PERGAMON

Omega 28 (2000) 37–50

omega
The International Journal of Management Science

www.elsevier.com/locate/orms

The Dempster–Shafer theory of evidence: an alternative approach to multicriteria decision modelling

Malcolm Beynon, Bruce Curry*, Peter Morgan

Cardiff Business School, Colum Drive, Cardiff, CF1 3EU, UK

Received 1 December 1998; accepted 1 June 1999

Abstract

The objective of this paper is to describe the potential offered by the Dempster–Shafer theory (DST) of evidence as a promising improvement on "traditional" approaches to decision analysis. Dempster–Shafer techniques originated in the work of Dempster on the use of probabilities with upper and lower bounds. They have subsequently been popularised in the literature on Artificial Intelligence (AI) and Expert Systems, with particular emphasis placed on combining evidence from different sources. In the paper we introduce the basic concepts of the DST of evidence, briefly mentioning its origins and comparisons with the more traditional Bayesian theory. Following this we discuss recent developments of this theory including analytical and application areas of interest. Finally we discuss developments via the use of an example incorporating DST with the Analytic Hierarchy Process (AHP). © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Dempster–Shafer theory; AHP; Probabilities; Multicriteria decision making; Belief functions; Evidence theory

1. Introduction: History of probability, uncertainty

The Dempster–Shafer theory (DST) of evidence originated in the work of Dempster [1] on the theory of probabilities with upper and lower bounds. It has since been extended by numerous authors and popularised, but only to a degree, in the literature on Artificial Intelligence (AI) and Expert Systems, as a technique for modelling reasoning under uncertainty. In this respect it can be seen to offer numerous advantages over the more "traditional" methods of Statistics and Bayesian decision theory. Hajek [2] remarked that real, practical applications of DST methods have been rare, but subsequent to these remarks there has been a marked increase in the applications incorporating the use of DST. Although DST is not in widespread use, it has been applied with some success to such topics as face recognition [3], statistical classification [4] and target identification [5]. Additional applications centred around multi-source information, including medical diagnosis [6] and plan recognition [7].

An exception is the paper by Cortes-Rello and Golshani [8], which although written for a computing science/AI readership does deal with the "knowledge domain" of forecasting and Marketing Planning. For those with even limited knowledge of these domains the paper appears rather naive, referring for example to rather venerable old editions of standard texts such as [9]. The aim of this paper is to suggest that there is a good deal of potential in the DST approach, which is as yet very largely unexploited.

The origins of the mathematical theory of probability date back at least to the work of the eighteenth century

* Corresponding author. Tel.: +44-1222-875668; fax: +44-1222-874419.
E-mail address: curry@cardiff.ac.uk (B. Curry)

0305-0483/00/$ - see front matter © 2000 Elsevier Science Ltd. All rights reserved.
PII: S0305-0483(99)00033-X

scholar, The Reverend Thomas Bayes [10], whose work was published posthumously in 1763. It provides the foundations for the theory of statistical inference (involving both estimation and the testing of hypotheses) and for techniques of decision making under uncertainty. The roots of decision analysis lie in the 1930s and 1940s. Wald [11], included the "complete class theorem", which stated that any procedure in a statistical decision problem can be beaten or at least matched in performance by Bayesian procedures, defined as procedures based on the adoption of some set of prior probabilities. The fact that numerous statistical principles and techniques may be developed without using prior and posterior probability distributions involves no loss of generality, given that the special case of a uniform or rectangular prior distribution may be adopted.

Decision analysis relies more on a subjectivist view of the use of probability, whereby the probability of an event indicates the degree to which someone believes it, rather than the alternative frequentist approach. The latter approach is based only on the number of times an event is observed to occur. As Savage [12,13] discusses, the subjectivists have been responsible for much of the theoretical work into statistical practice. He goes on to argue that the frequentists hold an uneasy upper hand over their Bayesian/Subjective colleagues in the domain of mathematical statistics. Bayesian statisticians may agree that their goal is to estimate objective probabilities from frequency data, but they advocate using subjective prior probabilities to improve the estimates [14]. French [15] questions Savage's theory of subjective expected utility, which suggests that each of us has within us an exact subjective probability for each possible event in the small world (model) under consideration.

For a much fuller discussion of subjective and frequentist approaches see the collection of papers in [16] who note that the three defining attributes of the Bayesian approach are;

1. Reliance on a complete probabilistic model of the domain or "frame of discernment".
2. Willingness to accept subjective judgements as an expedient substitute for empirical data.
3. The use of Bayes theorem (conditionality) as the primary mechanism for updating beliefs in light of new information.

However, the Bayesian technique is not without its critics, including among others Walley [17], as well as Caselton and Luo [18] who discussed the difficulty arising when conventional Bayesian analysis is presented only with weak information sources. In such cases we have the "Bayesian dogma of precision", whereby the information concerning uncertain statistical parameters, no matter how vague, must be represented by conventional, exactly specified, probability distributions.

Some of the difficulties can be understood through the "Principle of Insufficient Reason", as illustrated by Wilson [19]. Suppose we are given a random device that randomly generates integer numbers between 1 and 6 (its "frame of discernment"), but with unknown chances. What is our belief in "1" being the next number? A Bayesian will use a symmetry argument, or the Principle of Insufficient Reason to say that the Bayesian belief in a "1" being the next number, say $P(1)$ should be 1/6. In general in a situation of ignorance a Bayesian is forced to use this principle to evenly allocate subjective (additive) probabilities over the frame of discernment.

To further understand the Bayesian approach, especially with the regard to representation of ignorance, consider the following example, similar to that in [19]. Let a be a proposition that;

"I live in Kings Road, Cardiff".

How could one construct $P(a)$, a Bayesian belief in a? Firstly we must choose a frame of discernment, denoted by Θ and a subset A of Θ representing the proposition a; then would need to use the Principle of Insufficient Reason to arrive at a Bayesian belief. The problem is there are a number of possible frames of discernment Θ that we could choose, depending effectively on how many Cardiff roads can be enumerated. If only two such streets are identifiable, then $\Theta = \{x_1, x_2\}$, $A = \{x_1\}$. The "Principle of Insufficient Reason" then gives $P(A)$, to be 0.5, through evenly allocating subjective probabilities over the frame of discernment. If it is estimated that there are about 1000 roads in Cardiff, then $\Theta = \{x_1, x_2, \ldots, x_{1000}\}$ with again $A = \{x_1\}$ and the other x_i's representing the other roads. In this case the "theory of insufficient reason" gives $P(A) = 0.001$.

Either of these frames may be reasonable, but the probability assigned to A is crucially dependent upon the frame chosen. Hence one's Bayesian belief is a function not only of the information given and one's background knowledge, but also of a sometimes arbitrary choice of frame of discernment. To put the point another way, we need to distinguish between uncertainty and ignorance. Similar arguments hold where we are discussing not probabilities per se but weights which measure subjective assessments of relative importance. This issue arises in decision support models such as the Analytic Hierarchy Process (AHP), which requires that certain weights on a given level of the decision tree sum to unity, see [20].

