, pp. 203–206) predicted that such a reversal will not occur.

Phenomena as discussed above have been documented extensively for given probabilities. Quantitative empirical estimates of non-Bayesian decision weights for uncertainty have been virtually absent in the literature as yet. Our main empirical hypothesis is as follows.

Hypothesis 1 [Rank dependence of decision weights].  $\pi_U^w \neq \{\pi_U^{m,R}, \pi_U^{m,D}\} \neq \pi_U^b$ .

Our second empirical hypothesis concerns the nature of rank dependence.

Hypothesis 2 [Insensitivity and some pessimism].  $\pi_U^w \geq \pi_U^b \geq \{\pi_U^{m,R}, \pi_U^{m,D}\}$ .

We chose our experimental design to optimally test these two empirical hypotheses, sometimes at the cost of testing other questions.

To critically test prospect theory, we elicited decision weights in different situations that, according to prospect theory, should give the same results, but where violations of prospect theory are most likely to generate differences. For this purpose, we considered *degenerate* prospects. These are (“riskless”) prospects for which it is certain beforehand what the outcome will be. People have a special preference for degenerate prospects (the *certainty effect*). Expected utility explains preference for certainty through concave utility. The Allais paradox, however, showed that this explanation is not sufficient (Allais 1953). There are factors underlying the certainty effect that are beyond expected utility, i.e., beyond utility curvature. Prospect theory uses probability weighting (and loss aversion) as a further factor, besides utility curvature, to explain the special preference for certainty. Prospect theory can, indeed, accommodate the Allais paradox.

The most pronounced violations of classical theories have been found, indeed, when degenerate prospects are present (Birbaum and Thompson 1996; Humphrey 1995; McCord and de Neufville 1986; Starmer 2000). It is plausible that many psychological irregularities are effective in such situations, and that any theory, also prospect theory, will have difficulties there. We use the term *degeneracy effects* to designate factors underlying the certainty effect that are beyond prospect theory, i.e., beyond utility curvature and probability weighting (and loss aversion).

We elicited decision weights  $\pi_U^b$  of event U when in the best rank both with a degenerate prospect present, denoting the resulting decision weight by  $\pi_U^{b,d}$  and with no degenerate prospect present, denoting the resulting decision weight by  $\pi_U^{b,n}$ . Symbols such as  $\pi_U^{w,d}$ ,  $\pi_U^{w,n}$  and  $\pi_D^{b,d}$  are similar, and all these weights were elicited in our experiment. Elicitations of decision weights for middle rank positions are not possible with degenerate prospects. The experimental details will be explained later. Prospect theory predicts equalities  $\pi^d = \pi^n$ , but degeneracy effects will generate differences. Such differences entail a violation of prospect theory, and show if the degeneracy effects, factors beyond prospect theory, reinforce or weaken the certainty effect relative to prospect theory’s predictions.



### 3 Making decision tradeoffs transparent through de Finetti’s betting-odds system

Before reading on, the readers are invited to determine their preference between the following two prospects

$$(33, 46, 65) \text{ and } (19, 52, 71), \tag{3.1}$$

where the events U, D, R refer to the performance of the Dow Jones and Nikkei stock indexes on the day of reading. This choice concerns six outcomes and three levels of subjective likelihood. Such choices are complex and hard to evaluate (which the above question was intended to illustrate). For the experiment, we developed a special format of questions so as to induce subjects to make their tradeoffs along the lines of de Finetti’s betting-odds system, which makes choices transparent and gives the best assessment of the decision weights of the subjects.

Traditionally, de Finetti’s betting-odds system reveals indifferences such as  $(20,0,0) \sim (6,6,6)$ , to conclude that  $\pi_U = 6/20$  for subjective probability  $\pi_U$ . In our model, this preference reveals only  $\pi_U^b = 6/20$ , i.e. it reveals the decision weight of event U only when ranked best. To reveal, for instance, that  $\pi_U^w = 6/20$ , regarding the decision weight of U when in the worst rank, we add side payments, called reference prospects, to generate the desired rank ordering. In our experiment, for instance, we considered the traditional choice just mentioned with added the reference prospect  $(13,46,65)$ . Consider an indifference

$$(20, 0, 0) + (13, 46, 65) \sim (6, 6, 6) + (13, 46, 65),$$

i.e.,  $(33,46,65) \sim (19,52,71)$  (indifference in Eq. 3.1). Taking the reference prospect as point of departure, 20 more under U is equally good as 6 more for sure. In these considerations, event U is always ranked worst. After some algebraic manipulations, presented later, it follows that  $\pi_U^w = 6/20$ . An experimental layout to make these tradeoffs transparent to subjects (unlike Eq. 3.1 as presented above) will be described in the next section.

In general, we elicit indifferences of the form

$$(B, 0, 0) + (r_1, r_2, r_3) \sim (s, s, s) + (r_1, r_2, r_3), \tag{3.2}$$

where the left prospect is  $(B+r_1, r_2, r_3)$ , and the right one is  $(s+r_1, s+r_2, s+r_3)$ . The prospect  $(r_1, r_2, r_3)$  is the *reference prospect*, and  $B > s > 0$  ( $B$  designates big and  $s$  designates small or sure). If  $r_2$  or  $r_3$  exceed  $r_1$ , then they exceed it by so much that they also exceed  $r_1+B$ , for all the stimuli in our experiment. That is, we always chose  $B$  small relative to the differences between  $r_1$  and the other  $r$ s. In the preceding paragraph  $r_2=46$  and  $r_3=65$  exceed  $r_1=13$  by so much that, with  $B=20$ ,  $r_1+B$  is still worse than  $r_2$  and  $r_3$ . In this manner, the ranks of the events for both prospects are the same as for  $(r_1, r_2, r_3)$ .

