

UNDERSTANDING EVIDENTIAL REASONING

Technical Note No. 501

December 1990

By: Enrique H. Ruspini, Sr. Computer Scientist
John D. Lowrance, Program Director
Thomas M. Strat, Sr. Computer Scientist
Artificial Intelligence Center
Computing and Engineering Sciences Division

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

This work was supported in part by the Defense Advanced Research Projects Agency under Contract No. N00039-88-C-0248, and in part by the United States Army Research Office under Contract No. DAAL03-89-K-0156.

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE DEC 1990		2. REPORT TYPE		3. DATES COVERED 00-12-1990 to 00-12-1990	
4. TITLE AND SUBTITLE Understanding Evidential Reasoning				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) SRI International, 333 Ravenswood Avenue, Menlo Park, CA, 94025				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 24	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

UNDERSTANDING EVIDENTIAL REASONING

Enrique H. Ruspini

John D. Lowrance

Thomas M. Strat

Artificial Intelligence Center

SRI International

Menlo Park, California, U.S.A.

Abstract

We address recent criticisms of evidential reasoning, an approach to the analysis of imprecise and uncertain information that is based on the Dempster-Shafer calculus of evidence.

We show that evidential reasoning can be interpreted in terms of classical probability theory and that the Dempster-Shafer calculus of evidence may be considered to be a form of generalized probabilistic reasoning based on the representation of probabilistic ignorance by intervals of possible values. In particular, we emphasize that it is not necessary to resort to nonprobabilistic or subjectivist explanations to justify the validity of the approach.

We answer conceptual criticisms of evidential reasoning primarily on the basis of the criticism's confusion between the current state of development of the theory — mainly theoretical limitations in the treatment of conditional information— and its potential usefulness in treating a wide variety of uncertainty-analysis problems. Similarly, we indicate that the supposed lack of decision-support schemes of generalized probability approaches is not a theoretical handicap but, rather, an indication of basic informational shortcomings that is a desirable asset of any formal approximate reasoning approach. We also point to potential shortcomings of the underlying representation scheme to treat general probabilistic reasoning problems.

We also consider methodological criticisms of the approach, focusing primarily on the alleged counterintuitive nature of Dempster's combination formula, showing that such results are the result of its misapplication. We also address issues of complexity and validity of scope of the calculus of evidence.

1 Introduction

If artificially intelligent systems are to produce adequate assessments of the state and behavior of the real world, they must cope with information and knowledge that is characterized by varying degrees of uncertainty, ignorance, and correctness. To address this need, we have developed a technology called *evidential reasoning*. It is formally based upon the Dempster-Shafer [18] theory of belief functions; it has been implemented as a domain-independent automated reasoning system; and it has been successfully applied to a range of real-world problems [11]. Yet, its reliance on belief functions has drawn criticism.

Our choice of an approach based on the Dempster-Shafer theory was not arbitrary. We believe that that theory confers important methodological advantages, such as its ability to represent ignorance in a direct and straightforward fashion, its consistency with classical probability theory, its compatibility with Boolean logic, and its manageable computational complexity. At the same time, we recognize that other approaches may also complement and augment the assessments provided by evidential reasoning.

We examine several criticisms of belief functions that have appeared in the literature, discussing first the fundamental theoretical bases supporting the belief-function approach and justifying its use in terms of the requirements imposed by ignorance of certain probability distributions. We consider the nature of Dempster's rule of combination and argue that negative assessments either misinterpret the nature of the distributions being combined or ignore the basic independence assumptions that assure its validity. We stress also that it is not necessary to rely on explanations that are either nonprobabilistic or subjective to justify the validity of the Dempster-Shafer calculus of evidence.

Furthermore, we show that certain apparently counterintuitive properties of the approach (e.g., the "spoiled sandwich" paradox) are the natural consequence of considering families of possible probability distributions that solve an approximate reasoning problem. In the context of this discussion, we indicate also the inherent pitfalls of "axiomatic" approaches that accept or reject methodologies on the basis of their compliance with allegedly intuitive principles.

We also answer critiques based on the computational complexity of the belief-function approach. Such criticisms claim that the complexity of probabilistic knowledge representations grows exponentially with the size of the frame, thus making the theory unsuited for automated reasoning. Other comments addressed in our presentation center on limitations on the representational ability of belief functions and the lack of certain methodological capabilities (e.g., decision-making mechanisms).

Despite the criticism that belief functions have drawn, we believe that evidential reasoning is well-founded and that it may be effectively applied to the solution of a broad range of important practical problems.

Most of our comments will be made in direct reply to a recent criticism of the belief-function approach by Pearl [15], because we feel that his paper encompasses most of the

major worries and concerns expressed about the calculus of evidence. While most of the discussion of this paper consists of direct responses to issues raised by Pearl and others, our overall objective is considerably broader. Our answers are motivated by the remarks of DeGroot, quoted by Pearl at the conclusion of his work, about the need to use our methodological approaches "... with the utmost care and in accordance with the highest ethical standards." Our aim, like Pearl's, is to enlighten and clarify, through careful discussion of rather subtle and delicate issues, rather than to engage in dogmatic defense of one approach to the detriment of another. It is our earnest hope that this work, in conjunction with other evaluations of the belief-function approach, will lead to a better understanding of its foundations, capabilities, and limitations.

2 On Theoretical Soundness

The theory of belief functions was originated by Dempster [4] in the context of statistical research. The use of the term "belief," together with its subjectivist connotations, is due to Shafer [18], who first applied the theory to the analysis of imprecise and uncertain evidence.

Although much skepticism has been voiced about the naturality of belief functions and their agreement with conventional probabilistic approaches, its theoretical bases are provided by a simple consideration of the role of evidence as a basic information carrier.

In classical probabilistic treatments, it is assumed that, under certain evidential conditions \mathcal{E} ,¹ the value $\mathbf{P}(p|\mathcal{E})$ of the likelihood of a particular statement p is known. This view of evidence, while adequate to represent the informational conditions of most controlled experimental setups, fails, however, to adequately model the effects that acquiring similar information has on our state of knowledge when the state of the world can not be so readily manipulated.

In such circumstances, whenever the evidence \mathcal{E} is observed, three possible informational outcomes may result from examination of further information that later turns out to improve our state of knowledge: either p is found to be true, $\neg p$ is found to be true (i.e., p is false), or such information is insufficient to determine the truth value of p . Use of modal logic concepts, which are the bases of the formal model of Ruspini [17], suggests the use of the notation $\mathbf{K}p$, $\mathbf{K}\neg p$, and $\mathbf{I}p$ to identify these outcomes. Since these alternatives are exclusive, it is clear that

$$\mathbf{P}(\mathbf{K}p) + \mathbf{P}(\mathbf{K}\neg p) + \mathbf{P}(\mathbf{I}p) = 1.$$

Furthermore, since the probability of $\mathbf{I}p$ may be positive, it will be true, in general, that

$$\mathbf{P}(\mathbf{K}p) + \mathbf{P}(\mathbf{K}\neg p) \leq 1.$$

¹Throughout this paper, the symbol \mathcal{E} is used to denote available evidence, i.e., a collection of propositions about the real world that are known to be true either as the result of direct observation or as the consequences of applicable background knowledge.

