



# Evidence, knowledge and belief functions

Didier Dubois, Henri Prade

## ► To cite this version:

Didier Dubois, Henri Prade. Evidence, knowledge and belief functions. International Journal of Approximate Reasoning, 1992, 6 (3), pp.295–319. hal-04069114

**HAL Id: hal-04069114**

**<https://hal.science/hal-04069114>**

Submitted on 17 Apr 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Evidence, Knowledge, and Belief Functions

**Didier Dubois and Henri Prade**

*Institut de Recherche en Informatique de Toulouse – I.R.I.T.  
Université Paul Sabatier  
Toulouse Cedex, France*

---

## ABSTRACT

*This article tries to clarify some aspects of the theory of belief functions, especially with regard to its relevance as a model for incomplete knowledge. It is pointed out that the mathematical model of belief functions can be useful beyond a theory of evidence, for the purpose of handling imperfect statistical knowledge. Dempster's rule of conditioning is carefully examined and compared to upper and lower conditional probabilities. Although both notions are extensions of conditioning, they cannot serve the same purpose. The notion of focusing, as a change of reference class, is introduced and opposed to updating. Dempster's rule is good for updating, whereas the other form of conditioning expresses a focusing operation. In particular, the concept of focusing models the meaning of uncertain statements in a more natural way than updating. Finally, it is suggested that Dempster's rules of conditioning and combination can be justified by the Bayes rule itself. On the whole this article addresses most of the questions raised by Pearl in the 1990 special issue of the International Journal of Approximate Reasoning on belief functions and belief maintenance in artificial intelligence.*

**KEYWORDS:** *incomplete knowledge, evidential reasoning, conditioning, updating, focusing, probability, Bayes' rule, likelihood, upper and lower probabilities, Dempster's rule, belief function, possibility measure, three prisoners problem*

---

## INTRODUCTION

In the paper commented on in this special issue [1], Pearl gives a careful analysis of belief function theory from the point of view of artificial intelligence and Bayesian probability. He makes a number of claims about the inadequacy of belief functions for representing incomplete knowledge, updating beliefs, and pooling evidence. In this paper we try to show that these inadequacies are often due to an improper use of belief functions in dealing

---

*Address correspondence to Professor D. Dubois, I.R.I.T., Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France.*

with problems they were never meant to solve. More precisely we address the following issues:

- What are set functions that have the mathematical properties of belief functions able to model? We suggest that their use is certainly not restricted to the representation of subjective uncertain evidence.
- What becomes of probabilistic conditioning when it is applied to belief functions? We claim that it splits into two quite different notions, which we call updating and focusing; updating corresponds to Dempster's well-known rule of conditioning [2] used in belief function theory; focusing corresponds to upper and lower conditional probabilities advocated by Ruspini [3], De Campos et al. [4], and Fagin and Halpern [5], a concept actually introduced by Dempster himself [2]. It explains why Dempster's rule often gives results that are more precise than the upper and lower probability conditioning.
- How can we represent and handle generic knowledge with belief functions? We show that the focusing rule might address this problem. In particular, it solves the puzzle of the "spoiled sandwich paradox" [1].
- Does Bayes' rule of conditioning always give intuitive results in the belief updating process? We propose a new analysis of the three prisoners problem, and we claim that the generally accepted Bayesian solution is debatable, whereas the one obtained by Dempster's rule makes sense.
- Is Dempster's rule of combination the only reasonable way of pooling evidence? We claim that it applies only under some very specific assumptions about the reliability of the sources of evidence and that it also presupposes that the information supplied by each source can be revised, based on the contradictions existing with other sources.
- Are likelihood functions reasonable substitutes for belief functions in the combination process? We claim that it comes down to assuming that the belief functions involved in the combination process are consonant. In other words, likelihood functions are similar to Zadeh's possibility distributions.

---

## REPRESENTING UNCERTAINTY WITH BELIEF FUNCTIONS

---

Pearl [1] makes a very cogent distinction between knowledge and evidence. Knowledge is understood as "judgments about the general tendency of things to happen," whereas evidence refers to the description of a specific situation. In other words, knowledge summarizes the results of a lot of observations, things read in books, etc., often modeled in terms of if-then rules and refers to a set of situations actually encountered in the past or just hypothetical. Evidence describes one situation in its peculiarities.

There are several types of knowledge. Statistical knowledge corresponds to

the case when the set  $\Omega$  of considered situations is well defined and when a probability distribution over  $\Omega$  is available, so that  $\forall A, B \subseteq \Omega, P(B|A)$  is precisely known. This probability distribution basically comes from frequentist data. Only a small part of our knowledge takes this form. A weakest form of knowledge is incomplete statistical data, where  $\Omega$  can still be made precise but only bounds on  $P(B|A)$  for some subsets  $A, B \subseteq \Omega$  are available. In particular we may have no access to the actual probability distribution on  $\Omega$ , if any. Kyburg [6] insists that most of our knowledge takes this form. Still a weaker form of knowledge is when the set of situations  $\Omega$  itself is not well defined, and neither are the concerned subsets of situations that appear in the knowledge. The “birds fly” case is a good example of that kind of knowledge: The set of birds that support the sentence is ill defined (should we consider dead birds that did fly, birds that have been actually observed, bird species, etc.); as a consequence,  $P(B|A)$  is not well defined either, and all that can be said is that “birds fly, but there are exceptions.” The three above-mentioned types of knowledge can be respectively handled by probability theory, upper and lower probabilities, and nonmonotonic logic. The last two correspond to what Pearl calls “incomplete knowledge.” What about belief functions in that respect? We claim that the mathematics of belief functions can capture set-valued statistics and can serve as tools for approximate representation of incomplete statistical knowledge.

### Set-Valued Statistics

Any random experiment whose outcomes cannot be precisely observed lends itself to a treatment by belief functions. Let  $\{A_i | i = 1, n\}$  be the set of observations, where  $A_i \subseteq \Omega$ , and let  $m(A_i)$  be the frequency of making observation  $A_i$ . Making observation  $A_i$  means that some  $\omega \in A_i$  occurred but it has been impossible to locate it more precisely. Clearly,  $\{(A_i, m(A_i)) | i = 1, n\}$  is a (statistically induced) random subset of  $\Omega$  that is equivalent to a belief function.  $\text{Bel}(A) = \sum_{A_i \subseteq A} m(A_i)$  and  $\text{Pl}(A) = \sum_{A \cap A_i \neq \emptyset} m(A_i)$  can be viewed as upper and lower probabilities, since  $[\text{Bel}(A), \text{Pl}(A)]$  is the range of the probability  $P(A)$  that would have been obtained had the observations been precise. Set-valued statistics have been advocated as a way of defining fuzzy set membership functions, by considering  $\{\text{Pl}(\{\omega\}), \omega \in \Omega\}$  as membership grades (Wang [7]). Set-valued statistics can be very common in opinion polls where people can express partial indecisiveness:  $\Omega$  is the set of possible answers, and people are allowed to select a subset of possible answers among which they have not yet chosen (Dubois and Prade [8]).

### Approximations of Incomplete Statistical Knowledge

It is well known that belief and plausibility functions, viewed as upper and lower probabilities, are not the most general family thereof. Hence any subset

of probability measures cannot be represented by belief functions. Pearl [1] claims that “states of knowledge representable as belief functions are rather rare.” The example of set-valued statistics questions this claim. It can be further questioned by relaxing the meaning of the word “representable.” The interesting question is then, What families of upper and lower probabilities can be *approximately* represented by belief functions?

Let  $P_*$  be a lower probability function on  $\Omega$ . A belief function  $\text{Bel}$  will be called an outer approximation of  $P_*$  if and only if

$$\forall A \subseteq \Omega, \quad \text{Bel}(A) \leq P_*(A) \quad (1)$$

An inner approximation of  $P_*$  can be defined by reversing the inequality

$$\forall A \subseteq \Omega, \quad P_*(A) \leq \text{Bel}(A) \quad (2)$$

Clearly an outer approximation of  $P_*$  will be less specific than  $P_*$ . Indeed, let  $[P_*(A), P^*(A)]$  the probability interval for  $A$ , where  $P^*(A) = 1 - P_*(\bar{A})$ ,  $\bar{A}$  being the complement of  $A$ . (1) implies  $[P_*(A), P^*(A)] \subseteq [\text{Bel}(A), \text{Pl}(A)]$ . A *best* outer approximation of  $P_*$  is a belief function  $\text{Bel}_*$  that is an outer approximation of  $P_*$  such that no other outer approximation of  $P_*$  is an inner approximation of  $\text{Bel}_*$ . Best inner approximations of  $P_*$  can be defined likewise. We have introduced these notions of approximation (Dubois and Prade [9]) and thoroughly applied them to the problem of approximating general belief functions by means of consonant belief functions (Dubois and Prade [10]), equivalent to Zadeh's possibility measures [11]. We do not intend to solve this problem here. However, we can imagine ways of formulating the problem as an optimization problem. For instance, let  $\epsilon_A$  be a nonnegative variable such that  $\text{Bel}(A) = P_*(A) - \epsilon_A$ ; then a best outer approximation of  $P_*(A)$  can be found by finding  $\{m(A), A \subseteq \Omega\}$  and  $\epsilon_A; A \subseteq \Omega, A \neq \Omega\}$ , positive solution of the linear programming problem

$$\min \sum_{A \subseteq \Omega} \epsilon_A$$

Under the constraints

$$\sum_{B \subseteq A} m(B) = P_*(A) - \epsilon_A, \forall A \subseteq \Omega, A \neq \Omega,$$

$$\sum_{B \subseteq \Omega} m(B) = 1$$

The system of constraints is feasible because  $m(\Omega) = 1$ ,  $m(A) = 0$ , and  $\epsilon_A = P_*(A) \forall A \neq \Omega$  is a (very bad) solution to this system.

