Combining unreliable pieces of evidence

Frans Voorbraak*
Department of Mathematics, Computer Science,
Physics and Astronomy
University of Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam, The Netherlands
fransv@fwi.uva.nl

Abstract

The paper discusses the problem of combining unreliable pieces of evidence in a (generalized) probabilistic setting. The popular Dempster's rule for combining evidence in Dempster-Shafer theory is criticized on two accounts. Firstly, Dempster's rule is claimed to be invalid, and secondly, the rule may fail to support intuitively acceptable conclusions. The reason for this is that the application of Dempster's rule implicitly involves making unrealistic assumptions.

It is conceded that it in many situations it may be unavoidable to make some assumptions to reach conclusions that are of some practical use, but often intuitively more acceptable assumptions than those used in Dempster's rule can be argued for. We illustrate our position by discussing a concrete example of sensor fusion in robotics. Further, an abstract formalization of belief states is proposed in which assumptions are explicitly mentioned, and the processes conditioning and combining pieces of evidence in the presence of assumptions are discussed.

*The investigations were carried out as part of the PIONIER-project Reasoning with Uncertainty, subsidized by the Netherlands Organization of Scientific Research (NWO), under grant pgs-22-262.

1 Introduction

In [2] Arthur Dempster proposes a rule for combining pieces of evidence represented in a framework that in general only assigns lower and upper bounds to the probability values of propositions, thus allowing the expression of ignorance next to uncertainty. Dempster's framework has been reinterpreted and further developed by Glenn Shafer in [7], resulting in a mathematical formalism which is now widely known under the name *Dempster-Shafer theory* (*DS theory*) or *evidence theory*.

DS theory, with Dempster's combination rule as one of its prominent features, has become a popular formalism for handling uncertainty in knowledge-based systems, although several authors have pointed out that the combination rule is not valid, in the sense that it may result in tighter bounds on the probability values than those that can be derived in probability theory. See, for example, [3, 5, 10].

In this paper we argue that the bounds on probability values sanctioned by probability theory may very well be too wide to be of practical use, but that the tighter bounds obtained by Dempster's rule rely on (implicit) assumptions concerning the interaction of the pieces of evidence which cannot be justified in general. We use an informal discussion of a concrete problem to show that these assumptions underlying Dempster's rule may result in unhappy conclusions that can be improved if intuitively more acceptable assumptions are made.

We propose to distinguish between probabilistic constraints that are based on specific information and those that are justified by (plausible) assumptions. Incorporating new information is then not just a matter of (Bayesian) conditioning, since also the assumptions have to be reconsidered. Similarly, when combining two (unreliable) pieces of evidence one has to check the plausibility of old assumptions, and perhaps look for new reasonable assumptions.

Our position is informally introduced in the first part of the paper (sections 2 - 4), where the concrete problem of combining the readings of unreliable sensors is discussed. We admit that the discussion does no justice to the complexity of this problem in realistic situations, but we hope that it does illustrate the relevance of our criticism against Dempster's rule. At the end of the paper we make some comments on how more realistic sensor models may hinder the justified application of Dempster's rule even further.

In the second part of the paper (sections 5 - 9), first the relevant basic material of DS theory is summarized for the reader's convenience. (Although we refer to [7] for a proper introduction to the theory.) We then explain our alternative aproach, called Partial Probability Theory, in which a belief state is modelled by a set of constraints on probability measures representing the beliefs of an agent, possibly supplemented with a set of constraints representing the agent's assumptions. The processes of conditioning and combining pieces of evidence are discussed in this setting.

2 Round One

Arthur and his assistant Thomas work at the Theoretical Research Department of Cyber Inc., a large company specializing in mobile robots. One morning, while Arthur and Thomas are together checking the proof of some theorem, they are visited by Robert, the head of the Development Department.

Robert: Good morning, gentlemen. Sorry to disturb you, but I like to make use of your expertise to solve the following problem. We are building an autonomous vehicle for an important client. The vehicle is required to perform some subtle manoeuvres and for this it needs to know its distance to its nearest objects with an accuracy of 0.01 metre. We have excellent sensors for this task, but unfortunately, they all require the vehicle to come to a complete stop before they can obtain accurate measurements, and this leads to unacceptable delays. We have installed a number of sensors which measure the distance while the vehicle is moving, but they are rather unreliable. However, we have this feeling that, taken together, the readings of these unreliable sensors must be of some use. Now, I've been told that you have been studying some theory for combining pieces of evidence, so I was hoping that you could help me.

Arthur: Indeed, I have been studying a mathematical theory of evidence, called Dempster-Shafer theory, which includes a rule for combining evidence. Only yesterday, I explained the theory to Thomas.

Thomas: To be honest, I am still a bit puzzled, so I would like to see the theory in action.

Arthur: Okay, Robert, let us see what Dempster-Shafer theory has to say about your problem. I first need some more data. How unreliable are your sensors exactly?

Robert: We installed three sensors, let us call them S_1, S_2, and S_3. They each provide the vehicle's CPU with an integer between 0 and 999, representing the measurement in centimetres. They cannot detect objects which are removed 10 metre or more. The first sensor is reliable 20% of the time. That is, 20% of the time, the sensor is working properly and the returned number is the correct distance. The remaining 80% of the time, the sensor is in some way disturbed and the returned number is not properly related to the true distance, although the number may of course happen to be correct by shear luck. For the sensors S_2, and S_3 the same story holds, except that they are both only 10% reliable.

Arthur: Well, I guess it's your lucky day, because Dempster-Shafer theory is cut out for this kind of situation. Let us consider a typical reading of sensor S_1, say 454. If the sensor is reliable, then 454 is the actual distance. If the sensor is unreliable, then we know nothing. Since the sensor is reliable 20% of the time, the evidence justifies a 0.2 degree of belief in 454 and nothing more. This is modelled in Dempster-Shafer theory by a mass function m_1 assigning 0.2 mass to the singleton $\{454\}$ and 0.8 mass to the set $\Omega = \{0, 1, 2, \ldots, 999\}$, that is, the set of all possibilities, which is usually called *sample space* in probability theory. Mass functions generalize probability functions, since mass functions distribute the mass over the non-empty subsets of the sample space, whereas probability functions distribute the probability over the elements, or the singleton subsets, of the sample space.

Robert: I see. So if sensor S_2 also gives a reading of 454, then this evidence is represented by the mass function m_2 assigning 0.1 mass to 454 and 0.9 mass to the sample space Ω. But how do you combine the evidence of the two sensors?

Arthur: Dempster's rule of combination allows you to compute a new mass function out of two mass functions. This rule represents the pooling of evidence, provided some requirements are met. The most important requirement is that pieces of evidence to be combined derive from independent sources. So in order to apply Dempster's rule, I have to know whether your sensors are independent.

Robert: When do you call the sensors independent?

Arthur: There is some confusion about this issue. However, after careful study, I can put it in the following simple terms. Two sensors are independent if one sensor being reliable does not influence the fact whether the second sensor is reliable. This type of independence of the sources of pieces of evidence is called *DS independence* by Voorbraak [10].

Robert: In that case, I guess our sensors can be considered independent, but to be sure I will ask my people to check this.

Arthur: Excellent. Some authors have proposed generalizations of Dempster's rule for the case of dependent evidence, but for the time being we will assume that the sensors are independent. Let me write Dempster's rule on the blackboard.