This model, based on a combination of classical probability methods and the modal logic **S5** [8,12], essentially provides—through the logical notion of possible world—a meaning for the unary operator **K** as the representation of the state of knowledge of a statistician who is estimating the probability of truth of diverse propositions $\{p, q, \dots\}$ under evidential conditions \mathcal{E} .

This statistician estimates those distributions by considering multiple samples of the state or behavior of a real-world system. Using, for each sample, additional information collected through further experimentation, the statistician may then establish or not the validity of a proposition p . If he is rather lucky, our statistician will find himself in the ideal situation where he can actually “know”² or “prove” that the real world is in a state s that is described to the best level of detail that is necessary to understand its behavior (i.e., a “possible world”). This is the state of knowledge usually attained, under perfect laboratory conditions, when experimental samples are fully analyzed and when the outcome of such analyses is classified in terms of a set of exhaustive and mutually exclusive alternatives.

Under less desirable epistemological circumstances, however, the statistician will only be able to prove that a less specific proposition q is true. In the extreme case where no further information exists, he will be forced to say that his knowledge is limited to that provided by the evidence \mathcal{E} , or that it is “vacuous.”

All samples so analyzed, however, can be classified as to the “most specific knowledge” that could be determined in each case. The corresponding probability measure of the set $e(p)$ of samples where the proposition p was the most specific knowledge (called an *epistemic set* by Ruspini) corresponds, in Shafer’s framework, to the value $m(p)$ of a mass function m , i.e.,

$$m(p) = \mathbf{P}(e(p)).$$

Correspondingly, the probability that p was “known” to be true during statistical experimentation, corresponds to the value $\mathbf{Bel}(p)$ of Shafer’s belief function, i.e.,

$$\mathbf{Bel}(p) = \mathbf{P}(\mathbf{K} p).$$

The connection between the ability of our statistician to know that p was true and the belief and mass functions that he estimates through experimentation justifies both the expression *epistemic probability* introduced by Ruspini [17] to describe the underlying probabilities defined over a particular set of situations or scenarios $\mathbf{K}p$ (called the *epistemic universe*), and the description of the functions as being “probabilities of provability” or “probabilities of necessity” by Pearl [14], following a suggestion by Fagin and Halpern [6].

In short, all such interpretations are equivalent to the original model of Ruspini, where a rational agent was able to prove the truth of different propositions under different infor-

²Note that, in the context of epistemic logics such as **S5**, the operator **K** behaves as a logical necessity operator. “Knowing” a proposition simply means that observations logically imply such proposition, or that it is *necessarily* true.

mational circumstances that were found to prevail, during his statistical experiment, with different frequencies of occurrence.³

Since the ability to prove a proposition q entails the ability to prove any proposition p that is implied by q , it should be clear that

$$\mathbf{Bel}(p) = \sum_{q \Rightarrow p} m(q),$$

which is the fundamental equation relating the basic structures of the calculus of evidence. It is also true that

$$\mathbf{Bel}(p) \leq \mathbf{P}(p) \leq 1 - \mathbf{Bel}(\neg p),$$

providing bounds for the probability of p that may not be improved. This ability to manipulate probability intervals by means of the compact representation scheme of mass functions is the major reason for the appeal of the Dempster-Shafer methodology.

While the above discussion clarifies the nature of the statistician's knowledge modeled by belief and mass functions, doubts might still remain as to their utility to those who were not involved in their statistical estimation process. Such usage is, however, that made of any other probabilistic information. The analyst who observes \mathcal{E} does not have the luxury that was available to the statistician estimating epistemic probabilities, i.e., the ability to collect additional information that permits a more detailed characterization of the state of the world, for the same reasons that the user of statistical tables is unable to utilize the raw data of the estimating statistician. Under such circumstances, the analyst is forced to rely on the probabilistic estimates provided by the statistician, which are believed on the basis of the assumed regularity of the repetitive behavior of the system: the epistemological cornerstone of probabilistic reasoning.

In other words, the "probability of provability" is the best information that is available to the analyst; an observation that not only disposes of questions about its role in probabilistic reasoning, but also of Pearl's worries about its use in lieu of the obviously more desirable "probability of truth" [15]:

"why we should concern ourselves with the probability that the evidence implies A , rather than the probability that A is true, given the evidence?"

Clearly, we would prefer having the latter, but, unfortunately, we can only measure the former.

³Note, however, that while use of the terms "knowability," "provability," and "necessity" does much to provide adequate semantics to the calculus of evidence, its loose usage leads to unnecessary confusion. For example, in his recent criticism [15], Pearl takes some questionable semantic license with the term "necessity," mentioning, for example, the probability that a decision "will have to be made out of compelling necessity." Such "pragmatic" necessity does not have anything to do, of course, with the "logical necessity" that underlies the Dempster-Shafer theory, i.e., the necessary truth of a proposition given available evidence.

Our interpretation of the major evidential functions and structures also quickly disposes of erroneous arguments based on unintended interpretations of the intervals defined by belief functions. Each such interval represents ignorance of a single probability value for a proposition p under fixed evidential conditions \mathcal{E} . If critics choose, for example, to interpret such intervals as the possible values that conditional probabilities might attain when further evidence is collected, as suggested by Pearl [13], belief functions will not, indeed, behave according to such unintended semantics.

In closing this section, it is important to mention other alternative views of the structures of the calculus of evidence such as that recently proposed by Smets [19], which are based on a nonprobabilistic concept of belief. Although those models are interesting on the strength of their own virtues, we still emphasize that such interpretations are not required to reconcile the calculus of evidence with conventional probability theory.

In consideration of our ability to reconcile all structures and formulas of the calculus of evidence, including the Dempster's formula, with conventional probability structures, such as inner and outer probabilities, we do not feel strongly compelled to accept alternative epistemic interpretations. Our skepticism in this regard is further supported by the observation that, often, such epistemological alternatives are the result of misunderstandings about the role of certain evidential formulas and processes (e.g., normalization). For the same reasons, we remain unconvinced about the need to assign alternative interpretations to the structures of calculus of evidence or to its functions, as is the recent suggestion of Halpern and Fagin [7], which is echoed by Pearl [15].

3 On Decision Support

A criticism of a more fundamental nature of the calculus of evidence is often raised regarding the output of generalized interval-probability approaches. Since these methods often fail, because of basic knowledge deficiencies, to rank decision choices by the value of some measure that quantifies the desirability of each choice (e.g., expected utility), then it is said that they lack a decision-theoretic apparatus.

Although these arguments correctly point to the basic knowledge requirement that most decision problems entail—if a rational choice is to be made, then we must have a proper informational basis to do it—this obvious consideration is twisted to argue for the necessity to estimate unknown probability and utility values when they are not available. We do not think that this *pragmatic necessity* argument is either sound or compelling.

