A new approach to updating beliefs*

Ronald Fagin
Joseph Y. Halpern

IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120

email: fagin@ibm.com, halpern@ibm.com

Abstract: We define a new notion of *conditional belief*, which plays the same role for Dempster-Shafer belief functions as conditional probability does for probability functions. Our definition is different from the standard definition given by Dempster, and avoids many of the well-known problems of that definition. Just as the conditional probability $Pr(\cdot|B)$ is a probability function which is the result of conditioning on B being true, so too our conditional belief function $Bel(\cdot|B)$ is a belief function which is the result of conditioning on B being true. We define the conditional belief as the *lower envelope* (that is, the inf) of a family of conditional probability functions, and provide a closed-form expression for it. An alternate way of understanding our definition of conditional belief is provided by considering ideas from an earlier paper [FH91], where we connect belief functions with *inner measures*. In particular, we show here how to extend the definition of conditional probability to nonmeasurable sets, in order to get notions of *inner* and *outer conditional probabilities*, which can be viewed as best approximations to the true conditional probability, given our lack of information. Our definition of conditional belief turns out to be an exact analogue of our definition of inner conditional probability.

*This article is essentially the same as one that appears in *Uncertainty in Artificial Intelligence 6* (eds. P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer), 1991, pp. 347–374.

1 Introduction

How should one update one's belief given new evidence? If beliefs are expressed in terms of probability, then the standard approach is to use conditioning. If an agent's original estimate of the probability of A is given by $Pr(A)$, and then some new evidence, say B, is acquired, then the new estimate is given by $Pr(A|B)$. By definition, $Pr(A|B)$ is $Pr(A \cap B)/Pr(B)$.[1]

The Dempster-Shafer approach to reasoning about uncertainty [Sha76] has recently become quite popular in expert systems applications (see, for example, [Abe88; Fal88; LU88; LG83]). This approach uses *belief functions*, a class of functions that satisfy three axioms, somewhat related to the axioms of probability. In this paper, we consider how to define a notion of *conditional belief*, which generalizes conditional probability.

One definition for conditional belief was already suggested by Dempster [Dem67], and is derived using the *rule of combination*; hereafter we refer to Dempster's definition as *the DS definition of conditional belief*. Although the DS definition also generalizes conditional probability, it is well known to give counterintuitive results in a number of situations (see, e.g., [Ait68; Bla87; Dia78; DZ86; Hun87; Lem86; Pea88; Pea89; Zad84]). We provide here a new definition of conditional belief, which also generalizes conditional probability, but is different from the DS definition in general. We can show that our definition avoids many of the problems associated with the DS definition.

The motivation for our definition of conditional belief comes from probability theory. It is well known that a belief function Bel is the *lower envelope* of the family of all probability functions Pr consistent with Bel.[2] That is, $Bel(A)$ is the inf of $Pr(A)$, where the inf is taken over all probability functions Pr such that $Bel(A') \leq Pr(A')$ for all A'. We therefore define the conditional belief function $Bel(A|B)$ to be the lower envelope of the family of all functions $Pr(\cdot|B)$ where Pr is consistent with Bel (similarly to the situation with conditional probability, we assume that $Bel(B) > 0$, so that everything is well defined). Although we define $Bel(\cdot|B)$ in terms of a lower envelope, we show that there is an elegant closed form expression for it. Moreover, we can show that just as the conditional probability function is in fact a probability function, our conditional belief function is a belief function.

An alternate way of understanding our definition of conditional belief is provided by considering the approach to reasoning about uncertainty we advocated in an earlier paper [FH91]. There we observed that, although in the Bayesian approach to probability, it is assumed that all relevant events can be assigned a probability, this is an assumption that is *not* made in standard probability theory. Not every subset of a sample space S

[1] This definition is not completely uncontroversial (see, e.g., [DZ86] for a discussion and further references).

[2] We remark that some authors (e.g., [DP88]) have used the term *lower probability* to denote what we are calling lower envelopes. We have used the term lower envelope here to avoid confusion with Dempster's technical usage of the phrase lower probability in [Dem67; Dem68], which, although related, is not equivalent to what we are calling a lower envelope.

need be assigned a probability; some sets can be *nonmeasurable*, i.e., not in the domain of the probability function. We can view the measurable sets as those sets for which the agent has sufficient information to assign a probability. Although we cannot talk about the probability of a nonmeasurable event E, we can talk about $Pr_*(E)$ and $Pr^*(E)$—the *inner measure* and *outer measure* of E, respectively. If E is measurable, then $Pr_*(E) = Pr(E) = Pr^*(E)$; in general, we have $Pr_*(E) \leq Pr^*(E)$. Intuitively, $Pr_*(E)$ and $Pr^*(E)$ can be viewed as the best approximations from below and above to the "true probability" of E. This intuition is made precise by a well-known result, which says that if we start with an arbitrary probability function Pr and extend it to a probability function Pr' such that E is measurable with respect to Pr', then $Pr_*(E) \leq Pr'(E) \leq Pr^*(E)$. Moreover, there are extensions of Pr that achieve the inner and outer measure; i.e., there are extensions Pr_1 and Pr_2 of Pr such that E is measurable with respect to both Pr_1 and Pr_2, and $Pr_1(E) = Pr_*(E)$ and $Pr_2(E) = Pr^*(E)$.

