JOURNAL OF MULTI-CRITERIA DECISION ANALYSIS
J. Multi-Crit. Decis. Anal. **17**: 101–113 (2011)
Published online 10 March 2011 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/mcda.454

Comonotonic Proper Scoring Rules to Measure Ambiguity and Subjective Beliefs†

AMIT KOTHIYAL, VITALIE SPINU and PETER P. WAKKER*
Econometric Institute, Erasmus University, Rotterdam, The Netherlands

ABSTRACT

Proper scoring rules serve to measure subjective degrees of belief. Traditional proper scoring rules are based on the assumption of expected value maximization. There are, however, many deviations from expected value, primarily due to risk aversion. Correcting techniques have been proposed in the literature for deviations due to nonlinear utility. These techniques still assumed expected utility maximization. More recently, corrections for deviations from expected utility have been proposed. The latter concerned, however, only the quadratic scoring rule, and could handle only half of the domain of subjective beliefs. Further, beliefs close to 0.5 could not be discriminated. This paper generalizes the correcting techniques to all bounded binary proper scoring rules, covers the whole domain of beliefs and, in particular, can discriminate between all degrees of belief. Thus, we fully extend the properness requirement (in the sense of identifying all degrees of subjective beliefs) to virtually all models that deviate from expected value. Copyright © 2011 John Wiley & Sons, Ltd.

KEY WORDS: proper scoring rules; subjective beliefs; ambiguity; nonexpected utility

1. INTRODUCTION

Proper scoring rules are cleverly devised optimization problems that serve to efficiently measure subjective degrees of beliefs. Their original introduction, and almost exclusive use up to today, assumed expected value maximization. However, numerous deviations from expected value have been documented, due to risk aversion and other factors. Winkler and Murphy (1970) analysed deviations due to risk aversion, but still assumed expected utility (with risk aversion then captured by nonlinear, concave utility). In view of the many deviations from expected utility (Starmer, 2000; Gilboa, 2004), Offerman *et al.* (2009; abbreviated OSVW henceforth) extended Winkler and Murphy's analysis to also incorporate the latter deviations.

The purpose of OSVW was to analyse the most popular proper scoring rule, the quadratic one, as it is mostly applied, so as to clarify what problems arise in those applications according to modern decision theories. OSVW provided corrections for those existing applications to the extent possible. Some problems, however, are impossible to resolve for traditional proper scoring rules once expected utility is abandoned. These problems, pointed out by OSVW (Appendix A and p. 1483 penultimate paragraph), are discussed in detail in Sections 4 and 7. In brief, the first problem is that a function W (from which beliefs will be derived) can be measured only on half of its domain, being only the events that are more likely than their complements. Without further assumptions, we cannot observe W for events less likely than their complements. We call this first problem the *half-domain problem*, or the *H-problem*.

The second problem is that traditional proper scoring rules lose their discriminatory power for subjective beliefs around 0.5. That is, there is an interval of degrees of belief around 0.5 where the scoring rules give the same optimal reported probability ('fifty-fifty') for all those degrees of belief. The latter phenomenon, theoretically predicted by nonexpected utility, is confirmed by

*Correspondence to: Econometric Institute, Erasmus University, P.O. Box 1738, Rotterdam, 3000 DR, The Netherlands. E-mail: Wakker@ese.eur.nl

†This paper, especially Section 2, is based on the plenary lecture 'Extracting knowledge using decision principles: proper scoring rules reconsidered from the perspective of modern decision theories,' by Peter P. Wakker, 21 October 2009, ADT (First International Conference on Algorithmic Decision Theory), Venice, Italy. A referee made very helpful comments.

Copyright © 2011 John Wiley & Sons, Ltd.

Received 12 April 2010
Accepted 24 November 2010

empirical studies. A common finding there is that a large proportion of reported beliefs is exactly 0.5. It is not plausible that there will be so many true beliefs of exactly and precisely 0.5 at exactly the event considered. It is more plausible that people with true beliefs in a considerable region around 0.5 all report 0.5. The level of 0.5 serves as a magnet so to say (Andersen *et al.*, 2009). Then a reported probability of 0.5 relates to many degrees of true belief, and cannot be used to uniquely find back the true belief. If properness is taken to mean that all degrees of belief can be identified, then traditional scoring rules are not even proper under theories more general than expected value, strictly speaking, because they cannot discriminate between the aforementioned degrees of belief around 0.5. We call the second problem the *identification-problem*, or the *I-problem*.

Regarding the H-problem, OSVW specified a symmetry condition for beliefs (binary additivity, p. 1484) under which measurements from the half of the domain that can be covered give complete information about the other half. OSVW used this condition in their experiment (p. 1470, 2nd paragraph; see also the symmetry about 0.5 in their Figure 1, reproduced as the risk corrected curve in our Figure 4). If binary additivity is violated, however, then half of the domain remains unobservable (see OSVW's p. 1484). Regarding the I-problem, in OSVW's experiment, the lack of discriminatory power around 0.5 did not occur prominently (end of Section 8.1.1). OSVW suggested a modification that can avoid the loss of half of the domain also if binary additivity is violated (their Appendix A), but left its examination to future studies. This paper provides an analysis of their suggestion and gives generalizations.

Unlike OSVW, we do not focus on some traditional versions of quadratic proper scoring rules that have been popular in applications so far. Instead, our analysis applies to all bounded proper scoring rules. We extend the theoretical results of OSVW for quadratic scoring rules to general proper scoring rules, and nontraditional scoring rules are presented that avoid the aforementioned problems. Consequently, the latter scoring rules extend properness from expected value to general decision models. They are truly proper in the sense of fully identifying subjective beliefs (including ambiguity) over their whole domain, for most of the currently popular decision models.

The main contribution of OSVW was to propose a correction of traditional proper scoring rules for violations of expected value (risk neutrality) that works on the half of the domain where their scoring rule is proper. They called this correction the risk correction. The cause underlying the aforementioned H- and I-problems is that traditional proper scoring rules do not satisfy Schmeidler's (1989) condition called comonotonicity. We extend OSVW's correction to the whole domain of beliefs mainly by ensuring that Schmeidler's comonotonicity is satisfied. The latter can be seen to preserve ranks of events. Hence, our method will be called the rank correction.

Section 2 describes some history of proper scoring rules from a personal perspective. Basic definitions are in Section 3, with an example illustrating the difficulties of traditional proper scoring rules under nonexpected utility in Section 4. This example also suggests the direction of solution. Section 5 presents the main theorems of this paper, generalizing the results of OSVW. Section 6 shows how these theorems can be applied to measure subjective beliefs in a tractable manner. Similar to OSVW, we need no utility measurement but, unlike OSVW, we obtain full discrimination. A discussion is in Section 7, and Section 8 concludes.