> The combination of the mass functions m_1 and m_2 over a sample space Ω is the mass function $m_1 \oplus m_2$ over Ω, defined as follows.
>
> $$m_1 \oplus m_2(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ \frac{\sum_{A_i \cap B_j = A} m_1(A_i) \cdot m_2(B_j)}{\sum_{A_i \cap B_j \neq \emptyset} m_1(A_i) \cdot m_2(B_j)} & \text{if } \emptyset \neq A \subseteq \Omega. \end{cases}$$

Let us apply the rule to our two mass functions m_1 and m_2. One easily checks that in this case the denominator of the above definition equals 1, so $m_1 \oplus m_2(\{454\}) = m_1(\{454\}) \cdot m_2(\{454\}) + m_1(\Omega) \cdot m_2(\{454\}) + m_1(\Omega) \cdot m_2(\{454\}) = 0.2 \cdot 0.9 + 0.2 \cdot 0.1 + 0.8 \cdot 0.1 = 0.28$. Further, $m_1 \oplus m_2(\Omega) = m_1(\Omega) \cdot m_2(\Omega) = 0.8 \cdot 0.9 = 0.72$, and for all remaining sets $A \subseteq \Omega$, $m_1 \oplus m_2(A) = 0$.

Robert: The rule combines *two* mass functions. How do you combine *three* pieces of evidence?

Arthur: Since $m_1 \oplus m_2$ is again a mass function, you can simply use Dempster's rule again to combine it with a third mass function. The rule is commutative and associative, so the order in which you combine pieces of evidence does not matter. Suppose our third sensor also returns the number 454. Then the resulting mass function m_3 is the same as m_2, and you can easily check that $m_1 \oplus m_2 \oplus m_3$ assigns a mass 0.352 to 454 and the remaining 0.648 to Ω.

Robert: I trust you on that, but what does this mean for our project? Suppose all three sensors return the number 454. Can I then depend on this result?

Arthur: That depends. How certain do you have to be about the result before you are willing to trust it?

Robert: Let us say that a 5% failure rate would still be acceptable. So if there is at least a 95% chance that a reading is correct, I would be willing to accept it.

Arthur: Then I regret to tell you that you cannot use your unreliable sensors. From a mass function you can obtain a degree of belief in a proposition A by adding all masses assigned to the propositions which imply A. The function measuring these degrees of belief is called a *belief function*. Thus, the degree of belief in 454 is only 0.352. If you are more comfortable with probabilities than with degrees of belief, Philippe Smets has proposed a method, called pignistic transformation, to transform a mass function into a probability function, on which your decision can be based. According to this pignistic transformation, the mass function $m_1 \oplus m_2 \oplus m_3$ gives a probability 0.352648 that 454 is the actual distance, whereas every other element of Ω is assigned a probability 0.000648.

Thomas: This may be a stupid remark, but it seems to me that it is very unlikely for unreliable sensors to produce the same result, so if the sensors agree, isn't it then very likely that they are reliable and that 454 is the actual distance? I seem to recall that Bayes' theorem can be used to calculate the probabilities in such situations.

Arthur: Most of us, and that includes you, Thomas, have been indoctrinated by probability theory. If you want to represent ignorance next to uncertainty, as it is done in Dempster-Shafer theory, you have to be suspicious of your probabilistic "intuitions".

Robert: Okay, Arthur, what should I do? Increase the number of sensors?

Arthur: I'm afraid that is a rather impractical solution, since to obtain the desired certainty, you would have to add about 25 sensors with a 10% reliability, and then you would still only be sufficiently certain in the very rare case that all the sensors give identical readings. The way to go is to try to increase the reliability of the sensors. For example, if your sensors S_2 and S_3 both indicate 454, but your more reliable S_1 indicates 673, which result would you rather believe?

Robert: Let me see, on the one hand, if it is not 454, then *two* sensors have to be unreliable, as opposed to *one* unreliable sensor, in case 673 is wrong. On the other hand, this one sensor is S_1, which is more reliable than the other two. I give up. What's the answer?

Arthur: Actually, you are right to hesitate, since there is not much difference between the support for 454 and for 673. A simple calculation shows that in this case $m_1 \oplus m_2 \oplus m_3(\{454\}) \approx 0.16$, and $m_1 \oplus m_2 \oplus m_3(\{673\}) \approx 0.17$. Thus, the more reliable sensor beats the joint result of the two less reliable sensors.

Thomas: I agree that the prior probability of S_2 and S_3 being both unreliable is slightly larger than the prior probability that S_1 is unreliable. But I still believe that the fact that S_2 and S_3 agree makes it more likely that S_2 and S_3 are reliable.

Robert: To be honest, Arthur, I think Thomas might have a point there. But I think it is best to let you two alone to discuss such theoretical matters. I have to rush back to my department anyway. However, I'll be back to continue this discussion. In the meantime, I will follow Arthur's advice to concentrate on the possibility to improve the reliability of the sensors.

3 Round Two

Arthur and Thomas have become seriously interested in the problem of combining unreliable pieces of evidence. They frequently discuss this problem, but they still differ in opinion when Robert visits them again, a few months after his first call.

Robert: Hello friends, how are you? I have good news about our project. My people have been able to really boost the reliability of the installed sensors.

All three of them are now reliable 50% of the time. At the moment, my people are running a large number of tests, so I thought this might be a good opportunity to get some theoretical predictions from you guys. For example, we were wondering whether to use a plain majority decision rule, that is, to trust a result if at least two of the three sensors agree on it, or a more cautious unanimous decision rule, that is, to trust only those results on which all three sensors agree. Can you shed some light on that?

Arthur: We can certainly try. But before we start our computations, I like to know if you had the independence of the sensors checked, as you promised.

Robert: Yes I did. I cannot give absolute guarantees, but we are pretty sure that one sensor's reliability is independent from the reliability of the other sensors. Up to now, we have not found a significant correlation.

Arthur: Okay, that is good enough for me. Let us see what we get if we combine your sensor readings with Dempster's rule. Suppose all three sensors S_1, S_2 and S_3 return the same number, say 454. Then we obtain three identical mass functions m_1, m_2 and m_3. Each of them assigns 0.5 mass to $\{454\}$ and 0.5 mass to $\Omega = \{0, 1, 2, \ldots, 999\}$. Combining them leads to $m_1 \oplus m_2 \oplus m_3$, given by $m_1 \oplus m_2 \oplus m_3(\{454\}) = 0.875$, and $m_1 \oplus m_2 \oplus m_3(\Omega) = 0.125$. So I am afraid you still do not get your required 95% guarantee that 454 is correct, even if all sensors return this number. For that, you need to boost the reliability of the sensors still further to about 65%.

Robert: That's disappointing news. Just out of curiosity, how reliable should the sensors be to use the simple majority rule with a 95% guarantee?

Arthur: Wait a moment, I need my calculator for that. ...And the answer is ...94.8%. Um, well, I guess the simple majority rule is out.

Thomas: Or perhaps Dempster's rule is out?

Arthur: O, here we have that Bayesian again.

Robert: No need to call each other names. Please, Arthur, let Thomas explain. I am interested in what he has to say, since, to be honest, I do not share your unshakeable faith in this combination rule of Dempster.

