

## A BEHAVIORAL FOUNDATION FOR FUZZY MEASURES

Peter WAKKER\*

*University of Nijmegen, Nijmegen Institute for Cognition research and Information technology (NICI), Nijmegen, The Netherlands*

Received February 1989

Revised May 1989

**Abstract:** In Savage [41] a ‘behavioral foundation’ was given for subjective probabilities, to be used in the maximization of expected utility. This paper analogously gives a behavioral foundation for fuzzy measures, to be used in the maximization of ‘Choquet-expected utility’. This opens the way to empirical verification or falsification of fuzzy measures, and frees them of their ‘ad hoc’ character.

**Keywords:** Fuzzy measures; Choquet integral; decision making; Choquet-expected utility; comonotonic; possibility measure.

### 1. History

One of the most important scientific achievements of this century may be the result laid down in the first five chapters of Savage [41]. There a *behavioral foundation* was given for ‘subjective probabilities’, to be used in an expected utility criterion. That is to say, sufficient conditions were given for decision making to be representable by subjective probabilities, through expected utility. This demonstrated how to verify or falsify the appropriateness of subjective probabilities through observations, and made subjective probabilities operational. The meaning of subjective probabilities changed from obscure-ad-hoc into scientifically well-founded. Indeed, a foundation of (Bayesian) statistics had been laid. With this scientifically well-founded meaning, discussion and research could start about the question whether or not subjective probabilities are suited to represent (lack of) knowledge. In decision theory, expected utility is used to model all kinds of phenomena, related to risk and uncertainty. For instance, risk aversion is modeled through concavity of utility, and minus the quotient of the second and first derivative of utility has been found to be a good tool to compare attitudes towards risk, in Pratt [39] and Arrow [4]. (The works of de Finetti and Lindley are related to Savage [41] and will be discussed in Section 2.)

Also in knowledge representation devices like subjective probabilities are called for. Good linguistic tools are needed to describe the processing of information and the reasoning with uncertainty and inexactitude. Modeling knowledge only with ‘true’, ‘untrue’, or ‘unknown’, or only with one exactly

\* The research has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

determined value for any quantity, is too rough for most applications. Tools are searched by means of which an intelligent system can express its state of knowledge in a more refined way. Subjective probabilities might be such a tool.

Still, subjective probabilities have not become the sole tool used by scientists to numerically model (lack of) knowledge. As the axioms of Savage give empirical meaning to subjective probabilities, they (primarily the 'sure-thing principle') are mostly taken to show that subjective probabilities have too many drawbacks; this was already argued for instance in the early Allais [1]. The most influential result in recent decision literature to deviate from expected utility has been Machina [34]. There expected utility has been deprived of its most celebrated results, the results on risk aversion of Pratt and Arrow. Machina showed that these results, extended in a natural way, hold in the general (differentiable) case, without needing the assumption of expected utility maximization. Further Machina gave an extensive discussion of the 'sure-thing principle'. Nowadays decision scientists usually deviate from expected utility. See for instance the survey in Machina [35], and Fishburn [18]. Our paper builds on one of these deviations, the 'Choquet-expected utility approach', initiated by Schmeidler [42].

Another impulse for the deviation from expected utility stems from the recent developments in artificial intelligence. The prevailing view is nowadays that (subjective) probabilities are too restricted and intractable to describe reasoning with uncertainty, or for the updating of knowledge. See the introductions in Pearl [38], Halpern and Rabin [22], or Dubois and Prade [14]; or many contributions in Kanal and Lemmer [28, 29]. For example, it is usually felt that expressing subjective probabilities in exact real numbers is too specified to be sensible. This also hinders the application of the formula of Bayes to the incorporation of new information. The formula of Bayes requires exact probabilities (also concerning the event of the acquisition of information) to be available at the outset (compare Pearl [38], p. 242).

For the above reasons, ways of modeling knowledge have become popular which are less restrictive. For instance one may want to work with 'upper and lower probabilities'. This is central in the 'belief functions' (and 'plausibility functions') of Dempster [11] and Shafer [44], as shown in Kyburg [31]. See also Shafer and Logan [45] (in particular Section 1.3), Shortliffe and Buchanan [46], Bolaños, Lamata and Moral [5], and the refinement in Jaffray [26, 27]; further Gordon and Shortliffe [21] and Nilsson [37] and the references therein. It made possible a new, often discussed way to model learning from new information: 'Dempster's rule for combining evidence'.

We shall consider another example, still more general than belief functions: 'Fuzzy measures'. These were introduced by Sugeno [48], as a variation upon the fuzzy sets of Zadeh [60]. In other contexts set functions with identical mathematical properties had been known for longer times, such as the 'capacities' in Choquet [7], 'monotonic normalized' characteristic games in cooperative game theory (see for instance von Neumann and Morgenstern [51] or Driessen [13]), or the 'nonadditive probabilities' in decision making under uncertainty (see Schmeidler [42]).

Fuzzy measures, and the multitude of other functions used to model knowledge, have often been criticized for being 'ad hoc' (see for instance several places in Kanal and Lemmer [28, 29]). Mathematical operations are defined and pushed forward only on the basis of intuitive appealingness, without empirically-founded meaning. In applications where the 'min'-operation, used by 'possibility measures' for AND, did not give a satisfactory result, Zadeh proposed to use a multiplication operation instead, only on an ad hoc basis, in Zadeh [61] (footnote p. 31) and Zadeh [62] (p. 34). The desirability of an empirical basis has been mentioned in fuzzy set literature, see for instance Zadeh [61] (p. 7, l. 32–35), Dubois and Prade [14] (Introduction), Giles [20] (p. 299, especially lines 35–38), and Hisdal [23] (especially p. 329).

Such an empirical basis will be given in this paper. We use recent developments in decision theory, contribute to these, and derive from these a behavioral foundation for fuzzy measures. Thus the edge is taken off the criticisms as described above. Let us emphasize that the behavioral foundation by itself does not entail a defense or criticism of fuzzy measures. All it does is 'translate' the meaning of fuzzy measures into empirical terms. It shows how to derive testable implications, without choosing sides. Again, this is analogous to Savage's behavioral foundation of expected utility. Allais [1] used it (mainly the sure-thing principle, or its analogue 'independence') to devise his famous example *against* expected utility.

Let us consider Theorem 11 below. The 'theoretical' statement (i) describes the model with fuzzy measures and Choquet-expected utility (i.e., the integral in that statement). Statement (ii) 'translates' statement (i) into empirical terms. Fuzzy measures to be used through Choquet-expected utility are appropriate and can be justified/verified if and only if the empirical conditions in statement (ii) of Theorem 11 are/can be; alternatively, they are not appropriate and can be criticized/falsified if and only if the empirical conditions in the statement (ii) are/can be. If one starts from a fuzzy measure and derives preferences from it, then statement (ii) shows which preferences are excluded through this procedure, and which are available. Also the procedure may be reversed. One may start, in a 'descriptive vein', from observed preferences, and then derive from these the fuzzy measure, as described in the proof in Section 5 (see (9), (10), and (11)). Statement (ii) shows for which preferences indeed this procedure works.

The main intuitive tool to obtain our results will be the 'hedging' idea as described for risk theory in Yaari [58]. The main mathematical tool will be the work laid down in Schmeidler [42, 43]. Making the work of Schmeidler more easily accessible has been an additional goal of this paper; his result, for finite state spaces, is obtained below as a corollary of Theorem 11.

All theorems in the sequel will be formulated so that they can be understood immediately, without consultation of the text, with the exception of the involved definitions of course.

## 2. Elementary definitions of Savage

First let us shortly discuss the works of de Finetti [8, 9] and Lindley [33], which are related to Savage [41]. These works also obtain behavioral foundations for

subjective probabilities, by means of 'coherence', and are more often quoted in artificial intelligence literature. Everything said in this paper concerning Savage [41] applies to these works as well. Wakker [54] (Section A2) argues that the work of de Finetti (this applies to the derived work of Lindley as well) implies a linearity of utility, in a way often neglected in literature; a quantitative nature of consequences is essentially needed. Linearity of utility (w.r.t. the quantitative nature of consequences) is mostly considered too strong a restriction in decision theory.

Next we give the definitions needed to sketch, roughly, the work of Savage [41]. By  $S = \{s_1, \dots, s_n\}$  we denote a (for convenience finite) set ('space') of ('crisp') (*possible states (of nature)*) (in other contexts called 'possible worlds', 'elementary outcomes', 'elementary propositions', etc.). Exactly one state is the true state, the other states are untrue. For example one may think of a horse race that will take place, with  $n$  participating horses;  $s_j$  is the 'state' that horse  $j$  will win. A person (intelligent system, robot, animal, ...), called *decision maker*, does not have enough information to be certain which state is the true one, and has not any influence upon the truth of the states. Question is at this moment whether the knowledge of the decision maker can be modeled through a 'subjective' probability distribution  $P$  over  $S$ , with  $P(A)$  measuring the 'degree of belief' of the decision maker in the truth of  $A$ . A basic idea, and one of the great contributions, of Savage's set-up is the idea that a probability measure by itself has no meaning. Only when it can be related in some sense to (observable) behavior with consequences that matter, does it get meaning. The observable behavior of the decision maker is modeled in Savage's decision-theoretic approach as the choosing, in any situation, between the so-called 'acts' available to him (or her). The term 'act' may be somewhat misleading, the behavior of the decision maker only consists of *choosing* one of the available acts. The resulting consequence of a chosen act depends upon which state is the true one.