2. PROPER SCORING RULES AS AN APPEALING MULTICRITERIA OPTIMIZATION PROBLEM; HISTORY AND EXAMPLE

This section is less formal than the rest of this paper, and is close to the lecture referred to in the opening footnote. It can be skipped by readers only interested in theory or applications. To take an example examined in this paper, assume that you want to know very precisely how likely it is that there will be rain tomorrow according to a weather forecaster. The event of rain tomorrow is denoted E. You can simply ask directly, such as 'What is your subjective probability of E?' Problems are, first, that many people cannot or will not relate to subjective probabilities, especially for unique one-shot events as rain on the calendar day that tomorrow is. Second, even if the forecaster understands and accepts the concept of subjective probability, there may be no clear incentive for her to give a meaningful and truthful answer.

To find out about the ideas of the weather forecaster, we turn to a multicriteria optimization

Copyright © 2011 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

problem that will turn out to have very appealing properties. Imagine that we ask the weather forecaster to choose a number r between 0 and 1, telling her that she will be rewarded by $1-(1-r)^2$ if E (rain tomorrow), and by $1-r^2$ if no rain. This rewarding scheme is called the quadratic scoring rule. The problem presented to the weather forecaster entails a two-attribute optimization problem. The weather forecaster would like to choose r so as to maximize both the outcome $(1-(1-r)^2)$ under E and the outcome $(1-r^2)$ under the complementary event E^c. The former outcome increases in r and the latter one decreases in r, so that tradeoffs between the two attributes have to be made, as in all nontrivial multicriteria optimization problems. What is an optimal choice of r for the weather forecaster?

Imagine that the weather forecaster has a subjective probability p of rain tomorrow, consciously or subconsciously. Assume that she maximizes expected value; i.e. expected value is her goal function (possibly in an 'as if' sense). Then r is chosen so as to maximize

$$p(1-(1-r)^2)+(1-p)(1-r^2).$$

The first-order optimality condition is as follows:

$$2p(1-r)-2r(1-p)=0,$$

implying

$$r=p. \tag{1}$$

The algebra is so simple that the readers may wonder why we claim that proper scoring rules are among the most appealing multicriteria optimization problems presently existing. Yet such is our claim. The conceptual implications of the above finding are breathtaking, as is explained next.

Equation (1) shows that it is in the self-interest of the weather forecaster (assuming expected value for now) to truthfully report her subjective probability. Subjective probabilities describe the subjective degree of belief in cases where a person is not sure and has lack of information regarding whether the event E will happen. We all lack information in virtually all of our decisions. No one knows for sure what the weather will be the next day. Still we have to make decisions contingent on it. That we have degrees of belief, even precise and quantitative degrees of belief, may seem to be like science fiction. According to many today, including many frequentist statisticians, subjective probabilities are nonexisting and meaningless. Proper scoring rules can serve as a powerful antidote against such views. Not only do subjective probabilities exist for rational decision makers (being their reported probabilities in proper scoring rules) but even more, they are easy to measure and to observe. To the extent that a strict frequentist statistician will not report her true subjective probability in a proper scoring rule, she is only harming herself.

Proper scoring rules are, essentially, a devise for reading the minds of people. Brier (1950) introduced proper scoring rules. Bruno de Finetti, one of the greatest thinkers of the past century and the most important contributor to the concept of subjective probability (de Finetti, 1931, 1937), independently invented proper scoring rules (de Finetti, 1962). The independence of his discovery is credible given that he could not read or speak English in those days. This independence was confirmed by Savage (1971, p. 783). Further, de Finetti invented many great ideas independently and often before others who became more known for them later. For example, the famous index of risk aversion and concave utility, $-U''/U'$, commonly called the Pratt-Arrow index of risk aversion after Pratt (1964) and Arrow ([1965, 1971), first appeared in de Finetti (1952, p. 700/701). We can now add another point to de Finetti's score: He was, together with Brier, the first neuroeconomist. They could read the minds of people without needing expensive machines, and they could measure quantities (degrees of belief) more interesting than the neuronal activities in the brain studied by modern neuroeconomics.

In mainstream economics, private information of an agent that she has but that others do not have is often modelled as the type of the agent. The subjective probability of the weather forecaster can then be taken as her type. Incentive compatible mechanisms are games where it is in the self-interest of an agent to reveal her true type, and where the agent has no interest in manipulating and misrepresenting her private information. Such mechanisms are of great importance for efficiently organizing a society, for instance for taxation. Given this importance, Hurwicz (1960), Maskin, and Myerson received the 2007 'Nobel' prize in economics for this idea. However, Brier's (1950) proper scoring rule preceded Hurwicz (1960) by a decade.

Allowing to read the minds of people, supporting the Bayesian approach to statistics, dominating modern neuroeconomics historically, financially, and outputwise, and preceding a Nobel prize discovery by 10 years should be enough to

Copyright © 2011 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

qualify as one of the most appealing multi-criteria optimization problem presently existing. Algebraic simplicity for obtaining all this is an additional pro. When the third author, as a bachelor's student, and exposed to purely frequentist statistics teachers, discovered proper scoring rules through the inspiring text of de Finetti (1962), this determined his academic career.

An interesting application of proper scoring rules is in Tetlock (2005). For many years, he interviewed leading politicians, submitting their opinions to proper scoring rules. Then, years later, he could draw many interesting conclusions. Such an application had been suggested before by Hanson (2002) and Hofstee (1988).

3. BASIC DEFINITIONS

Let E denote an event of which it is uncertain whether or not it is true, such as rain tomorrow. E^c denotes the complementary event. A *prospect* $\alpha_E\beta$ is a function from $\{E, E^c\}$ to $\mathbb{R}^+$, assigning *outcome* α to E and outcome β to E^c. That is, in the example about rain tomorrow, it yields \$$\alpha$ if there will be rain tomorrow, and \$$\beta$ if there will be no rain tomorrow. Section 2 discussed proper scoring rules from a normative perspective, under the traditional assumption of expected value maximization. The rest of this paper analyzes proper scoring rules under descriptively more realistic models. We focus on binary scoring rules with the outcome depending on whether an event or its complement is true. We also focus on nonnegative outcomes. For negative outcomes, loss aversion and sign-dependent weighting have to be incorporated (Tversky and Kahneman, 1992; Wakker, 2010), but we leave this to future studies. In most applications, negative outcomes cannot be implemented so that this case is less important.

Virtually all presently existing decision theories (Appendix A), including expected value and expected utility, evaluate the prospect $\alpha_E\beta$ by:

$$\begin{aligned}
&\text{If } \alpha > \beta; \quad \text{then } \alpha_E\beta \mapsto W(E)U(\alpha) + (1 - W(E))U(\beta);\\
&\text{If } \alpha = \beta \quad \text{then } \alpha_E\beta \mapsto U(\alpha)(= U(\beta));\\
&\text{If } \alpha < \beta \quad \text{then } \alpha_E\beta \mapsto W(E^c)U(\beta) + (1 - W(E^c))U(\alpha).
\end{aligned} \tag{2}$$

Here $U : \mathbb{R} \to \mathbb{R}$ is the *utility function*; it is assumed continuously differentiable with positive derivative everywhere. W, the *weighting function,* is a set function with $0 \leqslant W(E) \leqslant 1$ for all E and $E \supset E' \Rightarrow W(E) \geqslant W(E')$. When choosing between different prospects, a decision maker chooses the one that maximizes the above evaluation. We call this decision model *binary rank-dependent utility*, abbreviated *binary RDU*, or just *RDU*. If $\alpha = \beta$, then we can also use the formula for the case $\alpha > \beta$ or for the case $\alpha < \beta$, and all these three formulas give the same result.