2. Dempster–Shafer theory

The origins of Dempster–Shafer theory go back to the work by A. P. Dempster [1,21] who developed a system of upper and lower probabilities. Following this his student G. Shafer [22], in his 1976 book "A Mathematical Theory of Evidence" added to Dempster's work, including a more thorough explanation of belief functions. Even though DST was not created specifically in relation to AI, the name Dempster–Shafer theory was coined by J. A. Barnett [23] in an article which marked the entry of the belief functions into the AI literature. In summary it is a numerical method for evidential reasoning (a term often used to denote the body of techniques specifically designed for manipulation of reasoning from evidence, based upon the DST of belief functions: see [24].

Following on from the example concerning Cardiff roads in the previous section, one of the primary features of the DST model is that we are relieved of the need to force our probability or belief measures to sum to unity. There is no requirement that belief not committed to a given proposition should be committed to its negation. As can be seen in the further analysis below, this allows us to construct and analyse our "frame of discernment" in a more flexible way. The total allocation of belief can vary to suit the extent of our knowledge.

The second basic idea of DST is that numerical measures of uncertainty may be assigned to overlapping sets and subsets of hypotheses, events or propositions as well as to individual hypothesis. To illustrate, consider the following expression of knowledge concerning murderer identification adapted from [25].

Mr. Jones has been murdered, and we know that the murderer was one of three notorious assassins, Peter, Paul and Mary, so we have a set of hypotheses i.e., frame of discernment, $\Theta = \{\text{Peter, Paul, Mary}\}$. The only evidence we have is that the person who saw the killer leaving is 80% sure that it was a man, i.e., $P(\text{man}) = 0.8$. The measures of uncertainty, taken collectively are known in DST terminology as a "basic probability assignment" (*bpa*). Hence we have a *bpa*, say m_1 of 0.8 given to the focal element {Peter, Paul} i.e., $m_1(\{\text{Peter, Paul}\}) = 0.8$, since we know nothing about the remaining probability it is allocated to the whole of the frame of the discernment i.e., $m_1(\{\text{Peter, Paul, Mary}\}) = 0.2$.

The key point to note is that assignments to "singleton" sets may operate at the same time as assignments to sets made up of a number of propositions. Such a situation is simply not permitted in a conventional Bayesian framework, although it is possible to have a Bayesian assignment of prior probabilities for groups of propositions (since conventional probability theory can cope with joint probabilities). As pointed out by Schubert [26], DST is in this sense a generalisation of the Bayesian theory. It avoids the problem of having to assign non-available prior probabilities and makes no assumptions about non-available probabilities.

2.1. Further analysis

The next question to consider is what can then be done with the *bpa*'s. The answer in the first instance is twofold. We may firstly collect *bpa*'s together in such a way as to express our overall belief. This is examined below. More importantly, we may then combine *bpa*'s from different sources. Continuing our example, if we gained further evidence that it was reported with confidence 0.6 that Peter was leaving on a jet plane when the murder occurred, in this case we have *bpa*, say m_2 $(\{\text{Paul, Mary}\}) = 0.6$. Since we know nothing about the remaining probability it is allocated to the whole of the frame of the discernment i.e., $m_2(\{\text{Peter, Paul, Mary}\}) = 0.4$.

Combination of conventional probabilities is achieved through the familiar multiplication rule. In the more generalised DST approach a more complex multiplication rule is required, which combines two pieces of evidence. We can see how this works in the context of the example (Table 1).

Put simply, the result of combining two assignments is that for any intersecting sets A and B, where A has mass M_1 from assignment m_1 (i.e., $m_1(\{A\}) = M_1$) and B has mass M_2 from assignment m_2, the belief accruing to their intersection is the product of M_1 and M_2. For example,

$$m_3(\{\text{Paul, Mary}\})$$

$$= m_1(\{\text{Peter, Paul, Mary}\}) \times m_2(\{\text{Paul, Mary}\}),$$

$$= 0.2 \times 0.6, \ = 0.12.$$

The new piece of evidence has a more spread-out allocation of probabilities to varying subsets of the frame

Table 1
Combination of two pieces of evidence

	$m_1(\{\text{Peter, Paul}\}) = 0.8$	$m_1(\{\text{Peter, Paul, Mary}\}) = 0.2$
$m_2(\{\text{Paul, Mary}\}) = 0.6$	$m_3(\{\text{Paul}\}) = 0.48$	$m_3(\{\text{Paul, Mary}\}) = 0.12$
$m_2(\{\text{Peter, Paul, Mary}\}) = 0.4$	$m_3(\{\text{Peter, Paul}\}) = 0.32$	$m_3(\{\text{Peter, Paul, Mary}\}) = 0.08$

of discernment. We can bring together this evidence to find some level of belief: the belief in any set is the sum of all the probabilities of all the subsets of that set Hence, for example:

$$Bel(\{\text{Peter, Paul}\})$$

$$= m_3(\{\text{Peter}\}) + m_3(\{\text{Paul}\}) + m_3(\{\text{Paul, Peter}\}),$$

$$= 0 + 0.48 + 0.32, \; = 0.8.$$

In the next section we give a more formal examination.

2.2. A more formal treatment

The above example has served to introduce the basic concepts. Let us discuss the terminology inherent within DST, which is slightly different from that used in probability theory. Let $\Theta = \{h_1, h_2, \ldots, h_n\}$ be a finite set of hypotheses (frame of discernment). A *basic probability assignment* (*bpa*) is a function $m{:}2^{\Theta} \to [0,1]$ such that:

$$m(\emptyset) = 0,$$

and

$$\sum_{x \in 2^{\Theta}} m(x) = 1.$$

We use the notation 2^{Θ} because of the need to allow for the number of elements in the power set, i.e., including all possible subsets, of the frame of discernment Θ. All of the assigned probabilities sum to unity and there is no belief in the empty set (which can be thought of as a hypothesis known to be false). Any subset x of the frame of discernment Θ for which $m(x)$ is non-zero is called a focal element and represents the exact belief in the proposition depicted by x. Possible propositions of interest we could consider are "*the true value of y lies in Y*", where $Y \subseteq \Theta$ and y is some quantity whose possible values comprise the frame of discernment, Θ. Thus propositions are subsets and vice-versa. The value $m(Y)$ represents our confidence that "*the true value of y lies in Y, and not in any proper subset of Y*", (A subset Z of Y is a proper subset of Y if $Z \subset Y$ but $Z \neq Y$).

Based on the *bpa* other measures of confidence are defined. A *belief* measure is a function $Bel{:}2^{\Theta} \to [0,1]$, and is drawn from the sum of probabilities that are subsets of the probabilities in question, defined by

$$Bel(A) = \sum_{B \subseteq A} m(B), \quad \text{for all } A \subseteq \Theta.$$

It represents our confidence that the value of y lies in A or any subset of A. A *plausibility* measure is a function $Pls{:}2^{\Theta} \to [0,1]$, defined by

$$Pls(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad \text{for all } A \subseteq \Theta.$$

Clearly $Pls(A)$ represents the extent to which we fail to disbelieve A. These measures are clearly related to one another, i.e.,

$$Bel(A) = 1 - Pls(\neg A) \quad \text{and} \quad Pls(A) = 1 - Bel(\neg A),$$

where $\neg A$ is referring to its compliment "not A", also $Bel(\neg A)$ is often called the *doubt* in A. Other notable relationships include;

$$Bel(A) + Bel(\neg A) \leqslant 1, \quad Pls(A) + Pls(\neg A) \geqslant 1.$$

The above two inequalities show a major departure from the more traditional simple probability function, used within the Bayesian approach. We also note that in the case of each of the focal elements being singletons then we return back to traditional Bayesian analysis incorporating normal probability theory, since in this case $Bel(A) = Pls(A)$.