Thomas: Actually, I do not consider myself a Bayesian, or at least not a strict Bayesian. For example, I agree with Arthur that it makes sense to distinguish between uncertainty and ignorance. But for this purpose I prefer to use *sets* of probability functions instead of the single probability function of the strict Bayesian or the mass functions used in Dempster-Shafer theory. In fact, this partial probability theory, as I call it, in a sense generalizes Dempster-Shafer theory, since belief functions are simply lower envelopes of particular sets of probability functions.

Arthur: I object to the identification of belief functions with lower probability functions. This may be true in Dempster's model of Dempster-Shafer theory, where belief functions were obtained by means of multi-valued mappings from a probability space, but this connection with probability theory is no longer assumed, and in fact explicitly denied, in Shafer's interpretation of the theory.

Robert: This philosophical discussion is quite interesting, but may I remind you that we have a concrete problem to solve. Our client wants to see a working prototype, and soon! So If you don't mind, I would like to hear what this partial probability theory of Thomas has to say about our sensors.

Thomas: Assume that all three sensors return the same number, say 454. Now, I agree with Arthur that this supports 454 to a degree equal to the probability that at least one of the sensors is reliable. Applying Dempster's rule involves the implicit assumption that this probability is equal to the prior probability, which is 0.875. In my opinion, this assumption is not plausible.

Robert: Why not?

Thomas: Let us write R_i for the proposition that sensor S_i is reliable, and r_i for the propositional variable denoting either R_i or $\overline{R_i}$. Further, $\langle x, y, z\rangle$ denotes the fact that the reading x is obtained from sensor S_1, reading y from S_2, and z from S_3. Then applying Dempster's rule involves the implicit assumption that $P(r_1, r_2, r_3 | \langle 454, 454, 454\rangle) = P(r_1, r_2, r_3)$. In the terminology of Voorbraak [10], it is assumed that each instantiation of r_1, r_2, r_3 is equally confirmed by the evidence $\langle 454, 454, 454\rangle$. This assumption is not plausible since agreeing sensors are more likely to be reliable (in which case they *have* to agree) than unreliable (in which case there is only a very small chance of agreeing). In fact, if a similar assumption is made for every reading $\langle x, x, x\rangle$ ($0 \leq x \leq 999$), then absurd conclusions can be derived.

Robert: Can you elaborate on this?

Thomas: The mentioned assumption implies that for every instantiation of the variables r_i, $P(\langle 454, 454, 454\rangle | r_1, r_2, r_3) = P(\langle 454, 454, 454\rangle)$. An analogous result holds for every $\langle x, x, x\rangle$ ($0 \leq x \leq 999$). Therefore,

$$\sum_{0 \leq x \leq 999} P(\langle x, x, x\rangle) = \sum_{0 \leq x \leq 999} P(\langle x, x, x\rangle | R_1, R_2, R_3) = 1.$$

In other words, the assumptions underlying Dempster's rule in case of agreeing pieces of evidence imply that the pieces of evidence could never have disagreed.

Arthur: So what can you conclude if you do *not* make the assumption of equal confirmation?

Thomas: If you do not make any additional assumption at all, you cannot conclude much. The available information about the sensors simply does not allow strong conclusions about the value of $P(r_1, r_2, r_3|\langle 454, 454, 454\rangle)$. However, in this case, relatively weak assumptions are sufficient to obtain at least as strong conclusions as DS theory. For example, one could assume that $P(\overline{R_1}, \overline{R_2}, \overline{R_3}|\langle 454, 454, 454\rangle) \leq P(\overline{R_1}, \overline{R_2}, \overline{R_3})$, which is much weaker and far more plausible than the correponding equality. What's more, I believe that one can argue for assumptions which justify both the unanimous decision rule and the majority rule.

Robert: I like to hear about that!

Thomas: I think the following assumption is quite reasonable. Given that a sensor is *unreliable*, its reading does not depend on whether the other sensors are reliable or on what their reading is. We can then conclude

$$P(\langle x, y, z\rangle|r_1, r_2, \overline{R_3}) = P(\langle x, y, \cdot\rangle|r_1, r_2, \overline{R_3}) \cdot P(\langle \cdot, \cdot, z\rangle|\overline{R_3}).$$

If we further assume that an unreliable sensor is modelled by a uniform probability distribution over the possible readings, then we obtain

$$P(\langle x, y, z\rangle|r_1, r_2, \overline{R_3}) = P(\langle x, y, \cdot\rangle|r_1, r_2, \overline{R_3}) \cdot \frac{1}{1000}.$$

Now we apply the following version of Bayes' theorem.

$$P(r_1, r_2, r_3|\langle x, y, z\rangle) = \frac{P(\langle x, y, z\rangle|r_1, r_2, r_3) \cdot P(r_1, r_2, r_3)}{\sum_{r_1, r_2, r_3} P(\langle x, y, z\rangle|r_1, r_2, r_3) \cdot P(r_1, r_2, r_3)}.$$

Since $P(r_1, r_2, r_3) = 0.125$ is a constant, we can eliminate this factor. Further notice that the probability of a reliable sensor giving a reading x is equal to the prior probability of x being the actual distance of the nearest object. Thus, for example, $P(\langle x, x, x\rangle|R_1, R_2, R_3) = P(x)$. It follows that

$$P(\overline{R_1}, \overline{R_2}, \overline{R_3}|\langle 454, 454, 454\rangle) = \frac{10^{-9}}{(1 + 3 \cdot 10^{-3} + 3 \cdot 10^{-6}) \cdot P(454) + 10^{-9}}.$$

A simple calculation shows that a 95% confidence in the unanimous reading 454 can already be obtained when the prior probability of 454 being the actual distance is assumed to be $1.9 \cdot 10^{-8}$, or higher. If this prior probability is assumed to be around 0.001, then $P(454|\langle 454, 454, 454\rangle) \approx 0.999999$.

Robert: But how about the majority rule?

Thomas: Let the evidence be $\langle 454, 454, 673\rangle$. Then the following five possibilities remain. (1) $R_1, R_2, \overline{R_3}$, (2) $R_1, \overline{R_2}, \overline{R_3}$, (3) $\overline{R_1}, R_2, \overline{R_3}$, (4) $\overline{R_1}, \overline{R_2}, R_3$, and (5) $\overline{R_1}, \overline{R_2}, \overline{R_3}$. The first three possibilities support 454, the fourth possibility supports 673, and the fifth possibility supports Ω. If, in addition to the previously mentioned assumptions, one assumes that the prior probabilities $P(454)$ and $P(673)$ are both approximately equal to 0.001, then $P(454|\langle 454, 454, 673\rangle) \approx 0.998$. Thus also in the case of *two* agreeing sensors one can have a high confidence in their reading. Of course, the confidence level is not as high as in the case *three* agreeing sensors, and one needs to assume that the prior probability of the distance indicated by the disagreeing sensor is not much larger than the prior probability of the distance indicated by the two agreeing sensors.

Robert: A confidence level of 0.998 is more than adequate for my purpose, and the mentioned assumptions seem quite reasonable. I would sure appreciate it if you would write a technical report on the matters we discussed.

4 Round Three

Arthur and Thomas are preparing a report on combining unreliable pieces of evidence. In the process, Thomas is clarifying his position to the still sceptical Arthur.

Arthur: I am still a bit confused concerning your representation of uncertainty. On the one hand, you use probabilistic reasoning. On the other hand, you say you do not represent a situation by a probability function, as strict Bayesians do.