For instance, consider the example recalled by Pearl [1], after Dalkey, where  $\Omega = \{\omega_1, \omega_2, \omega_3\}$ , and  $\{P \mid 0 \leq P(\{\omega_i\}) \leq 1/2\}$ . An outer approxima-

tion can be found as follows. Consider the set of belief functions on  $\Omega$  such that  $m(\{\omega_1\}) = m(\{\omega_2\}) = m(\{\omega_3\}) = x$ ,  $m(\{\omega_i, \omega_j\}) = y$ ,  $\forall i \neq j$ , and  $m(\Omega) = z$ . We will restrict to this family for obvious reasons of symmetry. Let us write the constraints

$$P_*(\{\omega_i\}) = 0 = x, \quad \forall i = 1, 2, 3$$

$$P^*(\{\omega_i\}) = 1/2 + \epsilon = x + 2y + z$$

$$P^*(\Omega) = 1 = 3x + 3y + z$$

It is easy to get  $y = 1/2 - \epsilon$ ,  $z = 3\epsilon - 1/2$ , which are positive only if  $1/2 \geq \epsilon \geq 1/6$ . Minimizing  $\epsilon$  leads to let  $\epsilon = 1/6$ . Hence there is a best outer approximation of the set of upper and lower probabilities that expands the probability intervals of  $\{\omega_i\}$  from  $[0, 1/2]$  to  $[0, 2/3]$ .

Of course, by acting so, we lose precision in the representation, but insofar as the loss of precision is minimized, and not too large in absolute value, belief functions could serve as tools for representing more classes of upper/lower probability bounds than Pearl [1] suggests. The loss in precision is counter-balanced by a gain in representational simplicity, an argument advocated by Pearl himself when he proposes likelihood functions as a substitute for basic probability assignments when pooling evidence [1, Sec. 4].

Best outer and inner approximations can be also searched for by taking advantage of the notions of inclusion between random sets (Moral [12], Dubois and Prade [9], Yager [13]) and cardinality of a belief function,  $|\text{Bel}| = \sum |A| m(A)$ , and more generally information measures (Yager [14], Klir and Folger [15]) that are the counterparts or extensions of entropy in the belief function setting. In particular, minimizing  $\sum |A| m(A)$ , instead of  $\sum \epsilon_A$ , may also lead to interesting solutions because  $\sum |A| m(A)$  is a measure of nonspecificity of the belief function (in fact, the expected cardinality of the random set), which is maximum ( $= |\Omega|$ ) for the vacuous belief function and minimum ( $= 1$ ) for probability measures; in other words, this criterion tries to assign masses to the greatest possible subsets of  $\Omega$ .

---

## REPRESENTING EVIDENCE

---

Let us turn to evidence. Strangely enough, Pearl [1] does not mention *incomplete* evidence in his typology. Yet, belief functions, as envisaged by Shafer [16] and Smets [17] are aimed directly at modeling incomplete evidence, but certainly not incomplete knowledge. This may be why both disavow the interpretation of belief functions in terms of upper and lower probabilities, and why Shafer [18] points out the inadequacy of his theory for dealing with statistical knowledge.

All canonical examples for belief function theory, that is, the theory where  $\text{Bel}(A)$  indeed models a degree of belief and where Dempster's rule is taken for granted, are about detective stories, testimonies, etc., in which the particulars of a situation are considered more important than the similarity of this situation to prototypical situations about which general knowledge is available. A simple piece of evidence  $\mathcal{E}$  is of the form " $x \in E$ ," where  $x$  is an attribute of some ill-described object  $\omega$  and  $E$  is a subset of the range  $X$  of  $x$ . Both Shafer and Smets acknowledge subjective probability as a proper tool for assessing the confidence of the person receiving the piece of evidence, in the reliability of that piece of evidence; this subjective probability assessment takes place on the simple frame  $\{\text{reliable}, \text{not reliable}\}$ , and the probability  $P(\text{reliable})$  is allocated to  $E$  while  $1 - P(\text{reliable})$  is allocated to  $X$  because when " $x \in E$ " is unreliable, it tells nothing. Then  $\forall A \subseteq X$ ,  $\text{Bel}(A)$  is indeed, as Pearl points out, equal to the probability that the piece of evidence  $\mathcal{E}$  proves  $A$ , that is,  $\text{Bel}(A) = P(\mathcal{E} \text{ implies } A)$ . Viewed this way there is no point in searching for a subset of probability measures whose lower envelope is  $\text{Bel}$ .

What is questionable in this construct is the fact that everything relies on the postulate that probability theory is relevant for expressing subjective uncertainty, independently of any frequentist interpretation. In particular, the question of finding measurement-theoretic foundations for belief functions has often been eluded on the grounds that subjective probabilities can be assigned on the basis of thought frequentist experiments and that belief functions are just a chapter of this view of probability theory. For instance, Shafer [19] insists on the unity of probability, where long-run frequencies, fair odds, and warranted beliefs cannot be told apart. The most controversial part of belief function theory, as a mathematical theory of evidence (not knowledge!), is not its account of incomplete evidence; it is the postulate that unreliable evidence can be described by means of subjective probability estimates.

---

## UPDATING AND FOCUSING

---

A large part of Pearl's critique [1, 20] of belief functions deals with Dempster's rule of conditioning and its counterintuitive behavior in celebrated examples such as the three prisoners problem. Pearl [1] also considers upper and lower probability conditioning and points out its deficiencies as an updating rule. Here we argue that part of the confusion that pervades the issue of using Dempster's rule of conditioning versus other rules has to do with the fact that in probability theory, Bayes' rule of conditioning serves two distinct purposes that no longer correspond to the same mathematical operation in the setting of upper and lower probabilities.

Consider the following two situations.

1. A die has been thrown 1 million times, and  $P(i)$  is the probability that

facet  $i$  comes out of a throw.  $P(i)$  is, of course, frequency-based. Consider the question of how often  $i$  occurs among odd facets. To do this we must, of course, compute  $P(i|\text{odd})$  using Bayes' rule.

2. A die has been thrown by a player, but he has not seen the result yet. He has ideas about the most probable outcomes in the form of a probability distribution  $P(i)$ ,  $i = 1, n$ , that expresses his subjective belief about the occurrence of each facet. Now before the official result is advertised, a friend tells him that the result is an odd number. How does the player update his belief? By computing again  $P(i|\text{odd})$ , the Bayesians say.

It should be clear to the reader that the two situations are quite different. In the first situation  $P(\text{even}) \neq 0$ , because in the 1 million throws, some results have been even. Moreover, no updating takes place. We always consider  $P$  the probability distribution to be used. We simply change the reference class of  $i$  to compute  $P(i|\text{odd})$ . That is what can be called *focusing*; we are answering the general knowledge question, How often does  $i$  occur when the outcome is odd with this die?

In the second situation, the player is computing a new probability distribution  $P'$  over the whole set  $\{1, 2, 3, 4, 5, 6\}$  by forcing  $P'(\text{even}) = 0$  because of the piece of evidence he received from his friend.  $P'$  has little to do with  $P$  except that it is the result of updating  $P$ . Moreover, the prior  $P$  does not necessarily derive from reading the results of the frequentist experiment. As a consequence, putting  $P'(i) = P(i|\text{odd})$  using Bayes' rule can be challenged. Other updating rules can be thought of, namely any rule that reassigns the mass  $P(\text{even})$  to odd  $i$ 's is a priori allowed.  $P'(i) = P(i|\text{odd})$  can be justified in the framework of probability kinematics (Domotor [21]) by means of a minimal change principle based on cross-entropy. Bayes' rule is also the only possible updating rule that does not change the degrees of probability in relative value; that is,  $\forall A, B, P'(A)/P'(B) = P(A)/P(B)$  for  $A, B$  such that  $A \cap C \neq \emptyset$ ,  $B \cap C \neq \emptyset$ , where  $C$  is the piece of evidence that leads to assuming  $P'(C) = 1$ . But Lewis's rule of imaging [22] is another updating rule that reassigns the mass of impossible outcomes to possible ones, based on proximity notions; namely,  $P(\bar{C})$  is allotted to  $\omega \in C$ , which is as close as possible to  $\bar{C}$ .