We apply Eq. 1.1 to Eq. 3.2. For  $v$  we take the identity (reflecting linear utility), and each event has the same decision weight with the same superscript for both prospects, which we suppress. The result is

$$\pi_U(B + r_1) + \pi_D r_2 + \pi_R r_3 = \pi_U(s + r_1) + \pi_D(s + r_2) + \pi_R(s + r_3).$$



Cancelling the prospect-theory value of the reference prospect yields

$$\pi_U B = \pi_U s + \pi_D s + \pi_R s = s,$$

where we used the unit summation  $\pi_U + \pi_D + \pi_R = 1$  (Observation 1.1). It follows that

$$\pi_U = s/B.$$

The decision weight of U equals the betting odds  $s/B$ , exactly as in the original betting-odds system of de Finetti where the reference prospect was (0,0,0) and where expected utility (with linear utility) was assumed. The reference prospect cancelled conveniently from the equations (except for the rank dependence that it generated) because of linearity of utility. For nonlinear utility, equalities of decision weights would result that, while still linear, would be considerably more complex to solve.

In modern theories, rank dependence is important. Our design has modified de Finetti’s design by reckoning with this rank dependence. Classical applications have elicited decision weights only for best ranking positions, but these need not reflect subjective beliefs more properly than decision weights in worst or middle ranking positions. We chose a layout, presented in the next section, so as to induce psychological processes that match the preceding algebraic derivation of the decision weight from Eq. 3.2.

#### 4 Experimental stimuli and layout that make de Finetti’s betting-odds system transparent to subjects

Figure 1 gives an example of the stimuli used in our experiment, and referred to in the preceding section. We confine our attention, for now, to the first table, i.e. the left matrix. The rest of Figure 1 will be explained in the next section. In the first table, the grey middle column depicts the reference prospect  $(r_1, r_2, r_3) = (13, 46, 65)$ , where event U is indicated by  $\uparrow\uparrow$ , event D by  $\downarrow\downarrow$ , and event R by  $\uparrow\downarrow$ .

The left column, indicated by a single large plus, designates the left side of Eq. 3.2, i.e. a gamble of  $B (=20)$  extra on event U. The right column, indicated by three small plusses, designates the right side of Eq. 3.2, yielding  $s=3$  more than the reference prospect with certainty. For all prospects, U is ranked worst with decision weight  $\pi_U^w$ , D is ranked middle with decision weight  $\pi_D^{m,R}$ , and R is ranked best with decision weight  $\pi_R^b$ .

In the instructions to the subjects, it was explained that the left prospects in the tables always result from the middle prospects through single big increases of the outcome for one event, and the right prospects always result from the middle ones

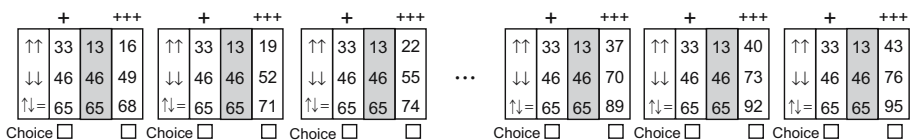


Fig. 1 Stimuli used in our experiment

through the same (small) increase of all three events. This layout and presentation of tables should make the tradeoffs transparent, of either getting B extra under event U or s extra for sure, as for de Finetti's betting odds. At the same time, the initial focus on the reference prospect should make the rank-ordering of the events salient. The layout of the first table in Figure 1 thus makes the relevant tradeoffs in Eqs. 3.1 and 3.2 transparent to the subjects.

## 5 Experiment

*Participants*  $N=186$  students, all from Tilburg University, took part. There were 62 psychology students divided into six groups. There also was a group of 124 students in general social sciences who participated in one big session. The average age of the participants was 20.1, and 32.8% were male.

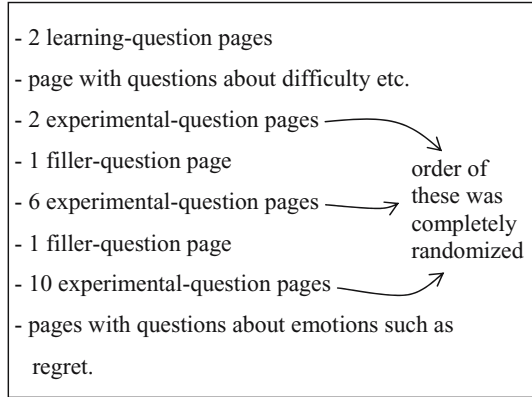
*Procedure* The experiment was carried out in classroom sessions. All items were administered using pencil-and-paper questionnaires. The subjects received brief verbal instructions, followed by detailed written instructions that took about 15 min to read (Appendix A). The stimuli and appendices are downloadable from the second author's homepage, at <http://people.few.eur.nl/wakker/pdfspubl/07.2dowjappdices.pdf>.

A transparency with a graph depicting the performances of the stock indexes during the last 2 months, up to the day of the experiment, was projected during the task (Appendix B). Such periods are commonly used because for periods further in the past the nature of the stock may be different (Hull 2005, Section 13.4). A brief text in the written instructions discussed the likelihood of the indexes increasing or decreasing, referring explicitly to the last 2 months. As different groups participated on different days, the information about the indexes varied from group to group. Then the participants were asked to fill out the questionnaire at their own pace. This usually took about 30 min.

*Stimuli; organization between pages* Besides a test-choice in the instructions, there were 22 pages with ten choice questions each. The experiment started with two learning-task pages, followed by 18 experimental-task pages and two filler-question pages. The filler-question pages were always on the third and tenth place after the learning-question pages, and served to discourage subjects from using heuristics such as switching in the same place on each page. Other than that, the order of the experimental-question pages was randomized. The randomization was generated through random permutations of 1, ..., 18 and a manual reordering of the pages (Figure 2).

After the two learning-question pages and before the first experimental-question page, there was a page with three questions about the difficulty of the other questions and about whether the participants paid attention to their perceived likelihoods of the events (Appendix C). The questions about likelihood served to focus subjects' attention on this aspect. Pilots had demonstrated that subjects were prone to using the heuristic of simply taking the sum of payments as their decision criterion, which ignores the different likelihoods of the events, leading to a loss of statistical power of our design.

**Fig. 2** Order of pages following the instructions



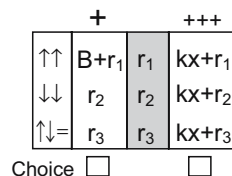
The outcomes used in the experiment ranged from Dfl. 10 (approximately €4.50) to Dfl. 99. At the end of the experiment, there were self-assessment questions about age and gender.

*Stimuli; organization within one page* Each page contained ten tables, reflecting ten choice questions. Figure 3 depicts the general format of a *k*th table on a page when the big increase *B* is obtained under event *U* (indicated by ↑↑).