In our view, the calculus of evidence may be used in a straightforward fashion to produce intervals of possible utility-values. When such intervals overlap and cannot be ordered, this fact simply reflects a basic deficiency in our knowledge. We look down upon “pragmatic justifications” with the same concern that any experimental scientist must show about proposals to guess what he has not measured: the ability to make decisions in the absence of knowledge is, in our view, a handicap rather than an advantage of any method.

Far from lacking a decision-theoretic methodology, our approach provides an understandable quantification of the undesirable effects that poor information has on our decision-making ability, ordering decisions whenever it is rationally possible but advising us that such ranking is not possible if our knowledge is insufficient. In brief, our approach not only supports decision-making but, through its built-in sensitivity-analysis features, helps us to determine what must be done to reach a happier epistemological state.⁴

4 On Dempster's Rule of Combination

The semantic model of the Dempster-Shafer theory also validates the so-called Dempster's rule of combination, which permits the combination of belief and mass functions corresponding to different evidential observations, made under certain conditions of independence. When such conditions are not valid, use of this formula leads, of course, to erroneous results, often, although incorrectly, considered to be an essential handicap of the evidential reasoning approach, rather than a consequence of its misapplication.

The Dempster formula is, currently, the principal evidence integration mechanism of the belief-function approach. It was derived in the context of a basic model of the effect of probabilistic evidence that correctly interprets such evidence as constraints on probability values rather than as the source of the actual values, which are typically undetermined. It may be described as an expression that, under certain conditions of independence, yields bounds for the conditional probability distribution $P(\cdot|\mathcal{E}_1, \mathcal{E}_2)$ on the basis of similar bounds for the probability distributions $P(\cdot|\mathcal{E}_1)$ and $P(\cdot|\mathcal{E}_2)$.

To understand the conceptual bases for the Dempster's formula of combination and its consistence with conventional probability, we resort to a generalization of the logical model used before to derive the basic relations of the calculus of evidence. Instead of considering a single epistemic operator, corresponding to a single statistician or observer, we will consider two such rational agents, with their knowledge modeled by means of two operators K_1 and K_2 . Each of these rational agents will be assumed to be ignorant of the knowledge possessed by the other, i.e., as if they were statisticians performing independent experiments under different evidential conditions \mathcal{E}_1 and \mathcal{E}_2 . Their common knowledge, however, will be modeled by means of a nonindexed operator K corresponding to a third reliable agent that aggregates the statistical knowledge gathered by the other two.

Clearly, in a given applicable situation (i.e., the first agent observes \mathcal{E}_1 and the second agent observes \mathcal{E}_2), the integrating agent, who does not add any knowledge of his own, will be able to prove (or to "know" the truth of) a proposition p , if the other agents provide individual items of information that, when combined (i.e., conjoined) imply p , as expressed by the basic combination axiom:

⁴For an example of an approach that incorporates decision-maker preferences into the framework of the belief-function calculus, the reader is referred to a recent paper by Strat [21].

$\mathbf{K}p$ is true if and only if there exist sentences p_1 and p_2 such that \mathbf{K}_1p_1 and \mathbf{K}_2p_2 are true, and such that $p_1 \wedge p_2 \Rightarrow p$.

Using our three operators to generate all possible (i.e., logically consistent) states of knowledge that may be attained by each of the three agents while assessing the state of a real system, we may say that each of them has, as was the case before, knowledge about the real world that may be represented by the “most specific”⁵ propositions p_1, p_2 , and p that each has been able to prove (with p being obviously more specific than either p_1 or p_2). In the terminology of Ruspini’s semantic model, each of the agents is in an epistemic state, denoted by $e(p)$, $e_1(p_1)$ and $e_2(p_2)$, respectively, each corresponding to the set of all conceivable states of the real world (i.e., possible worlds) having such knowledge characteristics.

The following important set-equation relating all of these types of epistemic sets as subsets of our enhanced epistemic universe is the basis for the derivation of various evidential combination formulas,

$$e(p) = \bigcup_{p_1 \wedge p_2 = p} (e_1(p_1) \cap e_2(p_2)),$$

of which the Dempster combination formula,

$$m(p) = \kappa \sum_{p_1 \wedge p_2 = p} m_1(p_1) m_2(p_2),$$

where

$$m(p) = \mathbf{P}(e(p)|\mathcal{E}_1, \mathcal{E}_2), \quad m_1(p_1) = \mathbf{P}(e_1(p_1)|\mathcal{E}_1), \quad m_2(p_2) = \mathbf{P}(e_2(p_2)|\mathcal{E}_2),$$

and where κ is a multiplicative factor, is the best known and used.

Before reviewing the actual process leading to the derivation of the Dempster’s formula, it is important to pause and reflect upon the nature of the above set-theoretic equation and its usefulness to derive evidence combination formulas.

We may first note that this equation has been derived as a relation between subsets of possible “epistemological states” that is valid regardless of any assumptions about probabilistic structures and their properties (e.g., independence). As such, it provides not only the bases for the derivation of the Dempster formula but actually for a variety of formulas that bound possible probability values within and outside the structures of the Dempster-Shafer theory.

Basically, this formula provides the basis to extend a probability function \mathbf{P} that is known over subsets of the form $e_1(p_1)$ and $e_2(p_2)$ (i.e., over two σ -algebras), to the set of unions of sets of the form $e_1(p_1) \cap e_2(p_2)$ (i.e., another σ -algebra). If such extension can be made uniquely—as is the case for Dempster’s formula—the resulting extension may be used to generate both the conditional probability $\mathbf{P}(\cdot|\mathcal{E}_1, \mathcal{E}_2)$ and its associated bounds \mathbf{Bel} and \mathbf{Pl} ,

⁵Note that such most-specific knowledge always exists and is unique but for logical equivalences, since the conjunction of all proved theorems is itself a theorem.

which are fully compliant with Shafer’s axioms. In other less fortunate cases (e.g., dependent evidence), such extension is not unique, and the lower envelope of the possible extensions, which is not a probability, will lead to bounds that do not satisfy the axioms of the calculus of evidence.

This equation is now being used to extend the evidential calculus approach by generalization of the notion of conditional probability by study of the probabilistic relations that define dependencies between the different types of epistemic sets (i.e., $e(p)$, $e_1(p_1)$ and $e_2(p_2)$). Pearl [15], however, believes, apparently as the result of his examination of the role of *compatibility relations* in the calculus of evidence, that this approach is essentially limited in its expressive ability to set-theoretic relations between epistemic sets, which correspond to classical logical conditional statements (i.e., material implications).

In fact, it may be easily seen from our epistemic identity that whenever the conditional probabilities $P(e_2(p_2)|e_1(p_1))$ and $P(e_1(p_1)|e_2(p_2))$ are restricted to take the values 0 or 1,⁶ this identity may be used to map one body of evidence into another, i.e., by means of the compatibility relations that such probabilities define.