When it comes to the case of upper and lower probabilities, the two operations, focusing and updating, may no longer coincide. Consider a belief function defined on a frame  $X$  that contains the possible values of attribute  $x$  of members of a population  $\Omega$ . Let us adopt a purely frequentist point of view for clarity;  $m(E)$  for  $E \subseteq X$  is the proportion of members of  $\Omega$  for which all we know is  $\mathcal{E}: x \in E$ .

Let us consider the focusing problem. Let  $B \subseteq X$  be a subset of  $X$ , and we ask the question, What is the proportion of the population  $\Omega$  that belong to  $A \subseteq X$  among those that belong to  $B$ ? The answer may not be precise;



indeed, all we know about  $\omega$  is that  $x(\omega) \in E$ , and we must decide whether or not  $x(\omega) \in B$ . Let  $P(B|E)$  be the proportion of  $\omega$  such that  $x(\omega) \in B$  given evidence  $\mathcal{E}$  for  $\omega$ . There are three situations (De Campos et al. [4]):

$$E \subseteq B, \quad \text{then } P(B|E) = 1$$

$$E \cap B = \emptyset, \quad \text{then } P(B|E) = 0$$

$$E \cap B \neq \emptyset \text{ and } E \cap \bar{B} \neq \emptyset, \quad \text{then } P(B|E) \text{ is unknown}$$

Assume that  $P(B|E)$  is known; then the answer to the question is

$$\forall A \subseteq X, \text{Bel}_{B, P(B|\cdot)}(A) = \sum_{F \subseteq A} \frac{\sum_{E: E \cap B = F} m(E) \cdot P(B|E)}{\sum_E m(E) \cdot P(B|E)} \quad (3)$$

It is clearly a belief function that depends upon  $P(B|E)$ ,  $\forall E \subseteq X$ . Since these quantities are unknown when  $E \cap B \neq \emptyset$  and  $E \cap \bar{B} \neq \emptyset$ , we can only compute

$$\text{Bel}_B(A) = \inf\{\text{Bel}_{B, P(B|\cdot)}(A) \mid P(B|E) \in [0, 1], \\ E \cap B \neq \emptyset, E \cap \bar{B} \neq \emptyset\}$$

As proved by De Campos et al. [4],

$$\text{Bel}_B(A) = \inf\left\{\frac{P(A \cap B)}{P(B)} \mid P(A) \geq \text{Bel}(A), \forall A \subseteq X\right\} \\ = \frac{\text{Bel}(A \cap B)}{\text{Bel}(A \cap B) + \text{Pl}(\bar{A} \cap B)} \quad (4)$$

that is, the conditioning rule already proposed by Dempster [2], Ruspini [3] and that preserve belief functions as proved by Fagin and Halpern [5] (but also Jaffray [23]).

Now let us assume that instead of focusing on the reference class of members of  $\Omega$  such that  $x(\omega) \in B$ , we get a piece of evidence that claims that *all* members of  $\Omega$  satisfy the condition  $x(\omega) \in B$ . Then we know that if  $x(\omega) \in E$ , and  $E \cap B \neq \emptyset$ , we are entitled to let  $P(B|E) = 1$ . In other words, the mass  $m(E)$  can be allotted to  $E \cap B$ . Then (3) becomes

$$\text{Bel}(A|B) = \sum_{F \subseteq A} \frac{\sum_{E: E \cap B = F} m(E)}{\sum_{E: E \cap B \neq \emptyset} m(E)} \\ = 1 - \frac{\text{Pl}(\bar{A} \cap B)}{\text{Pl}(B)} \quad (\text{Shafer [16]}) \quad (5)$$

which is Dempster's rule of conditioning. Again the normalization factor that reallocates  $\sum_{E \cap B = \emptyset} m(E)$  can be challenged in the scope of updating. Indeed,  $\text{Bel}(A | B)$  provides an answer to the question, What is the proportion of  $\omega$ 's with  $x(\omega) \in A$ , among those with  $x(\omega) \in B$ , given that  $x(\omega) \in B$  as soon as we know  $x(\omega) \in E$ , and  $E \cap B \neq \emptyset$ ? In other words, it is focusing with an optimistic assumption on  $P(B | E)$ . No surprise then if the interval  $[\text{Bel}(A | B), \text{Pl}(A | B)]$  is more narrow than  $[\text{Bel}_B(A), \text{Pl}_B(A)]$ . There is no "mystery of the vanished interval" (Pearl [1]).

If the piece of evidence is indeed that  $x(\omega) \in B, \forall \omega$ , then it is contradictory to the fact that  $\exists \omega, x(\omega) \in E$  and  $E \cap B = \emptyset$ . Hence, as an updating rule in the frequentist framework, (5) makes sense only if  $\sum_{E: E \cap B \neq \emptyset} m(E) = 1$ , that is, the piece of evidence  $B$  does not contradict the set-valued statistics on  $\Omega$ . In that case, updating is restricted to a set-theoretic operation (changing  $E$  into  $E \cap B, \forall E$ ) and does not touch the masses.

But this is not, of course, the regular setting for belief functions. In belief function theory á la Shafer,  $\Omega$  is a set of answers to a question (e.g., is the evidence reliable or not?),  $X$  is a set of answers to another question, and there is a compatibility relation that maps  $\omega \in \Omega$  to some subset  $E = x(\omega)$  of  $X$  of answers compatible with  $\omega$ . There is a subjective probability distribution  $p$  on  $\Omega$ , and  $m(E) = \sum_{x(\omega)=E} p(\omega)$ . Given a piece of evidence  $B \subseteq X$  that rules out answers  $x_i \notin B$ , the problem is to update  $p$  on  $\Omega$ ; that is, define  $p'$  such that

- $p'(\omega) = 0$  if  $x(\omega) \cap B = \emptyset$  (when only  $B$  is a possible answer,  $\omega$  becomes impossible).
- All masses given to  $E$  are given to  $E \cap B$  if  $E \cap B \neq \emptyset$ .

Adopting the Bayesian updating rule on  $\Omega$ , we have

$$p'(\omega) = p(\omega) / P(B^*)$$

where  $B^* = \{u \mid x(\omega) \cap B \neq \emptyset\}$  is the set of possible answers in  $\Omega$  after hearing that  $\bar{B}$  is impossible in  $X$ . Translated back to  $X$ , Bayes' rule justifies Dempster's rule of conditioning. So if we reject Dempster's rule of conditioning on  $X$ , we must reject Bayes' rule on  $\Omega$ .

However, it is clear that in that setting the updating problem and the focusing problem no longer coincide; that is, if one has no idea of the probability that  $B$  contains the answer in  $X$ , given that  $E$  contains the answer, it is not possible to compute the relative probability of  $\omega$  among the  $\omega'$  such that  $x(\omega') \in B$ , because we no longer know the subset  $\{\omega \mid x(\omega) \in B\}$ . All we can say on this relative probability is that it lies in the interval  $[p(\omega)/P(B^*), p(\omega)/P(B_*)]$  as long as  $\omega \in B_*$ , where  $B_* = \{\omega \mid x(\omega) \subseteq B\}$ . Again, focusing and updating behave quite differently. The focusing rule (4) in  $X$  corresponds to conditioning a probability by means of an imprecisely bounded set in  $\Omega$  (i.e.,  $B_* \subseteq B \subseteq B^*$ ).

---

## REPRESENTING INCOMPLETE KNOWLEDGE IN THE SETTING OF BELIEF FUNCTIONS

---

In his paper Pearl [1] displays all kinds of paradoxes of belief functions used to express rules with exceptions. He rightly points out the cause of the misbehavior, namely, the use of material implication for expressing uncertain if-then statements. An uncertain rule “if  $A$  then  $B(\alpha)$ ” is viewed as  $\text{Bel}(\bar{A} \cup B) = \alpha$ , which suggests as a basic assignment representing the rule

$$m(\bar{A} \cup B) = \alpha$$

$$m(\Omega) = 1 - \alpha, \quad \text{if } A, B \subseteq \Omega$$

It is the least specific belief function compatible with the constraint  $\text{Bel}(\bar{A} \cup B) = \alpha$ . We have suggested this mode of representation in the past (Chatalic et al. [24]) and criticized it with examples similar to Pearl’s in Dubois and Prade [25].

Smets’ view [17] of this problem relies on Dempster’s rule of conditioning, and he interprets the uncertain rule as  $\text{Bel}(B | A) = \alpha$ , that is,  $\text{Pl}(A \cap \bar{B}) = (1 - \alpha) \text{Pl}(A)$ . The principle of minimal specificity makes us notice that since nothing impinges upon  $A$ , whether  $A$  holds or not must remain free so we must let  $\text{Pl}(A) = 1$ ;  $\text{Pl}(A \cap \bar{B}) = 1 - \alpha$  follows, which is clearly equivalent to  $\text{Bel}(\bar{A} \cup B) = \alpha$ . That is, as Smets [17] puts it, material implication and conditioning are reconciled in the framework of belief functions. However, this reconciliation exacts a high price: unnatural chaining of rules, unnatural contraposition, counterintuitive results when reasoning by case, as shown by Pearl [1]. We have pointed out other counterintuitive results (Dubois and Prade [25]) with “birds fly” types of examples.