The different weighting of events under different orderings of outcomes is referred to as rank dependence. It was Gilboa's (1987) and Schmeidler's (1989, first version 1982) key idea for getting a sound decision theory for nonadditive beliefs. The same basic idea had been invented independently by Quiggin (1982) for the special case of risk (known probabilities). W gives the weight of events when they yield the best outcome. A pessimist will overweight the worst outcome and assign low W values to events. For an optimist it will be the other way around. For two events E, E^c, *pessimism* means that $W(E) + W(E^c) \leqslant 1$. It is psychologically plausible that people weigh events differently when the events yield favourable outcomes than when they yield unfavourable outcomes (Wakker, 2010 Section 6.4). This explains the empirical success of rank-dependent theories.

If W is a probability measure, then the traditional *expected utility* model results. Then the formula for $\alpha > \beta$ agrees with that for $\alpha < \beta$ in Equation (2), and both these formulas can be used for all cases. That is, we then need not invoke rank dependence. If, further, U is linear then *expected value* holds. In view of the many empirical violations of expected utility, several generalizations have been considered. Because most of these generalizations agree with binary rank-dependent utility on the domain of two-outcome nonnegative prospects considered here, our analysis is valid under all those theories.

Proper scoring rules concern prospects $S_E(r)_E S_{E^c}(r)$. A subject chooses a real number $0 \leqslant r \leqslant 1$, called *reported probability of E*, and then receives an outcome depending not only on r but also on whether or not event E obtains. The functions S_E and S_{E^c}(chosen by the experimenter) describe the dependence of the outcome received on E and E^c. This pair of functions, and the prospects $S_E(r)_E S_{E^c}(r)$ that they generate for each r, are called a *scoring rule*. We assume that S_E and S_{E^c} are continuously differentiable. We also assume that binary rank-dependent utility holds;

Copyright © 2011 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

i.e. r maximizes the RDU value of $S_E(r)_E S_{E^c}(r)$. Given continuity of the RDU functional and compactness of r's domain [0,1], an optimal r always exists. There may in general exist several optima, in which case one optimum is arbitrarily selected to be the reported probability r.

Definition 1

A scoring rule is *proper* if the following implication holds: If U is linear and W is a probability measure with $W(E)=p$ (expected value maximization), then the optimal value of r is unique and it is p.[1]

Section 1 presented an example of a proper scoring rule. We summarize the formal assumptions of our model.

Assumption 2 [Structural Assumption]

E denotes an event, E^c its complement, and $\alpha_E\beta$ a prospect ($\alpha \geqslant 0$ and $\beta \geqslant 0$). Prospects are evaluated by Equation (2) (binary rank-dependent utility), with U continuously differentiable with positive derivative everywhere, and W a weighting function. r is the reported probability, maximizing Equation (2) over prospects generated by the scoring rule $S_E(r)_E S_{E^c}(r)$. S_E and S_{E^c} are continuously differentiable. The scoring rule $S_E(r)_E\ S_{E^c}(r)$ is proper.

The following well-known result will often be used. Its proof, and all other proofs, are in Appendix B.

Lemma 3

$S_E(r)$ is strictly increasing in r and $S_{E^c}(r)$ is strictly decreasing in r

4. AN EXAMPLE ILLUSTRATING THE TWO PROBLEMS OF TRADITIONAL PROPER SCORING RULES

This section illustrates the main analytical difficulties of traditional proper scoring rules under nonexpected utility, and then suggests a solution. This section can be skipped by readers specialized in rank-dependent theories and by readers only interested in applications. We first give a numerical example, and then discuss it.

Example 4

Assume $U(\alpha)=\alpha$, and $W(E^c)<W(E)<0.5$. The weights $W(E)$ and $W(E^c)$ add to less than 1, which is typical of nonexpected utility with a nonadditive W and pessimism. $W(E)>W(E^c)$ suggests that E is subjectively more likely than its complement. Assume the *quadratic scoring rule* $(1-(1-r)^2)_E(1-r^2)$ that is discussed in Section 2. For $r=0.5$, it gives a constant score, 0.75 for both events. Its RDU value then is, accordingly, 0.75. The calculations given in the rest of this example show that $r=0.5$ gives the optimal value. In the main text following the example we discuss the result in intuitive terms, which may be more convenient for readers who want to skip algebra as given next.

If we increase r by a small $\varepsilon>0$ then the highest score is obtained under event E; i.e. for $r>0.5$ we have $S_E(r)=1-(1-r)^2>1-r^2=S_{E^c}(r)$. Accordingly, the first part of Equation (2) applies (with $\alpha=S_E(r)>\beta=S_{E^c}(r)$). The change in RDU value is approximately

$$W(E)S'_E(0.5)\varepsilon+(1-W(E))S'_{E^c}(0.5)\varepsilon \\ =(W(E)-(1-W(E))S'_E(0.5)\varepsilon<0.$$

The equality follows from substituting $S'_E(0.5)=-S'_{E^c}(0.5)>0$, and the inequality follows from substituting $W(E)-(1-W(E))<0$.

If we decrease r by a small $\varepsilon>0$ then the highest score is obtained under event E^c. Accordingly, the last part of Equation (2) applies. The change now is approximately

$$W(E^c)S'_{E^c}(0.5)(-\varepsilon)+(1-W(E^c))S'_E(0.5)(-\varepsilon) \\ =(W(E^c)-(1-W(E^c))S'_E(0.5)\varepsilon<0.$$

We substituted $W(E^c)-(1-W(E^c)<0$. The move away from certainty by decreasing r leads to a prospect with a considerably lower RDU value, both because E^c is less likely than E and because of pessimism.

It follows that $r=0.5$ is a local optimum. Because RDU is not differentiable at this r, first-order conditions do not apply. It can be seen that $r=0.5$ is in fact a global optimum. If we assume nonlinear utility, then the changes in RDU above have to be multiplied by $U'(0.75)$ which does not change signs. Hence $r=0.5$ then still is a local optimum. Concavity of utility reinforces risk

[1]Our formal definition is the same as the classical definition, with properness only referring to expected value maximization. This paper shows in fact how such rules can be extended to larger classes of decision models while still eliciting beliefs correctly. The most essential property of proper scoring rules is that all degrees of belief are uniquely related to reported probabilities through a known function, so that beliefs are identifiable from observed reported probabilities. We discussed this essential property in the introduction in the context of the I-problem; see also the end of Section 5.