Thomas: In my opinion, the representation of uncertain beliefs by means of a single probability function is not appropriate when the agent has to deal with ignorance next to uncertainty. This argument is also put forward by the proponents of DS theory. However, the original motivation for the formalism of DS theory in Dempster [2] can be clearly understood in probabilistic terms, provided one allows the probabilities of events to be only partially determined.

Arthur: But Shafer [7] reinterprets the formalism introduced by Dempster as a theory of evidence, and he explicitly denies belief functions to be lower bounds on probabilities.

Thomas: However, in [8] Shafer explains the philosophy behind belief functions as follows. An ideally rational agent obtains degrees of belief by judging uncertainties in a situation at hand to be similar to the uncertainties in canonical examples. If the situation is judged to be similar to drawing balls from an urn, then the degrees of belief can be modelled by a probability function. The situation can also be judged to be similar to the receipt of a coded message, where one knows for each possible code the probability that it is used. In that case, one obtains degrees of belief that can be modelled by a belief function. Notice that this interpretation of belief functions is just Dempster's model in disguise.

Arthur: Does this mean that you propose to use belief functions instead of probability functions?

Thomas: No, I propose to use something more general, namely partially specified probabity. In my opinion, one often lacks sufficient information for judging a situation to be similar to an urn experiment, but the available information may justify partial judgments. For example, in the case of a 50% reliable sensor, the uncertainty concerning the reliablity of the sensor can be judged to be similar to the uncertainty concerning the redness of a randomly drawn ball from an urn half filled with red balls and half filled with green balls. Using this probability judgment on the (un)reliability of the sensor, a particular reading of the unreliable sensor gives information on the actual distance, without determining a probability measure on the possible outcomes. In this example, one obtains a belief function, but in [5] Kyburg shows that partial probabilistic information not always generates a belief function.

Arthur: I remember this paper by Kyburg. He shows that there exists a set of probability measures such that its lower envelope is not a belief function. Does that mean that you propose to use such lower envelopes instead of belief functions?

Thomas: I do propose to use such lower envelopes for measuring lower degrees of belief. However, to formalize a belief state it is much better to use the sets of probability function themselves instead of their lower envelopes, since essentially different sets may have the same lower envelope. Similarly, one can argue that it is even better to use the constraints themselves instead of the set of probability functions that satisfy them.

Arthur: Thus your representation of a belief state is a set of constraints on probabilities.

Thomas: That is correct. However, the constraints that are supported by the available evidence may very well be too weak to justify any useful conclusions. See the example of combining the evidence of the unreliable sensors. In that case, one should consider the possibility of obtaining additional constraints by making some reasonable assumptions. Whether to formalize these assumptions as part of the belief state or not is a matter of taste, but it is essential to distinguish the constraints based on specific evidence and those modelling assumptions.

Arthur: I am glad you mention these assumptions, since I have a question concerning the way you combined the evidence of the unreliable sensors. I did some calculations, and I noticed that there is an easy way to obtain results similar to yours in DS theory. The trick is not to use Dempster's rule on the original belief functions, but on their pignistic transformations. That is, instead of the mass function m_i, given by $m_i(454) = m_i(\Omega) = 0.5$, use the mass function $Pig(m_i)$, given by $Pig(m_i)(454) = 0.5005$, and $Pig(m_i)(\omega) = 0.0005$, if $454 \neq \omega \in \Omega$. Why not use this trick?

Thomas: This trick has several disadvantages. There is no clear distinction between conclusions that are supported by the evidence and those that need additional assumptions. Moreover, since the belief functions are transformed into probability functions, there is no room left for representing ignorance, whereas adding additional constraints based on assumptions does not necessarily completely eliminate the ignorance.

Arthur: I see your points. Still, I would like to have a more simpler way to combine evidence, since I do not see how your general approach to representing belief states can be translated into a formalism that is simple and can be implemented.

Thomas: I agree there is still some work to be done here. However, notice that there have already been developed several formal languages for expressing probabilistic constraints. Moreover, even if the full theory cannot be implemented it may still be useful for analyzing used or proposed reasoning methods.

5 Dempster-Shafer Theory

Dempster-Shafer theory (DS theory) can be viewed as a generalization of, or an alternative to, probability theory. In DS theory uncertainty is represented by means of belief functions, which have several equivalent characterizations. Before we formally define belief functions, we first fix some basic terminology.

A sample space, or frame, Ω is an exhaustive set of mutually exclusive possibilities. We assume that sample spaces are finite. When a function is defined on the powerset 2^Ω of Ω we sometimes say that it is a function *over* Ω. A real-valued function F over Ω is called *monotone* iff $A \subseteq B \subseteq \Omega$ implies $F(A) \leq F(B)$.

A (normalized) *capacity* over a sample space Ω is a monotone function $F : 2^\Omega \to [0,1]$ such that $F(\Omega) = 1$. The class of capacities over Ω is denoted by $CAP(\Omega)$. In the literature, capacities are also known as *fuzzy measures* or *Sugeno measures*.

A *probability function* P over Ω can be defined to be a capacity over Ω which additionally satisfies the following property of *additivity*.

$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B) \qquad (A, B \subseteq \Omega). \tag{1}$$

As is well known, the monotonicity requirement is redundant when additivity is added. The class of probability functions over Ω is denoted by $PROB(\Omega)$.

Let F and G be functions on 2^Ω. The function F is called the *dual*, or conjugate, of G iff for every $A \subseteq \Omega$, $F(A) = 1 - G(\overline{A})$. (Here $\overline{A}$ denotes the complement of A with respect to Ω.) The dual of a capacity is also a capacity. If F is the dual of G, then G is the dual of F. If two functions are the dual of each other, then each one of them determines the other. A probability function is its own dual.

In [2] the following kind of situation is modelled. Suppose one is interested in the events of a sample space Ω, whereas the available information determines a probability measure P over a sample space Θ, which differs from, but is related to Ω in a way which is encoded by a mapping Γ from Θ to 2^Ω, where $\omega \in \Gamma(\theta)$ iff ω and θ are compatible. Such a Γ is called a multivalued mapping, and occasionally occurs in the literature in the guise of a compatibility relation $\subseteq \Theta \times \Omega$, or under the name random set.

From Γ and P over Θ one can derive constraints on the probabilities of the events of Ω, without, in general, being able to determine the exact probabilities. Consider the function which assigns to every $A \subseteq \Omega$ the best lower bound on the probability of A, given the available information. This function is called a *belief function* over Ω, and is denoted by $Bel_{\Gamma,P}$. In addition to this belief function, also the *mass function* $m_{\Gamma,P}$ and the *plausibility function* $Pl_{\Gamma,P}$ can be defined.