A further quantity of importance, used in [18], is known as *commonality*, in this paper we use the more recent definition given in [27], denoted as $Q(A)$, defined by

$$Q(A) = \sum_{A \subseteq B} m(B),$$

the commonality of A collects all the probability in the *bpa* that could potentially be committed to A from all supersets which include A. It is shown in [22] that all the above three measures of confidence and the *bpa* are equivalent, in the sense that each of them can be expressed as a function of any one of the rest.

Collectively the above measures provide DST with an explicit measure of ignorance about an event and its complement $\neg A$. The measure is defined as the length of the interval $[Bel(A), Pls(A)]$ ($Bel(A) \leqslant Pls(A)$, as shown by Yager [28]). It is known as the belief interval. $Pls(A) - Bel(A)$ was suggested as a measure by Dillard [29]. This can also be interpreted as the imprecision on the "true probability" of A. The mass assigned to Θ can be interpreted as the global ignorance since this weight of evidence is not discernible among the hypothesis [30]. Further theoretical work on the calculation of measures of uncertainty has been undertaken by numerous authors. For more information we direct the reader to George and Pal [27] and the references therein.

As mentioned earlier, DST also provides a method to combine the measures of evidence from different sources, using the Dempster's rule of combination. This rule assumes that the sources are independent. If

m_1 and m_2 are the *bpa*'s associated with Bel_1 and Bel_2 respectively and Bel_1 and Bel_2 are independent, then the function $m_1 \oplus m_2$:$2^\Theta \rightarrow [0,1]$, defined by:

$$[m_1 \oplus m_2](y) = \begin{cases} 0 & y = \varnothing, \\ \dfrac{\sum_{A \cap B = y} m_1(A) m_2(B)}{1 - \sum_{A \cap B = \varnothing} m_1(A) m_2(B)} & y \neq \varnothing, \end{cases}$$

is a *bpa*. Dempster's rule does not have a full formal justification, in the same way that one can derive all conventional probability theorems from basic axioms [4]. It does, however, possess the necessary commutative and associative properties. Bayes rule of conditioning is a special case of this rule [18]. The independence required by Dempster's rule is simply probabilistic independence, applied solely to the hypotheses or sets of elements for which we have probabilities, rather than direct.

An important feature in the above formula is in the denominator i.e., $1 - \sum_{A \cap B = \varnothing} m_1(A) m_2(B)$, denoted by k, which can be interpreted as a measure of conflict between the sources and is directly taken into account in the combination as a normalisation factor. The measure represents the mass which would be assigned to the empty set if masses were not normalised. It is very important to take this value into account for evaluating the quality of combination: when it is high (in the case of strong conflict:$k \approx 1$), the combination may not make sense and may lead to questionable decisions. For more discussion see [31] and [30]. Recognising this feature has led some authors, including Zadeh [32], to also criticise the use of normalisation since it can lead to counter intuitive results if most of the mass after a combination, but before normalisation is assigned to the empty set [25].

When we wish to combine several belief functions this is simply done by combining the first two and then combining the result with the third and so forth. To help us further understand the combination rule let us consider the following larger example presented in [33]. Suppose that the weather in Cardiff at noon tomorrow is to be predicted from the weather today. We assume that it is in exactly one of the three states: snowing, raining or dry. Hence the frame of discernment is represented by $\Theta = \{S, R, D\}$. Let us assume that two pieces of evidence have been gathered:

- The temperature today is below freezing.
- The barometric pressure is falling; i.e., a storm is likely.

These pieces of evidence are represented by the two basic probability assignments m_{freeze} and m_{storm} in Table 2:

Table 2
Allocation of two *bpa*'s m_{freeze} and m_{storm}

	∅	{S}	{R}	{D}	{S, R}	{S, D}	{R, D}	{S, R, D}
m_{freeze}	0.0	0.2	0.1	0.1	0.2	0.1	0.1	0.2
m_{storm}	0.0	0.1	0.2	0.1	0.3	0.1	0.1	0.1

By definition, neither m_{freeze} and m_{storm} can place any probability mass in the proposition ∅. m_{freeze} distributes its mass among the remaining elements of the power set of Θ, with extra weight on {S}, {S, R} and {S, R, D}. The function m_{storm} also distributes its weight over all the non-empty subsets of Θ, but it gives greatest weight to {S, R} and some emphasis to {R}. If we let m_{both} represent the basic probability assignment that represents the combined evidence of m_{freeze} and m_{storm}. Assuming that m_{freeze} and m_{storm} represent items of evidence which are independent of one another, m_{both} is given by Dempster's rule of combination; $m_{\text{both}} = m_{\text{freeze}} \oplus m_{\text{storm}}$, shown below (Table 3):

The values obtained by computation assigns more weight to each of the elements {S}, {R} and {D} than either of the starting basic probability assignments does. However, each of {S} and {R} has gained substantially more weight than {D} has. The other elements of the power set of Θ have net losses in weight.

The degree of belief in proposition {S, R} based on the combined evidence (0.282 + 0.282 + 0.180 = 0.744) is substantially higher than that based on m_{freeze}, 0.5 and m_{storm}, 0.6. The degree of belief in {S, R, D} is, of course unchanged and equal to 1.

When the *Bel* and *Pls* functions are equal this is an indicator of the "Bayesian" special case. As mentioned previously the difference between them is sometimes quoted as a measure of uncertainty. However, since in the Bayesian case we still have uncertainty in all cases except where one proposition has probability one, this is not quite appropriate. Perhaps we can better understand the difference between *Bel* and *Pls* as a measure of "singularity" i.e., the extent to which we approach the Bayesian special case.

We close this section with the cautionary words of Pearl [34], who states that:

> "Some people qualify any model that uses belief functions as Dempster–Shafer. This might be accept-

Table 3
Allocation of combined evidence from m_{freeze} and m_{storm}

	∅	{S}	{R}	{D}	{S, R}	{S, D}	{R, D}	{S, R, D}
m_{both}	0.0	0.282	0.282	0.128	0.180	0.051	0.051	0.026

able provided they did not blindly accept the applicability of both Dempster's rule of conditioning and combination. Such uncritical — and in fact often inappropriate — use of these two rules explain most of the errors encountered in the so called Dempster–Shafer literature."

3. A DST approach to the Analytical Hierarchy Process

The method we develop in this section incorporates DST with the philosophy behind the AHP, the well-known management tool developed by Thomas L. Saaty [20]. The AHP is designed to solve complex problems involving multiple criteria, having originally been applied to the Sudan transport study, Saaty [35]. It may be viewed as a systematic procedure for representing the elements of any problem hierarchically, breaking the problem into smaller and smaller constituent parts and subsequently guiding the decision maker through a series of pairwise comparison judgements. Research on applications of AHP continues to appear. For example, Andijani [36] uses it to examine the results of simulation experiments relating to the Kanban allocation problem.