Notice that implicit in the definition of the conditional probability $Pr(A|B)$ as $Pr(A \cap B)/Pr(B)$ is the assumption that both $A \cap B$ and B are measurable sets. If A and B are not measurable, then the conditional probability of A given B is not well defined. What we can consider instead are the *inner* and *outer conditional probability* of A given B; these are the best lower and upper bounds on the true conditional probability in the sense outlined above. We provide formal definitions and closed form expressions for these notions.

In [FH91] it is shown that every inner measure is a belief function, and that every belief function can be viewed as an inner measure. In view of this, it is perhaps not surprising that our definition of conditional belief is completely analogous to our definition of inner conditional probability. In fact, the definition of inner conditional probability was the inspiration for our definition of conditional belief.

The rest of this paper is organized as follows. In the next section, we define the notions of inner and outer conditional probability, and apply them to analyzing the well-known *three prisoners problem* [Gar61; Dia78]. In Section 3 we apply these ideas to belief functions, define our notion of conditional belief and compare it to the DS notion. Our comparison helps to explain the cause of the counterintuitive answers provided by the DS notion. In Section 4 we briefly consider other updating rules, such as *Jeffrey's rule* [Jef83]. In Section 5 we discuss the relationship between belief functions and sets of probability functions. We conclude in Section 6 with some discussion on the implications of our results to the use of belief functions.

2 Inner and outer conditional probability

We begin by reviewing basic definitions from probability theory. Our presentation follows that of [FH91]; the reader should consult a basic probability text such as [Fel57; Hal50] for more details.

A *probability space* $(S, \mathcal{X}, Pr)$ consists of a set S (called the *sample space*), a σ-

algebra $\mathcal{X}$ of subsets of S (i.e., a set of subsets of S containing S and closed under complementation and countable union, but not necessarily consisting of all subsets of S) whose elements are called *measurable sets*, and a *probability measure* $Pr: \mathcal{X} \rightarrow [0,1]$ satisfying the following properties:

P1. $Pr(X) \geq 0$ for all $X \in \mathcal{X}$

P2. $Pr(S) = 1$

P3. $Pr(\cup_{i=1}^{\infty} X_i) = \Sigma_{i=1}^{\infty} Pr(X_i)$, if the X_i's are pairwise disjoint members of $\mathcal{X}$.

Property P3 is called *countable additivity*. Of course, the fact that $\mathcal{X}$ is closed under countable union guarantees that if each $X_i \in \mathcal{X}$, then so is $\cup_{i=1}^{\infty} X_i$. If we restrict to finite spaces, it suffices to require *finite additivity*; i.e., we can restrict attention to finite rather than infinite unions.

In a probability space $(S, \mathcal{X}, Pr)$, the probability measure Pr is not necessarily defined on 2^S (the set of all subset of S), but only on $\mathcal{X}$. If $A \in \mathcal{X}$, then we say that A is *measurable with respect to* Pr. We can extend Pr to 2^S in two standard ways, by defining functions Pr_* and Pr^*, traditionally called the *inner measure* and *outer measure induced by* Pr [Hal50]. For an arbitrary subset $A \subseteq S$, we define

$$\begin{aligned} Pr_*(A) &= \sup\{Pr(X) \mid X \subseteq A \text{ and } X \in \mathcal{X}\} \\ Pr^*(A) &= \inf\{Pr(X) \mid X \supseteq A \text{ and } X \in \mathcal{X}\}. \end{aligned}$$

If there are only finitely many measurable sets (in particular, if S is finite), then it is easy to see that the inner measure of A is the measure of the largest measurable set contained in A, while the outer measure of A is the measure of the smallest measurable set containing A.

A subset $\mathcal{Y}$ of $\mathcal{X}$ is said to be a *basis* (of $\mathcal{X}$) if the members of $\mathcal{Y}$ are nonempty and disjoint, and if $\mathcal{X}$ consists precisely of countable unions of members of $\mathcal{Y}$. It is easy to see that if $\mathcal{X}$ is finite then it has a basis. Moreover, whenever $\mathcal{X}$ has a basis, it is unique: it consists precisely of the minimal elements of $\mathcal{X}$ (the nonempty sets none of whose nonempty subsets are in $\mathcal{X}$). Note that if $\mathcal{X}$ has a basis $\mathcal{Y}$, once we know the probability of every set in the basis, we can compute the probability of every measurable set by using countable additivity; i.e., for a measurable set A, we have

$$Pr(A) = \textstyle\sum_{X \subseteq A,\, X \in \mathcal{Y}} Pr(X).$$

We can also easily compute the inner and outer measure of an arbitrary set A. It is easy to check that

$$\begin{aligned} Pr_*(A) &= \textstyle\sum_{X \subseteq A,\, X \in \mathcal{Y}} Pr(X) \\ Pr^*(A) &= \textstyle\sum_{X \cap A \neq \emptyset,\, X \in \mathcal{Y}} Pr(X). \end{aligned}$$

Since, for a measurable set A, every basis set is either a subset of A or disjoint from A, these definitions again show that $Pr_*(A) = Pr^*(A)$ for a measurable set A.

It is easy to check that for any set A we have $Pr_*(A) \leq Pr^*(A)$; if A is measurable, then $Pr_*(A) = Pr^*(A) = Pr(A)$. The inner and outer measures of a set A can be viewed as our best estimate of the "true" measure of A, given our lack of knowledge. To make this precise, we say that a probability space $(S, \mathcal{X}', Pr')$ is an *extension* of the probability space $(S, \mathcal{X}, Pr)$ if $\mathcal{X}' \supseteq \mathcal{X}$, and $Pr'(A) = Pr(A)$ for all $A \in \mathcal{X}$ (so that Pr and Pr' agree on $\mathcal{X}$, their common domain). If $(S, \mathcal{X}', Pr')$ is an extension of $(S, \mathcal{X}, Pr)$, then we say that Pr' *extends* Pr. The following result is well known (a proof can be found in [Rus87]).