What is questionable with the above model of uncertain rules is not related to the setting of belief functions itself as Pearl would have us believe. The problem lies in the use of an updating operation in the representation of uncertain rules. “Birds almost certainly fly” does not mean “if all animals were birds they would almost certainly fly.” It rather means “if we focus our attention on the class of animals called birds, then most animals in this class fly.” This suggests that a belief function model for the uncertain rule “if  $A$  then  $B(\alpha)$ ” would rather be  $\text{Bel}_A(B) = \alpha$  using a focusing operation rather than an updating operation. Note that if one is reluctant to use this rule because of its upper-lower probability flavor, De Campos et al. [4] have shown that  $\text{Bel}_A$  is a lower belief function as much as a lower probability, so there is no heresy in using the focusing rule as long as it is not presented as an updating machinery. Anyway, Pearl [1] himself indicates how poorly the focusing rule behaves in terms of updating. It is not because it is a bad conditioning rule. It is just because it is *not* an updating rule (for instance, there is no point in

iterating the use of the rule for evidence accumulation; focusing is not a sequential process, in contrast to updating, and does not deal with integration of evidence).

In the following, let the rule “if  $A$  then  $B(\alpha)$ ” be represented, when  $\alpha > 0$ , as

$$\text{Bel}_A(B) = \frac{\text{Bel}(A \cap B)}{\text{Bel}(A \cap B) + \text{Pl}(A \cap \bar{B})} = \alpha$$

This is equivalent to letting  $\alpha \text{Pl}(A \cap \bar{B}) = (1 - \alpha) \text{Bel}(A \cap B)$  and  $\text{Pl}(A \cap \bar{B}) + \text{Bel}(A \cap B) \neq 0$ . It is interesting to exhibit an example of least specific belief function induced by this constraint. Note that some mass needs to be allotted to  $A \cap B$  as soon as  $\text{Pl}(A \cap \bar{B}) > 0$ , that is,  $\alpha \neq 1$ . The remaining mass can be allotted to the greatest possible subset of  $\Omega$  that intersects  $A \cap \bar{B}$ , and this is  $\Omega$  itself. It is easy to see that if we let  $x = m(A \cap B)$ ,  $y = m(\Omega)$ , and we put no other mass anywhere,  $x$  and  $y$  are solutions of

$$(1 - \alpha)x = \alpha y; \quad x + y = 1$$

which gives  $m(A \cap B) = \alpha$ ,  $m(\Omega) = 1 - \alpha$ , if  $\alpha \neq 1$ . When  $\alpha = 1$ ,  $\text{Pl}(A \cap \bar{B}) = 0$ , which forces  $m(\bar{A} \cup B) = 1$ ; that is, we recover the material implication. Note that here the uncertain if-then rule is modeled by material implication only if  $\alpha = 1$ . That is, it is a sure rule. And there is a discontinuity in meaning when  $\alpha = 1$  with respect to  $\alpha < 1$ . This is very much what happens when the uncertain rule is modeled by means of a conditional probability,  $P(B|A) \geq \alpha$ , as indicated by Pearl [20].

With this model of belief function based if-then rules, all paradoxes pointed out by Pearl [1] vanish:

- (i) CHAINING. Assume two rules “if  $A$  then  $B(\alpha)$ ” and “if  $B$  then  $C(\beta)$ ”,  $\alpha, \beta < 1$ , and let us prove that  $\text{Bel}_A(C) = 0$ . This boils down to solving the optimization problem

$$\text{Minimize} \quad \frac{\text{Bel}(C \cap A)}{\text{Bel}(C \cap A) + \text{Pl}(\bar{C} \cap A)}$$

Under the constraints

$$(1 - \alpha) \text{Bel}(A \cap B) = \alpha \text{Pl}(A \cap \bar{B}) > 0$$

$$(1 - \beta) \text{Bel}(B \cap C) = \beta \text{Pl}(B \cap \bar{C}) > 0$$

We have to prove that the minimum is attained for a belief function such that  $\text{Bel}(C \cap A) = 0$ . To see that it is true, just consider the belief

function

$$m(A \cap \bar{B} \cap \bar{C}) = k(1 - \alpha); \quad m(\bar{A} \cap B \cap \bar{C}) = (1 - k)(1 - \beta)$$

$$m(A \cap B \cap \bar{C}) = k\alpha; \quad m(\bar{A} \cap B \cap C) = (1 - k)\beta$$

with  $0 < k < 1$ . It is easy to check that

$$\text{Bel}(A \cap B) = k\alpha, \quad \text{Pl}(A \cap \bar{B}) = k(1 - \alpha)$$

$$\text{Bel}(B \cap C) = (1 - k)\beta, \quad \text{Pl}(B \cap \bar{C}) = (1 - k)(1 - \beta)$$

$$\text{Bel}(A \cap C) = 0$$

Hence from “students are adults” and “adults are employees” viewed as uncertain rules, there is no belief induced about “students are employees.”

- (ii) **CONTRAPOSITIVE LAW.** It is obvious to check that generally “if  $A$  then  $B(\alpha)$ ” is not equivalent to “if not  $B$  then not  $A(\alpha)$ .” Indeed,

$$\text{Bel}_A(B) \neq \frac{\text{Bel}(\bar{A} \cap \bar{B})}{\text{Bel}(\bar{A} \cap \bar{B}) + \text{Pl}(A \cap \bar{B})}$$

Again, we may have  $\text{Bel}_A(B) > 0$  and  $\text{Bel}_{\bar{B}}(\bar{A}) = 0$  simultaneously.

- (iii) **REASONING BY CASES.** The following result can be proved.

**PROPOSITION** *If  $\text{Bel}_A(B) \geq \alpha$  and  $\text{Bel}_{\bar{A}}(B) \geq \beta$ , then  $\text{Bel}(B) \geq \min(\alpha, \beta)$ .*

**Proof** Assume  $\alpha \geq \beta$  without loss of generality. Let us solve the optimization problem Minimize

$$\text{Minimize } \text{Bel}(B)$$

Under the constraints

$$\alpha \text{Pl}(A \cap \bar{B}) = (1 - \alpha) \text{Bel}(A \cap B)$$

$$\beta \text{Pl}(\bar{A} \cap \bar{B}) = (1 - \beta) \text{Bel}(\bar{A} \cap B)$$

Let  $x = m(A \cap B)$ ,  $y = m(\bar{A} \cap B)$ ,  $u = m(B)$ ,  $v = m(\bar{A}) + m(\bar{A} \cap \bar{B}) + m(\bar{A} \Delta B) + m(\bar{A} \cup B)$ , where  $\Delta$  denotes the symmetric difference  $w = m(A) + m(A \cap \bar{B}) + m(A \Delta B) + m(A \cup B)$ . Then we have to solve a linear programming problem:

$$\text{Minimize } x + y + u$$

Under the constraints

$$\begin{aligned}\alpha(1 - y - x - u - v) &= (1 - \alpha)x \\ \beta(1 - x - y - u - w) &= (1 - \beta)y \\ x + y + u + v + w &\leq 1; \quad x, y, u, v, w \geq 0\end{aligned}$$

Substituting

$$v = 1 - u - y - x/\alpha; \quad w = 1 - u - x - y/\beta$$

leads to the constraints

$$u + y + x/\alpha \leq 1 \quad (6)$$

$$u + x + y/\beta \leq 1 \quad (7)$$

$$u + x/\alpha + y/\beta \geq 1 \quad (8)$$

It is obvious that the solution to the minimization problem saturates (8), and in that case (6) and (7) are satisfied, so we are led to

$$\text{Minimize } x + y + u$$

under the constraint

$$u + x/\alpha + y/\beta = 1$$

and the optimal solution is attained for  $y = \beta$ ,  $u = x = 0$  since  $\beta \leq \alpha$ . Hence,  $\text{Bel}(B) \geq \beta = \min(\alpha, \beta)$ . ■

Note that the converse property  $\text{Bel}(B) \leq \max(\alpha, \beta)$  does not hold. For instance if  $\alpha = 0.8$ ,  $\beta = 0.5$ , a solution that satisfies the constraints is  $x = 4/7$ ,  $y = u = 1/7$ ,  $v = w = 0$ . Then  $\text{Bel}(B) = 6/7 > 0.8$ . Q.E.D.

To use Pearl's image [1], the focusing rule applied to the representation of generic knowledge "unspoils the sandwich" because we obtain

$$\begin{aligned}\min[\text{Bel}_A(B), \text{Bel}_{\bar{A}}(B)] &\leq \text{Bel}(B) \leq \text{Pl}(B) \\ &\leq \max[\text{Pl}_A(B), \text{Pl}_{\bar{A}}(B)].\end{aligned}$$

When using Dempster's rule it is not surprising that the above bracketing property does not hold, because  $\text{Bel}(B)$  is the degree of provability of  $B$  from the available evidence. As pointed out by Pearl himself, upon getting evidence  $A$  we might get a proof for  $B$  with some probability  $[\text{Bel}(B|A) > 0]$ , and upon getting evidence  $\bar{A}$ , we might get another proof for  $B$ , whereas none of these proofs may be available if no extra evidence comes in  $[\text{Bel}(B) = 0]$ . But again an uncertain rule "if  $A$  then  $B(\alpha)$ " does not mean that if  $A$  is true for all objects then I can prove  $B$  with some probability but rather specifies the probability that  $B$  is true in the reference class  $A$ .