For this paper we restrict ourselves to dominance hierarchies, which descend like an inverted tree as shown in the example (adapted from Anderson et al. [37]) in Fig. 1:

In this example, the decision involves buying a new car, from a choice of 3 known types of car (Decision Alternatives), say A, B and C. In the DST terminology used above, {A, B, C} is then the frame of discernment Θ. The criteria to help us judge each of these cars are, Price, Fuel, Comfort and Style, the overall objective (focus) is to decide which is the best car to buy. Since the AHP method is not the objective of this paper and for reasons of brevity we direct the reader to [37] and [20] for a full and detailed explanation of the above example and the AHP method. The decision maker is required to express an opinion on the relative importance of each criterion, comparing the criteria in pairs. For numeric criteria such as price, as long as the prices are known with certainty, no such comparisons are needed. For the "leaf nodes" of the tree, i.e., the Decision Alternatives, pairwise comparisons are needed vis a vis each of the 4 criteria. For each level, the pairwise comparisons are normalised to produce a vector of weights (summing to unity). The weights are then "synthesised" to produce an evaluation of the decision alternatives.

One criticism of the AHP method is the sheer number of pairwise comparisons to be performed before any rankings can be evaluated. In the above small example there would be 3 comparisons per criterion between the decision alternatives (D.A.'s) level, making 12 comparisons in all at that level. Another 6 comparisons at the criterion level, giving a total of 18 comparison judgements. The number of comparisons quickly rises as the number of alternatives and criteria rise, for example if there were a choice of 8 cars considered then a total of 118 prior comparison judgements would be required.

One of the reasons for the number of comparisons in the traditional AHP method is the need to compare each decision alternative with each other decision alternative. A further drawback is their consistency of these comparisons. For example, if car A is preferred to B, B preferred to C also C preferred to A, this would be understandably inconsistent. This understanding of consistency is measured and discussed within the AHP method. Additionally there is no allowance for ignorance with respect to types of car and available criteria. To summarise the concerns of the AHP method:

- The number of comparisons.
- The consistency of comparisons.
- The lack of a representation of ignorance within the model.

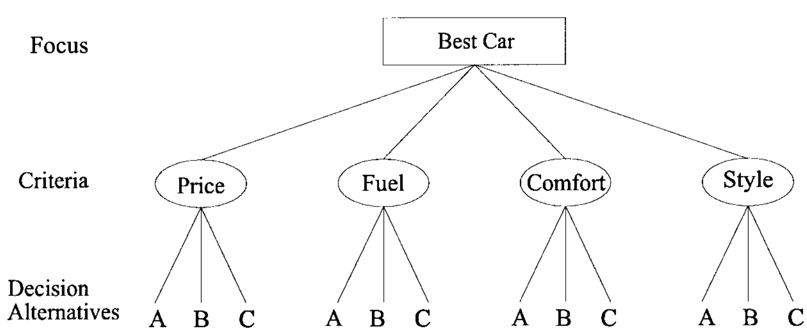

Fig. 1. Hierarchy of car choice AHP model.

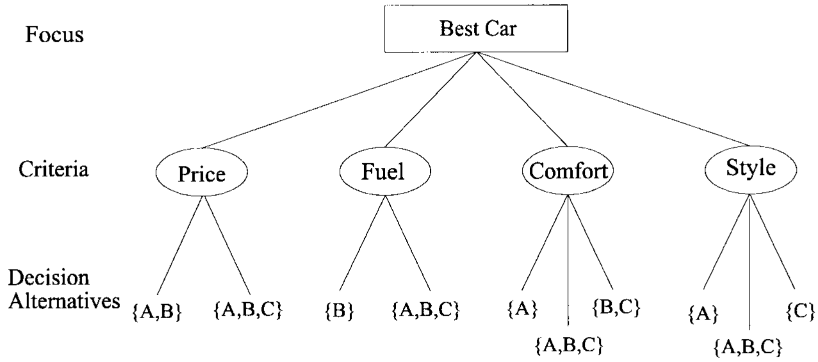

Fig. 2. Hierarchy of modified car choice model.

We consider next whether DST contribute to a method which might alleviate these difficulties. To reduce the number of comparisons would mean not considering certain D.A.'s, instead considering groups of D.A.'s. The method we suggest is based on a measure of favourability of knowledge we have about a group of D.A.'s compared with Θ the frame of discernment within the context of each specific criterion. This knowledge could come in the way of quantitative or qualitative evidence available for a certain collection of D.A.'s (as in the knowledge of the identity of the murderer example). An example of the resulting decision tree might be as follows (Fig. 2):

Here, for each criterion there are certain groups of D.A.'s, including Θ, about which the decision maker can express some degree of favourable knowledge. Following from our discussion of ignorance in knowledge of evidence, Θ (i.e., {A, B, C}) as a D.A. group allows for the opportunity of the allocation of ignorance. Since we are not performing pairwise comparisons of D.A.'s but relating groups of D.A.'s to Θ, there are no consistency problems within a criterion, as long as no two proper subsets of Θ considered in a criteria have a D.A. in common. For example, in the comfort criterion, if the sets {A, B} and {B, C} were compared in some favourable way to Θ, there would be inconsistency in our judgements, since we have included the car B in two groups, with each group having been given a different level of favourability.

The method we are developing aims to operate in a way essentially similar to those of the standard AHP method, however since already we have allowed for sets of D.A.'s to be used, our results in terms of the decisions made will also include knowledge assessments for sets of D.A.'s. Looking first at the decision alternative level and allowing for both quantitative and qualitative knowledge to aid the opinion of favourability of groups of D.A.'s to Θ, the following 5 unit scale (adapted from that in the AHP method) was adopted as a basis for discriminating levels of knowledge (Table 4):

The AHP is often implemented with a nine unit scale, to map the labels which indicate the decision maker's view to a numeric value. We have adopted a five unit scale for simplicity. It is important to note that we do not use the equally preferred rating of 1 (as in the AHP method), this being a consequence of evaluating groups of D.A.'s vis a vis the frame of discernment. Recent research on the AHP has included work on the particular scaling methods to be employed, Lootsma [38]. For our purposes, we employ linear scales for the sake of simplicity. As in the AHP method knowledge matrices are constructed for each of the criteria, a sample matrix for the comfort criterion being shown in Table 5:

In the above knowledge matrix, the values in the final column are the measures of favourability of certain groups of D.A. in each row with respect to Θ. For example, using the knowledge on the set of D.A.'s {B, C} is viewed as extremely favourable compared to

Table 4
Knowledge/Favourable scale

Opinion/Knowledgeable	Numerical rating	Opinion/Knowledgeable	Numerical rating
Extremely favourable	6	Moderately to strongly	3
Strongly to extremely	5	Moderately favourable	2
Strongly favourable	4		

Table 5
Initial knowledge matrix for comfort

Comfort	{A}	{B, C}	{A, B, C}
{A}	1	0	4
{B, C}	0	1	6
{A, B, C}	1/4	1/6	1

Θ. The zero's which appear in the knowledge matrix indicate no attempt to assert knowledge between groups of D.A.'s, (e.g., {A} to {B, C}); this assertion can be made indirectly through knowledge of the favourability of {A} to Θ and {B, C} to Θ relatively. In Table 5 the indirect knowledge is that {A} is not considered more favourable to {B, C} in relation to Θ.