All ten tables on one page had the same grey middle column, i.e. the same reference prospect. They also had the same left column, with the same single increase *B* (*B*=20 in Figure 1). The payoffs of the right prospects were increased stepwise with step size *x* (*x*=3 in Figure 1) up to 10*x*. We always had 10*x* ≥ *B*, so that the right prospect (10*x*+*r*<sub>1</sub>, 10*x*+*r*<sub>2</sub>, 10*x*+*r*<sub>3</sub>) in the last choice on each page always dominated the left one (10*x* + *r*<sub>1</sub> ≥ *B* + *r*<sub>1</sub>). In Figure 1, the right prospects dominate the left ones for all *k* ≥ 7 and, hence, for all three tables displayed on the right.

Table 2 gives the complete set of choice 10-tuples. To explain the layout of the table, we consider the row numbered 5. It serves to elicit the decision weight π<sub>U</sub><sup>w,n</sup> of event *U* when ranked worst and with no collapsing outcomes. The row concerns choices illustrated in Figure 1. Let us consider the first (left) matrix in Figure 1, with a choice between (33, 46, 65) and (16, 49, 68). The outcomes of the left prospect (indicated by a large plus sign in Figure 1) are given in Table 2 in the columns indicated by U, D, and R under the large plus sign. The outcomes of the right prospect (indicated by three small plus signs in Figure 1) are given in the columns

**Fig. 3** The general format of the *k*th table on a page



**Table 2** Stimuli and results

#		+			+++			mean (st.dev.) of $\pi$
		U	D	R	U	D	R	
1	$\pi_{U^{b,n}}$	(B+)r <sub>1</sub>	(B+)r <sub>2</sub>	(B+)r <sub>3</sub>	xk+r <sub>1</sub>	xk+r <sub>2</sub>	xk+r <sub>3</sub>	0.465 (0.22)
2	$\pi_{U^{b,d}}$	20+44	29	13	3k+44	3k+29	3k+13	0.414 (0.18)
3	$\pi_{U^{m,D}}$	30+24	24	24	3k+24	3k+24	3k+24	0.473 (0.23)
4	$\pi_{U^{m,R}}$	20+31	65	14	3k+31	3k+65	3k+14	0.485 (0.23)
5	$\pi_{U^{w,n}}$	20+23	14	59	3k+23	3k+14	3k+59	0.505 (0.23)
6	$\pi_{U^{w,d}}$	20+13	46	65	3k+13	3k+46	3k+65	0.430 (0.17)
7	$\pi_{D^{b,n}}$	30+16	46	46	3k+16	3k+46	3k+46	0.334 (0.18)
8	$\pi_{D^{b,d}}$	18	40+56	35	4k+18	4k+56	4k+35	0.352 (0.19)
9	$\pi_{D^{m,R}}$	37	30+37	37	4k+37	4k+37	4k+37	0.314 (0.19)
10	$\pi_{D^{m,U}}$	11	40+18	59	4k+11	4k+18	4k+59	0.310 (0.19)
11	$\pi_{D^{w,n}}$	59	40+15	10	4k+59	4k+15	4k+10	0.334 (0.19)
12	$\pi_{D^{w,d}}$	59	40+10	56	4k+59	4k+10	4k+56	0.355 (0.20)
13	$\pi_{R^{b,n}}$	58	30+28	58	4k+58	4k+28	4k+58	0.525 (0.20)
14	$\pi_{R^{b,d}}$	42	26	20+63	2k+42	2k+26	2k+63	0.506 (0.20)
15	$\pi_{R^{m,U}}$	17	17	20+17	2k+17	2k+17	2k+17	0.493 (0.21)
16	$\pi_{R^{m,D}}$	74	12	20+37	2k+74	2k+12	2k+37	0.513 (0.20)
17	$\pi_{R^{w,n}}$	16	61	20+27	2k+16	2k+61	2k+27	0.488 (0.20)
18	$\pi_{R^{w,d}}$	77	51	20+13	2k+77	2k+51	2k+13	0.488 (0.20)
L1		49	49	20+29	2k+49	2k+49	2k+29	0.367 (0.15)
L2		30+50	10	30	3k+50	3k+10	3k+30	0.367 (0.24)
F1		45	20+10	55	4k+45	4k+10	4k+55	0.347 (0.14)
F2		30+35	11	24	2k+35	2k+11	2k+24	0.271 (0.14)
		32	30+49	13	2k+32	2k+49	2k+13	

Each row refers to a numbered 10-tuple of choice questions (L1,L2: learning questions; F1,F2: filler questions). After the number follow: the decision weight ( $\pi$ ) measured, the three event-contingent outcomes of the single large-size plus prospect, the three event-contingent outcomes of the three-small-pluses prospects ( $k=1, \dots, 10$ ) and, finally the means and standard deviations of the elicited decision weights.

indicated by U, D, and R under the three small plus signs in Table 2, where  $k=1$  corresponds with the first matrix in Figure 1. The second matrix in Figure 1 has the same left prospect as the first matrix, and its right prospect is as for the first matrix but now with  $k=2$ . The page given to participants contained ten such matrixes, for  $k=1, \dots, 10$ . Figure 1 displays the cases for  $k=1, 2, 3, 8, 9,$  and  $10$ . In this way, the row numbered 5 in Table 2 concerns the choices depicted in Figure 1.

Elicitations of decision weights with best or worst ranks when there are degenerate prospects, indicated by superscript d, occurred for the 10-tuples number 2, 6, 8, 12, 14, and 18 in Table 2. Then either  $B + r_1 = r_2 = r_3$  or  $r_1=r_2=r_3$  in Figure 3. For example, the second 10-tuple first considers a choice between (54, 24, 24) and (27, 27, 27) (for  $k=1$ ), and then a choice between (54, 24, 24) and (30, 30, 30) (for  $k=2$ ); etc. For each choice in this 10-tuple the second option is degenerate.

Elicitations of decision weights when there are no degenerate prospects, indicated by superscript n, occurred for the 10-tuples 1, 5, 7, 11, 13, and 17. For example, the first 10-tuple first considers a choice between (64, 29, 13) and (47, 32, 16), and then a choice between (64, 29, 13) and (50, 35, 19); etc. Although there are no degenerate

prospects now, the rank of the first outcome (regarding the U event) is best, as it is for the second 10-tuple, and the same decision weight should result for event U from the first and the second 10-tuple according to prospect theory. Degeneracy effects will, however, generate differences between these decision weights.