Since under these assumptions, however, there can be only one proposition p_2 for every proposition p_1 such that $P(e_2(p_2)|e_1(p_1)) = 1$, and vice versa, then the compatibility relation that is so defined may be characterized by several implications of the form

$$e_1(p_1) \Rightarrow e_2(p_2) ,$$

and of the form

$$e_2(q_2) \Rightarrow e_1(q_1) ,$$

between knowledge states of one observer and knowledge states of the other which are useful to “transfer mass” between propositions. This correspondence must be contrasted with that following from the limited interpretation given by Pearl, who, from knowledge of

$$e_1(p_1) \Rightarrow e_2(p_2) ,$$

concludes (by contraposition), correctly but narrowly, that

$$\neg e_2(p_2) \Rightarrow \neg e_1(p_1) ,$$

and proceeds then to attach all material implication paradoxes (e.g., the “ravens paradox”) to the calculus of evidence as if they were an essential methodological bane. If that were to be the case—clearly it is not—the same concerns should be raised about the use of conditionals in conventional probability calculus.

The second observation that may be made about the nature of evidence combination, in general, and the role of our basic set identity to generate combination formulas, in particular,

⁶It may be shown from the definition of epistemic sets that, under such conditions, knowledge of $P(e_2(p_2)|e_1(p_1))$ suffices to derive $P(e_1(p_1)|e_2(p_2))$.

is that while the functions to be combined are conditional probabilities over two different evidential sets \mathcal{E}_1 and \mathcal{E}_2 (i.e., the evidence observed by two agents), the desired integrated probability is a distribution over $\mathcal{E}_1 \cap \mathcal{E}_2$ (since we know that both observations are correct). Except for unusual cases, however, computation of $\mathbf{P}(\cdot | \mathcal{E}_1, \mathcal{E}_2)$ entails a “normalization” operation that is fully consistent with the calculus of probability. Most of the normalization “paradoxes” are the result of misunderstanding about what is being combined: two different conditional probabilities rather than two different lower and upper bounds of the same probability function.⁷

Focusing now on the rationale for Dempster’s formula, we should notice first that the epistemic sets $e_1(p_1)$ and $e_2(p_2)$ are such that

$$e_1(p_1) \subseteq \mathcal{E}_1, \quad e_2(p_2) \subseteq \mathcal{E}_2,$$

i.e., the possible knowledge states of each statistician include awareness of the truth of the evidence that is observed by each. Furthermore,

$$\mathcal{E}_1 = \bigcup_{p_1} e_1(p_1), \quad \mathcal{E}_2 = \bigcup_{p_2} e_2(p_2),$$

where $p_1 \Rightarrow \mathcal{E}_1$ and $p_2 \Rightarrow \mathcal{E}_2$; i.e., each statistician knows something that implies that his evidential observation is true (otherwise he would not be “counting” that sample).⁸

Assume now that there exists a probability distribution \mathbf{P} defined over the space of all possible epistemic states for our observing statisticians and our “integrating” agent. Each such epistemic state is a possible world that corresponds to a possible state of the world and to a possible state of knowledge for each agent that, in addition, is consistent with the laws of logic. We will assume now that, whenever $p_1 \Rightarrow \mathcal{E}_1$ and $p_2 \Rightarrow \mathcal{E}_2$,

$$\mathbf{P}(e_1(p_1) \cap e_2(p_2)) = \begin{cases} \mathbf{P}(e_1(p_1)) \mathbf{P}(e_2(p_2)), & \text{if } p_1 \wedge p_2 \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases}$$

This assumption simply states that when \mathcal{E}_1 and \mathcal{E}_2 are both true the probability that a rational observer will be in a particular knowledge, or epistemic, state does not provide any information about the probability of the epistemic state of the other agent (i.e., beyond ruling out logical impossibilities). In purely formal terms, we may say that knowledge of values of \mathbf{P} over sets of the form $e_1(p_1)$ does not provide any indication, beyond exclusion of logical impossibilities, of the values of \mathbf{P} over sets of the form $e_2(p_2)$ and vice versa. The epistemic states of our two agents may be said, therefore, to be unrelated in that knowledge of the state of one of our observers (by our integrating agent) does not provide any information about the state of the other, save for elimination of logical impossibilities.

⁷It is fair to say that much of the skepticism raised by the normalization used in Dempster’s formula can be traced to the exposition given by Shafer [18], which suggests a nonprobabilistic method of evidence combination.

⁸Recall that our observers, or rational agents, are statisticians estimating properties of certain statistical distributions by classifying each sample using their evidence and additional sample-dependent knowledge.

Noting now that

$$\mathbf{P}(e_1(p_1)|\mathcal{E}_1) = \frac{\mathbf{P}(e_1(p_1))}{\mathbf{P}(\mathcal{E}_1)}, \quad \mathbf{P}(e_2(p_2)|\mathcal{E}_2) = \frac{\mathbf{P}(e_2(p_2))}{\mathbf{P}(\mathcal{E}_2)},$$

$$\mathbf{P}(e_1(p_1) \cap e_2(p_2)|\mathcal{E}_1, \mathcal{E}_2) = \frac{\mathbf{P}(e_1(p_1) \cap e_2(p_2))}{\mathbf{P}(\mathcal{E}_1 \cap \mathcal{E}_2)},$$

then, whenever $p_1 \wedge p_2 \neq \emptyset$,

$$\mathbf{P}(e_1(p_1) \cap e_2(p_2)|\mathcal{E}_1, \mathcal{E}_2) = \kappa \mathbf{P}(e_1(p_1)|\mathcal{E}_1) \mathbf{P}(e_2(p_2)|\mathcal{E}_2) = \kappa m_1(p_1) m_2(p_2),$$

from which the Dempster's formula readily follows.

The normalization factor

$$\kappa = \frac{\mathbf{P}(\mathcal{E}_1) \mathbf{P}(\mathcal{E}_2)}{\mathbf{P}(\mathcal{E}_1 \cap \mathcal{E}_2)},$$

has been the object of considerable concern on the part of both skeptics and proponents of the calculus of evidence. The above expression, however, provides the rationale for its usage while disposing of arguments about its alleged inconsistency with the probability calculus. In that expression, the denominator $\mathbf{P}(\mathcal{E}_1 \cap \mathcal{E}_2)$ appears as the consequence of the need to derive probability distribution estimates with respect to the intersection of the two observed evidences \mathcal{E}_1 and \mathcal{E}_2 . The numerator of that expression simply reflects the need to combine conditional distributions over the same reference set (i.e., the epistemic universe) while our probabilistic knowledge is expressed over two of its subsets (i.e., \mathcal{E}_1 and \mathcal{E}_2).

The essence of the conditions that lend validity to the Dempster formula may be summarized by saying that the formula's usefulness is confined to the limited, but rather important, cases where estimates of probabilistic likelihood have been formulated by two rational agents on the bases of independent observations, while ignoring the evidence available to each other.

If our integrating agent is thought of as being concerned with estimating the probabilities of certain events when both \mathcal{E}_1 and \mathcal{E}_2 are true, then we may say that, whenever the conditions validating the Dempster's formula hold, knowledge of the fact that a particular sample satisfies p_1 tells the agent nothing about the likelihood of p_2 (unless, of course, p_1 happens to be logically inconsistent with p_2). Furthermore, whenever our integrating agent is done with his job, he should find out that estimating this joint distribution (i.e., over $\mathcal{E}_1 \cap \mathcal{E}_2$) could have been accomplished in an easier fashion by estimating the marginal distributions over \mathcal{E}_1 and \mathcal{E}_2 and deriving the joint distribution by multiplication and normalization.