At the criteria level we still use the pairwise comparison matrix as in the AHP method. For the purpose of illustration we use the following sample priority values for a normalised weight vector for the criteria. These values came from a comparison matrix with consistency measure 0.069: see Saaty [20] (Table 6).

To use these priority values we depart from the traditional AHP method, and instead we bring them down from the criteria level, incorporating them into each of the decision knowledge matrices. We do this by multiplying the elements in the last column (except the last entry in that column) by the respective importance value for that criterion. If p is the importance value and x_{ij} is the favourability opinion for a particular group of D.A.'s, then the resultant value is $p \times x_{ij}$ (the resultant change in the bottom row of the matrix is similarly $(1/(p \times x_{ij}))$. It is important that the importance values do not affect the elements in the matrix which are either zero or unity. For the comfort criterion, which had an importance value $p = 0.2159$, we obtain (Table 7):

The approach taken below is to regard each criterion as generating an independent piece of evidence, with Dempster's rule of combination being employed to integrate them. We can apply the same method to the other criteria, the resultant sample matrices after the importance value has been brought down to the decision alternative level being (Table 8):

Using the knowledge matrices for each of the criteria we can produce normalised knowledge vectors, following the traditional AHP method. For simplicity we have adopted the basic normalisation method used in [37] as opposed to the method used in [20], whereby

Table 6
Criteria priority values

Criterion	Price	Fuel	Comfort	Style
Priority	0.3982	0.0851	0.2159	0.2988

Table 7
Knowledge matrix for comfort after influence of its priority rating

Comfort	{A}	{B, C}	{A, B, C}
{A}	1	0	0.8714
{B, C}	0	1	1.3072
{A, B, C}	1.1475	0.7650	1

the weights are derived as the eigenvectors of the Knowledge Matrix. The simpler method in fact produces approximate values for each of the criteria (Table 9).

The totals in these priority value columns sum to one, so we can think of these priority allocations as basic probability assignments (*bpa*'s). In the case of the price criterion, defining m_1 as the *bpa*, then:

$$m_1(\{A, B\}) = 0.7050, \quad m_1(\{A, B, C\}) = 0.2950.$$

In a similar way allowing the *bpa*'s for Fuel, Comfort and Style, say m_2, m_3 and m_4 respectively, then:

$$m_2(\{B\}) = 0.2034, \quad m_2(\{A, B, C\}) = 0.7966.$$

$$m_3(\{A\}) = 0.2466, \quad m_3(\{B, C\}) = 0.3259,$$
$$m_3(\{A, B, C\}) = 0.4275.$$

$$m_4(\{A\}) = 0.3608, \quad m_4(\{C\}) = 0.1891,$$
$$m_4(\{A, B, C\}) = 0.4501.$$

At the decision alternative level we have constructed basic probability assignments for each of the criteria in the model. The opinions we have made are on the favourability of knowledge we have towards groups of

Table 8
Knowledge matrices for price, fuel and style

Price	{A, B}	{A, B, C }	
{A, B}	1	2.3892	
{A, B, C}	0.4186	1	
Fuel	**{B}**	**{A, B, C}**	
{B}	1	0.2554	
{A, B, C}	3.9160	1	
Style	**{A}**	**{C}**	**{A, B, C}**
{A}	1	0	1.4940
{C}	0	1	0.5976
{A, B, C}	0.6693	1.6734	1

Table 9
Priority values for groups of D.A.'s and Θ

	Priority
Price	
{A,B}	0.7050
{A,B,C}	0.2950
Fuel	
{B}	0.2034
{A,B,C}	0.7966
Comfort	
{A}	0.2466
{B,C}	0.3259
{A,B,C}	0.4275
Style	
{A}	0.3608
{C}	0.1891
{A,B,C}	0.4501

D.A.'s compared with the whole frame of discernment. Within these opinions we have had to make the choice of which D.A.'s to include in the same set, correspondingly we have made decisions to purposely differentiate/dissipate our knowledge on different D.A.'s. We can now consider these criteria as independent pieces of evidence, each offering information on opinions towards the favourability of the particular D.A.'s. We are now in a position to combine these pieces of evidence using Dempster's rule of combination. The resulting *bpa*, defined as m_{car} is evaluated and shown in Table 10 below:

Using m_{car}, we can construct the interval of belief for subsets of the frame of discernment Θ, we do not necessarily have to restrict our choice to a single car, but we can see the belief in the best car coming from a subset of the cars within the available alternatives. Below is a table showing the belief (*Bel*) and plausibility (*Pls*) values (Table 11):

As mentioned previously the origins of the DST are in Dempster's work on imprecise probabilities. This imprecision is reflected in the differences between belief and plausibility. Standard formulations of the approach deal with an underlying set of mutually exclusive hypotheses and the evidence available concerning these hypotheses. Here, we regard the available car models as mutually exclusive choices.

Table 10
The *bpa* m_{car} after combining all evidence

	{A}	{B}	{C}	{A, B}	{B, C}	{A, B, C}
m_{car}	0.4309	0.2312	0.0511	0.1650	0.0527	0.0691

Table 11
Belief and plausibility values for subsets of cars

Cars	*Bel*	*Pls*
{A}	0.4309	0.6650
{B}	0.2312	0.5180
{C}	0.0511	0.1729
{A, B}	0.8271	0.9489
{A, C}	0.4820	0.7688
{B, C}	0.3350	0.5691

Lucas and van der Gaag [33] offer some suggestions as to how the belief and plausibility measures may be interpreted. Two examples they consider are,

- If $[Bel(Y),\ Pls(Y)] = [0.3,1]$, then there is some evidence in favour of the hypothesis Y.
- If $[Bel(Y),\ Pls(Y)] = [0.15,0.75]$, then we have evidence in favour as well as against Y.

Looking at Table 11, if we consider the focal element {A}, then there is a small amount of evidence in favour as well as against the hypothesis A being the best car. For {C} the table shows strong evidence against being the best choice. Interestingly the set {A, B} shows strong evidence in favour of including within its elements the best choice of car. Note that these last two indications are connected by the property $Pls(Y) = 1 - Bel(\neg Y)$, (Y is a subset of Θ).

As mentioned, the most striking point is for the belief interval for the set of cars {A, B}, which to the decision maker could indicate more research should be undertaken in finding out about differences between these particular cars. This may be confirmed by looking at the belief intervals for the cars A and B individually, which are seen to be somewhat similar, hence giving no indications concerning specific car choice. The decision maker may only want to reduce the number of types of car to be considered.

In the AHP method, a measure of consistency is used to indicate the confidence in the judgements made in each comparison matrix. For our modified version of the method, since the results in Table 11 are a consequence of Dempster's rule of combination, we can calculate a measure of conflict, namely $k = 1 - \sum_{A\cap B=\varnothing} m_1(A)m_2(B)$. In our example the level of conflict over all our pieces of evidence is equal to $k = 0.6893$, hence indicating some level of conflict existing between the judgements made on favourability of certain cars subject to different criteria (the larger the value of k the more conflict in our evidence).