*Motivating the participants* So as to avoid income effects, individual-choice experiments usually pay at most one of the choices made by each participant for real, where this choice is randomly selected from all choices. A theoretical problem, suggested by Holt (1986), was demonstrated not to occur empirically by Starmer and Sugden (1991). The random-lottery incentive system has since become the almost exclusively used incentive system for individual-choice experiments (Holt and Laury 2002; Harrison et al. 2002). We used a variation of the system where only one of every ten participants, randomly selected, played for real. Two studies examined whether there was a difference between this form of the random-lottery incentive system and the original form, and did not find a difference (Armantier 2006, p. 406; Harrison et al. 2007). The incentive system was explained to the participants beforehand. The participants collected the money gained the next morning, when the relevant uncertainties about the stock indexes had been resolved. In addition, the 62 psychology students received course credits, and each student of the large group of 124 received a flat payment of €11.

*Analysis* On each experimental page, we assessed the point at which the participants switched from a choice for the left column to a choice for the right column. This should happen in one place. Other choice patterns violate dominance and were coded as missing values. Assuming that indifference is halfway between the two choices where preferences switch, and assuming linear utility as discussed elsewhere, we could calculate the decision weights. For example, imagine that the switch for 10-tuple 5 in Table 2, also depicted in Figure 1, is from the second to the third table:

$$(6, 6, 6,) + (13, 46, 65) \prec (20, 0, 0) + (13, 46, 65) \prec (9, 9, 9) + (13, 46, 65).$$

We then assume that an approximate indifference

$$(7\frac{1}{2}, 7\frac{1}{2}, 7\frac{1}{2}) + (13, 46, 65) \sim (20, 0, 0) + (13, 46, 65)$$

holds, and estimate  $\pi_U^w = 7\frac{1}{2}/20 = 0.375$ .

We discuss the analysis of decision weights for event U in more detail. The analyses for the other events are similar. The “best” and “worst” decision weights were elicited with a degenerate prospect (denoted  $\pi_U^{b,d}$  and  $\pi_U^{w,d}$  see 10-tuples 2 and 6 in Table 2) present, and with no degenerate prospects present (denoted  $\pi_U^{b,n}$  and  $\pi_U^{w,n}$  see 10-tuples 1 and 5 in Table 2). The middle decision weights were elicited with event D yielding the best outcome ( $\pi_U^{m,D}$  10-tuple 3), and with event R yielding the best outcome ( $\pi_U^{m,R}$  10-tuple 4). This leads to six measurements of decision weights per event and, thus, to 18 measurements in total, given in the last column of Table 2. Before turning to the study of these separate decision weights, we first consider

overall estimates of the decision weights in the various ranks. For this purpose, we define averages

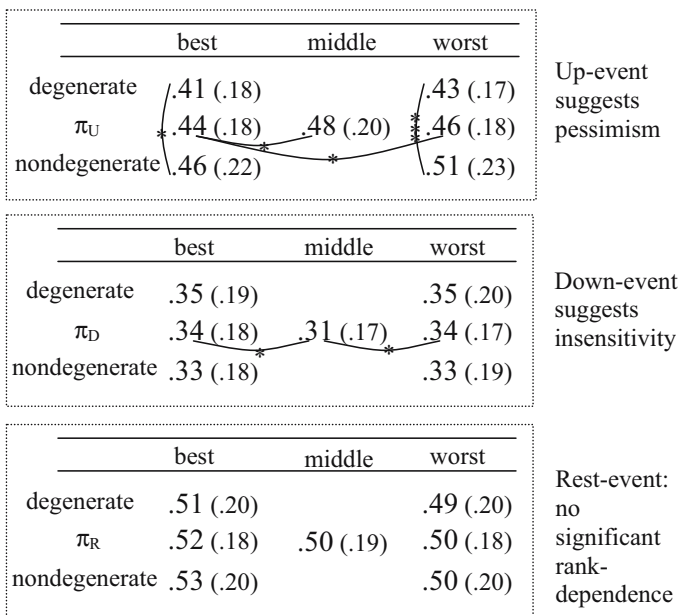
$$\pi_U^b = (\pi_U^{b,d} + \pi_U^{b,n}) / 2, \pi_U^w = (\pi_U^{w,d} + \pi_U^{w,n}) / 2, \pi_U^m = (\pi_U^{m,D} + \pi_U^{m,R}) / 2,$$

with averages for the events D and R defined similarly. According to prospect theory, we should have  $\pi_U^{b,d} = \pi_U^{b,n}$  and  $\pi_U^{w,d} = \pi_U^{w,n}$  so that  $\pi_U^b$  and  $\pi_U^w$  are estimations of single decision weights. Middle weights such as  $\pi_U^{m,D}$  and  $\pi_U^{m,R}$  may be different, so that  $\pi_U^m$  is an average of two different decision weights. All statistical tests will be two-tailed tests for normal distributions. Distribution-free tests usually gave the same results, and are reported only if deviating.

### 6 Results

The different subject groups exhibited the same patterns, and their data were pooled. Four subjects were dropped because a test question at the beginning of the experiment suggested that they did not understand the stimuli. Ten subjects were dropped because they had more than two incorrect choice-switches (from the right column to the left column), suggesting that they did not understand the stimuli. Dropping these subjects does not affect any of the main results hereafter.

Regarding our first hypothesis, analysis of variance with repeated measurements confirms rank dependence for event U with  $F(2,328)=9.44, p<0.001$  and for event D with  $F(2,322)=5.77, p=0.003$ . For event R the result is not significant, although it



\*: different ( $p \leq 0.05$ ); \*\*\*: different ( $p \leq 0.001$ ).

Fig. 4 Means (standard deviations) of decision weights

is marginally ( $F(2,334)=2.80, p=0.06$ ). To investigate the nature of rank dependence, we display averages of decision weights, and the results of pairwise  $t$ -tests, in Figure 4. The means and standard deviations of the average decision weights  $\pi_U, \pi_D,$  and  $\pi_R$  are given in the middle rows of the boxes in Figure 4. For example,  $\pi_U$ , the decision weight of event U, when its rank is best, is 0.44 on average (comprising both degenerate and nondegenerate measurements), with standard deviation 0.18. The average decision weight  $\pi_D$  of event D when ranked worst and when restricted to degenerate measurements is 0.35 with standard deviation 0.20. The decision weight of event R,  $\pi_R$ , when in a middle ranking position, is 0.50 on average with standard deviation 0.19.