Formally, the set of all consequences that may in any situation result from any act is denoted by  $\mathcal{C}$ . For simplicity of exposition we assume that  $\mathcal{C} = \mathbb{R}$ , designating amounts of money. In the appendix we shall show how the results can be generalized to more general consequence sets. So now we assume that a consequence is completely described through the resulting amount of money. An act  $f$  is just a function from  $S$  to  $\mathcal{C}$ , assigning to every state  $s$  the consequence  $f(s)$  which would result if  $s$  would be true and  $f$  would have been chosen.  $\mathcal{F}$  is  $\mathcal{C}^S$ , i.e., the set of all acts. Since the decision maker is uncertain about which state is true, he is uncertain about which consequence will result from a chosen act. For example if  $S$  describes a horse race, then 'acts' may consist of betting money on horses.

Formally the (dispositional) choice behavior of the decision maker is modeled through a binary relation (the *preference relation*)  $\succsim$  over the set  $\mathcal{F}$  of all acts. We write  $f \succsim g$  if the decision maker thinks  $f$  is at least as good as  $g$ , i.e., is willing to choose  $f$  if  $f, g$  are available to him. In any situation (with a finite number of acts available) the decision maker will choose the available act which is optimal according to  $\succsim$ . The choice behavior of the decision maker, modeled through  $\succsim$ , is considered to be the observable primitive of the model. Essential for the

meaning of subjective probabilities (or belief functions, fuzzy measures, . . . , etc., as well as utility functions) is their relation to the preference relation. Without that relation to the preference relation, the meaning of subjective probabilities etc. would be void. Note that in this set-up there is nothing fuzzy in the preferences. All fuzziness is modeled through the state space. The only vagueness (fuzziness, uncertainty, . . . etc.) concerns the true state; the preference relation describes how to make decisions while facing this vagueness.

Savage formulated a list of 'internal consistency' ('coherence') conditions for the preference relation, sufficient for the applicability of *expected utility*, i.e., sufficient for the *existence* of a probability measure  $P$  on  $S$ , and a 'utility function'  $U: C \rightarrow \mathbb{R}$ , so that

$$f \succcurlyeq g \Leftrightarrow \int_S U(f(s)) \, dP(s) \geq \int_S U(g(s)) \, dP(s).$$

We shall not repeat all the conditions of Savage [41]. Recent formulations have been provided in Fishburn [16] (Section 14), and Wakker [54] (Section A2), and on many other places. Let us only emphasize that all conditions are directly in terms of the preference relation, so that they have empirical meaning. The very idea of obtaining such a 'behavioral foundation' is one of the great achievements of Savage. Throughout this paper the behavioral foundations, given in statements numbered (ii) in theorems in the sequel, will also be directly in terms of conditions of the preference relation.

Savage's conditions can be split into two categories, the intuitively meaningful conditions which are also necessary for expected utility maximization, and the ('continuity-like') technical ones which are not fully necessary for expected utility maximization. The technical conditions in Savage are rather intractable, and for instance turned out to imply boundedness of utility in a 'hidden' way. Hence, for as far as we know, no scientific analysis has actually taken the set-up of Savage as point of departure. The most usual approach builds upon Anscombe and Aumann [3], and assumes that consequences consist of 'objective' probability distributions over 'prizes'. That induces a very convenient linear structure on the consequence space, and will also be the approach of this paper. In many contexts objective probabilities are not given, and other set-ups are required. This motivated the derivation of expected utility in Wakker [54] (Theorems IV.2.7 and V.6.1). There no objective probabilities are needed, and the only technical restriction left is continuity of utility. This is more tractable and suited for applications than Savage's technical conditions. The conditions for the preference relation are more complicated than those of Anscombe and Aumann [3], and the proofs much more complicated. Also for the characterization of fuzzy measures as provided in this paper a derivation which does not use objective probabilities as a tool may be desirable. Such a derivation (using the term 'capacity' instead of fuzzy measure), again with continuity of utility as only restriction left, has been obtained in Wakker [53] and Wakker [54] (Theorem VI.5.1). Again, the involved conditions are more complicated, and the derivations are much more complicated. A derivation of fuzzy measures with technical restrictions analogous to those of Savage, has been provided in Gilboa [19].

### 3. Elementary definitions of our set-up

$S$ ,  $\mathcal{F}$ , and  $\geq$  are as in Savage's approach above. We assume that the consequence space  $\mathcal{C}$  is the set  $\mathcal{P}$  of *money lotteries*. I.e., an element  $P = (p_1; x_1, \dots, p_m; x_m)$  from  $\mathcal{P}$  is a probability distribution with finite support over  $\mathbb{R}$ , resulting with probability  $p_1$  in the amount of money (real number)  $x_1, \dots$ , and with probability  $p_m$  in the amount of money (real number)  $x_m$ . Of course, to any  $A \subset \mathbb{R}$ ,  $P$  assigns probability  $\sum_{x_i \in A} p_i$ . Money lotteries can be 'mixed', for  $0 \leq \alpha \leq 1$ ,  $\alpha P^1 + (1 - \alpha)P^2$  assigns  $\alpha P^1(A) + (1 - \alpha)P^2(A)$  to each  $A \subset \mathbb{R}$ . For any function  $U: \mathbb{R} \rightarrow \mathbb{R}$ , by  $EU$  we denote the function assigning to each money lottery the expectation of  $U$  ('expected utility', if  $U$  is 'utility') under the lottery, i.e.,

$$EU: (p_1; x_1, \dots, p_m; x_m) \mapsto \sum_{i=1}^m p_i U(x_i). \quad (1)$$

An act  $f$  will result in the money lottery  $f(s_j)$  where  $s_j$  is the true state. Of course, next the probability mechanism of  $f(s_j)$  will determine the amount of money received by the decision maker. Note that we do not assume any probability measure given on  $S$ . Acts can also be mixed, in a 'pointwise manner', as  $\alpha f + (1 - \alpha)g: s \mapsto \alpha f(s) + (1 - \alpha)g(s)$ .

Like in Anscombe and Aumann [3] and Schmeidler [42], in our analysis the stochastic mechanism determining which amount of money will result from a money lottery is not the central interest. Rather is it an auxiliary structure to facilitate the analysis. It makes possible the application of techniques for mixture spaces as started in the Appendix of von Neumann and Morgenstern [51], and extensively used in Fishburn [17], and many other recent works in decision theory. We shall assume that the stochastic mechanism is well-established, and does not induce any vagueness. For instance, it may be constructed by an unbiased lottery wheel. Hence we shall, like Anscombe and Aumann [3] and Schmeidler [42], assume in the sequel that the decision maker uses expected utility to value money lotteries. Our central interest will be the way in which the decision maker processes the vagueness concerning the true state.

Let us give one more comment concerning Anscombe and Aumann [3]. They not only use probability distributions over consequences, but also over acts. However, they add a 'reversal of order' condition which entails that the only relevant aspect of a lottery over acts is the induced marginal lotteries (given any state) over consequences. Hence one may as well describe a lottery over acts by the induced marginal lotteries given states, leading to the set-up of this paper.

As usual, we write  $f > g$  if  $f \geq g$  and not  $g \geq f$ ,  $f \leq g$  if  $g \geq f$ ,  $f < g$  if  $g > f$ , and  $f = g$  if  $f \geq g$  and  $g \geq f$ . We call  $\geq$  a *weak order* if it is complete ( $f \geq g$  or  $g \geq f$  for all  $f, g$ ) and transitive, and  $\geq$  is *trivial* if  $f \geq g$  for all  $f, g$ . We identify money lotteries with constant acts, and write  $P^1 \geq P^2$  if  $f \geq g$  for the acts  $f$  constant  $P^1$  and  $g$  constant  $P^2$ ; analogous notations are used with  $>$ ,  $\leq$ ,  $<$ ,  $=$ . This induces preference relations over money lotteries. A function  $\phi: \mathcal{F} \rightarrow \mathbb{R}$  represents  $\geq$  if, for all acts  $f, g$ ,

$$f \geq g \Leftrightarrow \phi(f) \geq \phi(g).$$

We call  $\gg$  (mixture-)continuous if, for all  $f > g > h$ , there exist positive  $\alpha$  and  $\beta$  so ('close to 0') that  $(1 - \alpha)f + \alpha h > g > \beta f + (1 - \beta)h$ .

We call  $\geq$  weakly monotonic if, for all acts  $f, g$ ,

$$[f(s) \geq g(s) \text{ for all } s] \Rightarrow [f \geq g].$$

Before turning to the idea of hedging let us give some definitions. A function  $v: 2^S \rightarrow [0, 1]$  is a *fuzzy measure* if  $v(\emptyset) = 0$ ,  $v(S) = 1$ , and  $v$  is *monotonic w.r.t. set-inclusion*, i.e.,  $A \supset B \Rightarrow v(A) \geq v(B)$ . We call  $v$  a *probability measure* if it is a fuzzy measure which is *additive*, i.e.,  $v(A \cup B) = v(A) + v(B)$  for all disjoint  $A, B$ .

One of the most difficult steps in finding a behavioral foundation for fuzzy measures is to find a sensible way to integrate with respect to these (in general) nonadditive functions. In the context of (subclasses, usually 't-conorms', of) fuzzy measures, several functionals have been studied in the literature, and proposed as 'integrals'; see for instance Weber [56, 57], Suárez García and Gil Álvarez [47], Sugeno and Murofushi [49], and Hua [25]. The most well-known functional is the integral of Sugeno [48]. It is not linear (see Klement and Ralescu [30]) and for probability measures does not coincide with the usual integral. Sugeno did obtain, for probability measures, an upper bound for the difference between his integral and the usual integral; Murofushi and Sugeno [36] (Proposition A2.1) extend this. We follow the way of integration of Schmeidler [42], which later turned out to have been found already by Choquet [7]. Also Höhle [24] found this way of integration for fuzzy measures as did Quiggin [40] for the context of decision making under risk with given probabilities. We prefer this 'Choquet integral' because it is an extension of the usual integral for probability measures, is defined for any fuzzy measure, and is characterized by natural 'integral-like' additivity conditions in Anger [2] (Theorem 3) and Schmeidler [43]. Also Murofushi and Sugeno [36] observed these advantages. The applicability of the Choquet integral to fuzzy measures has been noted in Höhle [24] (his form is essentially equivalent to the Choquet integral), Weber [56], Sugeno and Murofushi [50], and Murofushi and Sugeno [36]. The usefulness for fuzzy measures of the decision-theoretic approach to Choquet integration has been noted in Wakker [52] (Example 7.8 and p. 21 lines 12–16). The present paper has grown out of an elaboration of the latter, and was announced in Wakker [55].