Theorem 2.1: *If $(S, \mathcal{X}', Pr')$ is an extension of $(S, \mathcal{X}, Pr)$ and $A \in \mathcal{X}'$, then $Pr_*(A) \leq Pr'(A) \leq Pr^*(A)$. Moreover, there exist extensions $(S, \mathcal{X}_1, Pr_1)$, $(S, \mathcal{X}_2, Pr_2)$ of $(S, \mathcal{X}, Pr)$ such that $A \in \mathcal{X}_1$, $A \in \mathcal{X}_2$, $Pr_1(A) = Pr_*(A)$, and $Pr_2(A) = Pr^*(A)$.*

Proof: Suppose $(S, \mathcal{X}', Pr')$ is an extension of $(S, \mathcal{X}, Pr)$, and assume $A \in \mathcal{X}'$. Since $\mathcal{X} \subseteq \mathcal{X}'$, it is clear that $Pr_*(A) \leq Pr'_*(A)$. Since $A \in \mathcal{X}'$, we have $Pr'_*(A) = Pr'(A)$. So $Pr_*(A) \leq Pr'(A)$. Similarly, $Pr'(A) \leq Pr^*(A)$.

We omit the proof of the second half of the theorem here. However, it is similar to (and simpler than) the proof of Theorem 2.3, which is sketched below. ∎

Intuitively, the first part of Theorem 2.1 tells us that if we acquire extra information enabling us to compute the probability of A, then it is bound to lie somewhere between the inner measure and outer measure of A. The second part of the theorem tells us that the inner measure and outer measure are the best estimates we can get.

We would like an analogue of inner and outer measure for conditional probability. The inspiration for our definition is Theorem 2.1. Let $(S, \mathcal{X}, Pr)$ be a probability space. We define the *inner conditional probability* $Pr_*(A|B)$ and the *outer conditional probability* $Pr^*(A|B)$ of A given B as follows:

$$Pr_*(A|B) = \inf\{Pr'(A|B) \,|\, (S, \mathcal{X}', Pr') \text{ extends } (S, \mathcal{X}, Pr) \text{ and } A, B \in \mathcal{X}'\}$$
$$Pr^*(A|B) = \sup\{Pr'(A|B) \,|\, (S, \mathcal{X}', Pr') \text{ extends } (S, \mathcal{X}, Pr) \text{ and } A, B \in \mathcal{X}'\}.$$

Since the inf and sup above are not well-defined unless $Pr_*(B) > 0$, we define $Pr_*(A|B)$ and $Pr^*(A|B)$ only if $Pr_*(B) > 0$.[3] We then have the following analogue to Theorem 2.1.

Theorem 2.2: *If $(S, \mathcal{X}', Pr')$ is an extension of $(S, \mathcal{X}, Pr)$ and $A, B \in \mathcal{X}'$, then*

$$Pr_*(A|B) \leq Pr'(A|B) \leq Pr^*(A|B).$$

Moreover, there exist extensions $(S, \mathcal{X}_1, Pr_1)$, $(S, \mathcal{X}_2, Pr_2)$ of $(S, \mathcal{X}, Pr)$ such that $A, B \in \mathcal{X}_1$, $A, B \in \mathcal{X}_2$, $Pr_1(A|B) = Pr_*(A|B)$, and $Pr_2(A|B) = Pr^*(A|B)$.

[3] If S is finite, then we could just as well take $\mathcal{X}'$ in the definitions of $Pr_*(A|B)$ and $Pr^*(A|B)$ to be 2^S, so that not only are A and B measurable under Pr', but also *every* subset of S. It is straightforward to verify that this new definition is equivalent to the old definition.

Proof: The first part follows immediately from the definitions of $Pr_*(A|B)$ as an inf and of $Pr^*(A|B)$ as a sup. The second part requires showing that the inf is actually attained (and similarly for the sup); this follows from the proof of Theorem 2.3 below. ∎

Our definition of inner and outer conditional probabilities as lower and upper envelopes does not give us much help in computing these expressions. We would like to have closed-form expressions for them. Since the (usual) conditional probability $Pr(A|B)$ of A given B is defined as $Pr(A \cap B)/Pr(B)$ (provided that A and B are measurable sets and $Pr(B) \neq 0$), we might guess that $Pr_*(A|B)$ is $Pr_*(A \cap B)/Pr^*(B)$, which is obtained by taking the smallest possible numerator and the largest possible denominator. Note that this gives the right answer when A and B are measurable. Similarly, we might guess that $Pr^*(A|B)$ is $Pr^*(A \cap B)/Pr_*(B)$, which is obtained by taking the largest possible numerator and the smallest possible denominator. However, we now show that these formulas are incorrect. For later reference, let us define $Pr^-(A|B) = Pr_*(A \cap B)/Pr^*(B)$, and $Pr^+(A|B) = Pr^*(A \cap B)/Pr_*(B)$. Suppose that A is not a measurable set and $Pr_*(A) < Pr^*(A)$. Then $Pr^-(A|A) < 1$. However, it is clear that $Pr_*(A|A) = 1$, since $Pr_*(A|A)$ is defined to be an inf of terms $Pr'(A|A)$, each of which equals 1. Similarly, under our assumptions $Pr^+(A|A) > 1$, whereas $Pr^*(A|A) = 1$.