We describe the results for the two events with significant rank dependence. For event U we find some pessimism, because  $\pi_U^b < \pi_U^m (t_{165} = -4.18, p < 0.001)$  and  $\pi_U^b < \pi_U^w (t_{168} = -3.24, p = 0.001)$ . Whereas pessimism suggests that  $\pi_U^m < \pi_U^w$  there is no significant difference in our data ( $t_{168}=0.93, ns$ ). Event D exhibits likelihood insensitivity. That is,  $\pi_D^b > \pi_D^m (t_{168} = 3.35, p = 0.001)$  and  $\pi_D^b > \pi_D^w (t_{169} = 3.12, p = 0.002)$ . There is no significant difference between  $\pi_D^b$  and  $\pi_D^w (t_{168} = 0.13, ns)$ .

None of our hypotheses predicts a relation between  $\pi_U^{m,D}$  and  $\pi_U^{m,R}$  or between  $\pi_D^{m,U}$  and  $\pi_D^{m,R}$  or between  $\pi_R^{m,U}$  and  $\pi_R^{m,D}$ . Insensitivity predicts most rank dependence at extreme outcomes, so that equalities between these pairs of decision weights are plausible. Indeed, none of these equalities is rejected statistically ( $t_{170}=-0.86, ns, t_{171}=-0.05, ns,$  and  $t_{171}=1.52, ns,$  respectively). Table 3 summarizes the significant inequalities and their support for properties of decision weights. Insensitivity is tested in fewer inequalities than optimism and pessimism, so that there are fewer plus signs among significant differences to be displayed. In return, the nonsignificant differences (not displayed) provide less evidence against insensitivity than against optimism and pessimism.

We now turn to critical tests of prospect theory. There are six tests of degeneracy-effects, given in Table 4. The inequalities given in the top row of the table are those found for means in our data, and are not statistical hypotheses. The statistical null hypothesis is always equality of the weights, the alternative hypothesis always claims inequality of the weights, and all tests are two-tailed. The first two equalities are rejected which suggests degeneracy-effects, in violation of prospect theory. The last four equalities are accepted, suggesting no degeneracy-effects and agreement with prospect theory.

We also analyzed pessimism at the individual level, taking  $\pi_U^{w,n} - \pi_U^{b,n}$  as index of pessimism for event U with noncollapsed decision weights, and taking  $\pi_D^{w,n} - \pi_D^{b,n}$  and  $\pi_R^{w,n} - \pi_R^{b,n}$  similarly. All correlations are positive. They are significant for events

**Table 3** Significant inequalities and their support for properties of decision weights

	$\pi_U^b < \pi_U^m$	$\pi_U^b < \pi_U^w$	$\pi_D^b > \pi_D^m$	$\pi_D^w > \pi_D^m$
Insensitivity	–	0	+	+
pessimism	+	+	–	+
optimism	–	–	+	–

+ the column-inequality supports the row-property, – it does not, 0 no support for or against.



**Table 4** Testing invariance of decision weights for same ranks of events

	$\pi_U^{b,n} > \pi_U^{b,d}$	$\pi_U^{w,n} > \pi_U^{w,d}$	$\pi_D^{b,n} < \pi_D^{b,d}$	$\pi_D^{w,n} > \pi_D^{w,d}$	$\pi_R^{b,n} > \pi_R^{b,d}$	$\pi_R^{w,n} > \pi_R^{w,d}$
<i>t</i> -statistic	$t_{170}=3.22$	$t_{172}=4.74$	$t_{170}=-1.81$	$t_{172}=-1.09$	$t_{170}=1.63$	$t_{171}=0.52$
<i>p</i> -value	$p=0.002$	$p=0.000$	$p=0.07$	ns	ns	ns

ns not significant,  $p=0.07$  marginally significant (significant,  $p=0.02$ , under distribution-free test).

U and D ( $r=0.174$ ,  $p=0.025$ ) and events U and R ( $r=0.214$ ,  $p=0.006$ ), and insignificant for events D and R ( $r=0.104$ ,  $p=0.178$ ). It suggests that individuals who are pessimistic for one event are pessimistic for another too, so that pessimism and optimism are individual traits to some extent. Similar correlational analyses for collapsed decision weights did not give significant results, suggesting that, in the presence of collapsing, effects other than rank dependence are dominant. A similar analysis for insensitivity is not well possible because middle-ranked events play a different role, without possibility of collapsing, in our design than events ranked worst or best.

Our final result concerns the prediction of rank dependence that the rows in Table 1 sum to the same, and that this sum is 1 (Observation 1.1). The average weights elicited did sum to approximately the same for each row, but this sum exceeded 1 and ranged from 1.24 to 1.34. Risk aversion (total number of right choices) correlated significantly with age ( $r=0.22$ ,  $p=0.007$ ), but not with gender.

## 7 Discussion

The analyses of variance detected strong deviations from expected utility, with decision weights affected by ranking. The paired *t*-tests found no clear overall patterns of rank dependence at the aggregate group level, with some support for pessimism and insensitivity. These findings together show that there are many violations of expected utility at the individual level, but there is also much heterogeneity between individuals. The size of the rank-dependent effects in our data will have been attenuated by some conservative aspects in our tests, discussed later.

Historical data suggest that the probability of the daily Dow Jones index going up is very close to 0.5, and so it is for the Nikkei index. These daily performances have a weak positive correlation (Hamao et al. 1990). Thus, R has a probability slightly below 0.5, and U and D have a probability slightly above 0.25. The decision weights of R obtained in our study are all close to 0.5. The decision weights of U and D exceed 0.25 considerably. These results reflect the general overweighting of decision weights in our data, in combination with the general regressive nature of judged probabilities. The latter implies that people overestimate small probabilities and underestimate moderate and high probabilities (Tversky and Fox 1995). The performances of the indexes in the months preceding the experiment were positive, with more movements up. Given that subjects received information about the two preceding months, it is natural that they weighted U more heavily than D.