For a fuzzy measure  $v$ , and a function  $\phi: S \rightarrow \mathbb{R}$ , the *Choquet integral* of  $\phi$  (with respect to  $v$ ), denoted  $\int_S \phi \, dv$ , or  $\int \phi \, dv$ , is

$$\int_0^\infty v(\{s \in S: \phi(s) \geq \tau\}) \, d\tau + \int_{-\infty}^0 [v(\{s \in S: \phi(s) \geq \tau\}) - 1] \, d\tau.$$

For probability measures the Choquet integral is identical to the usual integral. In this paper we shall exclusively deal with the case where  $\phi = EU \circ f$  with  $f$  an act, and  $EU$  a ('utility') function from  $\mathcal{P}$  to  $\mathbb{R}$ . Substituting that gives

$$\int_0^\infty v(\{s \in S: EU(f(s)) \geq \tau\}) \, d\tau + \int_{-\infty}^0 [v(\{s \in S: EU(f(s)) \geq \tau\}) - 1] \, d\tau.$$

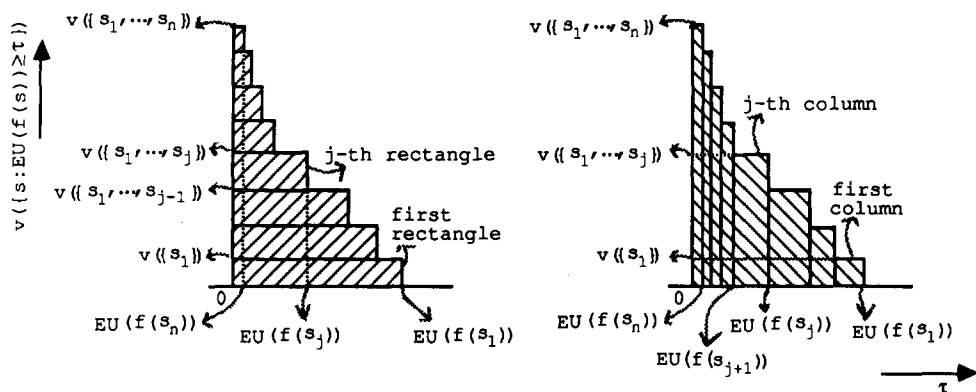


Fig. 1a. The Choquet integral of the function  $EU \circ f: S \rightarrow \mathbb{R}$  illustrated for the case  $EU(f(s_1)) \geq \dots \geq EU(f(s_n)) \geq 0$ , according to the definition. The Choquet integral is the area below the graph of the function  $\tau \mapsto v(\{s: EU(f(s)) \geq \tau\})$  on  $\mathbb{R}_+$ , i.e., the area dashed with upward or downward slope. It can be calculated as the sum of the areas of the rectangles dashed with upward slope. The height of the  $j$ -th rectangle is  $v(\{s_1, \dots, s_j\}) - v(\{s_1, \dots, s_{j-1}\})$ , the breadth is  $EU(f(s_j))$ . The product of these two is the  $j$ -th term in formula (4). Also one can calculate the sum of the areas of the columns dashed with downward slope. The area of the  $j$ -th column is the  $j$ -th term in formula (6).

See Figure 1a. As for additive  $v$  this is the usual expected utility of  $f$ , for general  $v$  we call it the *Choquet-expected utility* of  $f$ . The results concerning the Choquet integral, given below, and well-known, are elucidated in Wakker [54] (Chapter VI).

Let  $\pi: \{1, \dots, n\} \rightarrow S$  be a bijection so that

$$EU(f(\pi(1))) \geq EU(f(\pi(2))) \geq \dots \geq EU(f(\pi(n))). \quad (2)$$

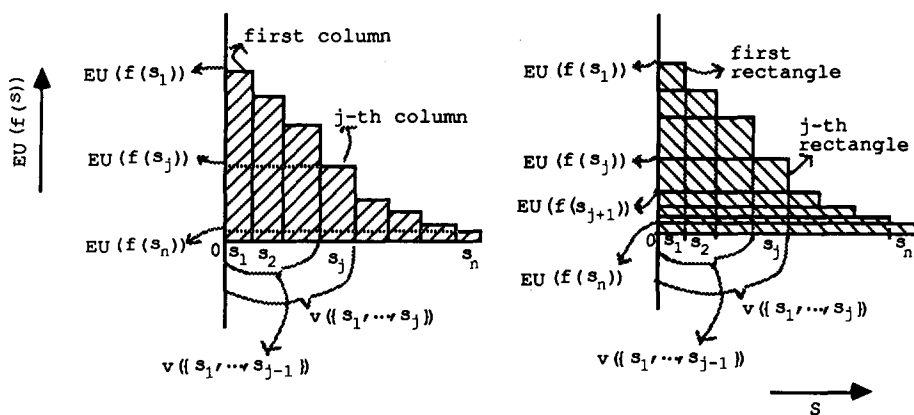


Fig. 1b. The same Choquet integral as in Figure 1a, but now obtained by flipping Figures 1a horizontally and rotating right. Rectangles thus become columns and vice versa. The resulting figures can be interpreted as the graph of the function  $EU \circ f: S \rightarrow \mathbb{R}$ . The height of the  $j$ -th column indicates the  $EU$ -value assigned to state  $s_j$ . The breadth of the first  $j$  columns, so of the  $j$ -th rectangle, is  $v(\{s_1, \dots, s_j\})$ .



I.e.,  $\pi$  orders states from most to least ‘favorable’, given function  $EU \circ f$ . Define

$$P^\pi(\{s\}) := v(\{t \in S: \pi^{-1}(t) \leq \pi^{-1}(s)\}) - v(\{t \in S: \pi^{-1}(t) < \pi^{-1}(s)\}). \quad (3)$$

So  $P^\pi(\{s\})$  is the ‘marginal fuzzy measure-contribution of  $s$  to the states more favorable than  $s$ ’. This induces a probability measure  $P^\pi$  on  $S$ . Now let  $f$  be an act, and  $EU: \mathcal{P} \rightarrow \mathbb{R}$  a function. It can be seen (see Figure 1) that

$$\int EU \circ f \, dv = \sum_{i=1}^n P^\pi(\{s_i\}) EU(f(s_i)). \quad (4)$$

From this one easily obtains, for any real  $\lambda$ ,

$$\int (\lambda + EU \circ f) \, dv = \lambda + \int EU \circ f \, dv. \quad (5)$$

Formula 4 is most suited for applications and intuition, since the numbers  $P^\pi(\{s_i\})$  can be interpreted as ‘act-dependent’ (through  $\pi$ ) probabilities. The next formula (see again Figure 1) is most suited for mathematical derivations, and will be used in Section 5. We write  $EU(f(\pi(n+1))) := 0$ .

$$\int EU \circ f \, dv = \sum_{i=1}^n [v(\{\pi(1), \dots, \pi(i)\})] \times [EU(f(\pi(i)) - EU(f(\pi(i+1)))]. \quad (6)$$

#### 4. Independence in the presence of hedging, and main theorems

Savage [41] (p. 68) wrote: “... the ... view sponsored here does not leave room for optimism or pessimism ... to play any role in the person’s judgement of probabilities”. Fuzzy measures generalize Savage’s expected utility approach by leaving such room for optimism or pessimism phenomena. Let us elucidate this through formula (4). The probabilities  $P^\pi(\{s_i\})$  (interpreting the marginal fuzzy-measure contributions as such) may now depend on the act  $f$  through the way in which  $f$  (viz.  $\pi$  in (4)) orders the states from most to least favorable (this interpretation has been introduced in Wakker [52]). A pessimistic decision maker now has the possibility to assign larger probabilities to states when they are unfavorable, than when they are favorable. Because of this, a pessimistic decision maker appreciates ‘hedging’ against vagueness as in the following example.

**Example 1 [Hedging].** (It can be seen that this example is a variation of the ‘Ellsberg paradox’.) Suppose there are only two states of nature,  $s_1$  and  $s_2$ . The decision maker does not have information which would make one state more likely than the other. The following table describes the  $EU$ -values of the money lotteries assigned to the states by acts  $f, g, h, f'$ . Act  $f$  is the  $\frac{1}{2}$ – $\frac{1}{2}$  mixture of  $g$  and  $h$ , act  $f'$  the  $\frac{1}{2}$ – $\frac{1}{2}$  mixture of  $f$  and  $h$ .

	$f$	$g$	$h$	$f'$
$s_1$	0	1	–1	– $\frac{1}{2}$
$s_2$	0	–1	1	$\frac{1}{2}$

A pessimistic decision maker will dislike the vagueness about the states. If he chooses act  $g$  he will give most weight to the unfavorable state  $s_2$ , if he chooses act  $h$  he will give most weight to the unfavorable state  $s_1$ . Act  $f$  does not involve any vagueness and will be preferred both to  $g$  and  $h$ . The effect of concern to us is the 'hedging' involved in mixing  $g$  and  $h$ . In the mixture  $(f=)\frac{1}{2}g + \frac{1}{2}h$ , the favorable outcome of  $g$  given  $s_1$  gives a *hedge* against the unfavorable outcome of  $h$  given  $s_1$ , and the favorable outcome of  $h$  given  $s_2$  gives a hedge against the unfavorable outcome of  $g$  given  $s_2$ . In this extreme case a complete hedge against vagueness occurs. Appreciation of the hedging can be modeled by Choquet-expected utility, by setting  $\lambda := v(\{s_1\}) = v(\{s_2\}) < \frac{1}{2}$ . Then both for  $g$  and  $h$  the most favorable state gets assigned, according to Formula (3), marginal fuzzy-measure contribution  $\lambda$ , the least favorable state gets assigned the higher marginal fuzzy-measure contribution  $1 - \lambda$ .