From our example this method which we call the DS/AHP method has been used to address the three concerns inherent within the AHP method. Also we can use the method as a way of reducing the number of alternatives we should look further at. Of special

Table 12
Knowledge matrices for MBA choice model

RS	{A, C, F}	{B, D, G, H}	Θ
{A, C, F}	1	0	3
{B, D, G, H}	0	1	2
Θ	1/3	1/2	1

GR	{D}	{A, C}	{B, F}	Θ
{D}	1	0	0	2
{G, C}	0	1	0	6
{B, F}	0	0	1	4
Θ	1/2	1/6	1/4	1

CL	{C}	{F}	{A, G}	{D, H}	Θ
{C}	1	0	0	0	4
{F}	0	1	0	0	2
{A, G}	0	0	1	0	3
{D, H}	0	0	0	1	6
Θ	1/4	1/2	1/3	1/6	1

LN	{C, D}	{B, F, H}	Θ
{C, D}	1	0	6
{B, F, H}	0	1	3
Θ	1/6	1/3	1

TG	{A, H}	{D, G}	{C, F}	Θ
{A, H}	1	0	0	2
{D, G}	0	1	0	6
{C, F}	0	0	1	5
Θ	1/2	1/6	1/5	1

RP	{A}	{H}	{D}	{C, F, G}	Θ
{A}	1	0	0	0	2
{H}	0	1	0	0	5
{D}	0	0	1	0	3
{C, F, G}	0	0	0	1	4
Θ	1/2	1/5	1/3	1/4	1

note is that in this case we have only performed 6 opinion judgements and 6 pairwise criteria comparisons.

4. A larger example: MBA course selection

The potential of this DS/AHP method may be illustrated by considering a larger example. Here we consider the problem of choosing which education establishment to attend to study for an MBA (Master in Business Administration). Within the UK there are over 100 education establishments which offer an MBA course. The focus of our choice is based on overall course/establishment satisfaction, while the goal is to reduce the number of education establishment that the decision maker may want to visit to generate more information.

The criteria we use to help us make our choice of education establishment are Research (RS), Location (LN), Growth (GR), Teaching Quality (TG), Colleagues (CL) and Reputation (RP). If we were to use the traditional AHP method with a choice of 100 establishments, then there would be a total of 29,715 pairwise comparisons to be made. If we reduce the number of education establishments to only 8 i.e., $\Theta = \{A, B, C, D, E, F, G, H\}$ we would need 183 comparisons. Using the DS/AHP method, examples of the initial knowledge matrices for decision alternatives within each of the criteria are shown below (Table 12):

The associated criteria priority values, which are a consequence of the pairwise criteria comparisons, are (Table 13):

We then apply the confidence rating for each criterion to the values in the knowledge matrices, as in the smaller car choice example. Within this DS/AHP model we have made 18 opinion judgements and 15 pairwise criteria comparisons. We are able to illustrate this example below (Fig. 3):

Using the knowledge matrices and criteria priority values in Tables 12 and 13 respectively, we can use Dempster's rule of combination to construct the overall *bpa*, say m_{course} as shown in Table 14 below:

We can then proceed to calculate the intervals of belief of all combinations of elements in the frame of discernment. If the decision maker is interested only in reducing the number of alternatives to say three education establishments (perhaps the number which they have decided to visit) then there are 56 possible outcomes. Table 14 shows the intervals of belief for 14 of these. It is important to note that subsets examined in this way do not need to be part of the original *bpa*'s. The belief measures in Table 15 are derived from the original expressions of knowledge.

Again following the example in Lucas and van der Gaag [33] we can make some interpretations of these belief intervals. All the intervals in the upper half of the table indicate evidence in favour and against them being the best choice of three education establishments to further look at. Within these groups of D.A.'s {C, D, F} and {C, D, H} offer the highest belief and

Table 13
Criteria priority values

Criteria	Research	Location	Growth	Teaching	Colleagues	Reputation
Priority	0.1051	0.0731	0.3965	0.4000	0.1279	0.2544

Table 14
The $bpam_{\text{course}}$ after combining all evidence

Subset	m_{course}	Subset	m_{course}	Subset	m_{course}
{A}	0.0482	{A, G}	0.0117	{D, G}	0.0821
{B}	0.0011	{A, H}	0.0378	{D, H}	0.0228
{C}	0.1177	{B, F}	0.0076	{A, C, F}	0.0177
{D}	0.1405	{B, H}	0.0018	{B, F, H}	0.0131
{F}	0.0634	{C, D}	0.0238	{C, F, G}	0.0270
{G}	0.0654	{C, F}	0.1075	{B, D, G, H}	0.0123
{H}	0.0974	{C, G}	0.0125	Θ	0.0886

Plausibility values respectively, depending on how cautious the decision maker is about their judgements these may be the two groups of D.A.'s to choose from, see further work. The lower half of the table illustrates groups of D.A.'s which give evidence against being the groups to choose for further investigation. We note the education establishment E appears in many of the groups of D.A.'s in the right hand column in Table 15, this is a direct consequence of the fact that E never appears within a group of D.A.'s for which there is favourable knowledge given within any criteria.

Lastly the decision maker can gain a view to how their opinion towards different criteria are in conflict with each other, by calculating the amount of conflict in the evidence. In this case $k = 0.7120$. The extent of this conflict may make the decision maker re-evaluate (relax) the various preferences.

Table 15
Belief and plausibility values for selection of subsets of Θ

Courses	*Bel*	*Pls*
{F, D, H}	0.3241	0.7435
{C, G, H}	0.2929	0.7392
{C, D, H}	0.4021	0.8026
{C, D, G}	0.4420	0.7297
{A, D, F}	0.2521	0.7042
{C, D, F}	0.4529	0.7367
{A, C, D}	0.3302	0.7503
{B, E, D}	0.1415	0.3937
{B, E, F}	0.0721	0.3402
{A, E, H}	0.1834	0.3514
{A, B, H}	0.1862	0.3601
{B, H, H}	0.1002	0.2825
{B, E, G}	0.0664	0.3233
{A, B, E}	0.0492	0.2399

The DS/AHP examples discussed here are for the purposes of illustration. As also mentioned the DS/AHP method is useful in that it could be an initial study of all the alternatives available, the results of which could lower the number of alternatives that fit the limited number of opinions given so far, with only a few opinions stated. Further research could be per-

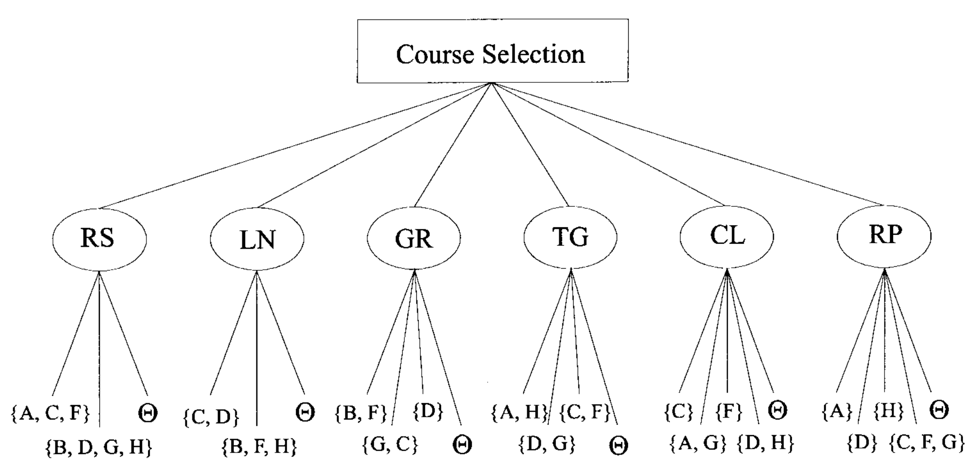

Fig. 3. Course selection model.

formed on this smaller set of alternatives, possibly using the traditional AHP method.