Copyright © 2011 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

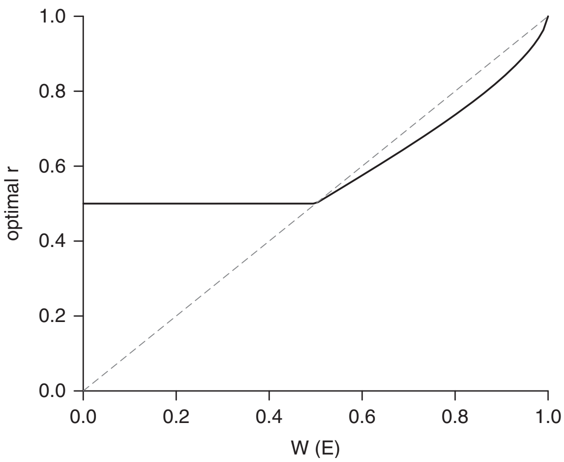

Figure 1. The solid line depicts the optimal reported probability r as a function of $W(E)$, assuming $U(x) = \sqrt{x}$ and $W(E^c) < W(E)$ or $W(E^c) < 0.5$.

aversion and a preference for $r = 0.5$. Hence, it readily follows that $r = 0.5$ is a global optimum for every concave utility function. Figure 1 depicts the optimal r as a function of $W(E)$ for square-root utility. The figure is valid whenever $W(E^c) < W(E)$ and, in fact, whenever $W(E^c) < 0.5$.

We continue to consider the quadratic scoring rule as in the above example. Under expected value and expected utility, an event that is more likely than its complement also has a weight exceeding 0.5, and then it is always favourable to increase r above 0.5. The above example demonstrates that things are different under nonexpected utility. Then for all events E that are more likely than their complements but that yet have their weight below 0.5 it is not favourable to increase r above 0.5. In this case, pessimism generates too much preference for certainty, and $r = 0.5$ is the optimal reported probability for all these events. Under pessimism, which is commonly found, there will be a range of events E like this (both $W(E) < 0.5$ and $W(E^c) < 0.5$), illustrated by the flat part in Figure 1. For all of them, $r = 0.5$ is optimal. The reported probability $r = 0.5$ does not discriminate between the degrees of beliefs of these events. The most essential property of properness, the possibility to discriminate between all levels of belief, then is lost. This loss of discrimination is the cause of the I-problem. The magnet-optimum at $r = 0.5$ that we have established is related to phenomena known as inertia and the bid–ask spread. All these phenomena can be explained by rank dependence through its implication that is known as first-order risk aversion (Wakker, 2010 Example 6.6.1).

With regard to the H-problem, it readily follows in the above example that $r > 0.5$ can be optimal only if $W(E) \geqslant W(E^c)$, and $r < 0.5$ can be optimal only if $W(E^c) \geqslant W(E)$.[2] Thus, the highest payment can only result under the most likely event (OSVW Observation A1). We can never observe the weight $W(E)$ of an event E that is less likely than its complement because $W(E)$ only plays a role if E receives the better outcome. This illustrates the H-problem.

Both problems just explained are implied by the rank-dependent nature of Equation (2), where the decision weights change in a drastic nonsmooth manner if the ordering of the scores for the two events changes. Such changes of orderings of outcomes occur for traditional proper scoring rules such as the quadratic scoring rule and, more generally, for all scoring rules that have $S_E(r) = S_{E^c}(r)$ for some $0 < r < 1$.[3] We can avoid these changes in weights, and the corresponding analytical problems, if we use scoring rules for which the ordering of outcomes is the same for all r; for instance, if $S_E(r) \geqslant S_{E^c}(r)$ for all r. In Schmeidler's (1989) terminology, we then consider a comonotonic set of prospects. This will be the plan of the rank correction defined later. We first present some theorems.

5. MAIN RESULTS

The following result generalizes Theorem 1 of OSVW.

Lemma 5

Suppose that Structural Assumption 2 holds. For all reported probabilities r with $S_E(r) > S_{E^c}(r)$, and also for $r = 0$ if $S_E(0) = S_{E^c}(0)$, we have

$$r = \frac{W(E)}{W(E) + (1 - W(E))\dfrac{U'(S_{E^c}(r))}{U'(S_E(r))}}. \tag{3}$$

The expression for r in Equation (3) is not explicit because r appears in both sides of the equality. It is, in general, possible that for one value of $W(E)$ several values of r satisfy Equation (3). The following example illustrates this point.

[2] If $W(E) < W(E^c)$ then each $r > 0.5$ is strictly dominated by $1 - r$.

[3] By Lemma 3, E then is ranked best for $r' > r$ but E^c is ranked best for $r' < r$.

Copyright © 2011 John Wiley & Sons, Ltd.

Example 6
Assume the quadratic scoring rule and expected utility. Assume that the utility function U is very convex, which implies strong risk seeking. Assume $W(E) = W(E^c) = 0.5$. Then the decision maker will dislike the certainty resulting from $r = 0.5$ (yielding a sure outcome 0.75), and will prefer the risk resulting from moving r up or down somewhat. For example, if $U(x) = e^{2.5x}$, then $r = 0.14$ and $r = 0.86$ are optimal (Figure 2; OSVW p. 1486). Then the reported probability r is selected to be one of these two numbers. Equation (3) then holds for both these numbers. It in fact also holds for $r = 0.5$, where there is however a local minimum.

We can obtain an explicit expression of W in terms of r, generalizing Corollary 2 of OSVW.

Corollary 7
Assume Structural Assumption 2, with r as in Lemma 5. Then we have for this r:

$$W(E) = \frac{r}{r + (1-r)\dfrac{U'(S_E(r))}{U'(S_{E^c}(r))}}. \tag{4}$$

Although the following result is logically weaker than Corollary 7, we present it as our main result because it is especially tractable for applications. It shows how properness can be extended to general decision models. The condition $S_E(0) \geqslant S_{E^c}(0)$ implies that $S_E(r) > S_{E^c}(r)$ for all $0 < r \leqslant 1$ because S_E is strictly increasing and S_{E^c} is strictly decreasing (Lemma 3). Event E always having the same rank in the sense of having the best outcome, means by definition that the set of prospects considered is *comonotonic*.[4] Schmeidler (1989) demonstrated the importance of this condition for nonexpected utility. Our analysis provides yet another illustration of this importance.

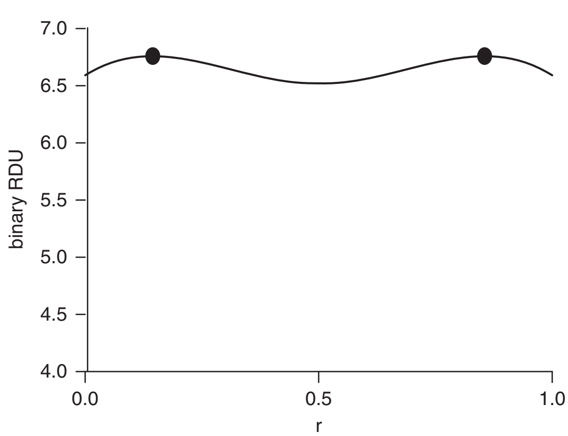

Figure 2. Multiple optima for $W(E) = W(E^c) = 0.5$ and convex utility in Example 6.

Theorem 8
[Extending properness from expected value to nonexpected utility]. Assume Structural Assumption 2 with, further, $S_E(0) \geqslant S_{E^c}(0)$. Then Equation (4) holds for all reported probabilities r.

One implication in the above theorem is that $W(E)$, being a function of r throughout its domain, can be uniquely inferred from r. This is essential for the identifiability of W from r.