For the sequel it is useful to note that, whereas  $f > g$ , mixing each act with  $\frac{1}{2}h$  leads to a reversion of preference, i.e.,  $\frac{1}{2}f + \frac{1}{2}h < \frac{1}{2}g + \frac{1}{2}h$ , or, rewritten,  $f' < f$ . Again, hedging effects explain this.

The point of the hedging in the above example was that acts were mixed which ordered states differently as regards 'favorableness': For act  $g$  state  $s_1$  was the most favorable, for act  $h$  state  $s_2$ . Such hedging effects are excluded if acts are 'comonotonic':

**Definition 2** (See Figure 2, left and middle). Acts  $f$  and  $g$  are *comonotonic* if there do not exist states  $s$  and  $t$  so that  $f(s) > f(t)$  and  $g(s) < g(t)$ .

If  $f$  and  $g$  are comonotonic, then by Wakker [54] (Lemma VI.3.3) there exists a

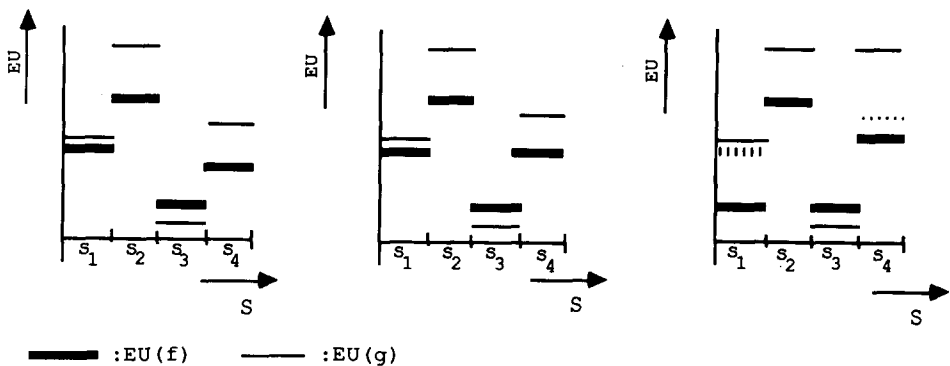


Fig. 2. *Comonotonicity and maxmin-relatedness*. Left:  $f$  and  $g$  are not comonotonic since for  $f$  state  $s_1$  is strictly more favorable than  $s_4$ , whereas for  $g$  it is reversed,  $s_4$  is strictly more favorable than  $s_1$ . Middle:  $f$  and  $g$  are comonotonic. For each act state  $s_2$  is at least as favorable as  $s_4$ ,  $s_4$  is at least as favorable as  $s_1$ ,  $s_1$  is at least as favorable as  $s_3$ .  $f$  and  $g$  are not maxmin-related, for state  $s_1$  (as well as  $s_4$ ) each act is neither maximal, nor minimal. Right: Extreme case of comonotonicity:  $f$  and  $g$  are maxmin-related. To change the acts from the middle figure to obtain maxmin-relatedness, EU-values had to be 'driven to the extreme' (old values indicated by dashed lines). In  $s_1$  the EU-value of  $f$  was lowered to the minimal EU-value of  $f$ , in  $s_4$  the EU-value of  $g$  was increased to the maximal EU-value of  $g$ . Now in  $s_1$  and  $s_3$  act  $f$  has its minimal EU-values, in  $s_2$  and  $s_4$  act  $g$  has its maximal EU-value.

bijection  $\pi: \{1, \dots, n\} \rightarrow S$  so that simultaneously

$$f(\pi(1)) \geq f(\pi(2)) \geq \dots \geq f(\pi(n))$$

and

$$g(\pi(1)) \geq g(\pi(2)) \geq \dots \geq g(\pi(n)).$$

This shows that two acts are comonotonic if and only if the favorability ordering of states, described by  $\pi$ , can be taken the same for the involved acts. A set of acts is *comonotonic* if every pair of acts in the set is comonotonic. As shown in Wakker [54] (Lemma VI.3.3) then a  $\pi$  as above can be obtained applying simultaneously to all acts in the set. In other words, on a comonotonic set the ‘act-dependent’ probabilities  $P^\pi$  in (4) can be taken constant, and the Choquet integral behaves as a usual additive integral with respect to a usual additive probability measure. Dellacherie [10] was the first to see this role of comonotonicity for nonadditive measures and the Choquet integral. The additive character of the Choquet integral was independently observed in Schmeidler [42, 43] and Murofushi and Sugeno [36]; the rank-ordering of outcomes as in Section 3, line 4 in the latter reflects comonotonicity.

The following condition was used in Anscombe and Aumann [3] to characterize expected utility in the present set-up, with money lotteries as consequences.

**Definition 3** (see Figure 3). We say  $\geq$  satisfies *independence* if, for all acts  $\{f, g, h\}$ , and  $0 < \alpha < 1$ ,

$$f > g \Rightarrow \alpha f + (1 - \alpha)h > \alpha g + (1 - \alpha)h.$$

The idea is that, if in a mixture of acts, one ‘ingredient’ ( $g$  above) is replaced by a better ‘ingredient’ ( $f$  above), then the mixture should improve by that. The condition does not reckon with attitudes towards hedging, and then, as in Example 1, is not appropriate if there is vagueness inducing hedging. Independence is the natural analogue of Savage’s famous ‘sure-thing principle’; about that principle analogous intuitive comments can be made.

The term *cardinal*, used in the following theorems, in this paper is no more than a convenient abbreviation of ‘unique up to a positive affine transformation’. A transformation is *positive affine* if it multiplies with a positive real number and/or adds up a real number. Section 5 proves Theorem 11, and derives Theorems 4 and 6 as corollaries.

**Theorem 4** (Anscombe and Aumann). *Let  $\mathcal{P}$  be the set of money lotteries,  $S$  a finite set of states,  $\mathcal{F}$  the set of functions from  $S$  to  $\mathcal{P}$  (acts), and  $\geq$  a nontrivial preference relation on  $\mathcal{F}$ . The following two statements are equivalent:*

(i) *There exist a probability measure  $P$  on  $S$ , and a ‘utility function’  $U: \mathbb{R} \rightarrow \mathbb{R}$ , so that  $f \mapsto \int_S EU \circ f \, dP$  represents  $\geq$ .*

(ii) *The preference relation  $\geq$  is a weakly monotonic mixture-continuous weak order, satisfying independence.*

*Further,  $P$  in (i) is uniquely determined, and  $U$  is cardinal.*

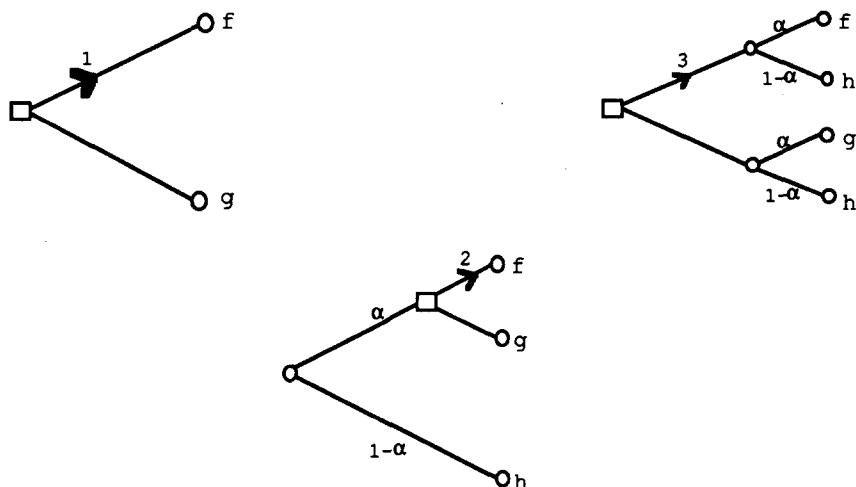


Fig. 3 (Independence). In the  $\square$ -points the decision maker must choose; in the  $\circ$ -points chance (say roulette wheel) chooses. The arrows  $>$  indicate preferences. According to independence the (very bold) preference 1 should imply the preference 3.

One justification argues that if in a mixture of  $g$  and  $h$  the ingredient  $g$  is replaced by the 'better' ingredient  $f$ , then the mixture as a whole should become better by that.

Another justification argues as follows. The preference 1 should imply the preference 2, the involved choice situations being identical. The preference 2 should imply the preference 3, the involved choice situations being 'strategically' equivalent.

Comonotonic independence requires the implication of preference 3 by preference 1 only if all involved acts are comonotonic, maxmin-independence requires the implication only if all involved acts are maxmin-related. The idea is that, because of vagueness, the mixing of acts may induce 'disturbing' hedging effects. Hence, only if hedging effects are excluded by comonotonicity or maxmin-relatedness, the implication is required.

So independence, in the presence of the other conditions, leads to the usual 'probabilized' expected utility, where no vagueness about the states is incorporated. In the presence of hedging-attitudes, the implication in independence is natural only when there is no 'disturbing' hedging effect in the involved mixings. This leads to the following definition, introduced in Schmeidler [42], and weaker than independence.

**Definition 5** (see Figure 3). We say  $\succsim$  satisfies *comonotonic independence* if, for all comonotonic  $\{f, g, h\}$ , and  $0 < \alpha < 1$ ,

$$f > g \Rightarrow \alpha f + (1 - \alpha)h > \alpha g + (1 - \alpha)h.$$

Schmeidler [42] proved, for the context of decision theory with 'nonadditive probabilities' instead of fuzzy measures, the following result. He indicated an extension of the result to infinite state spaces and 'bounded' acts (i.e.,  $EU \circ f$  is bounded).