5. Analysis of the required number of judgements

In this section we investigate the overall reduction in comparison between the traditional AHP method and the DS/AHP method we have developed in this paper. If we consider a two layer decision choice model with n criteria and m decision alternatives then for the normal AHP method there will be $n(n-1)/2$ criterion comparisons. For the worst case, with only qualitative information for each of the criteria, there would be $nm/2(m\text{-}1)$ decision alternative comparisons made giving a total of:

$$\frac{n}{2}(m^2 - m + n - 1),$$

pairwise comparisons.

When we consider the DS/AHP method with m decision alternatives available in each criterion, in the worst case we can consider there to be m judgements made in each criterion. The judgements can be thought of as deciding for each D.A. whether to include it in a particular favouribility group of D.A.'s or to not include it (as in the case of E in the large MBA example, indicating an ignorance on the favourability of this particular education establishment). Hence in the DS/AHP method we would still have $n(n\text{-}1)/2$ criterion and $\sum_{i=1}^{n} m = nm$ knowledge judgements. Thus the total number of judgements is:

$$\frac{n}{2}(2m + n - 1).$$

It follows that the overall maximum proportion of comparisons made in the DS/AHP method relative to the AHP method would be:

$$\frac{\frac{n}{2}(2m + n - 1)}{\frac{n}{2}(m^2 - m + n - 1)} = \frac{(2m + n - 1)}{(m^2 - m + n - 1)}.$$

Since at the criteria level we are performing the same number of comparisons then we can consider just what happens at the decision alternative level. The proportion of judgements made at the decision alternative level in the DS/AHP method compared to those in the AHP method would be:

$$\frac{nm}{\frac{nm}{2}(m-1)} = \frac{2}{m-1}.$$

So, for example, when $m = 8$ (as in the large MBA example) only 2/7, i.e., a maximum of 28.5% judgements would be made in the DS/AHP compared to the traditional AHP method. Fig. 4 shows how quickly this percentage decreases with m.

Fig. 4 shows at once that when the number of decision alternative is 11 or 12, then in the worst case the number of knowledge judgements at the decision level for the DS/AHP method is still 20% or less than for the standard AHP.

6. Conclusions

In this paper we have introduced and reviewed The DST as a generalisation of classical probability theory. Following this we have incorporated the theory in a modified version of the AHP, which we call the DS/AHP method.

The DST approach differs from more conventional methods in that there is no requirement that belief not committed to a given proposition should be committed to its negation. This offers a certain additional realism. For example, there is current debate as to the health risks which may be associated with mobile phones. Research does not at present offer conclusive support for the view that mobiles carry a health risk, but

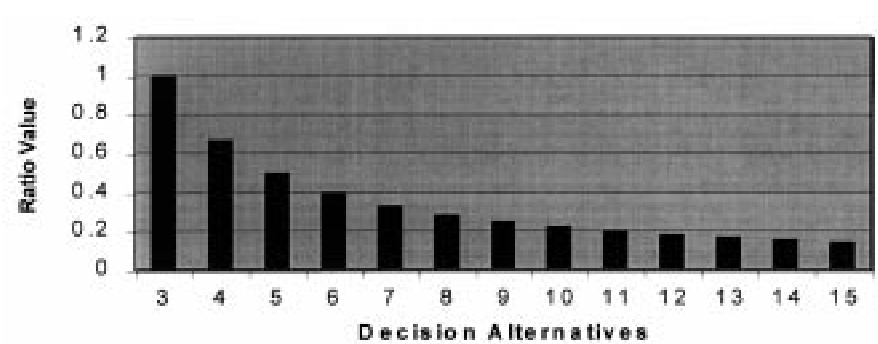

Fig. 4. Proportion of judgements in DS/AHP to AHP method.

neither can we say that the risk is negligible. The second basic idea of DST is that numerical measures of uncertainty may be assigned to overlapping sets and subsets of hypotheses, events or propositions as well as to individual hypothesis.

From a historical and theoretical background one may make a number of useful observations. In the first instance, although we have demonstrated that there are potential gains available through use of the DST approach, in another sense it is incomplete. For example, Cox [39] sets out 7 axioms which collectively define a proper or ideal subjective measure of uncertainty. These include completeness and complementarity. Bayesian probability measures satisfy the complete set of axioms, but DST measures do not. Our response to this is simply that Bayesian measures may in turn ask too much of a decision maker.

From a historical perspective, Krause and Clark [40] place the DST technique within a long tradition of "epistemic" (i.e., knowledge based) approaches to probability. In particular they show that Dempster's rule of combination, which is essential to the DST approach in that it brings together independent sources, can in turn be regarded as a generalisation of a much older rule. This rule can be traced back as far as Lambert [41].

The DS/AHP method allows opinions on sets of decision alternatives and addresses some of the concerns inherent within the "standard" AHP:

- The number of comparisons and opinions are at the decision makers discretion.
- There is no need for consistency checks at the decision alternative level.
- The allowance for ignorance/uncertainty in our judgements.

We remind the reader that the direction of this method is not necessarily towards obtaining the highest ranked decision alternative, but towards reducing the number of serious contenders. Although a major advantage of incorporating DST into the AHP lies in the reduction of the complexity of the problem, this is by no means the sole justification for employing it. It must also be recognised that our DS/AHP model provides an illustration of the potential of Dempster Shafer theory and does not incorporate more recent extensions of the original AHP technique.

7. Further work

Further work may be suggested on the scales to be employed for discriminating levels of knowledge, and also on employing the method for more complex hierarchical problems that exist. Consideration is needed on the best method for obtaining assessments of knowledge from the decision maker.

One important aspect of our method is the focus on belief for subsets of the frame of discernment. In the MBA example we restricted the decision maker to looking at the optimum set of three education establishments. We saw in Table 15 that no single subset stood out as being optimal (i.e., its belief and plausibility values being the largest of any subset of size 3). We need therefore to consider the possibility of enlarging the subset until an "optimum" subset can be seen. For the data in the MBA example we show below the largest belief and plausibility values for different sized subsets. It is again important to note that subsets examined in this way do not need to be part of the original *bpa*'s.

Table 16 shows that we can say with more confidence that the set {C, D, F, G, H} should be given more consideration, given that this shows the highest overall belief interval.