Other accounts supporting the validity of Dempster's formula and its consistence with the probability calculus have been advanced by several authors. A particularly compelling justification has been recently given by Wilson [22].

5 On “Paradoxes”

Criticisms of the Dempster formula may be broadly characterized as being the consequence of basic misunderstandings about either its meaning or its validity.

In this section, we examine three alleged paradoxes of the theory, showing that the purported inconsistencies are actually the results of conceptual misunderstandings or misrepresentations of the position of those who, while generally supporting the calculus of evidence, are concerned with its possible misapplication.

5.1 The “Three-Prisoner” Problem

Turning our attention first to concerns about the validity of the Dempster’s formula, we may note that, in general, such examples ignore its scope of applicability, producing counterintuitive results that are then used to dismiss the methodology as inadequate. Among those, the “three-prisoner” problem discussed by Diaconis and Zabell [5] has been perhaps the most quoted and discussed.

This problem is one of a variety of examples, in which the combination formula is used as a *conditioning formula* by assuming that one of the mass distributions being combined simply assigns all of its mass to a proposition p in the frame of discernment. Combination of such a simple support function with another mass function associated with a belief function $\text{Bel}(\cdot)$ leads to the conditioning formula

$$\text{Bel}(q|p) = \frac{\text{Bel}(q \vee \neg p) - \text{Bel}(\neg p)}{1 - \text{Bel}(\neg p)}.$$

In the particular case of the three-prisoner problem, concerned with the guilt or innocence of a prisoner that has been chosen (by the Warden) as the guilty party by random draw among three candidates A_1 , A_2 , and A_3 , our “logical space” or frame of discernment is simply the Boolean algebra induced by the three noncompatible propositions

“Prisoner A_i has been found guilty,”

where $i = 1, 2, 3$. Since only one of the three prisoners is chosen by the Warden, we clearly have

$$\mathbf{P}(p_i) = \frac{1}{3}, \quad i = 1, 2, 3.$$

(Note that \mathbf{P} is actually a classical, additive, probability distribution).

Prisoner A_1 now asks the Jailer to name one of the innocent prisoners (other than him) arguing that such information would clearly be of little help to him as an indicator of his potential fate. As Pearl notes, if q stands for the proposition “The Jailer names A_2 as one of the innocent,” then application of the conditioning rule leads to the result

$$\text{Bel}(p_1|q) = \mathbf{P}(p_1|q) = \frac{1}{2},$$

indicating that the conditional probability $\mathbf{P}(p_1|q)$ must be exactly $\frac{1}{2}$, instead of the “correct solution”

$$0 \leq \mathbf{P}(p_1|q) \leq \frac{1}{2},$$

while also saying, against the correct intuition of A_1 that his chances of guilt have been increased as the result of the irrelevant information provided by the Jailer. From such an observation, Pearl concludes that the formula is seriously flawed, both because of the counterintuitive result that it produces and for its “collapsing” of a family of solutions into a single value.

Before proceeding to the discussion of Pearl’s concerns, we may note, in passing, that this problem has been well known as a source of paradoxes and incorrect solutions within the scope of the conventional probability calculus [2] quite independently of any issues of validity of its treatment using the Dempster-Shafer calculus. The explanations given to describe the conceptual errors leading to incorrect classical treatments resemble to some extent those that shed light on the inapplicability of the Dempster’s formula.

Returning now to the role of the Dempster’s formula in this problem, we may first observe that, although, at first glance, the distributions representing the Jailer’s and Warden’s choices seem independent, it is actually impossible for the Jailer to tell to A_1 that A_2 is one of those to be spared if all he knew was that the Warden was choosing the guilty party by random draw (i.e., he needs to know exactly who is the one chosen for punishment). To use the terminology of Ruspini’s model, the probability of A_2 being named as one of the innocent depends on the epistemic state of the Warden, thus violating the independence assumptions of the Dempster’s formula. If all possible combinations of truth values for the propositions p_i , $i = 1, 2, 3$, and q are tabulated, together with their probabilities, as is done in Table 1, then it is clear that

$$\mathbf{P}(q|p_3) = 1, \quad \mathbf{P}(q) = \frac{1}{3}(1 + \alpha),$$

where $0 \leq \alpha \leq 1$ represents the unknown probability that the Jailer will choose to name A_2 rather than A_3 as innocent if A_1 is actually the one chosen by the Warden as guilty.

Possible World	Warden’s Choice	Jailer Identifies	Probability
W_1	A_1	A_2	$\frac{1}{3} \alpha$
W_2	A_1	A_3	$\frac{1}{3} (1 - \alpha)$
W_3	A_2	A_3	$\frac{1}{3}$
W_4	A_3	A_2	$\frac{1}{3}$

Table 1: Possible Worlds in the Three-Prisoner Problem

But then,

$$\mathbf{P}(q|p_3) \neq \mathbf{P}(q),$$

violating the assumptions, discussed above, that validate the utilization of Dempster’s formula (i.e. $\mathbf{P}(e_2(p_2)|e_1(p_1)) \neq \mathbf{P}(e_2(p_2))$). There is not, therefore, “total mystery,” as Pearl says, as to the incorrect results obtained using the Dempster’s formula. Because it fails to be applicable, there should be little wonder that it leads to apparent paradox.

Although, as clearly shown by this discussion, the incorrect treatment of the three-prisoner problem fails to invalidate the Dempster’s rule of combination, we share the concern of Pearl and others about its wide misapplication, particularly when it is used indiscriminately to generate conditional distributions. In our research, we are endeavoring to extend the original theory to produce expressions to produce and utilize conditional belief information [16] that incorporates known dependencies between evidential bodies. These formulas are intended to provide better interval estimates than the typically uninformative bounds that are supplied by strict derivation of bounds in the absence of additional information by the expression

$$\mathbf{Bel}(q|p) = \frac{\mathbf{Bel}(p \wedge q)}{\mathbf{Bel}(p \wedge q) + \mathbf{PI}(p \wedge \neg q)},$$

which is mentioned in Dempster’s original paper [4] and that has been the object of recent concern by several authors [3,7].

In closing, we believe it is important to address other concerns of Pearl, apparently going beyond the three-prisoner problem, about the counterintuitive nature of the “collapse” that usage of the Dempster formula often produces, which is manifested by production of a single conditional probability distribution when conditioning multiple members of a family \mathcal{P} of probabilities over some specific subset q . Just as it is true that all members of the family of distributions

$$\mathcal{P} = \{\mathbf{P}_t : t \text{ in } [0, 1]\}$$

defined in the set $X = \{a, b, c\}$ by the expression

$$\mathbf{P}_t(x) = \begin{cases} \frac{1}{2}t, & \text{if } x = a, \\ \frac{1}{2}(1-t), & \text{if } x = b, \\ \frac{1}{2}, & \text{if } x = c \end{cases}$$

are such that $\mathbf{P}_t(\{a, b\}) = \frac{1}{2}$, despite their variability over other subsets, it is also true that an extensive family of distributions may collapse into a single conditional probability without violating any rational or probabilistic principles. Such “invariants” are, in fact, desirable as elements that simplify the analysis of an otherwise complex probabilistic problem. For these reasons, we believe that, if the Dempster’s conditioning formula is applicable, its reduction of the variability of probability values should not be a particular cause for concern as to its validity.