**Theorem 6** (Schmeidler). *Let  $\mathcal{P}$  be the set of money lotteries,  $S$  a finite set of states,  $\mathcal{F}$  the set of functions from  $S$  to  $\mathcal{P}$  (acts), and  $\geq$  a nontrivial preference relation over  $\mathcal{F}$ . The following two statements are equivalent:*

(i) *There exist a fuzzy measure  $\nu$  on  $S$ , and a function  $U: \mathbb{R} \rightarrow \mathbb{R}$ , so that  $f \mapsto \int_S EU \circ f \, d\nu$  represents  $\geq$ .*

(ii) *The preference relation  $\geq$  is a weakly monotonic mixture-continuous weak order, satisfying comonotonic independence.*

*Further,  $\nu$  in (i) is uniquely determined, and  $U$  is cardinal.*

An extreme case of comonotonicity is the case where one act, say  $f$ , is constant. Then no optimism or pessimism is involved in the valuation of  $f$ , so, as one easily sees:

**Lemma 7.** *Every act  $g$  is comonotonic with every constant act  $f$ .  $\square$*

Restricting ‘comonotonicity’ in Definition 5 to the extreme case where one act (say  $h$ ) is constant, leads to a condition too weak for our purposes. The following, somewhat less extreme, case will however suffice. Note that the following definition is symmetric in  $f$  and  $g$ .

**Definition 8.** Acts  $f$  and  $g$  are *maxmin-related* if

either for every state  $s$ :  $[\forall t \in S: f(s) \geq f(t)]$  or  $[\forall t \in S: g(s) \leq g(t)]$ ,

or for every state  $s$ :  $[\forall t \in S: g(s) \geq g(t)]$  or  $[\forall t \in S: f(s) \leq f(t)]$ .

So one of the acts should be the ‘max-act’ ( $f$  in the ‘either-case’ above,  $g$  in the ‘or-case’), the other the ‘min-act’ ( $g$  in the ‘either-case’ above,  $f$  in the ‘or-case’), so that: *For every state: Either the max-act assigns its maximal value, or the min-act assigns its minimal value.* This explains the term. It should not appeal to max–min-like optimization techniques. Finally, a set of acts is *maxmin-related* if every pair of different acts in it is maxmin-related. (Note that an act  $f$  usually is not maxmin-related with itself!) Note that for maxmin-related acts  $f$ ,  $g$  never simultaneously  $f(s) > f(t)$  and  $g(s) < g(t)$ : Given these preferences, state  $s$  excludes the possibility that  $f$  were the min-act and  $g$  the max-act, state  $t$  excludes the possibility that  $f$  were the max-act and  $g$  the min-act. So maxmin-related acts are comonotonic:

**Lemma 9.** *A maxmin-related set is comonotonic.  $\square$*

The following definition replaces ‘comonotonic’ in Definition 5 by ‘maxmin-related’, thus leads to a definition which in general is weaker than comonotonic independence. We shall however see that in the presence of some other conditions maxmin-independence becomes equivalent to comonotonic independence. (Compare statement (ii) in Theorem 6 and statement (ii) in Theorem 11; each is equivalent to the same statement (i).)

**Definition 10** (see Figure 3). We say  $\succsim$  satisfies *maxmin-independence* if, for all maxmin-related  $\{f, g, h\}$ , and  $0 < \alpha < 1$ ,

$$f > g \Rightarrow \alpha f + (1 - \alpha)h > \alpha g + (1 - \alpha)h.$$

The following theorem strengthens the theorem of Schmeidler [42]. Recall that the notation *EU* has been introduced in (1).

**Theorem 11.** *Let  $\mathcal{P}$  be the set of money lotteries,  $S$  a finite set of states,  $\mathcal{F}$  the set of functions from  $S$  to  $\mathcal{P}$  (acts), and  $\succsim$  a nontrivial preference relation over  $\mathcal{F}$ . The following two statements are equivalent:*

(i) *There exist a fuzzy measure  $\nu$  on  $S$ , and a function  $U: \mathbb{R} \rightarrow \mathbb{R}$ , so that  $f \mapsto \int_S EU \circ f \, d\nu$  represents  $\succsim$ .*

(ii) *The preference relation  $\succsim$  is a weakly monotonic mixture-continuous weak order, satisfying maxmin-independence.*

*Further,  $\nu$  in (i) is uniquely determined, and  $U$  is cardinal.*

In the present context, where lotteries are for money, it is very natural that the function  $U$  in (i) in the theorems above is increasing; this can be obtained by adding in (ii) everywhere the requirement that  $\alpha \succsim \beta \Leftrightarrow \alpha \geq \beta$  for any degenerate lotteries  $\alpha (= (1; \alpha))$  and  $\beta (= (1; \beta))$ ; or, equivalently, by adding an assumption of ‘(strict) stochastic dominance’.

For the proof of the above theorem, given in Section 5, we could have closely followed Schmeidler [42, 43]. His proof can straightforwardly be adapted to our case of maxmin-relatedness. His proof is however not easily available. It is split into two parts. In Schmeidler [42] it is proved that a functional, analogous to *CU* below, satisfies a list of conditions. Next Schmeidler [43] is used to show in general that functionals satisfying the involved conditions are a Choquet integral. We take the opportunity to give an alternative presentation, with a ‘one-stage all-in proof’, without requiring knowledge of functional analysis, as Schmeidler’s proof does. Our proof uses ideas of Theorem 3 in Anger [2], who already, in a slightly different context, obtained the result of Schmeidler [43] with comonotonic independence weakened to ‘weak additivity’, an analogue to maxmin-relatedness. Also the ‘strong uncertainty aversion’ axiom in Chateauneuf [6] was inspiring.

## 5. Proof of theorems

### 5.1. Proof of Theorem 11

First suppose (i) holds. All conditions apart from maxmin-independence are well-known and straightforward. We shall derive, stronger than maxmin-independence, comonotonic independence. So let  $\{f, g, h\}$  be comonotonic. By Wakker [54] (Lemma VI.3.3) we can take  $\pi$  so that, for all  $f' \in \{f, g, h\}$ ,  $f'(\pi(1)) \geq f'(\pi(2)) \geq \dots \geq f'(\pi(n))$ . It is straightforward from (i) that this also holds for every mixture of  $f, g, h$ . So we can write, according to (4) (noting that *EU* represents  $\succsim$  on the constant acts, so that the  $\pi$  in (2) can be chosen the same

as the  $\pi$  above),  $\int EU \circ f' \, dv = \sum_{i=1}^n P^\pi(\{s_i\}) EU(f'(s_i))$ , for any mixture  $f'$  of  $f$ ,  $g$ ,  $h$ . Consequently

$$\int (EU \circ [\alpha f' + (1 - \alpha)h]) \, dv = \alpha \left[ \int (EU \circ f') \, dv \right] + (1 - \alpha) \left[ \int (EU \circ h) \, dv \right]$$

both for  $f' = f$ , and  $f' = g$ . From this and the fact that Choquet-expected utility represents  $\geq$ , comonotonic independence follows, hence also maxmin-independence.

Next we suppose that (ii) holds, and derive (i) and the uniqueness result at the end of the theorem. Since all constant acts are mutually maxmin-related,  $\geq$  restricted to the set of constant acts satisfies 'usual' independence; further it is a mixture-continuous weak order. Hence (see e.g. Fishburn [17] (Theorem 2.1)) there exists a cardinal function  $U: \mathbb{R} \rightarrow \mathbb{R}$  so that on the set of constant acts  $P$ ,  $\geq$  is represented by  $P \mapsto EU(P)$ .

Note that the function  $U$ , mentioned in (i), has to be like the  $U$  just defined; hence  $U$  now has been obtained (up to a positive affine transformation). The following lemma derives 'certainty equivalents'. Because the mixture set generated by one act together with all constant acts usually is not maxmin-related, the proof is less simple than in Schmeidler's case of comonotonic independence.

**Lemma 12.** *For every act  $f$  there exists a constant act  $P^f$  so that  $f \approx P^f$ .*

**Proof.** Let  $k, l$  be so that  $f(s_k) \geq f(s_i) \geq f(s_l)$  for all  $1 \leq i \leq n$ . By weak monotonicity,  $f(s_k) \geq f \geq f(s_l)$ . Suppose the preferences are strict (otherwise we are done immediately). Mixture-continuity implies that (within  $EU(\mathcal{P})$ ) the set of  $EU$ -values of the constant acts strictly preferred to  $f$  is open, as well as the set of  $EU$ -values of the constant acts to which  $f$  is strictly preferred. Since each of these open sets is nonempty, by convex-rangedness of  $EU$ , there must be a constant act  $P^f$  equivalent to  $f$ .  $\square$

Since all possible choices of  $P^f$  above are equivalent and have the same  $EU$ -value, we can define  $CU: f \mapsto EU(P^f)$ . Obviously  $CU$  represents  $\geq$ ; this will often be used without further mention. Further  $CU$  is identical to  $EU$  on the constant acts.

*The remainder of this proof will be devoted to the demonstration that  $CU$  is Choquet-expected utility, with capacity  $v$  uniquely determined.*

Firstly, a preparatory lemma establishes a kind of 'maxmin-additivity' of  $CU$  (and implies the condition of maxmin-independence with equivalences instead of strict preferences).