## A LOGICAL ANALYSIS OF THE THREE PRISONERS PARADOX AND ITS CONSEQUENCES

In his paper [1] as in his book [20], Pearl tries to convince his readers that Dempster's rule of conditioning is not a good updating rule compared to Bayes' rule even when some of the involved probabilities are not available. He uses the famous three prisoners example. Let us try to refute this claim. More specifically we shall argue the following.

- A purely logical solution to the problem is possible and gives the expected result.
- The solution usually put forward by Bayesian probability advocates is debatable. In particular, we are in a situation when reasonable prior probabilities cannot go along with updating via Bayes' rule.
- The solution given by Dempster's rule is alright and includes the logical solution as a special case.

Let us first recall the statement of the problem: The prisoners are  $A_1, A_2, A_3$ ; one of them will be executed, but prisoner  $A_1$  does not know who. He asks the guard to tell him one of  $A_2$  and  $A_3$  who will be saved, arguing that if he,  $A_1$ , is to be executed then this information is of no value anyway, since one of  $A_2$  and  $A_3$  will be saved. For some reason, the guard, who knows the result of the trial, tells him that  $A_2$  is not going to be executed. Should  $A_1$ 's opinion about his fate be modified?

Before going further we have to make it clear that the frame in which the problem is expressed contains six possibilities, which can be represented by

|        | $SA_2$   | $SA_3$   |
|--------|----------|----------|
| $EA_1$ | $A_{12}$ | $A_{13}$ |
| $EA_2$ | $A_{22}$ | $A_{23}$ |
| $EA_3$ | $A_{32}$ | $A_{33}$ |

where  $EA_i = "A_i \text{ will be executed}"$ ,  $SA_i = "guard \text{ says } A_i \text{ will be saved}"$  and  $A_{ij} = EA_i \cap SA_j$ ,  $i = 1, 2, 3$ ;  $j = 1, 2$ . Since the guard knows who will be executed and does not lie,  $A_{22}$  and  $A_{33}$  are impossible situations.

### Logical Analysis

The problem can be modeled by the following logical reasoning steps. Since  $SA_2$  is true,  $EA_2$  is false, and  $EA_1 \vee EA_3$  is true. Does  $A_1$  know more than before? Yes, he knows that  $A_2$  will be saved, so the uncertainty is between  $EA_1$  and  $EA_3$  only. Does  $A_1$  have a greater chance of being executed than the

other prisoner? At the beginning,  $A_1$  completely ignores this; all he knows is that one of  $A_1, A_2, A_3$  will be executed, which can be modeled by  $EA_1 \vee EA_2 \vee EA_3$ , and  $\forall i \neq j, \neg (EA_i \wedge EA_j)$  in logical terms. After the guard's reply,  $A_1$  remains in the same state of ignorance only with respect to  $A_3$  versus himself. In other words, the updating can be modeled as going from a lack of information on  $\{EA_1, EA_2, EA_3\}$  to a lack of information on  $\{EA_1, EA_3\}$ . Obviously, the situation has improved despite the residual ignorance. It is also obvious that this framework gives a reasonable solution to the problem that nobody can question.

Let us examine other solutions to this problem, as given by a Bayesian or belief function practitioner. A solution will be considered correct insofar as it is consistent with the above logical solution.

### Bayesian Analysis

Let us consider the Bayesian analysis that Pearl [20] considers the good one. We shall adopt the Bayesian representation of ignorance by means of equiprobability. The situation before the guard's answer is described as follows by Pearl:

$$P(EA_1) = P(EA_2) = P(EA_3) = 1/3 \quad \text{(lack of information a priori, expressed in Bayesian terms)}$$

$$P(A_{22}) = P(A_{33}) = 0 \quad \text{(the guard does not lie)}$$

$$P(SA_2 | EA_1) = P(SA_3 | EA_1) = 1/2 \quad \text{(the guard flips a fair coin when he knows } A_1 \text{ will be executed)}$$

These constraints completely determine the four remaining probabilities for  $A_{ij}$ :

$$P(A_{12}) = P(A_{13}) = 1/6; \quad P(A_{23}) = P(A_{32}) = 1/3$$

then

$$P(EA_1 | SA_2) = \frac{P(SA_2 | EA_1) \cdot P(EA_1)}{P(SA_2)} = \frac{1/2 \cdot 1/3}{1/6 + 1/3} = 1/3$$

This is considered the correct answer by Pearl because the probability that  $A_1$  will be executed has not changed. However, this result is not consistent



with the logical solution. Indeed,  $P(EA_3)$  has drastically changed, because  $P(EA_3|SA_2) = 2/3 = 2P(EA_3)$  since  $P(EA_1|SA_2) + P(EA_3|SA_2) = 1$ . This is perfectly unnatural. Why would  $A_1$  conclude that  $A_3$  has twice as many chances as himself to be executed given that the guard said  $A_2$  was to be saved? The expected solution, ignorance on  $\{EA_1, EA_3\}$ , cannot be expressed except by  $P(EA_1|SA_2) = P(EA_3|SA_2)$ , hence  $P(EA_1|SA_2) = 1/2$ . Indeed, it is not relevant to interpret an increase in probability as an increase of relative belief in some outcome when the set of outcomes itself has changed. So we agree with Pearl that in some sense  $A_1$ 's uncertainty should remain the same through the updating. But this stability is better expressed by a transition from  $\{1/3, 1/3, 1/3\}$  to  $\{1/2, 1/2\}$  (the incorrect solution following Pearl [20]) than by the transition from  $\{1/3, 1/3, 1/3\}$  to  $\{1/3, 2/3\}$ .

In this example, if we admit that the initial state of uncertainty is given by  $P(EA_i) = 1/3, \forall i$  what goes wrong is Bayes' rule. Indeed,  $P(EA_1|SA_2) = 1/3$  does not represent the result of an updating step when  $SA_2$  is known to be true.  $P(EA_1|SA_2) = 1/3$  evaluates, in frequentist terms, the proportion of times  $A_1$  would be executed among the cases when the guard replies  $SA_2$ , given that the guard flips a fair coin to do it. To make some sense out of this proportion, assume that  $A_1$  is a black person,  $A_2$  is a white person, and  $A_3$  is an Asiatic person. And  $P(EA_1|SA_2)$  reports on the statistics of blacks executed in a multiracial jail when the guard told the black guy that the white guy will be free. If  $A_1$  is indeed a black guy who has access to this general knowledge, then maybe he is right to increase his confidence in his salvation when the guard says the white guy will be saved [since  $P(EA_3|SA_2) = 2P(EA_1|SA_2)$ ]. But then, it is difficult to explain to him that the guard's answer is irrelevant to his beliefs because  $P(EA_1|SA_2) = P(EA_1)$ !

Let us consider the problem as an updating problem, and not as a focusing problem as above. Let us denote by  $P'$  the probability distribution after the updating step; learning  $SA_2$  leads to

$$\begin{aligned} P'(SA_2) &= 1 \Rightarrow P'(A_{13}) = P'(A_{23}) = 0 = P'(EA_2) \\ P'(A_{12}) &= P'(A_{32}) = 1/2 \quad (\text{symmetry of ignorance}) \end{aligned}$$

Let us examine what the underlying updating rule is in that case. We have to reallocate  $P(A_{23}) = 1/3$  and  $P(A_{13}) = 1/6$  to  $EA_1$  and  $EA_3$ .  $P(EA_2) = 1/3$  can be symmetrically reassigned to  $EA_1 \wedge SA_2$  and  $EA_3 \wedge SA_2$ . However, it would be strange to equally share  $P(A_{13})$  between  $EA_1$  and  $EA_3$  because  $A_{13}$  is incompatible with  $EA_3$  but perfectly compatible with  $EA_1$ . Hence we are inclined to reallocate  $P(A_{13})$  to  $EA_1$ . This is very similar to Lewis's imaging:  $A_{12}$  is "closer" to  $A_{13}$  than  $A_{32}$  and receives its weight. Hence, we get

$$\begin{aligned} P'(EA_1) &= P(A_{12}) + P(A_{13}) + (1/2)P(A_{23}) \\ &= P(EA_1) + (1/2)P(EA_2) = 1/2 \\ P'(EA_3) &= P(A_{32}) + (1/2)P(A_{23}) = P(EA_3) + (1/2)P(EA_2) = 1/2 \end{aligned}$$