Definition 5.1 *Assume that $\langle \Theta, 2^\Theta, P \rangle$ is a probability space, Ω is a sample space, and $\Gamma : \Theta \to 2^\Omega$. The functions $m_{\Gamma,P}$, $Bel_{\Gamma,P}$ and $Pl_{\Gamma,P}$ over Ω induced by Γ and P are defined as follows.*

$$m_{\Gamma,P}(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ \frac{P(\{\theta\in\Theta:\Gamma(\theta)=A\})}{P(\{\theta\in\Theta:\Gamma(\theta)\neq\emptyset\})} & \text{if } \emptyset \neq A \subseteq \Omega \end{cases} \tag{2}$$

$$Bel_{\Gamma,P}(A) = \frac{P(\{\theta \in \Theta : \Gamma(\theta) \subseteq A\})}{P(\{\theta \in \Theta : \Gamma(\theta) \neq \emptyset\})} \qquad (A \subseteq \Omega) \tag{3}$$

$$Pl_{\Gamma,P}(A) = \frac{P(\{\theta \in \Theta : \Gamma(\theta) \cap A \neq \emptyset\})}{P(\{\theta \in \Theta : \Gamma(\theta) \neq \emptyset\})} \qquad (A \subseteq \Omega) \tag{4}$$

The above can be somewhat simplified by assuming that the empty set is not in the range of Γ, since then the denominators disappear. The classes of belief and plausibility functions over Ω are denoted by $BEL(\Omega)$ and $PL(\Omega)$, respectively. It is easy to see that $Bel_{P,\Gamma}$ and $Pl_{P,\Gamma}$ are dual capacities.

Example 5.1 Consider one of the unreliable sensors discussed before. Let $\Omega = \{0, 1, 2, \ldots, 999\}$ be the sample space of distances to the nearest object. Further, the probability function P over $\Theta = \{R, \overline{R}\}$ given by $P(R) = 0.5$ formalizes that the sensor is reliable in 50% of the cases. Suppose that the sensor gives a reading of 454. This evidence is captured by the multivalued mapping $\Gamma : \Theta \to 2^{\Omega}$ defined by $\Gamma(R) = \{454\}$ and $\Gamma(\overline{R}) = \Omega$.

The function $m_{\Gamma,P}$ is given by $m_{\Gamma,P}(\{454\}) = 0.5$, and $m_{\Gamma,P}(\Omega) = 0.5$. The function $Bel_{\Gamma,P}$ is given by $Bel_{\Gamma,P}(\Omega) = 1$, $Bel_{\Gamma,P}(A) = 0.5$, if A is a proper subset of Ω containing 454, and $Bel_{\Gamma,P}(A) = 0$, if A does not contain 454.

Belief functions can be characterized as capacities of a certain type.

Definition 5.2 *Let Ω be a sample space. A function Bel over Ω is a* belief function *iff Bel is a capacity satisfying the following condition of* ∞-monotonicity.

$$\textit{For all } n \geq 1, P(\bigcup_{i=1}^{n} A_i) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} P(\bigcap_{i \in I} A_i). \tag{5}$$

Gustave Choquet [1]was probably the first to study the thus defined functions which he called *capacities monotone of order infinity*, or ∞*-monotone capacities*. Capacities satisfying equation (5) with the restriction that $n \leq N$ are called monotone of order N, or N-monotone.

Notice that this alternative characterization of belief functions over Ω as ∞-monotone capacities does not require any reference to a space other than Ω. Shafer [7] gives a third characterization of belief functions over a sample space Ω which also does not refer to a related probability space.

Definition 5.3 *A* mass function, *or* basic probability assignment (bpa), *over a sample space Ω is a function $m : 2^{\Omega} \to [0,1]$ which satisfies the following two conditions.*

$$m(\emptyset) = 0. \tag{6}$$

$$\sum_{A \subseteq \Omega} m(A) = 1. \tag{7}$$

The quantity $m(A)$ is a measure of the belief that is assigned to *exactly* the set A (and not to any proper subset of A). (If $m(A) = 0.4$, then a 40% portion of one's total belief is assigned to exactly A.) Equation (7) formalizes the convention that one's total belief has measure one, and equation (6) reflects the requirement that none of the belief of a rational agent is assigned to the impossible event.

A measure of the (total) belief in A should also take into account the measures of belief assigned to more specific propositions, i.e., to subsets of A. For this purpose belief functions induced by mass functions are introduced.

Definition 5.4 *Let m be a mass function over a sample space Ω. The* belief function Bel_m induced by m *is defined as follows.*

$$Bel_m(A) = \sum_{B \subseteq A} m(B) \qquad (A \subseteq \Omega). \tag{8}$$

The three given characterizations of belief functions are all equivalent.

Proposition 5.1 *Let Ω be a sample space. The following are equivalent.*

1. $Bel = Bel_{\Gamma,P}$, *for some probability space* $\langle \Theta, 2^{\Theta}, P \rangle$ *and* $\Gamma : \Theta \rightarrow 2^{\Omega}$.
2. Bel *is a* ∞*-monotone capacity over* Ω.
3. $Bel = Bel_m$, *for some (unique) mass function* m *over* Ω.

Proof. We show the equivalence of the first and third definition. On the one hand, if we assume $Bel = Bel_{\Gamma,P}$, then it is easy to see that $Bel = Bel_m$, where $m = m_{\Gamma,P}$. On the other hand, if we assume $Bel = Bel_m$, then $Bel = Bel_{\Gamma,P}$, where P is the probability function over $\Theta \stackrel{\text{def}}{=} 2^{\Omega}$ defined by $P((\{A\}) = m(A)$ $(A \subseteq \Omega)$, and $\Gamma : 2^{\Omega} \rightarrow 2^{\Omega}$ is the identity function on 2^{Ω}.

The equivalence of the second and third definition is shown in [7], whereas the proof of the equivalence of the first and second definition can essentially be found in [1]. ∎

If Bel is a belief function over Ω, then the unique mass function m over Ω such that $Bel = Bel_m$ is given by the following equation.

$$\text{For all } A \subseteq \Omega,\ m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \cdot Bel(B). \tag{9}$$

$Bel(A)$ is a measure of the (total) belief that is certainly assigned to A. There may also be belief that could possibly be assigned to A, namely belief assigned to some B which is consistent with A. Plausibility functions take also this belief into account.

Definition 5.5 *Let m be a mass function over a sample space Ω. The* plausibility function Pl induced by m *is defined as follows.*

$$\textit{For every } A \subseteq \Omega,\ Pl(A) = \sum_{A \cap B \neq \emptyset} m(B). \tag{10}$$

It is easy to check that Pl and Bel are dual. It follows that the functions m, Bel, Pl are interdefinable. We introduce some further notions. A subset A of a sample space Ω is called a *focal element* of a belief function Bel over Ω iff $m(A) > 0$. (Here it is assumed to be understood that the m denotes the mass function associated with Bel. Similar assumptions will be made frequently.) The union of all focal elements of a belief function is called its *core*.

The complete information about the measure of belief in A can be represented by the *belief interval* $[Bel(A), Pl(A)]$, where $Pl(A) - Bel(A)$ is a natural expression of the ignorance concerning A. Complete ignorance with respect to a sample space Ω is represented by the *vacuous belief function* over Ω, which is the belief function induced by the mass function m defined by $m(\Omega) = 1$ and for all $A \subset \Omega, m(A) = 0$.

For any subset A of a sample space Ω, we write 1_A for the mass function which assigns mass 1 to the set A (and zero mass to all other events). We will use the same notation for the associated belief function, plausibility function, et cetera, and let the context determine which function is meant. Thus, the vacuous belief function over Ω can be denoted by 1_Ω.

A belief function Bel over a sample space Ω is is a probability function over Ω iff all the focal elements of Bel are singletons. Such a belief function is called a *Bayesian belief function.*

It is tempting to consider $Bel(A)$, respectively $Pl(A)$, as lower, respectively upper, bound of the "true" probability of A. Although this interpretation is in accordance with the original intention of Arthur Dempster, it is explicitly rejected by Glenn Shafer [7]. In his view, DS theory is pre-eminently suited for situations in which there is not sufficient evidence available to speak of the true probabilities of all relevant propositions. It should be mentioned that in later publications Shafer comes quite close to Dempster's model. Presently, the *Transferable Belief Model* of Philippe Smets [9] is probably the best developed theory which takes Shafer's original interpretation serious.