**Lemma 13.** *Let acts  $f, g$  be maxmin-related. Then  $CU(\alpha f + (1 - \alpha)g) = \alpha CU(f) + (1 - \alpha)CU(g)$ .*

**Proof.** Let  $P^h \approx h$  for all acts  $h$ . We may assume  $0 < \alpha < 1$ . It suffices to show that

$$\alpha f + (1 - \alpha)g \approx \alpha P^f + (1 - \alpha)g,$$

because we can repeat this result to establish equivalence with  $\alpha P^f + (1 - \alpha)P^g$ . We shall only establish

$$\alpha f + (1 - \alpha)g \geq \alpha P^f + (1 - \alpha)g, \quad (7)$$

since all following preferences (and minimal/maximal) can be reversed. If  $P^f$  is *minimal* ( $P^f > Q$  for no  $Q$ ) then, since  $EU$  represents  $\geq$  on constant acts, the right-hand act in (7) assigns to every state a money lottery not better than the left-hand act. By weak monotonicity (7) follows. So let there be a money lottery  $Q$  so that  $P^f > Q$ . To complete the proof, we suppose that  $\alpha f + (1 - \alpha)g < \alpha P^f + (1 - \alpha)g$  and derive contradiction. Since  $Q < f$ , by maxmin-independence we may add

$$\alpha Q + (1 - \alpha)g < \alpha f + (1 - \alpha)g < \alpha P^f + (1 - \alpha)g.$$

By mixture-continuity there should be a mixture with nonzero weights of the left and right acts strictly preferred to the middle act. We are done if we show that the middle act is preferred to any such mixture. For any  $\beta > 0$ ,  $f > \beta Q + (1 - \beta)P^f$ , so that by maxmin-independence  $\alpha f + (1 - \alpha)g > \alpha(\beta Q + (1 - \beta)P^f) + (1 - \alpha)g$ .  $\square$

*Secondly*, we obtain the fuzzy measure  $v$ . By weak monotonicity and nontriviality we can fix some arbitrary money lotteries  $P^1 > P^0$ . Throughout this proof we use, for any  $P \in \mathcal{P}$  and  $A \subset S$ , the following notation:

$$(P^0_{-A}P) \text{ is the act assigning } P \text{ to } A, \text{ and } P^0 \text{ to } A^c. \quad (8)$$

Now, given  $U$ , if indeed  $CU$  is Choquet-expected utility, then the involved fuzzy measure  $v$  is straightforwardly seen to be uniquely determined as

$$v : A \mapsto [CU((P^0_{-A}P^1)) - EU(P^0)]/[EU(P^1) - EU(P^0)]. \quad (9)$$

While we have not yet established that  $CU$  is Choquet-expected utility, we can nevertheless see that the expression in (9) is invariant under application of a positive affine transformation on  $U$ , so that indeed  $v$  is independent of the particular choice of ('scale' and 'location' of) the cardinal  $U$ . Hence we define  $v$  as in (9), and have

$$v \text{ is well defined and uniquely determined.} \quad (10)$$

Mainly by weak monotonicity  $v$  is seen to be a fuzzy measure indeed. For simplicity of notation we normalize  $U$  so that

$$EU(P^0) = 0, \quad EU(P^1) = 1.$$

This gives the simpler

$$v : A \mapsto CU((P^0_{-A}P^1)). \quad (11)$$

*Thirdly*, we show:

**Lemma 14.** *For the set of acts of the form  $(P^0_{-A}P)$  with  $P \geq P^0$ ,  $CU$  is Choquet-expected utility.*



**Proof.** Twofold application of weak monotonicity (once with  $\geq$ , once with  $\leq$ ) shows

$$CU(f) = CU(g) \text{ as soon as } EU(f(s)) = EU(g(s)) \text{ for all } s. \quad (12)$$

Hence Lemma 13, with  $g = P^0$ , shows that, for  $0 \leq \alpha \leq 1$ ,  $CU(f) = \alpha CU(h)$  if  $EU(f(s)) = \alpha EU(h(s))$  for all  $s$ . Interchanging  $\alpha$  and  $1/\alpha$  and  $f$  and  $h$  if necessary, shows that this holds for all nonnegative  $\alpha$  (i.e.,  $CU$  is kind of positively homogeneous). This and (11) give the lemma.  $\square$

Fourthly, we show:

**Lemma 15.** *For the set of acts  $f$  so that  $f(s_1) \geq f(s_2) \geq \dots \geq f(s_n) \geq P^0$ ,  $CU$  is Choquet-expected utility.*

**Proof** (see Figure 4). We rewrite (6), with  $\pi(j) = s_j$  for all  $j$ , and with the order of terms reversed, as:

$$\begin{aligned} \int EU \circ f \, dv &= [v(\{s_1, \dots, s_n\}) \times EU(f(s_n))] \\ &\quad \vdots \\ &\quad + [v(\{s_1, \dots, s_{n-j+1}\}) \times [EU(f(s_{n-j+1})) - EU(f(s_{n-j+2}))]] \\ &\quad \vdots \\ &\quad + [v(\{s_1\}) \times [EU(f(s_1)) - EU(f(s_2))]]. \end{aligned} \quad (13)$$

The proof will use induction in a way that shows how the Choquet integral is ‘built up layer after layer’ by focusing during the induction step on the  $j$ -th term (‘layer’) of (13) (doubly dashed in Figure 4). We define, for  $j = 1, \dots, n$ ,  $f^j$  as the ‘truncation of  $f$  at its  $j$ -th lowest value ( $f(s_{n-j+1})$ )’, i.e.,

$$f^j \text{ is constant equal to } f(s_{n-j+1}) \text{ on } \{s_1, \dots, s_{n-j}\}, \quad f^j \text{ is } f \text{ on } \{s_{n-j+1}, \dots, s_n\}.$$

So  $f^j$  assigns (at most)  $j$  different consequences. It follows directly that  $CU$  of  $f^1$  is Choquet-expected utility,  $f^1$  being constant.

Next we suppose (induction hypothesis), for some  $j > 1$ , that  $CU$  of  $f^{j-1}$  is Choquet-expected utility (i.e., it is equal to the first  $j-1$  terms of (13)). We show that  $CU$  of  $f^j$  is Choquet-expected utility. By convex-rangedness of  $EU$  we can take  $P$  so that  $EU(P) = EU(f(s_{n-j+1})) - EU(f(s_{n-j+2}))$ . By (12) (see Figure 4b)

$$CU(\tfrac{1}{2}f^j + \tfrac{1}{2}P^0) = CU(\tfrac{1}{2}f^{j-1} + \tfrac{1}{2}P^0_{-\{s_1, \dots, s_{n-j+1}\}}P).$$

Applying Lemma 13 to both sides of the equality, interchanging terms and multiplying by 2 gives

$$CU(f^j) - CU(f^{j-1}) = CU(P^0_{-\{s_1, \dots, s_{n-j+1}\}}P) - CU(P^0).$$

Substituting (by Lemma 14) Choquet-expected utility for the  $CU$ -values at the right-hand side, and using the definition of  $P$ , results in

$$CU(f^j) - CU(f^{j-1}) = v(\{s_1, \dots, s_{n-j+1}\}) \times [EU(f(s_{n-j+1})) - EU(f(s_{n-j+2}))].$$

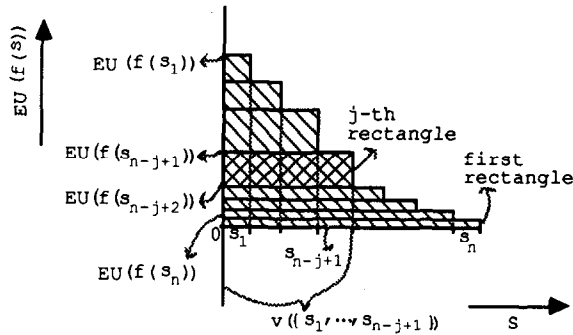


Fig. 4a. Illustration of Lemma 15. Formula (13) reverses the order of terms (or rectangles) as compared to formula (6) (or Figure 1b) in the calculation of the Choquet integral.

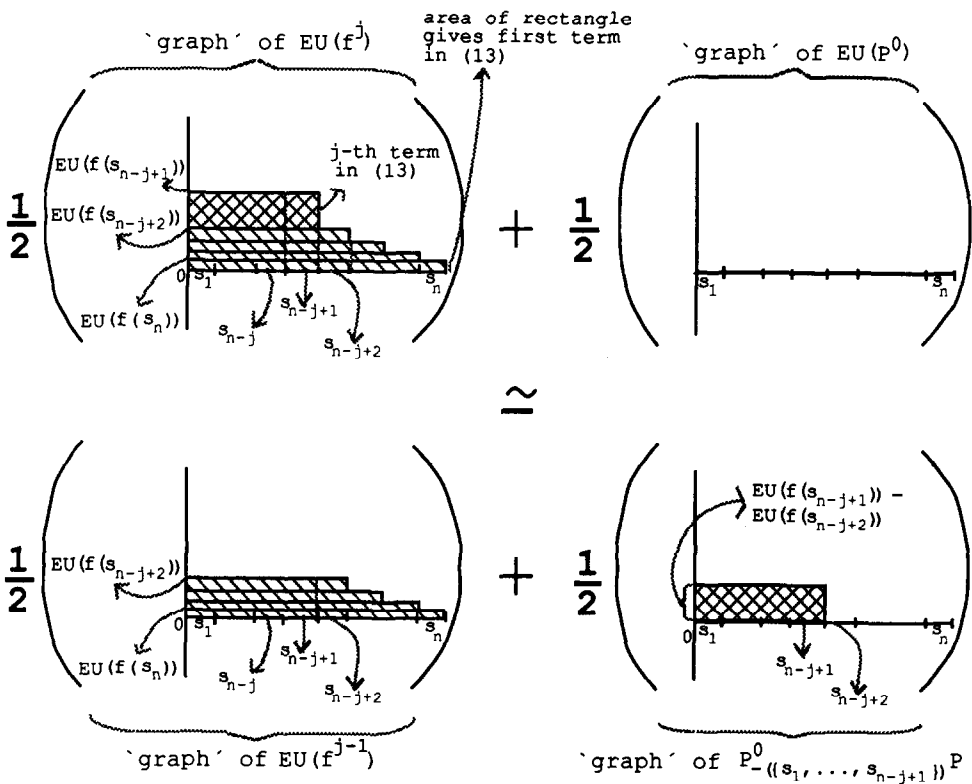


Fig. 4b. A useful equivalence. The above mixtures (mixing  $f^j$  with  $P^0$ , respectively mixing  $f^{j-1}$  with a two-valued act), give equivalent acts, by formula (12). Apply Lemma 13 to both mixtures, and interchange terms. It follows that the 'CU-difference' between the left-hand sides  $f^j$  and  $f^{j-1}$  is the same as between the right-hand sides, i.e., as the  $j$ -th term in formula (13). This term has been doubly dashed in Figures 4a and 4b.