Interestingly enough, the symmetric repartition of the mass of the impossible event is not compatible with Bayes' rule, and we have found a case where Bayes' rule does not apply. It can be checked that  $P'(EA_i) = P(EA_i | \neg EA_2)$ ,  $\forall i$ . But, as rightly claimed by Pearl [20], it is not reasonable to use Bayes' conditioning upon  $\neg EA_2$  here because the piece of evidence is  $SA_2 \neq \neg EA_2$  (indeed,  $SA_2 = A_{12} \vee A_{32}$ ,  $\neg EA_2 = A_{12} \vee A_{13} \vee A_{32}$ )  $\cdot P(EA_i | \neg EA_2)$  gives the correct answer by chance. The good result [ $P'(EA_1) = 1/2$ ] obtains by an updating rule different from Bayes' rule, because here, equally sharing  $P(EA_2) + P(A_{13})$  among  $A_{12}$  and  $EA_3$  is not suitable. If we want to get the good result by a proper use of Bayes' rule, we need to change the assignment of the prior. Namely, if we know, because of the guard's agreeing to reply, that the world contains four states ( $A_{12}, A_{13}, A_{23}, A_{32}$ ), then our ignorance state should be expressed by  $P''(A_{12}) = P''(A_{13}) = P''(A_{23}) = P''(A_{32}) = 1/4$ , according to the Bayesian principle that says we must first enumerate the set of possible states of the world and afterwards assign prior probabilities. Then, of course,  $P(EA_1 | SA_2) = 1/2 = P(EA_3 | SA_2)$  using Bayes' rule. But one is led to accept as prior probabilities  $P''(EA_1) = 1/2$ ,  $P''(EA_2) = 1/4$ ,  $P''(EA_3) = 1/4$ , a strange prior on a three-element set, in the absence of information. One might argue that on the set  $\{EA_1, EO\}$ , where  $EO = EA_2 \vee EA_3$  means "another one will be executed," that this is a genuine noninformative prior. But as can be seen, in order to get the correct answer  $P'(EA_1) = P'(EA_2) = 1/2$ , we are bound to use probability theory in either a dubious way or an unusual way:

- If we adopt a natural prior on  $\{EA_1, EA_2, EA_3\}$ , then we cannot use Bayes' rule for updating upon  $SA_2$ .
- If we adopt a natural prior on  $\{EA_1, EA_2, EA_3\}$  and we insist on using Bayes' rule, then conditioning must be applied to  $\neg EA_2$ , that is, not the piece of evidence we have, which is not satisfactory.
- If we want to apply Bayes' rule correctly by conditioning upon the available evidence, and get the good result, we are forced to use a strange prior probability on  $\{EA_1, EA_2, EA_3\}$ .

### The Cautious Bayesian Solution

Here we no longer know  $P(SA_2 | EA_1) = \alpha$ . Carrying on the uninformative prior  $P(EA_1) = P(EA_2) = P(EA_3) = 1/3$ , we get

- (i)  $P(EA_1 | SA_2) = \alpha / (1 + \alpha) \in [0, 1/2]$  (Bayes on the right evidence)
- (ii)  $P(EA_1 | \neg EA_2) = 1/2$  (Bayes on the wrong evidence)

Solution (i) described by Diaconis and Zabell [26] has been considered the good one by Pearl [20]. But again, it implies a dissymmetry between

$P(EA_1 | SA_2)$  and  $P(EA_3 | SA_2)$  ( $\in [1/2, 1]$ ) that is not in accordance with the logical solution. If we use, instead of Bayes' rule, the alternative updating rule that, upon knowing  $SA_2$  is true, reallocates  $P(A_{13}) = (1 - \alpha)/3$  to  $EA_1$ , we get  $\forall \alpha$ ,  $P'(EA_1) = 1/2$ , that is, a result equal to solution (ii) that agrees with the logical solution. Note that this solution also obtains if, by virtue of symmetry, the constraint  $P(EA_3 | SA_2) = P(EA_1 | SA_2)$  is added to refine solution (i). Hence the vanishing of the interval  $[0, 1/2]$  (reduced to  $1/2$ ) can be defended even in the context of an upper and lower probability analysis of the case.

Besides, it may be strange to simultaneously express ignorance by the principle of insufficient reason using a uniform probability distribution on  $\{EA_1, EA_2, EA_3\}$  and an unknown probability on  $\{SA_2, SA_3\}$ . One should use an unknown probability on both. And then all we would get is that  $P(EA_i | SA_2) \in [0, 1]$ ,  $i = 1, 3$ ,  $P(EA_2 | SA_2) = 0$ , a solution that is exactly the logical solution.

### The Belief Function Based Solution

The belief function based solution is well known (Diaconis and Zabell [26]). Namely, from the Bayesian prior on  $\{EA_1, EA_2, EA_3\}$  one builds the mass assignment  $m(\{A_{12}, A_{13}\}) = m(\{A_{23}\}) = m(\{A_{32}\}) = 1/3$ ; then Dempster's rule of conditioning leads to changing the focal elements  $A$  into  $A \cap SA_2$  upon learning the piece of evidence from the guard. Since  $\{A_{12}, A_{13}\} \cap SA_2 = \{A_{12}\}$ ,  $A_{23} \cap SA_2 = \emptyset$ ,  $A_{32} \cap SA_2 = \{A_{32}\}$ , we get  $m(A_{12}) = 1/3 / (1/3 + 1/3) = 1/2 = m(A_{32})$ , the solution that coincides with intuition, that is, a form of total uncertainty on  $\{EA_1, EA_3\}$  just like the logical solution.

To conclude, the claim that the three prisoners problem furnishes a counterexample for Dempster's rule can be challenged. Indeed, it is Bayes's rule that yields results that are counterintuitive to the logical solution when a uniform prior is assumed on  $\{EA_1, EA_2, EA_3\}$  in the presence of the constraints  $A_{22} = A_{33} = \emptyset$ . In contrast, Dempster's rule yields a proper update. Actually, in order to reconcile Bayes's uniform priors on  $\{EA_1, EA_2, EA_3\}$  and the logical solution, one may proceed as follows:

- Assume  $P(EA_i) = 1/3$  and  $P(SA_2) = P(SA_3) = 1/2$  without requiring any further knowledge of the guard's behavior. In particular, he may be allowed to lie so that  $P(EA_i \cap SA_j) = 1/6$ ,  $\forall i, \forall j$ . This is the assumed prior probability.
- Consider the two pieces of evidence "the guard does not lie" modeled by  $A_{22} \vee A_{33}$  and "the guard says  $A_2$  is saved" modeled by  $SA_2$ , and calculate  $P(EA_i | SA_2 \wedge \bar{A}_{22} \wedge \bar{A}_{33})$ . Again we find  $1/2$  as posterior probabilities attached to  $EA_1$  and  $EA_3$ .

This way of solving the problem brings us close to Dempster's rule of combination.

---

## DEMPSTER'S RULE OF COMBINATION

---

If we examine how Shafer [18] and Shafer and Tversky [27] present Dempster's rule of combination, we find the following elements:

- Two sets  $\Omega$ ,  $\Omega'$  are given independently, with a priori probability measures  $P$  and  $P'$  on them. In the random codes example (Shafer and Tversky [27]),  $\Omega$  and  $\Omega'$  are sets of codes used by two different persons that encode a given message,  $P(\omega)$  and  $P'(\omega')$  being the probabilities that one person uses code  $\omega$  and that the other uses  $\omega'$ , respectively.
- Two multivalued mappings  $\Gamma: \Omega \rightarrow X$  and  $\Gamma': \Omega' \rightarrow X$  are given.  $\Gamma(\omega)$  is the set of elements of  $X$  that are compatible with  $\omega$ , and the same with  $\Gamma'$ . Hence  $(P, \Gamma)$  and  $(P', \Gamma')$  induce belief functions on  $X$ .
- Independence between  $P$  and  $P'$  forces the joint a priori probability on  $\Omega \times \Omega'$  to be  $P''(\omega, \omega') = P(\omega) \cdot P'(\omega')$ .
- An updating step takes place by which  $\Gamma$  and  $\Gamma'$  refine each other, that is,  $P(\omega) \cdot P(\omega')$  is allotted to  $\Gamma(\omega) \cap \Gamma'(\omega')$  when not empty, then redistributing the mass  $\sum \{P(\omega) \cdot P(\omega') \mid \Gamma(\omega) \cap \Gamma'(\omega') = \emptyset\}$  over to the other pairs  $(\omega, \omega')$ .

This leads to the now celebrated formula on  $X$ ,

$$\forall A \subseteq X, \quad m''(A) = (m \oplus m')(A) = \frac{1}{k} \sum_{B \cap C = A} m(B) \cdot m'(C) \quad (9)$$

with

$$m(B) = \sum \{p(\omega) \mid \Gamma(\omega) = B\}$$

$$m'(C) = \sum \{p'(\omega') \mid \Gamma'(\omega') = C\}$$

$$k = \sum \{m(B) \cdot m'(C) \mid B \cap C \neq \emptyset\}$$

In the code example,  $\Gamma$  and  $\Gamma'$  are decoding functions such that  $\Gamma(\omega)$  contains the actual message that was coded on  $\Omega$ , and  $\Gamma'(\omega')$  likewise on  $\Omega'$ . The updating step corresponds to the integration of the following piece of evidence: The same message was encoded by the two persons. Hence  $m''$  is the belief function whose underlying probability measure on  $\Omega \times \Omega'$  is  $P''(\cdot \mid E)$ , where  $E = \{(\omega, \omega') \mid \Gamma(\omega) \cap \Gamma'(\omega') \neq \emptyset\}$ .