We can in fact give elegant closed-form expressions for the inner and outer conditional probabilities. Taking $\overline{A}$ to be the complement of A, we can show that

$$
\begin{aligned}
Pr_*(A|B) &= \frac{Pr_*(A \cap B)}{Pr_*(A \cap B) + Pr^*(\overline{A} \cap B)} \\
Pr^*(A|B) &= \frac{Pr^*(A \cap B)}{Pr^*(A \cap B) + Pr_*(\overline{A} \cap B)}
\end{aligned}
$$

The intuition behind the formula for $Pr_*(A|B)$ is that rather than having $Pr^*(B)$ in the denominator, we divide B up into two parts: $A \cap B$ and $\overline{A} \cap B$. On $A \cap B$ we use the inner measure, since we have already used the inner measure in estimating the likelihood to assign to $A \cap B$ in the numerator. It is only on $\overline{A} \cap B$ that we use the outer measure. A similar intuition holds for $Pr^*(A|B)$. Note that the formulas give us the correct value of 1 when $A = B$, even if A is not measurable.

Although we do not bother with the proof here, it is not hard to show that $Pr_*(A \cap B) + Pr^*(\overline{A} \cap B) \leq Pr^*(B)$. Thus we get that $Pr^-(A|B) \leq Pr_*(A|B)$, and similarly $Pr^*(A|B) \leq Pr^+(A|B)$. So the interval defined by $[Pr_*(A|B), Pr^*(A|B)]$ is nested within the interval defined by $[Pr^-(A|B), Pr^+(A|B)]$. As we saw, this nesting may be proper, even when $A = B$.

The expressions given above for the inner and outer conditional probability are very natural and well motivated. Not surprisingly, it turns out that other authors have discovered them as well. In particular, essentially these expressions appear in [Wal81], [SK89], and [dCLM90]. Indeed, it even appears (lost in a welter of notation) as Equation 4.8 in [Dem67]! (Interestingly, none of these papers references any other work as the source of the formula.)

For the sake of completeness, we now prove that the closed-form expressions do indeed work as claimed.

Theorem 2.3: *For any probability function Pr on S and subsets $A, B \subseteq S$ such that $Pr_*(B) > 0$,*

$$Pr_*(A|B) = \frac{Pr_*(A \cap B)}{Pr_*(A \cap B) + Pr^*(\overline{A} \cap B)}$$
$$Pr^*(A|B) = \frac{Pr^*(A \cap B)}{Pr^*(A \cap B) + Pr_*(\overline{A} \cap B)}$$

Proof: Let us denote the right-hand side of the first equality in the statement of the theorem by $P_*(A|B)$, and the right-hand side of the second equality in the statement of the theorem by $P^*(A|B)$. Assume that $Pr_*(B) > 0$ (so that $Pr_*(A|B)$ and $Pr^*(A|B)$ are defined). We need only show that Theorem 2.2 holds when we replace $Pr_*(A|B)$ by $P_*(A|B)$ and $Pr^*(A|B)$ by $P^*(A|B)$. Let Pr' be a probability function extending Pr such that A and B are measurable with respect to Pr'. We first show that $P_*(A|B) \leq Pr'(A|B)$. From Theorem 2.1, we have that (1) $Pr_*(A \cap B) \leq Pr'(A \cap B)$ and (2) $Pr^*(\overline{A} \cap B) \geq Pr'(\overline{A} \cap B)$. From (1), it follows that that

$$\begin{aligned} P_*(A|B) &= \frac{Pr_*(A \cap B)}{Pr_*(A \cap B) + Pr^*(\overline{A} \cap B)} \\ &\leq \frac{Pr'(A \cap B)}{Pr'(A \cap B) + Pr^*(\overline{A} \cap B)}. \end{aligned}$$

(The inequality is just an application of the general fact that $x/(x+y) \leq x'/(x'+y)$ if $x, y \geq 0$ and $x \leq x'$.) From (2), it follows that

$$\begin{aligned} \frac{Pr'(A \cap B)}{Pr'(A \cap B) + Pr^*(\overline{A} \cap B)} &\leq \frac{Pr'(A \cap B)}{Pr'(A \cap B) + Pr'(\overline{A} \cap B)} \\ &= \frac{Pr'(A \cap B)}{Pr'(B)} \\ &= Pr'(A|B). \end{aligned}$$

The proof that $Pr'(A|B) \leq P^*(A|B)$ is similar and left to the reader.

We next show that there exists some extension $(S, \mathcal{X}_1, Pr_1)$ of (S, X, Pr) such that $A, B \in \mathcal{X}_1$ and $Pr_1(A|B) = P_*(A|B)$. This is the probability function Pr_1 required to complete the proof of Theorem 2.2 We just briefly sketch the ideas here, leaving the details (as well as the construction of the corresponding probability function Pr_2 such that $Pr_2(A|B) = P^*(A|B)$) to the reader. We take $\mathcal{X}_1$ to consist of all sets of the form $(A \cap B \cap X_1) \cup (\overline{A} \cap B \cap X_2) \cup (A \cap \overline{B} \cap X_3) \cup (\overline{A} \cap \overline{B} \cap X_4)$, where $X_1, \ldots, X_4 \in \mathcal{X}$. It is straightforward to show that $\mathcal{X}_1$ is closed under complementation and countable union. Further, $\mathcal{X}_1 \supseteq \mathcal{X}$, since if $X \in \mathcal{X}$, then $X = (A \cap B \cap X) \cup (\overline{A} \cap B \cap X) \cup (A \cap \overline{B} \cap X) \cup (\overline{A} \cap \overline{B} \cap X)$. We define Pr_1 on $\mathcal{X}_1$ as follows:

$$\begin{aligned} & Pr_1((A \cap B \cap X_1) \cup (\overline{A} \cap B \cap X_2) \cup (A \cap \overline{B} \cap X_3) \cup (\overline{A} \cap \overline{B} \cap X_4)) \\ = \; & Pr_*(A \cap B \cap X_1) + Pr^*(\overline{A} \cap B \cap X_2) + Pr_*(A \cap X_3) - Pr_*(A \cap B \cap X_3) \\ & + Pr^*(\overline{A} \cap X_4) - Pr^*(\overline{A} \cap B \cap X_4). \end{aligned}$$

This is nonnegative because $Pr_*(A \cap X_3) - Pr_*(A \cap B \cap X_3) \geq 0$ (since $(A \cap X_3) \supseteq (A \cap B \cap X_3)$), and similarly for $Pr^*(\overline{A} \cap X_4) - Pr^*(\overline{A} \cap B \cap X_4)$.

It is easy to see that for a set $X \in \mathcal{X}$, we have

$$\begin{aligned} Pr_1(X) &= Pr_1((A \cap B \cap X) \cup (\overline{A} \cap B \cap X) \cup (A \cap \overline{B} \cap X) \cup (\overline{A} \cap \overline{B} \cap X)) \\ &= Pr_*(A \cap X) + Pr^*(\overline{A} \cap X) \\ &= Pr(X), \end{aligned}$$

so that Pr_1 and Pr agree on sets in $\mathcal{X}$. In particular, we have that $Pr_1(\emptyset) = 0$ and $Pr_1(S) = 1$. We leave it to the reader to check that if $X_1, X_2, \ldots$ are pairwise disjoint sets in $\mathcal{X}_1$, then $Pr_1(\cup_n X_n) = \sum_n Pr_1(X_n)$.

By definition of Pr_1, it follows that $Pr_1(A \cap B) = Pr_*(A \cap B)$ and $Pr_1(\overline{A} \cap B) = Pr^*(\overline{A} \cap B)$. Since $A \cap B$ and $\overline{A} \cap B$ are measurable with respect to Pr_1, we have that $Pr_1(B) = Pr_1(A \cap B) + Pr_1(\overline{A} \cap B)$. It is now immediate that $Pr_1(A|B) = P_*(A|B)$. ∎

Example 1 (The three prisoners problem): In order to illustrate our updating technique, we consider the well-known *three prisoners problem.*[4]

> Of three prisoners a, b, and c, two are to be executed but a does not know which. He therefore says to the jailer, "Since either b or c is certainly going to be executed, you will give me no information about my own chances if you give me the name of one man, either b or c, who is going to be executed." Accepting this argument, the jailer truthfully replies, "b will be executed." Thereupon a feels happier because before the jailer replied, his own chance of execution was two-thirds, but afterwards there are only two people, himself and c, who could be the one not executed, and so his chance of execution is one-half.

Note that in order for a to believe that his own chance of execution was two-thirds before the jailer replied, he seems to be implicitly assuming that the one to get pardoned

[4]For an excellent introduction to the problem as well as a Bayesian solution, see [Gar61]. Our description of the story is taken from [Dia78] and much of our discussion is based on that of Diaconis and Zabell [Dia78; DZ86].

is chosen at random from among a, b, and c. We make this assumption explicit in the remainder of our discussion.

Is a justified in believing that his chances of escaping have improved? It seems that the jailer did not give him any relevant extra information. Yet how could a's subjective probabilities change if he does not acquire any relevant extra information?

Following [DZ86], we model a possible situation by a pair (x, y), where $x, y \in \{a, b, c\}$. Intuitively, a pair (x, y) represents a situation where x is pardoned and the jailer says that y will be executed in response to a's question. Since the jailer answers truthfully, we cannot have $x = y$; since the jailer will never tell a directly that a will be executed, we cannot have $y = a$. Thus, the set of possible outcomes is $\{(a, b), (a, c), (b, c), (c, b)\}$. The event that a lives, which we denote *lives-a*, corresponds to the set $\{(a, b), (a, c)\}$. Similarly, we define the events *lives-b* and *lives-c*, which correspond to the sets $\{(b, c)\}$ and $\{(c, b)\}$, respectively. By assumption, each of these three events has probability 1/3.

The event that the jailer says b, which we denote *says-b*, corresponds to the set $\{(a, b), (c, b)\}$; the story does not give us a probability for this event. In order to do a Bayesian analysis of the situation, we will need this probability. Note that we do know that the probability of $\{(c, b)\}$ is 1/3; we just need to know the probability of $\{(a, b)\}$. This depends on the jailer's strategy in the one case that he has free choice, namely when a lives. He gets to choose between saying b and c in that case. We need to know the probability that he says b; i.e., $Pr(says\text{-}b|lives\text{-}a)$.

If we assume that the jailer chooses at random between saying b and c if a is pardoned, so that $Pr(says\text{-}b|lives\text{-}a) = 1/2$, then $Pr(\{(a, b)\}) = Pr(\{((a, c)\}) = 1/6$, and $Pr(says\text{-}b) = 1/2$. We can now easily compute that

$$Pr(lives\text{-}a|says\text{-}b) = Pr(lives\text{-}a \cap says\text{-}b)/Pr(says\text{-}b) = (1/6)/(1/2) = 1/3.$$

Thus, in this case, the jailer's answer does not affect a's probability.