6. THE RANK CORRECTION AND OTHER TRACTABLE APPLICATIONS OF THEOREM 8

The condition $S_E(0) \geqslant S_{E^c}(0)$ used in the preceding results can easily be established for bounded proper scoring rules by means of the following observation, whose proof is trivial.

Observation 9
If $S_E(r)_E S_{E^c}(r)$ is proper, then so is

$$(k + S_E(r))_E S_{E^c}(s) \quad \text{for all } k \in \mathbb{R} \text{ with } k + S_E(0)) \geqslant 0. \tag{5}$$

The requirement $k + S_E(0)) \geqslant 0$ ensures that $k + S_E(r)) \geqslant 0$ for all r. We can thus take any

$$k \geqslant S_{E^c}(0) - S_E(0) \tag{6}$$

to have the condition of Theorem 8 satisfied. For example, for the quadratic scoring rule defined in Section 2, we can take $k = 1$ (Figure 4 in Section 7).

[4]For a general state space S, a set of functions from S to $\mathbb{R}$ is *comonotonic* if there are no two functions f,g in the set and two states s,t in S with $f(s) > f(t)$ and $g(s) < g(t)$. Then there is a complete ordering (ranking) of S such that all functions in the set assign (weakly) higher values to better ranked states (Wakker, 2010 Exercise 10.12.2). For a comonotonic set of functions, the effects of optimistically overweighting best ranked events, pessimistically overweighting worst ranked events, and other forms of rank dependence work the same for all functions. For our comonotonic scoring rules, event E is always ranked better in the sense of always yielding the best outcome. Hence, pessimists will over- or underweight E the same for all prospects considered.

Copyright © 2011 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

For applying Observation 9 it is necessary that $S_E(0)$ is a real number. For instance, (in generalizations of our results to negative outcomes) $S_E(0)$ should not be $-\infty$. Thus our comonotonic approach cannot be used for the logarithmic proper scoring rule.[5] It can be used for all bounded proper scoring rules, such as the classical spherical rule $(r/\sqrt{r^2+(1-r)^2})_E((1-r)/\sqrt{r^2+(1-r)^2})$. For example, we can then take $k=1$ in Equation (5). Using this scoring rule instead of the quadratic scoring rule in Figure 4 leads to virtually the same correction curve.

An alternative rank correction can be obtained by arranging $S_E(1) \leqslant S_{E^c}(1)$, implying $S_E(r) < S_{E^c}(r)$ for all $r<1$. Then all conclusions of Theorem 8, and the following applications, remain valid by interchanging E^c and E.

At first, our main results, based on Equation (4), may seem to be intractable for practical purposes. To apply it, it seems that we have to infer ratios of derivatives of utility, and measuring utility can be as difficult as measuring beliefs. Before discussing the general solution, we mention a convenient result that can be obtained if we may assume linear utility. Then the ratio of utility derivatives is simply 1. Linear utility is plausible for moderate stakes as often used in proper scoring rules (Luce, 2000, p. 86; Pigou, 1920, p. 785; Rabin, 2000; Ramsey, 1931, p. 176). The following observation concerns this case, and provides the most efficient way to empirically measure nonadditive measures W presently available in the literature.

Observation 10

Assume that Structural Assumption 2 holds, that $S_E(0) \geqslant S_{E^c}(0)$, and that U is linear. Then

$$W(E) = r.$$

We next turn to a general method, introduced by OSVW for their quadratic scoring rule, and now adapted to our general setup, to avoid measuring utility (and probability weighting) also if utility is not linear. We assume that events Q with objective probabilities p are available, for instance generated by flipping symmetric coins. It is commonly assumed that $W(Q)$ is the same for all events Q with the same probability p, so that we can define $w(p) = W(Q)$ for a function w, called the *probability weighting function*.[6] OSVW considered the function $B = w^{-1}(W)$, arguing that this is a better candidate for an index of belief than W.[7] Substitution in Equation (4) shows the following: Let probability p have the same reported probability r as event E, and let Q be the event with objective probability $P(Q) = p$. Then $W(E)$ and $W(Q)$ have the same right-hand side and, hence, the same left-hand side in Equation (4):

$$W(E) = W(Q) = w(p) \tag{7}$$

which implies

$$B(E) = w^{-1}(W(E)) = p. \tag{8}$$

Thus, an easy way results for measuring subjective beliefs $w^{-1}(W(E))$:

Step 1. For event E, observe the reported probability r.

Step 2. For r, find an event Q with objective probability p such that Q has the same reported probability r as E.

Step 3. Then $B(E) = p = w^{-1}(W(E))$ is the subjective belief in E.

OSVW measured the reported probability of a sufficiently dense set of objective probabilities p, taking all values $p \in \{0, 1/20, \ldots, 20/20\}$. With this, straightforward, work carried out once, they could for each event E in their analysis immediately infer its subjective belief from the reported probability r. OSVW called this way of filtering out w from W the *risk correction*.

Comonotonic scoring rules described in Theorem 8 provide a better way to implement the above procedure than the traditional quadratic scoring rule of OSVW because the three-step procedure then is valid for all r. This also holds if binary additivity is violated and if $r = 0.5$. We call the resulting procedure the *rank correction*,

[5]It can be used for the logarithmic scoring rule if we, in generalizations to negative outcomes, can restrict attention to a subdomain of events whose belief exceeds some positive threshold.

[6]In this way, Observation 10 in particular provides the most efficient way presently existing to empirically measure the probability weighting function w. We then have $w(p) = r$ with p the objective probability of event E.

[7]In the same way as we can define decision weight $w(p)$ for objective probability p, we can define decision weight $W(E) = w(B(E))$ for general events E. This illustrates that B rather than W is the natural analog of probabilities. The decision component w should be removed from $W = woB$ before an interpretation as index of belief can be considered.

Copyright © 2011 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

and we recommend its use for applications. Comonotonicity is not only useful for the analytical purposes explained before but also for psychological reasons. It controls for the psychologically realistic effects of optimism and pessimism in rank dependence, keeping them constant throughout the measurement of beliefs.

7. ILLUSTRATIONS AND DISCUSSION

Figure 3 depicts the modification of Figure 1 with a rank correction instead of a risk correction. The rank correction curve is based on our method in Theorem 8, using Equation (5) with $k=1$. It has no flat part and discriminates between all levels $W(E)$. The I-problem has been resolved. The H-problem has been resolved too. We will discuss it in detail for Figure 4, which is similar.

Figure 4 replicates the risk correction curve of OSVW's Figure 1 for quadratic scoring rules, and adds the rank correction curve. This figure concerns the special case where a probability p is given for event E, objective or subjective, and $W(E)=w(p)$, with w and U specified in the figure. Now we can let the x-axis designate probability p rather than the subjective weight $W(E)$, which would be $w(p)$ in this case. The curves now give the optimal reported probabilities $r=R(p)$ as a function of p. The risk correction curve is based on the traditional quadratic scoring rule, defined in Section 2 and used by OSVW. Again, this curve has the I- and H-problems, but the rank-corrected curve does not. The rank correction curve is based on our method in Theorem 8, using Equation (5) with $k=1$. It has no flat part and discriminates between all probabilities. Thus, the I-problem has been solved. With regard to the H-problem, in the risk correction curve replicated from OSVW, the part for $p<0.5$ has simply been obtained from flipping and rotating the part for $p>0.5$. This can be justified only under a condition called binary additivity (OSVW, p. 1484). Such a procedure was not needed for the rank correction. This whole curve is derived from data, and it is valid irrespective of whether or not binary additivity holds. Thus, the H-problem has been solved too.