6 Dempster's Combination Rule

In this section we discuss a rule, introduced by Dempster [2], for combining two belief functions into a new belief function which is supposed to represent the combined evidence.

Assume that Bel_1 and Bel_2 are belief functions over Ω induced by mass functions m_1 and m_2 such that $\sum_{A_i \cap B_j \neq \emptyset} m_1(A_i) \cdot m_2(B_j) \neq 0$. The *combination of Bel_1 and Bel_2 by Dempster's rule*, is the belief function $Bel_1 \oplus Bel_2$ induced by the mass function $m_1 \oplus m_2$ over Ω, defined as follows.

$$m_1 \oplus m_2(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ \frac{\sum_{A_i \cap B_j = A} m_1(A_i) \cdot m_2(B_j)}{\sum_{A_i \cap B_j \neq \emptyset} m_1(A_i) \cdot m_2(B_j)} & \text{if } \emptyset \neq A \subseteq \Omega. \end{cases} \tag{11}$$

The function $Bel_1 \oplus Bel_2$ is sometimes called the *orthogonal sum of Bel_1 and Bel_2*. The factor $[\sum_{A_i \cap B_j \neq \emptyset} m_1(A_i) \cdot m_2(B_j)]^{-1}$ is called the *renormalizing constant* of Bel_1 and Bel_2. If $\sum_{A_i \cap B_j \neq \emptyset} m_1(A_i) \cdot m_2(B_j) = 0$, then $Bel_1 \oplus Bel_2$ is undefined, and Bel_1 and Bel_2 are called *not combinable.*

The rationale for the rule is that assigning $m_1(A_i)$ to A_i and $m_2(B_j)$ to B_j results in an assignment of $m_1(A_i) \cdot m_2(B_j)$ to the intersection of A_i and B_j. To avoid that mass is assigned to the empty set, the mass $m_1(A_i) \cdot m_2(B_j)$ is discarded whenever $A_i \cap B_j = \emptyset$. The remaining masses are then renormalized to enforce that the masses sum to 1.

The following proposition lists some properties of Dempster's combination rule. Proofs of the properties can be found in [7]. The notation $x \simeq y$ is used to abbreviate "x and y are both defined and equal, or x and y are both undefined".

Proposition 6.1 *Let Bel_1, Bel_2, and Bel_3 be belief functions over Ω.*

1. $Bel_1 \oplus Bel_2 \simeq Bel_2 \oplus Bel_1$.
2. $Bel_1 \oplus (Bel_2 \oplus Bel_3) \simeq (Bel_1 \oplus Bel_2) \oplus Bel_3$.
3. *If $Bel_1 \oplus Bel_2$ is defined and Bel_1 is Bayesian, then $Bel_1 \oplus Bel_2$ is also Bayesian.*
4. $Bel_1 \oplus 1_\Omega = Bel_1$.

It follows that the result of Dempster's combination rule does not depend on the order in which pieces of evidence are combined, since the rule is commutative (first property above) and associative (second property). The third property states that the combination of a belief function with a Bayesian belief function is again Bayesian (provided the belief functions are combinable). The fourth property states that combining with the vacuous belief function does not change one's degrees of belief.

Dempster's combination rule gives intuitively appealing results in some simple examples (see [7]), but its justification in general is problematic. This is perhaps most clear when we consider the effect of the rule for belief functions as lower probabilities induced by multivalued mappings.

Proposition 6.2 *Let P_1 and P_2 be a probability functions over Θ_1 and Θ_2, respectively, and let $\Gamma_1 : \Theta_1 \to 2^\Omega$ and $\Gamma_2 : \Theta_2 \to 2^\Omega$. Define P to be the product probability function of P_1 and P_2 over $\Theta_1 \times \Theta_2$, and let $\Gamma : \Theta_1 \times \Theta_2 \to 2^\Omega$ be defined by $\Gamma(\langle\theta_1, \theta_2\rangle) = \Gamma_1(\theta_1) \cap \Gamma_2(\theta_2)$. Then*

$$Bel_{\Gamma_1,P_1} \oplus Bel_{\Gamma_2,P_2} = Bel_{\Gamma,P}.$$

The problem with this result is the following. Assuming (DS) independence of the sources of the pieces of evidence to be combined, it is quite reasonable to assume that the prior probability of the behaviour of the two sources is the product probability of the probabilities of their respective behaviour. However, this prior probability has to be updated since the combination of the two pieces of evidence may provide evidence concerning the behaviour of the underlying sources. For example, if two sensors give contradictory readings, then they cannot both be reliable, even though there may be a positive prior probability of both sensors being reliable.

In fact, Dempster's rule does take into account such clear effects of the combination as illustrated above. This is done by discounting the mass assigned to the empty set and by renormalizing. However, the rule does not take into account more subtle effects such as the increase in the likelihood of sensors being reliable when the agree. As mentioned before, Dempster's rule implicitly assumes that every possible behaviour of the two sources is equally confirmed by the evidence. To make the above more concrete, we consider the example of the unreliable sensors in more detail.

Example 6.1 Let $\Omega = \{0, 1, 2, \ldots, 999\}$, $\Theta_1 = \{R_1, \overline{R_1}\}$ and $\Theta_2 = \{R_2, \overline{R_2}\}$. The probability function P_1 over Θ_1 given by $P_1(R_1) = 0.5$ formalizes that the first sensor is reliable in 50% of the cases. Similarly, we assume P_2 over Θ_2 to be defined by $P_2(R_2) = 0.5$. Suppose that both sensors give a reading of 454. This evidence is captured by the multivalued mappings $\Gamma_i : \Theta_i \to 2^\Omega$ defined by $\Gamma_i(R_i) = \{454\}$ and $\Gamma_i(\overline{R_i}) = \Omega$ $(i = 1, 2)$.

Let P be the product probabilty function of P_1 and P_2 over $\Theta_1 \times \Theta_2$. According to Dempster's rule there is no need to update the prior probability, since $\Gamma_1(r_1) \cap \Gamma_2(r_2)$ is always non-empty, and therefore no mass threatens to become assigned to the empty set. However, intuitively, the fact that the sensors agree provides support for the proposition that the sensors are reliable. In the first part of the paper it is already explained that this intuition can be justified by making some reasonable assumptions.

The above argument against Dempster's combination rule can essentially already be found in [10].

7 Partial Probability Theory

Let β be the belief state of an ideally rational agent. This belief state can be formalized on different levels of abstractions. A very abstract formalization of β would be a logical theory, representing the set of propositions that are fully believed by the agent. Equivalently, one could use a set of possible worlds, representing the situations that are not excluded by the agent.

A more refined formalization is obtained when propositions with degrees of belief less than one are also represented. This can for example be achieved by defining a probability measure over the possibilities or possible worlds. However, in our opinion, an agent does not always have sufficient information to be able to completely determine a probability measure. That is why we propose a more general formalization of belief states where the probability measure is only required to be partially determined.

When modelling a belief state one not only has to choose between allowing degrees of belief or restricting oneself to full beliefs, but one should also answer the question whether it is necessary to keep track of the justifications for the beliefs or not. The approach that answers this question positively is called the foundations theory by Harman [4], whereas the other approach is called the coherence theory.