Suppose more generally that $Pr(says\text{-}b|lives\text{-}a) = \alpha$, for $0 \leq \alpha \leq 1$. Then straightforward computations show that

$$\begin{array}{c} Pr(\{(a, b)\}) = Pr(lives\text{-}a) \times Pr(says\text{-}b|lives\text{-}a) = \alpha/3, \\ Pr(says\text{-}b) = Pr(\{(a, b)\}) + Pr(\{(c, b)\}) = (\alpha + 1)/3, \text{ and} \\ Pr(lives\text{-}a|says\text{-}b) = \frac{\alpha/3}{(\alpha+1)/3} = \alpha/(\alpha + 1). \end{array}$$

This says that if $\alpha \neq 1/2$ (i.e., if the jailer had a particular preference for answering either b or c when a was the one pardoned), then a would learn something from the answer, in that he would change his estimate of the probability that he will be executed. For example, if $\alpha = 0$, then if a is pardoned, the jailer will definitely say c. Thus, if the jailer actually says b, then a knows that he is definitely not pardoned; i.e., that $Pr(lives\text{-}a|says\text{-}b) = 0$. Similarly, if $\alpha = 1$, then a knows that if either he or c is pardoned, then the jailer will say b, while if b is pardoned the jailer will say c. Given that the jailer says b, then from a's point of view the one pardoned is equally likely to be him or c; thus, $Pr(lives\text{-}a|says\text{-}b) = 1/2$. As α ranges from 0 to 1, it is easy to check that $Pr(lives\text{-}a|says\text{-}b)$ ranges from 0 to 1/2.

Rather than assuming that there is some unknown probability α that the jailer will say b given that a is pardoned, suppose we instead capture this situation using the idea of non-measurable sets. Thus, we take *lives-a*, *lives-b*, and *lives-c* as a basis for the measurable sets; in particular, neither of the singleton sets $\{(a,b)\}$ and $\{(a,c)\}$ is measurable, since we are not given the probability that the jailer will say b (resp. c) if a is pardoned. An easy computation shows that (1) $Pr_*(lives\text{-}a \cap says\text{-}b) = Pr_*(\{(a,b)\}) = 0$ (since there are no nonempty measurable subsets of $\{(a,b)\}$), (2) $Pr^*(lives\text{-}a \cap says\text{-}b) = Pr^*(\{(a,b)\}) = 1/3$, and (3) $Pr_*(\overline{lives\text{-}a} \cap says\text{-}b) = Pr^*(\overline{lives\text{-}a} \cap says\text{-}b) = Pr(\{(c,b)\} = 1/3$. It follows that

$$Pr_*(lives\text{-}a|says\text{-}b) = \frac{Pr_*(lives\text{-}a \cap says\text{-}b)}{Pr_*(lives\text{-}a \cap says\text{-}b) + Pr^*(\overline{lives\text{-}a} \cap says\text{-}b)} = 0,$$

$$Pr^*(lives\text{-}a|says\text{-}b) = \frac{Pr^*(lives\text{-}a \cap says\text{-}b)}{Pr^*(lives\text{-}a \cap says\text{-}b) + Pr_*(\overline{lives\text{-}a} \cap says\text{-}b)} = 1/2.$$

Notice that the range 0 to 1/2 is precisely that obtained in the Bayesian analysis by letting α range from 0 to 1. This should be no surprise; it is a consequence of Theorem 2.2. ∎

3 Updating belief functions

The Dempster-Shafer theory of evidence [Sha76] provides another approach to attaching likelihoods to events. This approach is meant to be an alternative to probability theory. The theory starts out with a *belief function*. For every event (i.e., set) A, the belief in A, denoted $Bel(A)$, is a number in the interval $[0,1]$ that places a lower bound on likelihood of A. We have a corresponding number $Pl(A) = 1 - Bel(\overline{A})$, called the *plausibility* of A, which places an upper bound on the likelihood of A. Thus, to every event A we can attach the interval $[Bel(A), Pl(A)]$. Like a probability measure, a belief function assigns a "weight" to subsets of a set S, but unlike a probability measure, the domain of a belief function is always taken to be *all* subsets of S. Formally, a belief function Bel on a set S is a function $Bel: 2^S \rightarrow [0,1]$ satisfying:

B0. $Bel(\emptyset) = 0$

B1. $Bel(A) \geq 0$

B2. $Bel(\mathrm{S}) = 1$

B3. $Bel(A_1 \cup \ldots \cup A_k) \geq \sum_{I \subseteq \{1,\ldots,k\}, I \neq \emptyset} (-1)^{|I|+1} Bel(\bigcap_{i \in I} A_i)$.

As long as we restrict to finite spaces, there is another formulation of belief functions that is perhaps more intuitive. A *mass function* is simply a function $m: 2^S \rightarrow [0,1]$ such that

M1. $m(\emptyset) = 0$

M2. $\sum_{A \subseteq S} m(A) = 1$.

Intuitively, $m(A)$ is the weight of evidence for A that has not already been assigned to some proper subset of A. With this interpretation of mass, we would expect that an agent's belief in A is the sum of the masses he has assigned to all the subsets of A; i.e., $Bel(A) = \sum_{B \subseteq A} m(B)$. Indeed, this intuition is correct.