The second term in the difference in the left-hand side, by the induction hypothesis, is the Choquet-expected utility of  $f^{j-1}$ , i.e., equals the first  $j-1$  terms in (13). The right-hand side equals the  $j$ -th term in (13). Hence the first term in the difference in the left-hand side equals the first  $j$  terms in (13), i.e., the Choquet-expected utility of  $f^j$ . By induction, for  $f = f^n$ ,  $CU(f^n)$  equals (13), i.e., the Choquet-expected utility of  $f$ .  $\square$

*Fifthly* and finally, we show for arbitrary  $f$  that  $CU(f)$  is the Choquet-expected utility of  $f$ , i.e., the Choquet integral of  $f \circ EU$ . Showing it for any  $f$  with  $f(s) \geq P^0$  for all  $s$  is fully analogous to the above lemma, by a simple permutation of states. So we finally turn to acts  $f$  so that  $P^0$  is strictly preferred to some of the assigned money lotteries. Let for such an, arbitrary but now fixed, act  $f$ ,  $P^{0'}$  be a money lottery so that  $P^{0'} \leq f(s_j)$  for each  $s_j$ , e.g.,  $P^{0'}$  is the ‘minimal’ money lottery assigned by  $f$ . Define the ‘location constant’  $l_c$  by  $l_c = -EU(P^{0'})$ . Replace the cardinal  $U$  by  $U' = U + l_c$ , resulting in  $EU' = EU + l_c$  instead of  $EU$ , and  $CU' = CU + l_c$  instead of  $CU$ . We can repeat the above exposition with  $P^{0'}$  instead of  $P^0$ , and  $P^1$  instead of  $P^1$  so that  $EU'(P^1) = 1 = EU(P^1) + l_c$ . The result is that  $CU'(g)$  is the Choquet integral of  $g \circ EU'$  for any  $g$  with  $g(s) \geq P^{0'}$  for all  $s$ , w.r.t. a capacity  $v'$  ( $v' = v$  will be shown). The result is also, by (5) and the role of  $l_c$  above, that  $CU(g)$  is the Choquet integral, w.r.t. the capacity  $v'$ , of  $g \circ EU$  for any  $g$  with  $g(s) \geq P^{0'}$  for all  $s$ , so certainly for  $f$  (so that showing  $v = v'$  will complete the proof), and certainly for any  $g$  with  $g(s) \geq P^0$  for all  $s$ . The latter, together with (9), shows that  $v = v'$ .

## 5.2. Proof of Theorems 6 and 4

The implication (i)  $\Rightarrow$  (ii) in Theorem 6 can be inferred from the beginning of Subsection 5.1. The implication (ii)  $\Rightarrow$  (i) follows from Theorem 11, since comonotonic independence implies maxmin-independence. The implication (i)  $\Rightarrow$  (ii) in Theorem 4 is straightforward. Finally, suppose (ii) there holds. This implies (ii) of Theorem 11, hence (i) there. Remains to be shown that the fuzzy measure  $v$  is additive. The result of Lemma 13 can now completely analogously be derived for all  $f, g$ . Applying it, for disjoint  $A, B$ , to the  $CU$ -values of the left- and right-hand side of

$$\frac{1}{2}(P_{-A \cup B}^0 P^1) + \frac{1}{2}P^0 = \frac{1}{2}P_{-A}^0 P^1 + \frac{1}{2}P_{-B}^0 P^1$$

and using (11), gives additivity of the fuzzy measure.

## 6. A behavioral foundation for Zadeh’s possibility measures

In their full generality, fuzzy measures do not seem to be specified enough to be useful. This was argued for fuzzy set theory in Dubois and Prade [14] (above formula (15)), and several other places. Because of this there is interest in specified forms of fuzzy measures.

Hence the results in Theorem 6 and its supplement Theorem 11 seem to be too general for direct applications. Still we hope these results will prove useful.

Because of their generality, they provide a convenient starting point for the derivation of more specified results. (A short survey of specified kinds of fuzzy measures and integrals is provided in Dubois and Prade [15].) As an example we show how to obtain a behavioral foundation of 'possibility measures', introduced in Zadeh [63]. It is our hope to find a behavioral foundation of the belief functions of Dempster and Shafer, building upon Theorem 11.

**Definition 16.** A fuzzy measure  $v$  is a *possibility measure* if, for all subsets  $A, B$  of  $S$ ,

$$v(A \cup B) = \max\{v(A), v(B)\}.$$

Below we use the notation (8).

**Lemma 17.** Let the assumptions, and statement (i), in Theorem 11, hold. Let  $A, B$  be subsets of  $S$ . Then

$$v(A \cup B) = \max\{v(A), v(B)\}$$

if and only if, for some money lotteries  $P^1 > P^0$ ,

$$(P^0_{-A}P^1) \simeq (P^0_{-A \cup B}P^1) \quad \text{or} \quad (P^0_{-B}P^1) \simeq (P^0_{-A \cup B}P^1). \quad \square \quad (14)$$

The proof of the above lemma is obtained by substituting Choquet-expected utility, and is left to the reader. Let us next combine the above lemma, and Theorem 11, to formulate a behavioral foundation of possibility measures:

**Theorem 18.** Let  $\mathcal{P}$  be the set of money lotteries,  $S$  a finite set of states,  $\mathcal{F}$  the set of functions from  $S$  to  $\mathcal{P}$  (acts), and  $\succsim$  a nontrivial preference relation over  $\mathcal{F}$ . The following two statements are equivalent:

(i) There exist a possibility measure  $v$  on  $S$ , and a function  $U: \mathbb{R} \rightarrow \mathbb{R}$ , so that  $f \mapsto \int_S EU \circ f \, dv$  represents  $\succsim$ .

(ii) The preference relation  $\succsim$  is a weakly monotonic mixture-continuous weak order, satisfying maxmin-independence; further for any subsets  $A, B$  of  $S$ , formula (14) holds.

The uniqueness results are as in Theorem 11.  $\square$

Note that (14), the characterizing condition for possibility measures in the presence of the other conditions, reflects the intuitive idea of possibility measures that one of the subsets should be 'decisive', and that a kind of interaction between disjoint sets should be excluded. Zadeh [63] (formula (2.26)) defined two events  $A, B$  to be *noninteractive* if  $v(A \cap B) = \min\{v(A), v(B)\}$ . (Also the max-operation as applied to unions above is sometimes taken as noninteractiveness of events.) The following lemma shows how to characterize noninteractiveness. It is analogous to Lemma 17, and again the proof is left to the reader.

**Lemma 19.** Let the assumptions, and statement (i), in Theorem 11 hold. Subsets  $A, B$  of  $S$  are noninteractive if and only if, for some money lotteries  $P^1 > P^0$ ,

$$(P^0_{-A}P^1) \simeq (P^0_{-A \cap B}P^1) \quad \text{or} \quad (P^0_{-B}P^1) \simeq (P^0_{-A \cap B}P^1). \quad \square \quad (15)$$

## 7. Conclusion

Measures for valuing knowledge are usually judged by their intuitive appeal-  
ingness, computational efficiency, inner coherence, and/or performance in some  
applications. For one kind of such measures, fuzzy measures, this paper has given  
another kind of criteria, criteria derived from the decision-theoretic appropriate-  
ness for implementing decisions. This has been based on a theorem introduced by  
Schmeidler [42] into decision theory, and a strengthening of that theorem. While  
the criteria seem to be too mild, i.e., fuzzy measures seem to be too general, to  
be useful in applications, the obtained results may be useful as a starting point for  
other results. One example of that has been given, by deriving a behavioral  
foundation of possibility measures.

Hopefully the results of this paper will help to value fuzzy measures and their  
many specified forms by more than ad hoc criteria.

## 8. Appendix: Mathematical generalizations

To make the message of this paper as clear and accessible as possible, we have  
assumed that consequences are real numbers, designating amounts of money, or  
are money lotteries, and we have assumed that the state space is finite. In fact,  
the only mathematical structure of the consequence space  $\mathcal{P}$  that we used was  
that this set is a 'mixture space' (see for instance Fishburn [17] (Section 2.1), Fish-  
burn [18] (Definition 8.3), Wakker [54] (Definition VII.2.1), etc.). This applies  
for instance to any convex subset of a linear space, a special case of which is the case  
of any convex set of probability distributions, such as considered in Schmeidler  
[42]. The only property of  $EU$  that we used was that it is affine. Hence our results  
can for instance be applied to the case considered in Yaari [59], Chateauneuf [6]  
and Denneberg [12], with consequences (certain) amounts of money and utility  
linear.

If indeed  $\mathcal{P}$  is a convex subset of a linear space, then the generalization of  
Theorem 6 to the case of continuous (instead of affine) utility, (even when  $\mathcal{P}$  is  
any connected topological space) has been carried out in Wakker [53] and  
Wakker [54] (Chapter VI). The extension of these results to infinite state spaces  
has been given in Wakker [52]. All the references just mentioned have been  
formulated for the context of decision theory.

The following observation can be established by rereading the proof, and  
checking that nowhere we used more of maxmin-independence than the  
case described in the lemma. In particular Lemma 13 was needed only for two-  
valued  $g$ .