7.1. Noise in observations

For practical purposes we do not only want to avoid completely flat parts, but also shallow parts in Figure 4, because they give little discriminatory power and they are prone to errors due to noise in the data. Finding proper scoring rules that provide the optimal discriminatory power in the part of the domain of our maximal interest is a topic for future research. Sometimes we may deliberately restrict our attention to events for which particular beliefs can be excluded on *a priori* grounds, and require properness and the restrictions of Lemma 5 only on that part of the domain. For example, if we believe beforehand that E will be judged considerably more likely than its complement, so much that we are also sure to be at a safe distance from the flat

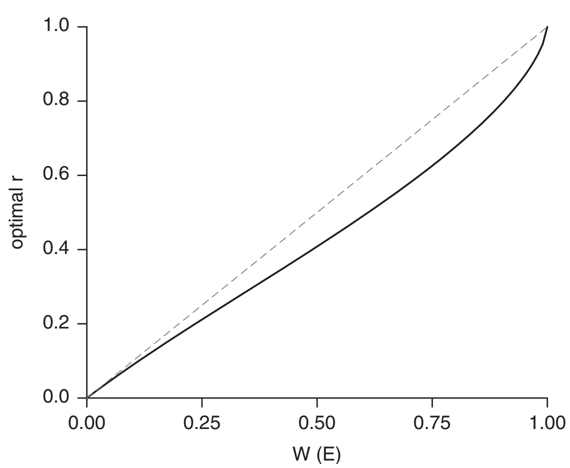

Figure 3. The solid line gives the optimal reported probability r as a function of $W(E)$, assuming $U(x)=\sqrt{x}$, for the comonotonic quadratic scoring rule modified using Equation (5) with $k=1$: $(2-(1-r)^2)_E(1-r^2)$. The figure is valid for all values of $W(E^c)$.

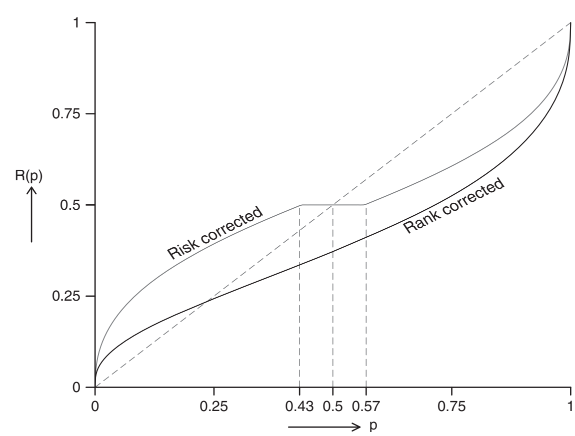

Figure 4. Reported probability $r=R(p)$ as a function of probability p. We assume binary RDU for known probabilities, with $U(\alpha)=\alpha^{0.5}$ and $W(E)=w(p)=\exp^{-((-\log(p))^{0.65})}$.

Copyright © 2011 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

part of the risk correction curve in Figure 4, and in the half of the domain where the H-problem does not occur, then we can again consider using the traditional quadratic scoring rule of OSVW.

7.2. Interpretations of subjective beliefs and ambiguity

Reported probabilities are purely choice-based concepts. Any psychological concept derived from them is based on interpretations. Under expected utility, (subjective) probabilities are usually interpreted as cognitive indexes reflecting degrees of beliefs. This interpretation is commonly followed in the literature on proper scoring rules. Schmeidler (1989) interpreted his nonadditive generalization W of subjective probabilities also as an index of belief but assumed expected utility for risk. As explained before, OSVW suggested that $w^{-1}(W)$ is a good index if expected utility for risk is violated. In Schmeidler's approach, with expected utility holding for risk so that w is the identity function, we have $B = W$ so that OSVW's interpretation is consistent with Schmeidler's. As did OSVW, we follow these generalized interpretations to stay as close to the traditions in the literature on proper scoring rules as possible.

Further generalizations and different interpretations can be considered. Nonadditive measures are often assumed to capture the effects of ambiguity (unknown probabilities). It can be debated to what extent ambiguity concerns the cognitive component of belief or, differently, components of decision–attitude, or, possibly, a mix/interaction of these two components. No consensus has yet been reached in decision theory on these questions. We leave the interpretations and further decompositions and disentangling of various components to future studies.

8. CONCLUSION

Proper scoring rules are important tools for measuring subjective degrees of beliefs. They have many appealing and valuable properties. Unfortunately, they have mostly been analysed under the assumption of expected value maximization, whereas many empirical studies have demonstrated violations of expected value. OSVW extended one proper scoring rule, the quadratic one, to general decision theories, but had no discriminatory power for events with likelihood close to fifty–fifty (the identification problem). Further, they could not provide observations for events less likely than their complement (the half-domain problem). This paper generalizes the results of OSVW to all proper scoring rules and degrees of belief, and shows how the problems mentioned can be resolved. Thus, we have provided a general method for measuring subjective beliefs.

APPENDIX A: SPECIAL CASES OF BINARY RANK-DEPENDENT UTILITY

As explained in Appendix C of OSVW, many existing theories are special cases of Equation (2). If an objective probability p is given of E and $W(E) = p = 1 - W(E^c)$, then we deal with expected utility for decision under risk (von Neumann and Morgenstern, 1944). If no objective probability p is given of E, but $W(E)$ is a subjective probability, then we deal with subjective expected utility for decision under uncertainty (Savage, 1954). In both cases considered, $W(E) + W(E^c) = 1$.

If an objective probability p is given of E, but $W(E) = w(p)$ for a nonlinear function w then we deal with rank-dependent utility (Quiggin, 1982) and, for gains, with prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). Studies with rich domains of events E have found that W often cannot be obtained as a transformation of probabilities, not even of subjective probabilities (ambiguity). Then general nonadditive set functions W are used (Denneberg, 1994), as in Choquet expected utility (Gilboa, 1987; Schmeidler, 1989). For prospects with two outcomes as considered here, other theories are also special cases of Equation (2). These theories include multiple priors with maxmin expected utility (Chateauneuf, 1991; Gilboa and Schmeidler, 1989; Luce and Raiffa, 1957; Wald, 1950) and with α-maxmin (Eichberger *et al.*, 2010; Ghirardato *et al.*, 2004; Hurwicz, 1951; Jaffray, 1994 Section 3.4; Luce and Raiffa, 1957 Section 13.5). For gains, prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) is included. Other models include Einhorn and Hogarth (1985) and Pfanzagl (1959).