Several decisions in the design of our study served to maintain tractability for the subjects. For example, we did not counterbalance for the following two order effects.

First, in all pages, sure payments were increasing from left to right. Given a natural tendency against changes, a bias can be expected of switching from left to right too late, generating a systematic bias upwards in the measured decision weights. This bias is enhanced by the presence of pages where the final choices are governed by dominance, e.g. in Figure 1, so that there is more room for overestimation than for underestimation. Thus, biases upwards in our measurements will have been generated, explaining the violations of unit summation in Table 1. Second, the location of events on pages was always the same, with event U on top, D in the middle, and R at the bottom.

Both order effects just discussed do not affect the presence of rank dependence or its direction, to the best of our knowledge. The mentioned biases do affect the exact quantitative elicitation of decision weights. For such elicitation, the mentioned biases could have been avoided by counterbalanced elicitation and averaging. We did not carry out such corrective procedures so as to avoid a lengthy experiment and so as to reduce the cognitive burden for the subjects. For this paper, we optimized our design for the testing of general hypotheses about existence and direction of rank dependence.

Our design may have encouraged subjects to focus on the big and small changes  $B$  and  $s$ , and to ignore the reference prospects. This effect reduces rank dependence, and the power of our design to find it, so that our statistical conclusions become conservative.

Different groups of subjects participated in the experiment at different times and, accordingly, received different information about the stock indexes during the preceding 2 months. This difference leads to an additional variation between individuals, and again leads to a loss of power. We, however, do not consider it to be a bias. In real life it is only natural that different individuals have different information about events and, hence, different attitudes towards them. The difference mentioned does not distort our findings on rank dependence because these findings are all based on within-subject differences.

The inequality  $\pi_U^{b,d} < \pi_U^{b,n}$  that we found, with decision weights lower and more choices for right (= safe = risk averse) columns in matrices with degenerate prospects present, suggests that the degeneracy-effects enhanced the certainty effect beyond prospect theory. The inequality  $\pi_U^{w,d} < \pi_U^{w,n}$  that we found, however, suggests the opposite (Wu et al. 2005, p. 120, also report findings opposite to the certainty effect). There is, therefore, no clear direction in the violation of prospect theory that we found here. The other violation of prospect theory in our data, concerning the sums in Table 1 exceeding 1, was discussed before. Many other violations of prospect theory have been found, including Barron and Erev (2003), Birnbaum (2006), Goeree et al. (2002), González-Vallejo et al. (2003), Harbaugh et al. (2002), Lopes and Oden (1999), Neilson and Stowe (2001), and Starmer (1999). However, to date there is no more successful and tractable theory available for decision under risk or uncertainty. Many phenomena remain unpredictable in this domain, with only post-hoc heuristic explanations conceivable.

We extensively tested many alternative layouts and framings in pilot studies, where subjects were asked to give feedback. The layout of the stimuli chosen for the experiment was found to be most suited for making the decision-relevant tradeoffs transparent to the subjects. In the pilot studies, we found that grouping

the 10-tuples by events, rather than completely randomizing the order of presentation, and some other changes in design, induced participants to resort to heuristics equivalent to expected value maximization instead of expressing subjective preferences. We believe that experimental choices, derived from a transparent design where the relevant tradeoffs are clear to the participants, will be more representative of choices made in significant real-life decisions than choices derived from non-transparent designs.

The most problematic heuristic to be avoided was the one of simply adding the outcomes, completely ignoring the uncertain events and taking them all as if equally likely. This is an extreme case of Viscusi's (1989) model of biases toward uniform prior distributions that in our setup enhance expected value maximization. Hence, we used events that were very clearly not symmetric or equally likely. We did not want to use events relating to one continuous variable so as to avoid distorting different perceptions of convex versus nonconvex unions. We neither wanted to use events that would arouse emotions, such as events pertaining to soccer, the most popular sport in our country.

Had we assumed concave instead of linear utility, even if utility were assumed known such as in  $v(x)=x^{0.88}$ , then the calculations would have been considerably more intricate. Then decision weights do not cancel from equations any more so that an indifference only gives an equation with several decision weights as unknowns. Then solutions and approximative solutions to complex linear equalities would have been required, with complex data fittings. Convenient techniques for parametric fitting for uncertainty, when weighting functions need not be transforms of given probabilities, are yet to be developed. Hence we propose our method at present only when it is reasonable to assume linear utility. Many references have argued for linear utility for small stakes (Edwards 1955; Fox et al. 1996; Lopes and Oden 1999 p. 290; Luce 2000 p. 86; Rabin 2000; Ramsey 1931 p. 176; Savage 1954 p. 91). According to modern theories, risk aversion for moderate stakes (such as in Holt and Laury 2002), will be caused by factors other than utility curvature, such as the decision weights studied in this paper. With concave utility, the resulting decision weights would have been higher than in our calculations, so that utility curvature cannot account for the sums in Table 1 exceeding 1.

To further clarify the distentanglement of utility and risk attitude that is central in our extension of de Finetti's method, we refer to Chateauneuf and Cohen (1994, Corollary 2 on p. 86) and Abdellaoui et al. (2007). The former demonstrated that it is theoretically possible to have risk aversion with strictly *convex* utility, and the latter found this phenomenon empirically in an experiment with loss outcomes.

Many empirical studies have found that the local curvature of utility is most nonlinear around zero (Tversky and Kahneman 1992), a phenomenon incorporated in the most commonly used parametric utility family, the CRRA (logpower) family, which usually has infinite derivative at zero. Our outcomes are all remote from zero, with minimal outcome €4, so as to have approximately linear utility. An additional reason for avoiding the zero outcome is that it induces several biases in the evaluation of prospects (Birbaum et al. 1992). Preference foundations of de Finetti's betting-odds system for prospect theory with linear utility are in Chateauneuf (1991) and Diecidue and Wakker (2002).

## 8 Conclusion

Using de Finetti's betting-odds system, we developed a method for eliciting decision weights for prospect theory and rank-dependent utility. We found evidence for rank dependence of the decision weights. This finding constitutes a deviation from the classical Bayesian model for eliciting subjective beliefs in ambiguous events. Regarding the direction of the deviation, there was much individual variation, with as much evidence for likelihood-insensitivity as for pessimism. Both effects seem to play a role. Given the widespread use of belief elicitation, virtually always based on Bayesian principles, it is important that the deviations from Bayesianism be identified, so that more accurate estimations can be obtained of the beliefs of financial experts, players in games, and so on.