The statement of the three prisoners problem given at the end of the previous section was exactly based on this model:  $\Omega = \{EA_1, EA_2, EA_3\}$ ,  $\Omega' =$

$\{SA_2, SA_3\}$ ,  $\Gamma(EA_i) = \{A_{i2}, A_{i3}\}$ ,  $i = 1, 2, 3$ ;  $\Gamma(SA_j) = \{A_{1j}, A_{2j}, A_{3j}\}$ ,  $j = 2, 3$ . The piece of evidence is "The guard is not a liar;" so it excludes  $\{(EA_2, SA_2); (EA_3, SA_3)\}$ . Dempster's rule then coincides with Bayes' rule and leads to  $P(A_{ij}) = 1/4$ ,  $\forall i, j$ , except  $P(A_{22}) = 0$ ,  $P(A_{33}) = 0$ , before a new piece of evidence ( $SA_2$ ) is entered.

Again we find the coherence between Bayes's rule of conditioning on  $\Omega \times \Omega'$  and Dempster's rule of combination on  $X$ . What is very important here is that if we project back  $m''$  over to  $\Omega$  and  $\Omega'$  we do not recover  $P$  and  $P'$ , owing to the updating step. It means that  $P$  and  $P'$  are elicited prior to knowing anything about the coherence between  $\Gamma(\omega)$  and  $\Gamma'(\omega')$ . Maybe this is the meaning of independence between belief functions to be combined by Dempster's rule. The requested coherence between  $\Gamma$  and  $\Gamma'$  must be taken as a piece of evidence.

If it were not the case,  $P$  and  $P'$  would already account for the fact that  $P''(\omega, \omega') = 0$  for some pairs  $(\omega, \omega')$  due to  $\Gamma(\omega) \cap \Gamma'(\omega') = \emptyset$ . And then the two basic assignments  $m$  and  $m'$  would no longer be stochastically independent; there should exist a joint mass function over  $X \times X$ , say  $\bar{m}$ , such that

$$m(A) = \sum_{B \subseteq X} \bar{m}(A \times B) \quad (10)$$

$$m'(B) = \sum_{A \subseteq X} \bar{m}(A \times B) \quad (11)$$

and the combination of  $m$  and  $m'$  that results from constructing  $\Gamma''(\omega, \omega') = \Gamma(\omega) \cap \Gamma'(\omega')$  leads to a formula different from Dempster's rule:

$$m''(A) = \sum_{B \cap C = A} \bar{m}(B \times C).$$

We suggested this type of combination earlier (Dubois and Prade [28]). In particular, if the constraints linking  $\Omega$  and  $\Omega'$  through the compatibility relation  $\Gamma$  and  $\Gamma'$  have been taken into account when specifying  $P$  and  $P'$ , the infeasibility of Eqs. (10), (11) linking the joint basic assignment and its marginals indicates a strong conflict between the sources of evidence. For instance, in the three prisoners example, if the uniform prior on  $\{EA_1, EA_2, EA_3\}$  is obtained while prisoner  $A_1$  knows that the guard is not a liar, then the joint assignment of probabilities on  $\{EA_1, EA_2, EA_3\} \times \{SA_2, SA_3\}$  must respect the constraints  $P(EA_1) = P(A_{12}) + P(A_{13}) = 1/3$  and  $P(A_{23}) = P(A_{32}) = 1/3$ , and we obtain Pearl's solution [20] to the puzzle.

Note that another assumption in the setting of Dempster's rule of combination can be challenged: the one that allows the refinement operation  $\Gamma''(\omega, \omega') = \Gamma(\omega) \cap \Gamma'(\omega')$  to be performed. It comes down to questioning whether the two persons in the random code experiment did encode the same message.

Failure of this assumption may be due to the fact that, for instance, one forgot to make sure that they dealt with the same message. Or maybe they did, but one of them distorted the message for some reason. Hence we may question the reliability of the sources of evidence as sensors (not only as transmitters).

If we take a weak assumption, namely, that one at least of the two persons encoded the good message, then from decoding  $\omega$  and  $\omega'$  we can conclude only that the good message  $x$  belongs to  $\Gamma(\omega) \cup \Gamma(\omega')$ .

This gives the following disjunctive counterpart to Dempster's rule:

$$m_{\cup}(A) = \sum_{A=B \cup C} m(A) \cdot m'(C) \quad (12)$$

We proposed this rule [9] in 1986, and it appears in Smets's Ph.D. thesis (Smets [29]) when he defines a counterpart to Bayes' rule of pooling evidence [i.e., computing  $P(\cdot | \mathcal{E}_1 \vee \mathcal{E}_2)$  from  $P(\cdot | \mathcal{E}_1)$  and  $P(\cdot | \mathcal{E}_2)$ ] under the assumption of conditional independence. Namely, if  $X$  and  $Y$  are two sets,  $\text{Bel}(\cdot | x)$  are belief functions on  $Y$  given  $x$ ,  $x \in X$ , then  $\text{Bel}(\cdot | \{x, x'\})$  is computed by applying the disjunctive rule on  $\text{Bel}(\cdot | x)$  and  $\text{Bel}(\cdot | x')$ .

A noticeable property of the disjunctive rule is that it corresponds to the product of the belief functions based on  $m$  and  $m'$ . This property, proved by Dubois and Prade [9], also appears in Berres [30]. Strangely enough, Shafer [18], in his overview article, criticizes the alternative rules we proposed [9] (namely the disjunctive rule) as generalizations of Dempster's rule lacking a rationale but finds Berres's proposal [30] of computing the product of belief functions interesting "in the context of discounting belief functions" and never mentions Smets's [29] pioneering contribution on this rule. However, from the above explanation and other discussions (Dubois and Prade [25, 31]), we do agree that the disjunctive rule makes sense for a mutual discounting of sources of evidence when only one of them is assumed to be a good sensor.

---

## LIKELIHOOD FUNCTIONS AS SUBSTITUTES FOR BELIEF FUNCTIONS

---

Pearl [1] points out that the plausibility function restricted to singletons can be viewed as a likelihood function. Given a body of evidence  $\mathcal{E}$  described by a belief function on  $\Omega$ , we can postulate the identity

$$\forall \omega \in \Omega, \quad P(\mathcal{E} | \omega) = \text{Pl}(\{\omega\}) = \sum_{A: \omega \in A} m(A) \quad (13)$$

Shafer calls  $\text{Pl}(\{\omega\})$  the contour function of  $m$ . Let us denote it by  $\mu$ . Pearl motivates his remark by noticing that when combining  $m$  with a probability function  $P$  that assigns mass  $p(\omega)$  to  $\omega$ , only this function is necessary to compute the result, that is,

$$(m \oplus p)(\omega) = kp(\omega) \cdot \mu(\omega) \quad (14)$$

And (14) is nothing but Bayes' formula again. This remark suggests several comments. First it is well known that the contour function of  $m \oplus m'$ , for any belief functions with basic assignments  $m$  and  $m'$ , is proportional to the product of contour functions of  $m$  and  $m'$ , that is (e.g., Shafer [32]),

$$\mu''(\omega) = k\mu(\omega) \cdot \mu'(\omega) \quad (15)$$

Interestingly enough, while Pearl [1] uses these properties as a reason for advocating probability theory as a sufficient tool for pooling evidence, the definition of a contour function has been used by others (e.g., Kampé de Fériet [33], Goodman and Nguyen [34], Dubois and Prade [35]) to bridge the gap between fuzzy set and probability theory, and (15) as a departure point to justify fuzzy set-theoretic operations from a random set-theoretic point of view (Goodman and Nguyen [34], Dubois and Prade [8]). Indeed  $\mu$  is formally a fuzzy set membership function (Zadeh [36]), and it can be viewed as a possibility distribution (Zadeh [11]) because  $\text{Pl}(\{\omega\}) = 0$  means that  $\omega$  is impossible and  $P'(\{\omega\}) = 1$  means that  $\omega$  is completely possible. The analogy between a fuzzy set membership function and a likelihood function, noted by Cheeseman [37], is not new and has been used for the definition of psychometric experiments with fuzzy sets (Hisdal [38]).

Instead of using these mathematical similarities as a reason for remaining within a regular Bayesian framework, one may use them as well to find new motivations for developing alternative uncertainty calculi in better agreement with probability theory (Dubois and Prade [39]). For instance, the discovery of the mathematical equivalence between likelihoods, plausibility of singletons, and possibility distributions leads to bridging the gap between Bayes' theorem and fuzzy set-theoretic operations (Dubois and Prade [40]). It suggests possibility theory as a good framework for handling likelihood functions. It is also consistent with the attempt to work with consonant approximations of belief functions (Dubois and Prade [10]). Indeed, the contents of a belief function are equivalent to the contents of its contour function if and only if it is consonant. Hence belief functions can be equivalently represented by likelihood functions only in the framework of possibility theory.