5.2 The Spoiled Sandwich

While discussing the suitability of the calculus of evidence either as a form of generalized probabilistic calculus or as a new theory that intends to capture a novel notion of belief, Pearl [15] again faults the approach for failing to satisfy the following rationality principle originally stated by Aleilunas [1]:

“If two diametrically opposed assumptions yield two different degrees of belief in a proposition Q , then the unconditional degree of belief merited by Q should be somewhere between the two.”

As natural as such a principle might look at first, the following simple and clever example from Wilson [23] clearly shows that it is neither intuitive nor appealing but points, instead, to the pitfalls of creating or supporting one’s favorite scheme on the strength of supposedly rational axioms.

Let $X = \{a, b, c, d\}$ with $A = \{a, b\}$ and $B = \{a, c\}$, so that $\bar{B} = \{b, d\}$. Consider the family of probability distributions in X

$$\mathcal{P} = \{ \mathbf{P}_t : t \text{ in } [0, 1] \},$$

indexed by a parameter t in $[0, 1]$ and defined by

$$\begin{aligned} \mathbf{P}_t(\{a\}) &= \frac{1}{2}t, \\ \mathbf{P}_t(\{b\}) &= \frac{1}{2}(1-t), \\ \mathbf{P}_t(\{c\}) &= \frac{1}{4}, \\ \mathbf{P}_t(\{d\}) &= \frac{1}{4}, \end{aligned}$$

and let

$$\mathbf{P}_* = \inf_t \{ \mathbf{P}_t \} .$$

Then, clearly,

$$\mathbf{P}_t(A) = \frac{1}{2}t + \frac{1}{2}(1-t) = \frac{1}{2},$$

and, therefore, $\mathbf{P}_*(A) = \frac{1}{2}$. The conditional probabilities $\mathbf{P}_t(A|B)$ and $\mathbf{P}_t(A|\bar{B})$ are given by the expressions

$$\begin{aligned} \mathbf{P}_t(A|B) &= \frac{\mathbf{P}_t(\{a\})}{\mathbf{P}_t(\{a, c\})} = \frac{\frac{1}{2}t}{\frac{1}{4} + \frac{1}{2}t}, \\ \mathbf{P}_t(A|\bar{B}) &= \frac{\mathbf{P}_t(\{b\})}{\mathbf{P}_t(\{b, d\})} = \frac{\frac{1}{2}(1-t)}{\frac{1}{4} + \frac{1}{2}(1-t)}, \end{aligned}$$

from which the lower bounds

$$\begin{aligned} \mathbf{P}_*(A|B) &= \inf_t \mathbf{P}_t(A|B) = 0, \\ \mathbf{P}_*(A|\bar{B}) &= \inf_t \mathbf{P}_t(A|\bar{B}) = 0, \end{aligned}$$

are easily derived. It is clear, however, that

$$\frac{1}{2} = \mathbf{P}_*(A) > \mathbf{P}_*(A|B) = \mathbf{P}_*(A|\bar{B}) = 0,$$

showing that the sandwich principle is violated even within the confines of conventional probability theory.

5.3 Other Ways to Spoil the Sandwich

Although such simple examples should suffice to dispose of concerns about spoiled sandwiches, we feel that Pearl’s discussion of the problem deserves a more detailed analysis, mainly because of its philosophical implications to rational thinking. This is particularly important because loose use of such terms as “assured winnings,” “support,” or “belief” in the absence of a sound, formal interpretive framework may quickly mislead those engaged in the comparison of alternative methodologies.

In an example, called “the Peter, Paul, and Mary Sandwich problem,” Pearl presents a betting situation in which Mary prepares either a ham or a turkey sandwich, promising to pay Paul \$1000 should he guess correctly the type of sandwich that she has prepared. Not having a clue as to Mary’s choice, Paul then flips a coin, guessing “ham” if the coin turns up heads and guessing “turkey” if it comes up tails. Paul, as Pearl notes, behaves like an “incurable Bayesian,” reckoning that

$$\begin{aligned} \mathbf{P}(\text{win}) &= \mathbf{P}(\text{win} \mid \text{turkey}) \mathbf{P}(\text{turkey}) + \mathbf{P}(\text{win} \mid \text{ham}) \mathbf{P}(\text{ham}) \\ &= \mathbf{P}(\text{tails} \mid \text{turkey}) \alpha + \mathbf{P}(\text{heads} \mid \text{ham}) (1 - \alpha) = \frac{1}{2}, \end{aligned}$$

regardless of the value α of the probability that Mary has actually prepared a turkey sandwich. Thus, in spite of not being “assured” a win or having “supporting evidence,” Paul can invoke the rationality (doubtful, as we already saw) of the sandwich principle and argue that he does not need to engage in unnecessary knowledge acquisition or experimentation [15]:

“If every possible outcome of an experiment would lead you to choose the same action, then you ought to choose that action without running the experiment.”

From such an observation, Pearl proceeds to fault the philosophical underpinnings of the evidential reasoning approach, eventually going as far as to suggest that, should Bayesian orthodoxy be unapplicable, the Dempster’s formula—which, he freely admits, does not play any role in this example—be replaced by other formulas such as the well-known bounds recently rediscovered by Halpern and Fagin [7].

In the light of our previous example about the rather inconvenient ability of conventional probability families to spoil sandwiches, all of these pronouncements look increasingly suspicious: What, however, may we say is wrong? This question may be answered in two equivalent ways.

We may say first, keeping ourselves at the informal discussion level, that, often, the experiments may interact with probabilities in complex ways that, obviously, Pearl has not considered. Nothing in Pearl’s formalism suggests, for example, that the sandwich has already been prepared and that it may not be artfully substituted by Mary to assure that Paul always loses, thus invalidating his hopes of having at least a 50 percent chance of winning.

The second, more formal, rendering of this observation is again based on the semantic model of Ruspini. In this, and in other similar problems, we have several agents that deliberate about the state of the world on the basis of their knowledge and knowledge of the knowledge of others. If the unary operator \mathbf{K} represents the state of knowledge of one of these agents, then, as observed before, our agent is always in one of three possible epistemological states with respect to the validity of a proposition p : either he knows that p is true (denoted $\mathbf{K}p$), or he knows that p is false (denoted $\mathbf{K}\neg p$), or he may be ignorant of such truth (i.e., $\neg\mathbf{K}p \wedge \neg\mathbf{K}\neg p$, denoted $\mathbf{I}q$).