When the individual has picked the number of D.A.'s they are willing to spend more time on investigating, this may mean that no one subset stands out as "optimum". Hence we need to pick out a particular subset in the face of inconclusive evidence, this implies the use of some form of expected value within an interval needs to be constructed.

Although an essential part of the DST approach is that it deals with belief intervals, Strat [42] has suggested the possibility of producing point probabilities, at the end of the analysis. This suggestion is quite interesting, and warrants further work. It would mean exploiting the benefits of the DST approach, but with more precise end results. The idea is to produce point probabilities by interpolation.

Table 16
Largest belief and plausibility values for subsets of Θ

Size of subset	Largest belief value	Largest plausibility value
1	{D}, 0.1405	{C}, 0.3949
2	{C, F}, 0.2886	{C, D}, 0.6526
3	{C, D, F}, 0.4529	{C, D, H}, 0.8026
4	{C, D, F, G}, 0.6399	{C, D, G, H}, 0.8797
5	{C, D, F, G, H}, 0.7601	{C, D, F, G, H}, 0.9508

References

[1] Dempster AP. Upper and lower probabilities induced by a muilti-valued mapping. Ann Math Stat 1967;38:325–39.

[2] Hajek P. Systems of conditional beliefs in Dempster–Shafer theory and expert systems. Int J General Systems 1994;22:113–24.

[3] Ip HHS, Ng JMC. Human face recognition using Dempster–Shafer theory. In: ICIP. 1st International Conference on Image Processing, vol. 2, 1994. p. 292–5.

[4] Denoeux T. A *k*-nearest neighbour classification rule based on Dempster–Shafer theory. IEEE Transactions on Systems, Man and Cybernetics 1995;25(5):804–13.

[5] Buede DM, Girardi P. A target identification comparison of Bayesian and Dempster–Shafer multisensor fusion. IEEE Transaction on Systems, Man and Cybernetics-Part A: Systems and Humans 1997;27(5):569–77.

[6] Yen J. GERTIS: A Dempster–Shafer approach to diagnosing hierarchical hypotheses. Commun ACM 1989;32(5):573–85.

[7] Bauer M. A Dempster–Shafer approach to modeling agent preferences for plan recognition. User Modeling and User-Adapted Interaction 1996;5:317–48.

[8] Cortes-Rello E, Golshani F. Uncertain reasoning using the Dempster–Shafer method: an application in forecasting and marketing management. Expert Systems 1990;7(1):9–17.

[9] Kotler P. Marketing management: analysis, planning and control. Englewood Cliffs, NJ: Prentice Hall, 1980.

[10] Bayes T. An essay toward solving a problem in the doctrine of chances. Phil Trans Roy Soc (London) 1763;53:370–418.

[11] Wald A. Statistical decision functions. New York: Wiley, 1950.

[12] Savage LJ. The foundations of statistics. New York: Wiley, 1954 (2nd rev. ed., 1972 Dover).

[13] Savage LJ. The foundation of statistics reconsidered. In: Proceedings of the Fourth Berkeley Symposium on Mathematics and Probability 1. Berkeley: University of California Press, 1961.

[14] Good IJ. Good thinking: the foundations of probability and its applications. Minneapolis: University of Minnesota Press, 1983.

[15] French S. Uncertainty and imprecision: modelling and analysis. J Opl Res Soc 1995;46:70–9.

[16] Shafer G, Pearl J. Readings in uncertain reasoning. San Mateo, CA: Morgan Kaufman, 1990.

[17] Walley P. Belief-function representations of statistical evidence. Ann Stat 1987;10:741–61.

[18] Caselton WF, Luo W. Decision making with imprecise probabilities: Dempster–Shafer theory and applications. Water Resources Research 1992;28(12):3071–83.

[19] Wilson PN. Some theoretical aspects of the Dempster–Shafer theory. PhD Thesis, Oxford Polytechnic, 1992.

[20] Saaty TL. The Analytic Hierarchy Process: planning, priority setting, resource allocation. New York: McGraw-Hill, 1980.

[21] Dempster AP. A generalization of Bayesian inference (with discussion). J Roy Stat Soc, Series B 1968;30(2):205–47.

[22] Shafer G. A mathematical theory of evidence. Princeton: Princeton University Press, 1976.

[23] Barnett JA. Computational methods for a mathematical theory of evidence. In: Proceedings 7th International Joint Conference on Artificial Intelligence (IJCAI), Vancouver, vol. II, 1981. p. 868–75.

[24] Lowrance JD, Garvey TD, Strat TM. A framework for evidential-reasoning systems. In: Proceedings of the 5th National Conference on Artificial Intelligence (AAAI-86), Philadelphia, 1986. p. 896–901.

[25] Parsons S. Some qualitative approaches to applying the Dempster–Shafer theory. Information and Decision Technologies 1994;19:321–37.

[26] Schubert J. Cluster-based specification techniques in Dempster–Shafer theory for an evidential intelligence analysis of multiple target tracks. Department of Numerical Analysis and Computer Science Royal Institute of Technology, S-100 44 Stockholm, Sweden, 1994.

[27] George T, Pal NR. Quantification of conflict in Dempster–Shafer framework: a new approach. Int J General Systems 1996;24(4):407–23.

[28] Yager RR. On the Dempster–Shafer framework and new combination rules. Information Science 1987;41(2):93–137.

[29] Dillard RA. The Dempster–Shafer theory applied to tactical data fusion in an inference system. In: Proceedings of the Fifth MIT/ONR Workshop, 1982. p. 170–4.

[30] Bloch B. Some aspects of Dempster–Shafer evidence theory for classification of multi-modality images taking partial volume effect into account. Pattern Recognition Letters 1996;17:905–19.

[31] Dubois D, Prade H. Representation and combination of uncertainty with belief functions and possibility measures. Computational Intelligence 1988;4:244–64.

[32] Zadeh LA. Review of Shafer's "A mathematical theory of evidence". The AI Magazine 1984;81–3.

[33] Lucas P, Gaag van der L. Principles of expert systems. Reading, MA: Addison-Wesley, 1991.

[34] Pearl J. Reasoning with belief functions: An analysis of compatibility. Int J Approx Reasoning 1990;4:363–90.

[35] Saaty TL. The Sudan transport study. Interfaces 1977;8(1):37–57.

[36] Andijani AA. A Multi-criterion approach for Kanban allocations. Omega 1998;26(4):483–94.

[37] Anderson DR, Sweeney DJ, Williams TA. An introduction to management science. New York: West, 1998.

[38] Lootsma FA. Multicriteria decision analysis in a decision tree. Eur J Opl Res 1997;101:442–51.

[39] Cox R. Probability, frequency and reasonable expectation. American Journal of Physics 1946;14:1–13.

[40] Krause P, Clark D. Representing uncertain knowledge: an artificial intelligence approach. London: Intellect Books, 1993.

[41] Lambert JH. Philosophiesche schriften. Reprinted in 1965 by Olms of Hildesheim, 1764.

[42] Strat TM. Decision analysis using belief functions. In: Yager RR, Fedrizzi M, editors. Advances in the Dempster–Shafer theory of evidence. New York: Wiley, 1994.