## References

- Abdellaoui, Mohammed. (2000). "Parameter-free Elicitation of Utilities and Probability Weighting Functions," *Management Science* 46, 1497–1512.
- Abdellaoui, Mohammed, Han Bleichrodt, and Olivier L'Haridon. (2007). *A Tractable Method to Measure Utility and Loss Aversion under Prospect Theory*. iMTA/iBMG. Rotterdam, The Netherlands: Erasmus University.
- Abdellaoui, Mohammed, Frank Vossman, and Martin Weber. (2005). "Choice-based Elicitation and Decomposition of Decision Weights for Gains and Losses under Uncertainty," *Management Science* 51, 1384–1399.
- Allais, Maurice. (1953). "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Américaine," *Econometrica* 21, 503–546.
- Armantier, Olivier. (2006). "Do Wealth Differences Affect Fairness Considerations," *International Economic Review* 47, 391–429.
- Barron, Greg, and Ido Erev. (2003). "Small Feedback-based Decisions and their Limited Correspondence to Description-based Decisions," *Journal of Behavioral Decision Making* 16, 215–233.
- Birnbaum, Michael H. (2006). "Evidence against Prospect Theories in Gambles with Positive, Negative, and Mixed Consequences," *Journal of Economic Psychology* 27, 737–761.
- Birnbaum, Michael H., Gregory Coffey, Barbara A. Mellers, and Robin Weiss. (1992). "Utility Measurement: Configural-weight Theory and the Judge's Point of View," *Journal of Experimental Psychology: Human Perception and Performance* 18, 331–346.
- Birnbaum, Michael H., and Laura A. Thompson. (1996). "Violations of Monotonicity in Choices between Gambles and Certain Cash," *American Journal of Psychology* 109, 501–523.
- Bleichrodt, Han and José Luis Pinto. (2000). "A Parameter-free Elicitation of the Probability Weighting Function in Medical Decision Analysis," *Management Science* 46, 1485–1496.
- Borel, Émile. (1924). "A Propos d'un Traité de Probabilités," *Revue Philosophique* 98, 321–336.
- Camerer, Colin F., and Martin Weber. (1992). "Recent Developments in Modelling Preferences: Uncertainty and Ambiguity," *Journal of Risk and Uncertainty* 5, 325–370.
- Chateauneuf, Alain. (1991). "On the Use of Capacities in Modeling Uncertainty Aversion and Risk Aversion," *Journal of Mathematical Economics* 20, 343–369.
- Chateauneuf, Alain, and Michèle Cohen. (1994). "Risk Seeking with Diminishing Marginal Utility in a Non-expected Utility Model," *Journal of Risk and Uncertainty* 9, 77–91.
- Chateauneuf, Alain, Jürgen Eichberger, and Simon Grant. (2005). "Choice under Uncertainty with the Best and Worst in Mind: NEO-additive Capacities," CERMSEM, CEM, University of Paris I.
- Chew, Soo Hong, and William S. Waller. (1986). "Empirical Tests of Weighted Utility Theory," *Journal of Mathematical Psychology* 30, 55–72.
- Clemen, Robert T., and Kenneth C. Lichtendahl. (2005). *Debiasing Expert Overconfidence: A Bayesian Calibration Model*. Durham, NC: Fuqua School of Business, Duke University.
- de Finetti, Bruno. (1931). "Sul Significato Soggettivo della Probabilità," *Fundamenta Mathematicae* 17, 298–329. Translated into English as "On the Subjective Meaning of Probability." In Paola Monari and Daniela Cocchi (eds.), "*Probabilità e Induzione*," Clueb, Bologna, 1993.