---

## CONCLUSION

---

In this paper it has been pointed out that the mathematical theory of evidence currently developed by Shafer and Smets has perhaps been misleading by interpreted as a theory of uncertain or incomplete generic knowledge by people who have tried to solve approximate reasoning problems in artificial intelligence (including ourselves). We have tried to suggest that the set functions commonly referred to as belief functions may serve for purposes other than

representing and pooling evidence, especially regarding set-valued statistics, and as approximations of sets of probability measures.

Two conditioning rules have been discussed, one that can serve for focusing on a reference class, the other for updating upon the arrival of a piece of evidence. The conditioning rule derived from sensitivity analysis on imprecise conditional probabilities never appears in evidence theory because it is not suitable for the integration of evidence. However, it has been suggested as a good candidate for the representation of uncertain if-then rules. On the other hand, it has been stressed that Dempster's rule of conditioning and combination are nothing but projections of Bayes' rule of conditioning through multi-valued mappings. Hence, as updating rules they should be accepted by probability theory proponents. It has also been suggested that the normalization step is acceptable insofar as the constraints that lead to such a normalization (i.e., the empty intersection of focal elements stemming from combined belief functions) can count as evidence as well. The existence of a disjunctive counterpart to Dempster's rule has been recalled and motivated.

We hope that this paper has contributed to a better understanding of the areas where belief functions can be applied, and that we have addressed some of the questions that Pearl [1] very legitimately asked in his position paper. Other issues would require more investigation—for instance the position, with respect to Dempster's rule, of Jeffrey's rule of updating in the presence of uncertain evidence discussed by Shafer [41] in the scope of belief functions and by Pearl [20] in the scope of Bayesian probability. This is a topic for further research.

---

## References

---

1. Pearl, J., Reasoning with belief functions: an analysis of compatibility, *Int. J. Approx. Reasoning* 4(5/6), 363–389, 1990.
2. Dempster, A. P., Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Stat.* 38, 325–339, 1967.
3. Ruspini, E. H., Approximate deduction in single evidential bodies, *Proceedings of the 2nd Workshop on Uncertainty in AI* (AAAI-86), Univ. Pennsylvania, Aug. 8–10, 1986, 215–222, 1986.
4. De Campos, L. M., Lamata, M. T., and Moral, S., The concept of conditional fuzzy measure, *Int. J. Intell. Syst.* 5(3), 237–246 (1990).
5. Fagin, R., and Halpern, J. Y., A new approach to updating beliefs, Research Rep. RJ 7222, IBM Research Division, San Jose, Calif., 1989.
6. Kyburg, H. Jr., Knowledge, in *Uncertainty in Artificial Intelligence*, Vol. 2 (J. F. Lemmer and L. N. Kanal, Eds.), Elsevier, Amsterdam, 1988, pp. 263–272.
7. Wang, P. Z., From the fuzzy statistics to the falling random subsets, in *Advances*



- in *Fuzzy Sets, Possibility and Applications* (P. P. Wang, Ed.), Plenum, New York, 1983, pp. 81–96.
8. Dubois, D., and Prade, H., Fuzzy sets, probability and measurement, *Eur. J. Op. Res.* **40**, 133–154, 1989.
  9. Dubois, D., and Prade, H., A set-theoretic view of belief functions—logical operations and approximation by fuzzy sets, *Int. J. Gen. Syst.* **12**, 193–226, 1986.
  10. Dubois, D., and Prade, H., Consonant approximations of belief functions, *Int. J. Approx. Reasoning* **4**(5/6), 419–449, 1990.
  11. Zadeh, L. A., Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets Syst.* **1**, 3–28, 1978.
  12. Moral, S., Información difusa—relaciones entre probabilidad y posibilidad, Tesis Doctoral, Univ. Granada, Spain, 1985.
  13. Yager, R. R., The entailment principle for Dempster–Shafer granules, *Int. J. Intell. Syst.* **1**, 247–262, 1986.
  14. Yager, R. R., Entropy and specificity in a mathematical theory of evidence, *Int. J. Gen. Syst.* **9**, 249–260, 1983.
  15. Klir, G., and Folger, T., *Fuzzy Sets, Uncertainty and Information*, Prentice-Hall, Englewood Cliffs, N.J., 1987.
  16. Shafer, G., *A Mathematical Theory of Evidence*, Princeton Univ. Press, Princeton, N.J., 1976.
  17. Smets, P., Belief functions, in *Non-Standard Logics for Automated Reasoning* (P. Smets, A. Mamdani, D. Dubois, and H. Prade, Eds.), Academic, New York, 1988, pp. 253–286.
  18. Shafer, G., Perspectives on the theory and practice of belief functions, *Int. J. Approx. Reasoning* **4**(5/6), 323–362, 1990.
  19. Shafer, G., The unicity of probability, in *Acting Under Uncertainty: Multidisciplinary Conceptions* (G. Von Furstenberg, Ed.), Kluwer, Boston, 1990, pp. 95–126.
  20. Pearl, J., *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Mateo, Calif., 1988.
  21. Domotor, Z., Probability kinematics—conditionals and entropy principles, *Synthese* **63**, 74–115, 1985.
  22. Lewis, D., Probabilities of conditionals and conditional probabilities, *Phil. Rev.* **85**, 297–315, 1976.
  23. Jaffray, J. Y., Bayesian updating and belief functions, *Proceedings of the 3rd Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems* (IPMU'90), Paris, July 1990, ENSTA, Paris, 449–451, 1990.
  24. Chatalic, P., Dubois, D., and Prade, H., An approach to approximate reasoning based on the Dempster rule of combination, *Int. J. Expert Syst.* **1**, 67–85, 1987.

25. Dubois, D., and Prade, H., Representation and combination of uncertainty with belief functions and possibility measures, *Comput. Intell.* **4**, 244–264, 1988.
26. Diaconis, P., and Zabell, S. L., Some alternatives to Bayes's rule, in *Information Pooling and Group Decision Making* (B. Grofman and O. Guillermo, Eds.), JAI Press, Greenwich, Conn., 1986, pp. 25–38.
27. Shafer, G., and Tversky, A., Languages and designs for probability judgment, *Cogn. Sci.* **9**, 309–339, 1985.
28. Dubois, D., and Prade, H., On the unicity of Dempster's rule of combination, *Int. J. Intell. Sys.* **1**, 133–142, 1986.
29. Smets, P., Un modèle mathématico-statistique simulant le processus du diagnostic médical, Doctoral Dissertation, Free Univ. of Brussels, Presses Universitaires de Bruxelles, 1978.
30. Berres, M., On a multiplication and a theory of integration for belief and plausibility functions, *J. Math. Anal. Appl.* **121**, 487–505, 1987.
31. Dubois, D., and Prade, H., Une approche ensembliste de la combinaison d'informations imprécises et incertaines, *Revue Intell. Artif.* **1**(4), 23–42, 1987.
32. Shafer, G., Belief functions and possibility measures, in *The Analysis of Fuzzy Information*, Vol. 1 (J. C. Bezdek, Ed.), CRC Press, Boca Raton, Fla., 1987, pp. 51–84.
33. Kampé de Fériet, J., Interpretation of membership functions of fuzzy sets in terms of plausibility and belief, in *Fuzzy Information and Decision Processes* (M. M. Gupta and E. Sanchez, Eds.), North-Holland, Amsterdam, 1982, pp. 93–98.
34. Goodman, I. R., and Nguyen, H. T., *Uncertainty Models for Knowledge-Based Systems*, North-Holland, Amsterdam, 1985.
35. Dubois, D., and Prade, H., On several representations of an uncertain body of evidence, in *Fuzzy Information and Decision Processes* (M. M. Gupta and E. Sanchez, Eds.), North-Holland, Amsterdam, 1982, pp. 167–181.
36. Zadeh, L. A., Fuzzy sets, *Inf. Control* **8**, 338–353, 1965.
37. Cheeseman, P., Probabilistic versus fuzzy reasoning, in *Uncertainty in Artificial Intelligence*, Vol. 1 (L. N. Kanal and J. F. Lemmer, Eds.), Elsevier, Amsterdam, 1986, pp. 85–102.
38. Hisdal, E., Are grades of membership probabilities?, *Fuzzy Sets Syst.* **25**, 325–348, 1988.
39. Dubois, D., and Prade, H., Modeling uncertain and vague knowledge in possibility and evidence theories, in *Uncertainty in Artificial Intelligence*, Vol. 4 (R. D. Shachter, T. S. Leviitt, L. N. Kanal, and J. F. Lemmer, Eds.), North-Holland, Amsterdam, 1990, pp. 303–318.
40. Dubois, D., and Prade, H., On the combination of evidence in various mathematical frameworks, in Tech. Rep. LSI 312, IRIT, Univ. Paul Sabatier, Toulouse, France, 1988.
41. Shafer, G., Jeffrey's rule of conditioning, *Phil. Sci.* **48**, 337–362 (1987).