In standard accounts, assuming that *knowledge* of the truth of one proposition does not affect the likelihood of *truth* of other propositions,⁹ we are simply concerned with a single form of conditional probability: that measuring the likelihood of p being true when q is true. In more complex epistemological situations, we may need to be concerned with such quantities as $\mathbf{P}(\mathbf{K}p \mid \mathbf{K}q)$, $\mathbf{P}(\mathbf{K}p \mid q)$, $\mathbf{P}(\mathbf{K}p \mid \mathbf{I}q)$, and the like. In other words, $\mathbf{Bel}(p \mid q)$ measures the support that knowledge of the truth of q provides to the truth of p , rather than the support provided by the truth of q to the truth of p .

In the Peter, Paul, and Mary sandwich problem, Pearl implicitly assumes that

$$\begin{aligned} \mathbf{P}(\mathbf{K}_{\text{MARY}}\text{heads}) &= 0, \\ \mathbf{P}(\mathbf{K}_{\text{MARY}}\text{tails}) &= 0, \\ \mathbf{P}(\text{turkey} \mid \mathbf{I}_{\text{MARY}}\text{heads}) &= \alpha, \\ \mathbf{P}(\text{ham} \mid \mathbf{I}_{\text{MARY}}\text{heads}) &= 1 - \alpha, \end{aligned}$$

concluding correctly, by application of the total probability law, over the exhaustive and exclusive set of possibilities

$$\{\mathbf{K}_{\text{MARY}}\text{heads}, \mathbf{K}_{\text{MARY}}\text{tails}, \mathbf{I}_{\text{MARY}}\text{heads}\},$$

that Paul has at least a 50 percent chance of winning.

This correct use of the total probability law does not mean that, by contrast, one should assume that the full extent of the conditional information provided by belief functions is limited to the conditional support functions

$$\mathbf{Bel}(p \mid q) = \mathbf{P}(p \mid \mathbf{K}q), \quad \mathbf{Bel}(p \mid \neg q) = \mathbf{P}(p \mid \mathbf{K}\neg q),$$

⁹The relations between knowledge and truth are more evident if “knowing” is thought of as sensing or observing, and if independence is understood as a lack of relationship between the errors of the sensors.

as Pearl evidently does. In short, not knowing p is not the same as knowing $\neg p$. The example of the Peter, Paul, and Mary sandwich shows that one needs to consider states of ignorance that, when properly accounted for, spoil even the best-conceived principles of rationality.

To fully appreciate the complexity of the problem, suppose that we change Pearl’s implicit assumptions, bringing the previously absent Peter into the scene as a spy acting on behalf of Mary. In this new scenario, still consistent with Pearl’s explicit statement of the problem, Peter, spying on Paul’s coin flipping experiment, alerts Mary, who, being rather artful and deft of hand, substitutes the sandwich so as to make sure that Paul always loses. In this case,

$$\mathbf{P}(\text{ham} \mid \mathbf{K}_{\text{MARY}}\text{tails}) = 1, \quad \mathbf{P}(\text{turkey} \mid \mathbf{K}_{\text{MARY}}\text{heads}) = 1;$$

and, most importantly,

$$\mathbf{P}\left(\left(\mathbf{K}_{\text{MARY}}\text{heads}\right) \cup \left(\mathbf{K}_{\text{MARY}}\text{tails}\right)\right) = 1,$$

i.e., Mary is never ignorant as to what Paul will bet.

The Peter, Paul, and Mary sandwich example does not, in our view, invalidate the applicability of the evidential approach, but rather highlights the need to make necessary discriminations between propositional truth, knowledge of that truth, and the interplay between such conditions that are likely to be glossed over by cursory analyses based on conventional approaches.

5.4 The Disagreeing Experts

Another common misunderstanding regarding the role of Dempster’s combination formula is that provoked by an example of Zadeh [24], which is often described as an indication of theoretical inadequacy.

This example concerns two reliable experts that assess, in a rather conflicting fashion, the likelihood of three, noncompatible, events A , B , and C as shown in Table 2. Representation of each of the expert’s assessments as a mass distribution followed by their combination with the Dempster’s rule yields $\mathbf{P}(B) = 1$, indicating that the “true” event is B , an alternative considered to be rather unlikely by either of the assessors.

Observer	$\mathbf{P}(A)$	$\mathbf{P}(B)$	$\mathbf{P}(C)$
1	0.99	0.01	0
2	0	0.01	0.99

Table 2: Experts Disagree on the State of the World

Although this example is often quoted as an example of the failure of the Dempster’s rule, it is clear that each of the rows in Table 2 defines a conventional probability distribution, thus suggesting that the problem is likely to lie elsewhere. While one may be tempted to defend any method of evidence combination by saying that the evidence, however peculiar, indicates that Observer 1 is ruling out alternative *C* while Observer 2 is excluding alternative *A*, thus leaving only *B* as the sole possible answer, it is clear, upon further examination, that the rows of Table 2 cannot possibly be evaluations of the same probability distribution. If that were the case, then at least one of the experts must be wrong, since there can only be one correct probability distribution, contradicting the assumption that they are both reliable.

Clearly, if the example is to make any sense—under any type of probabilistic interpretation—each row must correspond to a different conditional probability where the conditions correspond to different observations available to each expert. A simple example, suggested by a recent example used by Kyburg [9] to address other probabilistic reasoning issues, will help to clarify matters.

In this example we are being asked to reason, on the basis of available evidence, about the taste and edibility of certain berries that may be either small or large; and red or blue; have good or bad taste; or be safe or poisonous to eat. We will assume that the berries in question are distributed according to the distribution shown in Table 3.

Color	Size	Taste/Edibility	Probability
Red	Small	Good/Edible	99/199
Blue	Large	Bad/Edible	99/199
Red	Large	Poisonous	1/199

Table 3: The Berries Probability Distribution

If now a berry is picked up and found by an expert to be large, he will correctly conclude from such evidence that

$$P(\text{Good}|\text{Large}) = 0, \quad P(\text{Poisonous}|\text{Large}) = 0.01, \quad P(\text{Bad Taste}|\text{Large}) = 0.99.$$

Another expert, noticing that the berry is red, will conclude, on the other hand, that

$$P(\text{Good}|\text{Red}) = 0.99, \quad P(\text{Poisonous}|\text{Red}) = 0.01, \quad P(\text{Bad Taste}|\text{Red}) = 0.$$

Clearly the evidential implications of these two separate observations are identical to the situation summarized in Table 2. Examination of Table 3, however, reveals that

$$P(\text{Poisonous}|\text{Red, Large}) = 1,$$

a correct solution that must be rationally expected from any reasoning method that purports to be valid.

The solution to the puzzle of the disagreeing experts lies on recognizing that there is, in fact, no disparity of opinion among them. Each is providing quantitative measures of likelihood with respect to *different* reference classes. The Dempster formula, should never be applied to pool partial information about the same probability distribution. Furthermore, as shown by a sensitivity analysis of the results of its application to the berries example, its usage in situations where there is considerable disparity between reference classes (as suggested by the large normalization factor) should be discouraged on the basis of practical rather than conceptual considerations.