**Observation 20.** *In Statement (ii) in Theorem 11, maxmin-independence can  
further be weakened by requiring it only for the case where, in Definition 10, the  
act  $h$  is two-valued, and  $f$  or  $g$  is constant.  $\square$*

## References

- [1] M. Allais, Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école américaine, *Coll. Internat. Centre National Rech. Sci. (Econométrie)* **40** (1953) 257–332. Translated into English, with additions, as: The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American school, in: M. Allais, O. Hagen, Eds., *Expected Utility Hypotheses and the Allais Paradox* (Reidel, Dordrecht, 1979) 27–145.
- [2] B. Anger, Representations of capacities, *Math. Ann.* **229** (1977) 245–258.
- [3] F.J. Anscombe and R.J. Aumann, A definition of subjective probability, *Ann. Math. Statist.* **34** (1963) 199–205.
- [4] K.J. Arrow, *Aspects of the Theory of Risk-Bearing* (Academic Bookstore, Helsinki, 1965) Elaborated as K.J. Arrow, *Essays in the Theory of Risk-Bearing* (North-Holland, Amsterdam, 1971).
- [5] M.J. Bolaños, M.J. Lamata and S. Moral, Decision making problems in a general environment, *Fuzzy Sets and Systems* **25** (1988) 135–144.
- [6] A. Chateauneuf, Uncertainty aversion and risk aversion in models with nonadditive probabilities, in: B.R. Munier, Ed., *Risk, Decision and Rationality* (Reidel, Dordrecht, 1988) 615–629.
- [7] G. Choquet, Theory of capacities, *Ann. Inst. Fourier (Grenoble)* **5** (1953–4) 131–295.
- [8] B. de Finetti, La prévision: Ses lois logiques, ses sources subjectives, *Ann. Inst. Henri Poincaré* **7** (1937) 1–68. Translated into English by H.E. Kyburg, Foresight: Its logical laws, its subjective sources, in: H.E. Kyburg, H.E. Smokler, Eds., *Studies in Subjective Probability* (Wiley, New York, 1964) 53–118.
- [9] B. de Finetti, *Probability, Induction and Statistics* (Wiley, New York, 1972).
- [10] C. Dellacherie, Quelques commentaires sur les prolongements de capacités, *Seminaire de Probabilités V Strasbourg*, Lecture Notes in Mathematics No. 191 (Springer-Verlag, Berlin, 1970).
- [11] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Statist.* **38** (1967) 325–339.
- [12] D. Denneberg, On non-expected-utility preferences, Paper presented at 4th FUR Conference, Budapest 1988).
- [13] T.S.H. Driessen, *Cooperative Games, Solutions and Applications* (Kluwer Academic Publishers, Dordrecht, 1988).
- [14] D. Dubois and H. Prade, Modelling uncertainty and inductive inference: A survey of recent non-additive probability systems, *Acta Psychologica* **68** (1988) 53–78.
- [15] D. Dubois and H. Prade, Fuzzy measures: Fuzzy integral approach, in: M.G. Singh, Ed., *Systems & Control Encyclopedia; Theory, Technology, Applications* (Pergamon, New York, 1988) 1821–1822.
- [16] P.C. Fishburn, *Utility Theory for Decision Making* (Wiley, New York, 1970).
- [17] P.C. Fishburn, *The Foundations of Expected Utility* (Reidel, Dordrecht, 1982).
- [18] P.C. Fishburn, *Nonlinear Preference and Utility Theory* (Johns Hopkins University Press, Baltimore, MD, 1988).
- [19] I. Gilboa, Expected utility with purely subjective non-additive probabilities, *J. Math. Econom.* **16** (1987) 65–88.
- [20] R. Giles, The concept of grade of membership, *Fuzzy Sets and Systems* **25** (1988) 297–323.
- [21] J. Gordon and E.H. Shortliffe, A method for managing evidential reasoning in a hierarchical hypothesis space, *Artificial Intelligence* **26** (1985) 323–357.
- [22] J.Y. Halpern and M.O. Rabin, A logic to reason about likelihood, *Artificial Intelligence* **32** (1987) 379–405.
- [23] E. Hisdal, Are grades of membership probabilities?, *Fuzzy Sets and Systems* **25** (1988) 325–348.
- [24] U. Höhle, Integration with respect to fuzzy measures, *Proceedings IFAC Symposium on Theory and Application of Digital Control*, New Delhi, (Jan. 1982) 35–37.
- [25] W. Hua, The properties of some non-additive measures, *Fuzzy Sets and Systems* **27** (1988) 373–377.

- [26] J.Y. Jaffray, Linear utility theory for belief functions, *Operations Research Letters* **8** (1989) 107–112.
- [27] J.Y. Jaffray, Généralisation du critère de l'utilité espérée aux choix dans l'incertain régulier', *RAIRO Oper. Res.* (1989) (forthcoming).
- [28] L.N. Kanal and J.F. Lemmer, *Uncertainty in Artificial Intelligence*, Machine Intelligence and Pattern Recognition, Vol. 4 (North-Holland, Amsterdam, 1986).
- [29] L.N. Kanal and J.F. Lemmer, *Uncertainty in Artificial Intelligence 2*, Machine Intelligence and Pattern Recognition, Vol. 5 (North-Holland, Amsterdam, 1987).
- [30] E.P. Klement and D. Ralescu, Nonlinearity of the fuzzy integral, *Fuzzy Sets and Systems* **11** (1983) 309–315.
- [31] H.E. Kyburg, Jr., Bayesian and non-Bayesian evidential updating, *Artificial Intelligence* **31** (1987) 271–293.
- [32] H.E. Kyburg, Jr., Addendum to Bayesian and non-Bayesian evidential updating, *Artificial Intelligence* **36** (1988) 265–266.
- [33] D.V. Lindley, Scoring rules and the inevitability of probability, *Internat. Statist. Rev.* **50** (1982) 1–26.
- [34] M.J. Machina, 'Expected utility' analysis without the independence axiom, *Econometrica* **50** (1982) 277–323.
- [35] M.J. Machina, Choice under uncertainty: Problems solved and unsolved, *Economic Perspectives* **1** (1987) 121–154.
- [36] T. Murofushi and M. Sugeno, An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure, *Fuzzy Sets and Systems* **29** (1989) 201–227.
- [37] N.J. Nilsson, Probabilistic logic, *Artificial Intelligence* **28** (1986) 71–87.
- [38] J. Pearl, Fusion, propagation, and structuring in belief networks, *Artificial Intelligence* **29** (1986) 241–288.
- [39] J.W. Pratt, Risk aversion in the small and in the large, *Econometrica* **32** (1964) 122–136.
- [40] J. Quiggin, A theory of anticipated utility, *J. Econom. Behaviour and Organization* **3** (1982) 323–343.
- [41] L.J. Savage, *The Foundations of Statistics* (Wiley, New York, 1954). (Second edition: Dover, New York, 1972).
- [42] D. Schmeidler, Subjective probability, and expected utility without additivity, *Econometrica* **57** (1989) 571–587.
- [43] D. Schmeidler, Integral representation without additivity, *Proc. Amer. Math. Soc.* **97** (1986) 255–261.
- [44] G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, NJ, (1976).
- [45] G. Shafer and R. Logan, Implementing Dempster's rule for hierarchical evidence, *Artificial Intelligence* **32** (1987) 271–298.
- [46] E.H. Shortliffe and B.G. Buchanan, A model of inexact reasoning in medicine, *Math. Biosci.* **23** (1975) 351–379.
- [47] F. Suárez García and P. Gil Álvarez, Two families of fuzzy integrals, *Fuzzy Sets and Systems* **18** (1986) 67–81.
- [48] M. Sugeno, Theory of fuzzy integrals and their applications, Ph.D. Thesis, Tokyo Institute of Technology (1974).
- [49] M. Sugeno and T. Murofushi, Pseudo-additive measures and integrals, *J. Math. Anal. Appl.* **122** (1987) 197–222.
- [50] M. Sugeno and T. Murofushi, Choquet's integrals as an integral form for the general class of fuzzy measures, *Preprints of 2nd IFSA Congress* (1988) 408–411.
- [51] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, NJ, 1944/1947/1953).
- [52] P.P. Wakker, Nonadditive probabilities and derived strengths of preferences, Internal Report 87 MA 03, University of Nijmegen, Department of Mathematical Psychology, Nijmegen, The Netherlands (1987).
- [53] P.P. Wakker, Continuous subjective expected utility with nonadditive probabilities, *J. Math. Econom.* **18** (1989) 1–27.

- [54] P.P. Wakker, *Additive Representations of Preferences, A New Foundation of Decision Analysis* (Kluwer Academic, Dordrecht, 1989).
- [55] P.P. Wakker, A decision-theoretic foundation for fuzzy measures (Extended abstract), *Comm. IFSA Math. Chapter*, submitted (1989).
- [56] S. Weber,  $\perp$ -Decomposable measures and integrals for Archimedean  $t$ -conorms  $\perp$ , *J. Math. Anal. Appl.* **101** (1984) 114–138.
- [57] S. Weber, Two integrals and some modified versions – Critical Remarks, *Fuzzy Sets and Systems* **20** (1986) 97–105.
- [58] M.E. Yaari, Some remarks on measures of risk aversion and on their uses, *J. Econom. Theory* **1** (1969) 315–329.
- [59] M.E. Yaari, The dual theory of choice under risk, *Econometrica* **55** (1987) 95–115.
- [60] L.A. Zadeh, Fuzzy sets, *Inform. and Control* **8** (1965) 338–353.
- [61] L.A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Trans. Systems Man Cybernet.* **3** (1973) 28–44.
- [62] L.A. Zadeh, Calculus of fuzzy restrictions, in: L.A. Zadeh, K.S. Fu, K. Tanaka, M. Shimura, Eds., *Fuzzy Sets and their Applications to Cognitive and Decision Processes* (Academic Press, New York, 1975) 1–39.
- [63] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets and Systems* **1** (1978) 3–28.