Many papers considered nonadditive belief measures W without explicitly linking them to decisions, as in belief functions (Chateauneuf and Jaffray, 1989; Dempster, 1967; Denoeux, 2008; Grabisch *et al.*, 2001; Rota, 1964; Shafer, 1976; Stanley, 1986), fuzzy measures (Grabisch *et al.*,

Copyright © 2011 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

2000; Zadeh, 1965), imprecise probabilities (Walley, 1991), support functions (Tversky and Koehler, 1994), and upper and lower probabilities (Kyburg, 1983).

APPENDIX B: PROOFS

Proof of Lemma 3

Assume, for contradiction, that $S_E(r) = S_E(r')$ for some $r \neq r'$. Then r is (at least weakly) preferable to r' for all $W(E)$ if $S_{E^c}(r) \geqslant S_{E^c}(r')$. r' is (at least weakly) preferable to r for all $W(E)$ if $S_{E^c}(r) \leqslant S_{E^c}(r')$. r and r' are equivalent for all $W(E)$ if $S_{E^c}(r) = S_{E^c}(r')$. That is, the comparison between $S_{E^c}(r)$ and $S_{E^c}(r')$ determines a choice between r and r' irrespective of $W(E)$. It then is impossible that r is strictly preferable at $p = r$ and r' is strictly preferable at $p = r'$ under expected value, and a contradiction with properness has resulted. Hence the continuous S_E must be either strictly increasing or strictly decreasing, and S_{E^c} must also be one of these two. Because for $p = 1$ the value r is chosen so as to maximize S_E and this value of r, by properness, is $r = 1$, we see that S_E is maximal at $r = 1$. Similarly, inspecting $p = 0$ we see that S_{E^c} is maximal at $r = 0$. Hence, the only possibility is that S_E is strictly increasing and S_{E^c} is strictly decreasing. □

We next establish Equation (3) for all optimal r that satisfy a first-order condition, dealing with boundary solutions only after. We denote the reported probability r that optimizes the value of the scoring rule by $\tilde{r}$. By r we denote general values in [0,1]. From classical properness, it follows that the expected value under $W(E) = p$ is optimal at $r = \tilde{r} = p$ for all $p \in [0,1]$ Hence, we have, for all $r \in [0,1]$ (for interior r as first-order condition, and then for the boundary r's by continuity)

$$pS'_E(p) + (1-p)S'_{E^c}(p) = 0 \quad \forall p \in [0,1]. \tag{B1}$$

It implies

$$S'_{E^c}(p) = -\frac{pS'_E(p)}{1-p} \quad \forall p \in [0,1). \tag{B2}$$

We now turn to general W,U. We write π for the decision weight $W(E)$. If $\tilde{r} > 0$, then, by continuity, $S_E(r') > S_{E^c}(r')$ for an $r' < \tilde{r}$. If $\tilde{r} = 0$, then we define $r' = \tilde{r} = 0$. The *comonotonic region* is $[r',1]$. By Lemma 3, $S_E(r) \geqslant S_{E^c}(r)$ for all r in the comonotonic region. For all prospects here, E yields the best outcome, which by Schmeidler's (1989) definition means that this set of prospects is comonotonic. We restrict our attention to the comonotonic region in what follows. There the prospect $S_E(r)_E(S_{E^c}(r))$ is evaluated by:

$$V(r) := \pi U(S_E(r)) + (1-\pi)U(S_{E^c}(r))$$

with first derivative

$$V'(r) = \pi S'_E(r)U'(S_E(r)) + (1-\pi)S'_{E^c}(r)U'(S_{E^c}(r)). \tag{B3}$$

Substituting Equation (B2) for $r = p$, the first derivative is as follows:

$$V'(r) = \frac{S'_E(r)}{1-r}[(1-r)\pi U'(S_E(r)) - r(1-\pi)U'(S_{E^c}(r))] \quad \forall r \in [0,1). \tag{B4}$$

For interior optimal $\tilde{r}$ $(0 < \tilde{r} < 1)$, $V'(\tilde{r}) = 0$, and Equations (3) and (4) follow from the following equalities:

$$(1-\tilde{r})\pi U'(S_E(\tilde{r})) - \tilde{r}(1-\pi)U'(S_{E^c}(\tilde{r})) = 0 \tag{B5}$$

which follows from Equation (B4) because $S_E(r)$ is strictly increasing.[8] We now get

$$\frac{\tilde{r}}{1-\tilde{r}} = \frac{\pi}{1-\pi}\frac{U'(S_E(\tilde{r}))}{U'(S_{E^c}(\tilde{r}))}.$$

Algebraic manipulations now give Equations (3) and (4) for $\tilde{r}$.

We finally turn to $r = 1$ and $r = 0$; i.e. we consider boundary solutions r. The case of $\tilde{r} = 1$ is similar to the case $\tilde{r} = 0$.[9] In the rest of the proof we consider the latter case. If $\pi = 0$ then Equations (3) and (4) follow immediately, and we are done. We assume $\pi > 0$ henceforth, and derive a contradiction.

We have '>0' in Equation (B5) because $U' > 0$, $\pi > 0$, and $\tilde{r} = 0$. By continuity, we have '>0' for

[8]This is direct if $S'_E(\tilde{r}) > 0$. It is also possible that $S'_E(\tilde{r}) = 0$, and we assume this henceforth. Assume, for contradiction, that we have '<0' in Equation (B5). By continuity, this holds on an interval $[r', \tilde{r}]$ for some $r' < \tilde{r}$ that can be assumed to be in the comonotonic region. $S'_E \geqslant 0$ implies $V' \leqslant 0$ on $(r', \tilde{r}]$. Because S_E is strictly increasing, S'_E takes positive values at some values in $(r', \tilde{r}]$, implying that $V' < 0$ there and V is strictly larger there than at $\tilde{r}$, contradicting optimality at $\tilde{r}$. '>0' in Equation (B5) similarly leads to a contradiction, with V larger at some value $r' > \tilde{r}$.

[9]There is a duality between E and E^c, r and $1-r$, and p and $1-p$, because of which the case $\tilde{r} = 1$ follows from the case $\tilde{r} = 0$.

Copyright © 2011 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda

all r in $[0, r')$ for some $r' > 0$. The interval $[0, r']$ can and will be assumed to be contained in the comonotonic region. Because $S'_E(r) \geqslant 0$, Equation (B4) would imply $V'(r) \geqslant 0$ on the open interval $(0, r')$. By continuity, this would hold on $[0, r']$. V, being maximal at $r = 0$, then would be constant on $[0, r']$. Because S_E is strictly increasing it must have positive derivate at some points in $[0, r']$. V', the product in Equation (B4), is positive there, and V cannot be constant on $[0,r']$, so that a contradiction has resulted. The proof is done.

REFERENCES

Andersen S, Fountain J, Harrison GW, Rutstöm EE. 2009. *Estimating Subjective Probabilities*. Center for the Economic Analysis of Risk (CEAR), Georgia State University: Atlanta, GA.

Arrow KJ. 1965. *Aspects of the Theory of Risk-Bearing*. Academic Bookstore: Helsinki.

Arrow KJ. 1971. *Essays in the Theory of Risk-Bearing*. North-Holland: Amsterdam.

Brier GW. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **78**: 1–3.