Proposition 3.1: ([Sha76, p. 39])

1. *If m is a mass function on S, then the function $Bel: 2^S \rightarrow [0,1]$ defined by $Bel(A) = \sum_{B \subseteq A} m(B)$ is a belief function.*

2. *If Bel is a belief function on 2^S and S is finite, then there is a unique mass function m on 2^S such that $Bel(A) = \sum_{B \subseteq A} m(B)$ for every subset A of S.*

Note: From here on, we restrict for convenience to *finite* sample spaces S, both when we discuss belief functions and probability functions. Because of this assumption, (a) there is always a mass function for every belief function, as in part (2) of Proposition 3.1, and (b) every probability space we shall consider has a basis. We note that many of our results would still go through without the assumption of finiteness.

The interval defined by $[Bel(A), Pl(A)]$ may strike the reader as somewhat reminiscent of the interval $[Pr_*(A), Pr^*(A)]$ defined by the inner and outer measure. This similarity is not accidental. As is shown in [FH91], every inner measure induced by a probability function is a belief function. In fact, if Pr is a probability function defined on a set $\mathcal{X}$ of measurable subsets of a finite set S, and $\mathcal{Y}$ is a basis of $\mathcal{X}$, let m be the mass function such that

$$m(A) = \begin{cases} Pr(A) & \text{if } A \in \mathcal{Y} \\ 0 & \text{otherwise,} \end{cases}$$

and let Bel be the belief function corresponding to m. Then it is easy to show that $Bel(B) = Pr_*(B)$ for all $B \subseteq S$. Thus, Bel agrees with Pr on the measurable sets and, more generally, is equal to the inner measure on arbitrary subsets. The corresponding plausibility function Pl also agrees with Pr on the measurable sets, and is equal to the outer measure on arbitrary subsets. We call Bel (resp. Pl) *the belief* (resp. *plausibility*) *function corresponding to Pr.*

As shown in [FH91], there is also a strong sense in which every belief function can be viewed as an inner measure induced by a probability function.

Theorem 3.2: [FH91] *Given a belief function Bel defined on a finite set S, there is a probability space $(S', \mathcal{X}, Pr)$ and a surjection $f : S' \rightarrow S$ such that for each $A \subseteq S$, we have $Bel(A) = Pr_*(f^{-1}(A))$.*

Given a set $\mathcal{P}$ of probability functions all defined on a sample space S, define the *lower envelope* of $\mathcal{P}$ to be the function f such that for each $A \subseteq S$, we have $f(A) = \inf\{Pr(A) : Pr \in \mathcal{P}$ and A is measurable with respect to $Pr\}$. We have the corresponding definition of *upper envelope* of $\mathcal{P}$. Theorem 2.1 says that the inner measure induced by a probability function Pr can be viewed as the lower envelope of the family of probability functions extending Pr; the outer measure is the corresponding upper envelope. Since by Theorem 3.2, a belief function is essentially an inner measure, this suggests that a belief function can also be viewed as a lower envelope. This is true, and was already known to Dempster [Dem67]. Let Bel be a belief function defined on S, and let $(S, \mathcal{X}, Pr)$ be a probability space with sample space S. We say that Pr is *consistent with* Bel if $Bel(A) \leq Pr(A) \leq Pl(A)$ for each $A \in \mathcal{X}$. Intuitively, Pr is consistent with Bel if the probabilities assigned by Pr is consistent with the intervals $[Bel(A), Pl(A)]$ given by the belief function Bel. It is easy to see that Pr is consistent with Bel if $Bel(A) \leq Pr(A)$ for each $A \in \mathcal{X}$ (that is, it follows automatically that $Pr(A) \leq Pl(A)$ for each $A \in \mathcal{X}$). This is because $Pl(A) = 1 - Bel(\overline{A}) \geq 1 - Pr(\overline{A}) = Pr(A)$. Let $\mathcal{P}_{Bel}$ be the set of all probability functions defined on 2^S consistent with Bel. The next theorem tells us that the belief function Bel is the lower envelope of $\mathcal{P}_{Bel}$, and Pl is the upper envelope.

Theorem 3.3: [Dem67] *Let Bel be a belief function on S. Then for all $A \subseteq S$,*

$$Bel(A) = \inf_{Pr \in \mathcal{P}_{Bel}} Pr(A)$$
$$Pl(A) = \sup_{Pr \in \mathcal{P}_{Bel}} Pr(A).$$

Proof: The theorem follows from Theorem 2.1 (which says that each inner measure is an infimum of a family of probability functions) and from Theorem 3.2 (which says that belief functions are essentially inner measures). The details, which are omitted, are in the same spirit as in the proof of Theorem 3.4 below. ∎

Using techniques similar to those of Theorem 2.3, we can show that the infimum and supremum are actually attainable, i.e., for each $A \subseteq S$, there are probability functions Pr_1 and Pr_2 in $\mathcal{P}_{Bel}$ such that $Bel(A) = Pr_1(A)$ and $Pl(A) = Pr_2(A)$. We remark that the converse to Theorem 3.3 does not hold: not every lower envelope is a belief function. Counterexamples are well known [Dem67; Kyb87; Bla87]. We return to this issue in Section 5.