6 On Complexity and Generality

The potential complexity of the belief-function approach to represent and manipulate interval constraints on a family of probability distributions has been often mentioned as a handicap of the evidential reasoning methodology. In spite of such misgivings, two major empirical observations have indicated that the approach is applicable to a wide variety of practical problems.

First, our experience shows that, notwithstanding criticisms based on unrealistic worst-case scenarios, the approach is computationally efficient. In particular, we have found that representation of belief functions in terms of mass functions results in a storage and manipulation scheme that is both economical and easy to understand. In addition, we have successfully implemented tools, such as summarization and coarsening operators, which may be effectively utilized to limit representational complexity.

Second, our current functional operators have been chosen to guarantee that the manipulation of evidential knowledge results also in knowledge that may be represented in the evidential framework (i.e., the operators are closed).

The lack of generality of the belief-function approach to represent general lower-upper probability constraints is well known [10]. Our reliance on the methodology is primarily the result of practical considerations: although we would prefer to manipulate more general constraints on probability values, compelling computational efficiency arguments force us to limit the scope of the problems considered to those capable of being at least approximately solved by a belief-function treatment.

Being, in general, partial toward interpretations of evidential structures that are fully compatible with probability theory, our current research is being directed toward the development of more general, yet efficient, representation and manipulation methods.

Our current concerns with the manipulation of conditional and dependent evidence (i.e., the evidential counterpart of conditional probabilities) show, for example, that, for some important problems, the results of evidential combination fall outside the scope of its representational capabilities. In our experience, these methodological limitations are more worri-

some than any of the supposedly paradoxical results arising from its misuse or its claimed lack of a decision-making apparatus.

Preliminary results [16] indicate, on the other hand, that the belief-function approach may be used to approximate the results of these evidential combination operations and that extended representation mechanisms [20] may yet be developed to treat more general evidential problems. This research also shows the basic errors inherent in criticisms that regard the belief-function approach as a fully developed methodology incapable of sustaining further enhancement and modification. Because it has been studied in depth for only 15 years, its technological status is that of a young discipline, being both capable of enhancement on its own and of combination with other approaches to produce more general tools for probabilistic reasoning. Far from proving that we have reached a technological plateau, our investigations indicate that much is yet to be gained from such a development and integration process.

7 Acknowledgment

The authors want to thank Aviv Bergman, Aaron Bobick, Piero Bonissone, Luis de Campos, Allan Chateaufneuf, Mike Clarke, Didier Dubois, Tom Garvey, Joe Halpern, Andrew Hanson, Y.T. Hsia, Jean-Yves Jaffray, Henry Kyburg, María Teresa Lamata, Abe Mamdani, Serafín Moral, Henri Prade, Alessandro Saffiotti, Glen Shafer, Prakash Shenoy, Philippe Smets, Marcus Spies, Nic Wilson, Llorenç Valverde, Leonard Wesley, John Yen, and Lotfi Zadeh for useful exchanges and insights.

This work was supported in part by the Defense Advanced Research Project Agency under contract N00039-88-C-0248, in part by the Air Force Office of Scientific Research under Contract No. F49620-89-K-0001, and in part by the United States Army Research Office under Contract No. DAAL03-89-K-0156. The views, opinions, and/or conclusions contained in this paper are those of the authors and should not be interpreted as representative of the official positions, decisions, or policies, either express or implied, of the supporting agencies or the United States Government.

References

- [1] R.A. Aleilunas. A new normative theory of probabilistic logic. In *Proceedings of the 7th Biennial Conference of the Canadian Society for Computational Studies of Intelligence (CSCSI-88)*, pp. 67–74, Edmonton, Alberta, 1988.
- [2] Y. Bar-Hillel and R. Falk. Some teasers concerning conditional probabilities. *Cognition*, 11: 109–122, 1982.
- [3] L.M. de Campos, M.T. Lamata, and S. Moral. The concept of conditional fuzzy measure. *International Journal of Intelligent Systems* (forthcoming).

- [4] A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Statistics*, 38:325–339, 1967.
- [5] P. Diaconis and S.L. Zabell. Some alternatives to Bayes’s Rule. In B. Grofman and O. Guillermo, editors, *Information Pooling and Group Decision Making*, pp. 25–38. JAI Press, Greenwich, Connecticut, 1986.
- [6] R. Fagin and J.Y. Halpern. Uncertainty, belief and probability. Research Report RP. 619/(60901), IBM Almaden Research Center, Almaden, California, 1989.
- [7] J. Halpern and R. Fagin. Updating beliefs versus combining beliefs. Unpublished report, 1989.
- [8] G.E. Hughes and M.J. Creswell. *An Introduction to Modal Logic*. New York: Methuen, 1972.
- [9] H.E. Kyburg. Jerry’s berries. Unpublished report, 1985.
- [10] H.E. Kyburg. Interval valued probabilities. In P. Smets, M.R.B. Clarke, C. Froidevaux and J. Campbell, editors, *Positions and Debates in Nonstandard Logics*, New York: Academic Press (forthcoming).
- [11] J. Lowrance, T. Strat and T. Garvey. A framework for evidential-reasoning systems. *Proc. National Conference on Artificial Intelligence*, 1986, pp. 896–903.
- [12] R. Moore. *Reasoning about Knowledge and Action*. Technical Note 191, Menlo Park, California: SRI International, 1980.
- [13] J. Pearl. On probability intervals. *International Journal of Approximate Reasoning*, 2:211–216, 1988.
- [14] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, California: Morgan Kaufmann, 1988.
- [15] J. Pearl. Reasoning with Belief Functions: An Analysis of Compatibility. *International Journal of Approximate Reasoning*, 4: 363–390, 1990.
- [16] E.H. Ruspini. Approximate Deduction in Single Evidential Bodies. *Proceedings 1986 AAAI Uncertainty Workshop*, Philadelphia, 1986.
- [17] E.H. Ruspini. *The Logical Foundations of Evidential Reasoning*. Technical Note No. 408, Artificial Intelligence Center, SRI International, Menlo Park, California, 1987.
- [18] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

- [19] P. Smets. Belief functions. In P. Smets, M.R.B. Clarke, C. Froidevaux and J. Campbell, editors, *Positions and Debates in Nonstandard Logics*, New York: Academic Press (forthcoming).
- [20] M. Spies. Combination of evidence with conditional objects and its application to cognitive modeling. To appear in I. Goodman and H. Nguyen ,editors, *Conditional Logic in Expert Systems*. North Holland, Amsterdam, 1990.
- [21] T.M. Strat. Decision analysis using belief functions. *International Journal of Approximate Reasoning*, 4: 391-418, 1990.
- [22] N. Wilson. *Justification, Computational Complexity, and Generalization of the Dempster Combination Formula*. Unpublished report, 1989.
- [23] N. Wilson. Private communication, 1990.
- [24] L.A. Zadeh. *On the validity of Dempster's rule of combination of evidence*. Memorandum No. UCB/ERL M79/24, Electronics Research Laboratory, University of California, Berkeley, California, 1979.