Chateauneuf A. 1991. On the use of capacities in modeling uncertainty aversion and risk aversion. *Journal of Mathematical Economics* **20**: 343–369.

Chateauneuf A, Jaffray J-Y. 1989. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences* **17**: 263–283.

de Finetti B. 1931. Sul Significato Soggettivo della Probabilità. *Fundamenta Mathematicae* **17**, 298–329. Translated into English as 'On the subjective meaning of probability.' *In Probabilità e Induzione*, Monari P, Cocchi D (1993, eds), Clueb, Bologna; 291–321.

de Finetti B. 1937. La Prévision: Ses Lois Logiques, ses Sources Subjectives. *Annales de l'Institut Henri Poincaré* 7, 1–68. Translated into English by Henry E. Kyburg, 'Foresight: its logical laws, its subjective sources.' In *Studies in Subjective Probability*, Kyburg HE, Smokler HE (1964, eds). Wiley: New York; 53–118; 2nd edition 1980, Krieger: New York.

de Finetti B. 1952. Sulla Preferibilitá. *Giornale degli Economistii e Annali di Economia* **11**: 685–709.

de Finetti B. 1962. Does it make sense to speak of 'good probability appraisers'? *In The Scientist Speculates: An Anthology of Partly-Baked Ideas*, Isidore J. Good (ed.) William Heinemann Ltd., London. Reprinted as Ch. 3 in Bruno de Finetti (1972), '*Probability, Induction and Statistics*.' Wiley: New York.

Dempster AP. 1967. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics* **38**: 325–339.

Denneberg D. 1994. *Non-Additive Measure and Integral*. Kluwer Academic Publishers: Dordrecht.

Denoeux T. 2008. Conjunctive and disjunctive combination of belief functions induced by non distinct bodies of evidence. *Artificial Intelligence* **172**: 234–264.

Eichberger J, Grant S, Kelsey D, Koshevoy GA. 2010. The alpha-Meu model: a comment. *Journal of Economic Theory*, forthcoming.

Einhorn HJ, Hogarth RM. 1985. Ambiguity and uncertainty in probabilistic inference. *Psychological Review* **92**: 433–461.

Ghirardato P, Maccheroni F, Marinacci M. 2004. Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory* **118**: 133–173.

Gilboa I. 1987. Expected utility with purely subjective non-additive probabilities. *Journal of Mathematical Economics* **16**: 65–88.

Gilboa I. 2004. (ed.). *Uncertainty in Economic Theory: Essays in Honor of David Schmeidler's 65th Birthday*. Routledge: London.

Gilboa I, Schmeidler D. 1989. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics* **18**: 141–153.

Grabisch M, Marichal J-L, Roubens M. 2001. Equivalent representations of set functions. *Mathematics of Operations Research* **25**: 157–178.

Grabisch M, Murofushi T, Sugeno M. 2000. (eds). *Fuzzy Measures and Integrals: Theory and Applications*. Physica-Verlag: Berlin.

Hanson R. 2002. Piece entitled 'Wanna Bet?' *Nature* **420**: 354–355.

Hofstee WKB. 1988. *De Empirische Discussie, Theorie van het Sociaal-Wetenschappelijk Onderzoek*. Boom: Meppel.

Hurwicz L. 1951. Some specification problems and applications to econometric models (Abstract). *Econometrica* **19**: 343–344.

Hurwicz L. 1960. Optimality and informational efficiency in resource allocation. In *Mathematical Methods in the Social Sciences*, Arrow KJ, Karlin S, Suppes P (eds). Stanford University Press: Stanford, CA; 17–46.

Jaffray J-Y. 1994. Dynamic decision making with belief functions. In *Advances in the Dempster-Shafer Theory of Evidence*, Ronald R, Yager MF, Kacprzyk J (eds). Wiley: New York; 331–352.

Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* **47**: 263–291.

Kyburg Jr HE. 1983. *Epistemology and Inference*. University of Minnesota Press: Minneapolis, MN.

Luce RD. 2000. *Utility of Gains and Losses: Measurement—Theoretical and Experimental Approaches*. Lawrence Erlbaum Publishers: London.

Luce RD, Raiffa H. 1957. *Games and Decisions*. Wiley: New York.

Offerman T, Sonnemans J, van de Kuilen G, Wakker PP. 2009. A truth-serum for non-Bayesians:

Copyright © 2011 John Wiley & Sons, Ltd.

correcting proper scoring rules for risk attitudes. *Review of Economic Studies* **76**: 1461–1489 (OSVW).
Pfanzagl J. 1959. A general theory of measurement—applications to utility. *Naval Research Logistics Quarterly* **6**: 283–294.
Pigou AC. 1920. *The Economics of Welfare* (edn. 1952). MacMillan: London.
Pratt JW. 1964. Risk aversion in the small and in the large. *Econometrica* **32**: 122–136.
Quiggin J. 1982. A theory of anticipated utility. *Journal of Economic Behaviour and Organization* **3**: 323–343.
Rabin M. 2000. Risk aversion and expected-utility theory: a calibration theorem. *Econometrica* **68**: 1281–1292.
Ramsey FP. 1931. Truth and probability. In *The Foundations of Mathematics and Other Logical Essays*. Routledge and Kegan Paul: London; 156–198. Reprinted *in Studies in Subjective Probability*, Kyburg HE, Smokler HE (1964, eds). Wiley: New York; 61–92.
Rota GC. 1964. On the foundations of combinatorial theory I. Theory of Möbius Functions. *Zeitschrift für Warscheinlichkeitstheorie und Verwandte Gebiete* **2**: 340–368.
Savage LJ. 1954. *The Foundations of Statistics*. Wiley: New York (second edition 1972, Dover Publications: New York.)
Savage LJ. 1971. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association* **66**: 783–801.
Schmeidler D. 1989. Subjective probability and expected utility without additivity. *Econometrica* **57**: 571–587.
Shafer G. 1976. *A Mathematical Theory of Evidence*. Princeton University Press: Princeton, NJ.
Stanley RP. 1986. *Enumerative Combinatorics*, Vol. I. Wadsworth & Brooks/Cole: Monterey, CA.
Starmer C. 2000. Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* **38**: 332–382.
Tetlock PE. 2005. *Expert Political Judgment*. Princeton University Press: Princeton, NJ.
Tversky A, Kahneman D. 1992. Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5**: 297–323.
Tversky A, Koehler DJ. 1994. Support theory: a nonextensional representation of subjective probability. *Psychological Review* **101**: 547–567.
von Neumann J, Morgenstern O. (1944, 1947, 1953), *Theory of Games and Economic Behavior*. Princeton University Press: Princeton NJ.
Wakker PP. 2010. *Prospect Theory: For Risk and Ambiguity*. Cambridge University Press: Cambridge, UK.
Wald A. 1950. *Statistical Decision Functions*. Wiley: New York.
Walley P. 1991. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall: London.
Winkler RL, Murphy AH. 1970. Nonlinear utility and the probability score. *Journal of Applied Meteorology* **9**: 143–148.
Zadeh LA. 1965. Fuzzy sets. *Information and Control* **8**: 338–353.

Copyright © 2011 John Wiley & Sons, Ltd.

DOI: 10.1002/mcda