Perhaps the simplest foundational representation of a belief state is a set of propositional sentences which is not deductively closed. Such a (base) set of sentences can be considered to justify the (full) belief in all the sentences in its deductive closure. In the literature on belief revision such a base set representation is typically used in so-called syntax-based approaches to belief revision. However, base set representations are not restricticted to the syntactic domain, since one could view any particular sentence in the base set as a constraint on what the actual world might be, and represent each sentence by the set of possible worlds satisfying the constraint. One then obtains a collection of sets of possible worlds as a representation of the justifications for believing that the intersection of the sets in the collection contains the actual world.

The probabilistic analogue of a base set of propositional sentences (justifying a deductively closed set) is a set of sentences in some (formal) language which allows the expression of probabilistic information. Semantically, each sentence expressing probabilistic information can be viewed as a constraint on possible probability measures. Since formal language for probabilistic reasoning is less standardized than in the case of propositional reasoning, we prefer to use semantic notions in our general formalization of a probabilistic belief state.

Definition 7.1 *A* (partial) probabilistic belief state (pbs) *is a pair* $\langle \Omega, \mathcal{B} \rangle$*, where* Ω *is a sample space and* $\mathcal{B}$ *is a set of constraints on probability measures over* Ω*.*

A pbs $\langle \Omega, \mathcal{B} \rangle$ is a quite general representation of a belief state. If $\mathcal{B}$ is empty, then one obtains a full belief set represented by the set of possibilities Ω. If $\mathcal{B}$ determines a single probability measure over Ω, then one has a simple foundational representation of the classical Bayesian formalization of a belief state. In general, $\langle \Omega, \mathcal{B} \rangle$ naturally determines the set $\mathcal{B}(\Omega)$ of all probability measures over Ω satisfying the constraints of $\mathcal{B}$.

We write $\mathcal{B}(\Omega)_{low}$ for the lower envelope of $\mathcal{B}(\Omega)$, that is, the function defined by

$$\mathcal{B}(\Omega)_{low}(A) = \inf\{P(A) : P \in \mathcal{B}(\Omega)\} \qquad (A \subseteq \Omega).$$

Similarly, the upper envelope $\mathcal{B}(\Omega)_{up}$ is defined by

$$\mathcal{B}(\Omega)_{up}(A) = \sup\{P(A) : P \in \mathcal{B}(\Omega)\} \qquad (A \subseteq \Omega).$$

The functions $\mathcal{B}(\Omega)_{low}$ and $\mathcal{B}(\Omega)_{up}$ are dual capacities and they can be viewed as functions measuring lower and upper degrees of belief. Any belief function of DS theory can be obtained as a lower envelope induced by a pbs, but not every such lower envelope is a belief function.

Thus the notion of pbs generalizes the formalization of a belief state in (propositional) logic, (Bayesian) probability theory, and DS theory. However, we want to generalize the formalization one step further. In our opinion, the available probabilistic information is not only often insufficient to determine a single probability measure, it may also be insufficient to justify conclusions that are of practical use. In that case, one might try to strengthen one's conclusions by making some reasonable assumptions. This suggests the following.

Definition 7.2 *A* (partial) probabilistic belief state with assumptions (pbsa) *is a triple $\langle \Omega, \mathcal{B}, \mathcal{A} \rangle$, where Ω is a sample space and $\mathcal{B}$ and $\mathcal{A}$ are sets of constraints on probability measures over Ω. The set $\mathcal{B}$ is called the set of base beliefs, and $\mathcal{A}$ is called the set of base assumptions.*

If $\langle \Omega, \mathcal{B}, \mathcal{A} \rangle$ is a pbsa, then $\langle \Omega, \mathcal{B} \rangle$ is a pbs. We write $\mathcal{B}_{\mathcal{A}}(\Omega)$ for the set of all probability measures over Ω which satisfy all the constraints in $\mathcal{A} \cup \mathcal{B}$. By definition, $\mathcal{B}_{\mathcal{A}}(\Omega) \subseteq \mathcal{B}(\Omega)$. Therefore, the conclusions warranted by $\mathcal{B}_{\mathcal{A}}(\Omega)$ are (weakly) stronger than those warranted by $\mathcal{B}(\Omega)$. However, it is intended to be understood that the conclusions warranted by $\mathcal{B}_{\mathcal{A}}(\Omega)$ depend on the assumptions represented in $\mathcal{A}$.

The following example illustrates the expressive power of pbsa's.

Example 7.1 Using pbsa's, one can distinguish between the following three belief states concerning the outcome of a coin toss.

1. One has reliable information saying that the coin is fair. (This belief state can be represented by $\langle \{H, T\}, \{P(H) = P(T) = 0.5\}, \emptyset \rangle$)

2. One has no information concerning the fairness of the coin, but one considers it reasonable to assume that heads and tails are equally likely outcomes (based on some method for choosing least informative priors in the presence of insufficient information). ($\langle \{H, T\}, \emptyset, \{P(H) = P(T) = 0.5\} \rangle$)

3. One has no information concerning the fairness of the coin, and one does not consider it reasonable or appropriate to make assumptions concerning the outcome. ($\langle \{H, T\}, \emptyset, \emptyset \rangle$)

In probability theory, the first two belief states have identical representations, whereas the third belief state cannot be represented at all. If one uses pbs's (without assumption), then it is not clear how to represent the second belief state.

By partial probability theory (PPT) we mean the approach to (probabilistic) reasoning in which belief states are represented by pbsa's. Notice that PPT generalizes both probability theory and DS theory. In the following section we discuss how in PPT the process of conditioning is not only complicated by the presence of assumptions, but also by the fact that belief states have a foundational representation.

8 Conditioning

In Bayesian probability theory, acquiring new evidence to the effect that the truth is in a subset A of the sample space Ω is modelled by conditioning the current probability measure P on A. This is called the Bayesian rule of conditioning. Let us write P_A for the thus obtained probability measure.

In DS theory, the analogous problem is solved by combining the current belief function Bel with the belief function 1_A which assigns mass 1 to A. This is known as Dempster's rule of conditioning. However, one sometimes considers an alternative approach based on the fact that a belief function Bel over Ω naturally induces a set of probability measures, namely the set $\langle Bel \rangle$ of probability measures P over Ω such that for every $B \subseteq \Omega$, $P(B) \geq Bel(B)$.

For any set Π of probability functions over Ω, it is natural to define the result of conditioning on A to be the set $\Pi_A = \{P_A : P \in \Pi, P(A) \neq 0\}$. This is called the generalized Bayesian rule of conditioning. It follows that one could define Bel conditioned by A to be the lower envelope of $\langle Bel \rangle_A$. However, the thus defined lower envelope does not necessarily agree with $Bel \oplus 1_A$. Therefore, there is some disagreement on how to condition a belief function.

In PPT, the notion of conditioning is even less clear. Let $\langle \Omega, \mathcal{B}, \mathcal{A} \rangle$ be a pbsa, and let $A \in \Omega$ be the new evidence. Then several possibilities of incorporating the new evidence in the pbsa suggest themselves.