- Diecidue, Enrico, and Peter P. Wakker. (2002). “Dutch Books: Avoiding Strategic and Dynamic Complications, and a Comonotonic Extension,” *Mathematical Social Sciences* 43, 135–149.
- Dow, James, and Sérgio R.C. Werlang. (1992). “Uncertainty Aversion, Risk Aversion and the Optimal Choice of Portfolio,” *Econometrica* 60, 197–204.
- Edwards, Ward. (1955). “The Prediction of Decisions Among Bets,” *Journal of Experimental Psychology* 50, 201–214.
- Einhorn, Hillel J., and Robin M. Hogarth. (1985). “Ambiguity and Uncertainty in Probabilistic Inference,” *Psychological Review* 92, 433–461.
- Ellsberg, Daniel. (2001). *Risk, Ambiguity and Decision*. New York: Garland. Original Ph.D. dissertation: Ellsberg, Daniel. (1962). *Risk, Ambiguity and Decision*. Cambridge, MA: Harvard University.
- Fox, Craig R., Brett A. Rogers, and Amos Tversky. (1996). “Options Traders Exhibit Subadditive Decision Weights,” *Journal of Risk and Uncertainty* 13, 5–17.
- Fox, Craig R., and Amos Tversky. (1998). “A Belief-based Account of Decision under Uncertainty,” *Management Science* 44, 879–895.
- Gilboa, Itzhak, and David Schmeidler. (1989). “Maxmin Expected Utility with a Non-unique Prior,” *Journal of Mathematical Economics* 18, 141–153.
- Goeree, Jacob K., Charles A. Holt, and Thomas R. Palfrey. (2002). “Quantal Response Equilibrium and Overbidding in Private-value Auctions,” *Journal of Economic Theory* 104, 247–272.
- Gonzalez, Richard, and George Wu. (1999). “On the Shape of the Probability Weighting Function,” *Cognitive Psychology* 38, 129–166.
- Gonzalez, Richard, and George Wu. (2003). “Composition Rules in Original and Cumulative Prospect Theory,” mimeo.
- González-Vallejo, Claudia C., Aaron A. Reid, and Joel Schiltz. (2003). “Context Effects: The Proportional Difference Model and the Reflection of Preference,” *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29, 942–953.
- Hamao, Yasushi, Ronald W. Masulis, and Victor Ng. (1990). “Correlations in Price Changes and Volatility across International Stock Markets,” *The Review of Financial Studies* 3, 281–307.
- Harbaugh, William T., Kate Krause, and Lise Vesterlund. (2002). “Risk Attitudes of Children and Adults: Choices over Small and Large Probability Gains and Losses,” *Experimental Economics* 5, 53–84.
- Harrison, Glenn W., Morten I. Lau, and E. Elisabet Rutström. (2007). “Estimating Risk Attitudes in Denmark: A Field Experiment,” *Scandinavian Journal of Economics* (forthcoming).
- Harrison, Glenn W., Morten I. Lau, and Melonie B. Williams. (2002). “Estimating Individual Discount Rates in Denmark: A Field Experiment,” *American Economic Review* 92, 1606–1617.
- Holt, Charles A. (1986). “Preference Reversals and the Independence Axiom,” *American Economic Review* 76, 508–513.
- Holt, Charles A., and Susan K. Laury. (2002). “Risk Aversion and Incentive Effects,” *American Economic Review* 92, 1644–1655.
- Hull, John C. (2005). *Options, Futures, and Other Derivatives* (6th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Humphrey, Stephen J. (1995). “Regret Aversion or Event-Splitting Effects? More Evidence under Risk and Uncertainty,” *Journal of Risk and Uncertainty* 11, 263–274.
- Kahneman, Daniel, and Amos Tversky. (1979). “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica* 47, 263–291.
- Karni, Edi. (1996). “Probabilities and Beliefs,” *Journal of Risk and Uncertainty* 13, 249–262.
- Keynes, John Maynard. (1921). *A Treatise on Probability*. (2nd ed. London: McMillan, 1948.)
- Knight, Frank H. (1921). *Risk, Uncertainty, and Profit*. New York: Houghton Mifflin.
- Lopes, Lola L., and Gregg C. Oden. (1999). “The Role of Aspiration Level in Risky Choice: A Comparison of Cumulative Prospect Theory and SP/A Theory,” *Journal of Mathematical Psychology* 43, 286–313.
- Luce, R. Duncan. (2000). *Utility of Gains and Losses: Measurement-Theoretical and Experimental Approaches*. London: Erlbaum.
- Luce, R. Duncan, and Peter C. Fishburn. (1991). “Rank- and Sign-dependent Linear Utility Models for Finite First-order Gambles,” *Journal of Risk and Uncertainty* 4, 29–59.
- MacCrimmon, Kenneth R., and Stig Larsson. (1979). “Utility Theory: Axioms versus ‘Paradoxes.’” In Maurice Allais and Ole Hagen (eds.), *Expected Utility Hypotheses and the Allais Paradox* 333–409, Reidel, Dordrecht: The Netherlands.
- McCord, Mark R., and Richard de Neufville. (1986). “‘Lottery Equivalents’: Reduction of the Certainty Effect Problem in Utility Assessment,” *Management Science* 32, 56–60.
- Nau, Robert F. (1995). “Coherent Decision Analysis with Inseparable Probabilities and Utilities,” *Journal of Risk and Uncertainty* 10, 71–91.

- Neilson, William S., and Jill Stowe. (2001). "A Further Examination of Cumulative Prospect Theory Parameterizations," *Journal of Risk and Uncertainty* 24, 31–46.
- Nyarko, Yaw, and Andrew Schotter. (2002). "An Experimental Study of Belief Learning Using Elicited Beliefs," *Econometrica* 70, 971–1005.
- Pfanzagl, Johann. (1959). "A General Theory of Measurement-Applications to Utility," *Naval Research Logistics Quarterly* 6, 283–294.
- Quiggin, John. (1981). "Risk Perception and Risk Aversion Among Australian Farmers," *Australian Journal of Agricultural Economics* 25, 160–169.
- Rabin, Matthew. (2000). "Risk Aversion and Expected-utility Theory: A Calibration Theorem," *Econometrica* 68, 1281–1292.
- Ramsey, Frank P. (1931). "Truth and Probability." In *The Foundations of Mathematics and other Logical Essays*. London: Routledge.
- Savage, Leonard J. (1954). *The Foundations of Statistics*. New York: Wiley.
- Schmeidler, David. (1989). "Subjective Probability and Expected Utility without Additivity," *Econometrica* 57, 571–587.
- Shiller, Robert J., Fumiko Kon-Ya, and Yoshiro Tsutsui. (1996). "Why Did the Nikkei Crash? Expanding the Scope of Expectations Data Collection," *The Review of Economics and Statistics* 78, 156–164.
- Starmer, Chris. (1999). "Cycling with Rules of Thumb: An Experimental Test for a New Form of Non-transitive Behavior," *Theory and Decision* 46, 141–158.
- 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, and Robert Sugden. (1991). "Does the Random-lottery Incentive System Elicit True Preferences? An Experimental Investigation," *American Economic Review* 81, 971–978.
- Tversky, Amos, and Craig R. Fox. (1995). "Weighing Risk and Uncertainty," *Psychological Review* 102, 269–283.
- Tversky, Amos, and Daniel Kahneman. (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5, 297–323.
- Tversky, Amos, and Peter P. Wakker. (1995). "Risk Attitudes and Decision Weights," *Econometrica* 63, 1255–1280.
- Viscusi, W. Kip. (1989). "Prospective Reference Theory: Toward an Explanation of the Paradoxes," *Journal of Risk and Uncertainty* 2, 235–264.
- Viscusi, W. Kip, and William N. Evans. (2006). "Behavioral Probabilities," *Journal of Risk and Uncertainty* 32, 5–15.
- Wald, Abraham. (1950). *Statistical Decision Functions*. New York: Wiley.
- Weber, Elke U., and Britt Kirsner. (1997). "Reasons for Rank-dependent Utility Evaluation," *Journal of Risk and Uncertainty* 14, 41–61.
- Winkler, Robert L. (1972). *An Introduction to Bayesian Inference and Decision Theory*. New York: Holt, Rinehart and Winston.
- Wu, George, and Richard Gonzalez. (1999). "Nonlinear Decision Weights in Choice under Uncertainty," *Management Science* 45, 74–85.
- Wu, George, Jiao Zhang, and Mohammed Abdellaoui. (2005). "Testing Prospect Theories Using Tradeoff Consistency," *The Journal of Risk and Uncertainty* 30, 107–131.