Perhaps the most obvious possibility is to add the constraint $P(A) = 1$ to $\mathcal{B}$. However, this procedure is only well-defined if this additional constraint is consistent with $\mathcal{B}$. More importantly, this procedure does not generalize the ordinary notion of Bayesian conditioning. For example, if $\mathcal{B}$ determines a probability measure P such that $0 \neq P(A) \neq 1$, then the mentioned procedure is not defined, in contrast to ordinary Bayesian conditioning.

Another possibility is to apply the generalized Bayesian rule of conditioning to the set $\mathcal{B}(\Omega)$. One could then for example define $\{P(B) \geq (\mathcal{B}(\Omega)_A)_{low}(B) : B \subseteq A\}$ to be the new set of belief constraints. However, this approach is rather conservative, as illustrated by the following example.

Example 8.1 Let $\Omega = \{0, 1, 2, \ldots, 999\}$ and consider the pbsa $\langle \Omega, \{P(\{454\}) \geq 0.5\}, \emptyset \rangle$ modelling the 50% reliable sensor S returning the number 454. Now assume we obtain the new evidence $A = \{454, 545\}$. Intuitively, this new evidence lends a high degree of support to the sensor's reliability, but applying the generalized Bayesian rule does not result in an increase of the lower bound on the probability of 454.

This remains true if one uses the finer sample space $\{\langle R, 454 \rangle\} \cup (\{\overline{R}\} \times \Omega)$ and the pbsa $\langle \{\langle R, 454 \rangle\} \cup (\{\overline{R}\} \times \Omega), \{P(\{\langle R, 454 \rangle\}) = 0.5\}, \emptyset \rangle$.

However, if one adds to this pbsa the assumptions $P(\langle \overline{R}, \omega \rangle) = 0.0005$ $(\omega \in \Omega)$, then applying the generalized Bayesian rule to the set of probability functions satisfying both sets of constraints warrants the intuitively correct conclusion that $P(\{454\}) \approx 0.99$.

The above example illustrates the use of additional assumptions in order to get nontrivial probability bounds after receiving new evidence. Moreover, it illustrates that to express these assumptions it may be necessary to use a finer sample space than the set of possibilities one is primarily interested in.

The general procedure we propose is the following. By applying the generalized Bayesian rule to $\mathcal{B}(\Omega)$ one obtains the new beliefs. In the presence of assumptions, stronger conclusions can be obtained by applying the generalized Bayesian rule to $\mathcal{B}_{\mathcal{A}}(\Omega)$. The new assumptions can be inferred from the difference between $(\mathcal{B}(\Omega))_A$ and $(\mathcal{B}_{\mathcal{A}}(\Omega))_A$.

9 Combining Evidence

In Bayesian probability theory there exists no rule for combining evidence. That is, there is no general rule for computing $P(A|B,C)$ from the values of $P(A|B)$, $P(A|C)$, $P(A)$, $P(B)$, $P(C)$, $P(A,B)$, et cetera. More precisely, we have the following result, which is proved in Neapolitan [6].

Proposition 9.1 (Neapolitan) *Let Σ be the algebra generated by the sets A, B, and C. There is in general no function F which computes for any probability function P on Σ the value $P(A|B,C)$ from the values of P on the algebras generated by at most two elements from $\{A,B,C\}$.*

Neapolitan's proof shows that the result still holds if the events A, B, and C are pairwise independent and if the function F is only required to approximate the value of $P(A|B,C)$. One can conclude from the result that the value of $P(A|B,C)$ is only guaranteed to be derivable if some information is available (or assumed) concerning the probability of some event involving all three elements from $\{A,B,C\}$.

In DS theory, there does exist a rule for combining evidence, namely Dempster's rule. However, we already mentioned that the rule not only requires the evidence to be DS independent, but also implicitly involves rather unrealistic assumptions.

In the light of the above, it is to be expected that there is no simple rule for combining evidence in PPT. The example of the unreliable sensors discussed in section 3 illustrates that without any additional assumptions the result of combining two pieces of evidence in PPT is typically highly indecisive. One simply needs some information or assumptions concerning the interaction of the pieces of evidence. In the case of the unreliable sensors one can argue for such assumptions which are more reasonable than those underlying Dempster's rule.

The general procedure for combining evidence is to merge the pbsa's representing the two pieces of evidence into one pbsa with a sample space which is a common refinement of the two original sample spaces, with the induced belief constraints and assumptions of both the original pbsa's, and with as much additional belief constraints and assumptions concerning the interaction of the evidence as possible.

In the absence of specific information concerning the interaction of the evidence, it may still be possible to argue for some additional assumptions which together with the original constraints justify useful conclusions.

10 Discussion

We argued that using Dempster's rule to combine evidence implicitly involves making unrealistic assumptions. It should be pointed out that in our argument we used a very idealized example of sensor fusion, and that for more realistic sensor models the application of Dempster's rule is even less justified.

For example, the error in a sensor reading usually depends on the particular environment the sensor is used in. Therefore, two sensors used in the same environment are typically not DS independent since their reliability is influenced by a common factor. Another example is that in practice one may have some prior knowledge about the actual distances of a robot to its nearest object, whereas in DS theory the readings of the range sensors are represented under the assumption that no such prior knowledge is available.

In the presence of ignorance next to uncertainty it may be impossible to achieve useful conclusions on the basis of the available probabilistic information only. Therefore, it may be unavoidable to use additional assumptions. However, these assumptions should be clearly distinguished from constraints based on specific information.

We proposed to use a pbsa as a general formalization of a belief state in which belief constraints and assumptions are clearly separated, and we discussed the notions of conditioning and evidence combination in this general framework.

The introduced Partial Probability Theory has still a long way to go before it can be considered to be a useful alternative to probability theory or DS theory. For example, it should be investigated which formal languages for expressing probabilistic constraints are the most suitable. It is very well possible that, in its present form, PPT is too general to be a serious alternative to the mentioned elegant theories, and should be viewed as a tool for analysis rather than a formalism that may be implemented.

In the future, we plan to study the reasoning of real (resource-bounded) agents, instead of ideally rational agents. Belief states of such non-ideal agents are likely to contain base beliefs modelling the explicit beliefs and to contain many assumptions due to the fact that the limited resources may hinder the acquisition of information. Therefore, we believe that our pbsa's are closer to the kind of formalization of belief states we need for real agents, than the Bayesian probability functions or the belief functions of DS theory.

References

[1] G. Choquet, Theory of capacities, *Ann. Inst. Fourier* 5 (1953-54) 131-295.

[2] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics* 38 (1967) 325-339.

[3] J.Y. Halpern and R. Fagin, Two views of belief: belief as generalized probability and belief as evidence, *Artificial Intelligence* 54 (1992) 275-317.

[4] G. Harman, *Change in View: Principles of Reasoning*, (MIT Press Cambridge Mass., 1986).

[5] H.E. Kyburg, Bayesian and non-Bayesian evidential updating, *Artificial Intelligence* 31 (1987) 271-293.

[6] R.E. Neapolitan, A note of caution on combining certainties, *International Journal of Pattern Recognition and Artificial Intelligence* 1 (1987) 427-433.

[7] G. Shafer, *A Mathematical Theory of Evidence* (Princeton U.P., Princeton, 1976).

[8] G. Shafer, Constructive probability, *Synthese* 48 (1981) 1-60.

[9] P. Smets and R. Kennes, The transferable belief model, *Artificial Intelligence* 66 (1994) 191-234.

[10] F. Voorbraak, On the justification of Dempster's rule of combination, *Artificial Intelligence* 48 (1991) 499-515.