Theorem 3.3 suggests how we might update a belief function to a *conditional belief function* (and a plausibility function to a *conditional plausibility function*):

$$Bel(A|B) = \inf_{Pr \in \mathcal{P}_{Bel}} Pr(A|B)$$
$$Pl(A|B) = \sup_{Pr \in \mathcal{P}_{Bel}} Pr(A|B).$$

It is not hard to see that the inf and sup above are not well-defined unless $Bel(B) > 0$; therefore, we define $Bel(A|B)$ and $Pl(A|B)$ only if $Bel(B) > 0$. It is straightforward to check that if Pr is a probability function, Bel is the belief function corresponding to Pr, and A and B are measurable sets with respect to Pr, then $Bel(A|B) = Pr(A|B)$. Thus, our definition of conditional belief generalizes that of conditional probability. Note that taking $B = true$ in the preceding definition, we get as a special case that $Bel(A) = \inf_{Pr \in \mathcal{P}_{Bel}} Pr(A)$ and $Pl(A) = \sup_{Pr \in \mathcal{P}_{Bel}} Pr(A)$, which is Theorem 3.3 above.

Because of the close analogy between our definitions of conditional inner measures and conditional belief functions, and the fact that inner measures and belief functions are essentially the same, we might suspect that a closed-form formula for the conditional belief function can be obtained by replacing inner measures in Theorem 2.3 by belief functions and outer measures by plausibility functions. The next theorem says that this is indeed the case.

Theorem 3.4: *If Bel is a belief function on S such that $Bel(B) > 0$, then*

$$Bel(A|B) = \frac{Bel(A \cap B)}{Bel(A \cap B) + Pl(\overline{A} \cap B)}$$

$$Pl(A|B) = \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(\overline{A} \cap B)}.$$

Proof: We consider $Bel(A|B)$ here. The proof for $Pl(A|B)$ is similar and left to the reader. The result follows from Theorem 3.2. Let $(S', \mathcal{X}, Pr)$ be the probability space guaranteed to exist by the theorem and f the surjection from S' onto S such that for each $A \subseteq S$, we have $Bel(A) = Pr_*(f^{-1}(A))$. Now it is easy to show that every probability function Pr' extending Pr defined on all the sets of the form $f^{-1}(A)$, $A \subseteq S$, can be projected to a probability function Pr'_S on S extending Bel (where $Pr'_S(A) = Pr'(f^{-1}(A))$); conversely, for every probability function Pr'_S extending Bel we can find a probability function Pr' defined on all the sets $f^{-1}(A)$ such that Pr' projects to Pr'_S. The result now follows from Theorem 2.2. ∎

It is well known that the conditional probability function is a probability function. That is, if we start with a probability function Pr defined on a σ-algebra $\mathcal{X}$ of subsets of S and if $B \in \mathcal{X}$ and $Pr(B) > 0$, then the function $Pr(\cdot|B)$ defined on $\mathcal{X}$ is a probability function. We might hope that the same situation holds with belief functions, so that the conditional belief and plausibility functions are indeed belief and plausibility functions. Given our definitions of conditional belief and plausibility as lower and upper envelopes, it is not clear that this should be so, since lower and upper envelopes of arbitrary sets of probability functions do not in general result in belief and plausibility functions. Fortunately, as the next result shows, in this case they do. Thus, we have a way of updating belief and plausibility functions to give us new belief and plausibility functions in the light of new information.

Theorem 3.5: *Let Bel be a belief function defined on S, and Pl the corresponding plausibility function. Let $B \subseteq S$ be such that $Bel(B) > 0$. Then $Bel(\cdot|B)$ is a belief function and $Pl(\cdot|B)$ is the corresponding plausibility function.*

The proof of Theorem 3.5 is somewhat difficult; details can be found in the appendix, where we also discuss further technical properties of these definitions. We remark that this result—which we view as the main technical result of the paper—appears in none of the papers cited above that contain the expression for conditional belief that appears in Theorem 3.4. In [dCLM90] the question of whether $Bel(\cdot|B)$ is a belief function is discussed, but left unanswered. Theorem 3.5 was proven independently by Jaffray [Jaf90], with a somewhat different proof. In response to an early draft of this paper, Zhang [Zha89] constructed another proof along very different lines. Very recently, Sundberg and Wagner [SW91] proved a stronger result. Let us say that a function f is *k-monotone* if

$$f(A_1 \cup \ldots \cup A_k) \geq \sum_{I \subseteq \{1,\ldots,k\},\, I \neq \emptyset} (-1)^{|I|+1} f(\bigcap_{i \in I} A_i),$$

for each choice of $A_1, \ldots, A_k$. A belief function is said to be ∞-monotone, since it is k-monotone for each k. Sundberg and Wagner showed the interesting result that if a function f is k-monotone and if $f(B) > 0$, then $f(\cdot|B)$ is also k-monotone. In particular, it follows immediately that if f is ∞-monotone, then so is $f(\cdot|B)$. In the case when $k = 2$, Sundberg and Wagner's result was already proved by Walley [Wal81] (the $k = 1$ case is trivial).

As we mentioned in the introduction, our definition is quite different from that given by Dempster. Given a belief function Bel, Dempster defines a conditional belief function $Bel(\cdot||B)$ as follows [Sha76, p. 97]:[5]

$$Bel(A||B) = \frac{Bel(A \cup \overline{B}) - Bel(\overline{B})}{1 - Bel(\overline{B})}.$$

The corresponding plausibility function is shown to satisfy:

$$Pl(A||B) = \frac{Pl(A \cap B)}{Pl(B)}.$$

If Bel is the belief function corresponding to a probability function Pr, and both A and B are measurable sets with respect to Pr, then it is easy to show $Bel(A||B) = Pr(A|B)$. Thus, the DS definition of conditional belief generalizes the idea of conditional probability, just as our definition of conditional belief does. However, a brief glance at the DS definition compared with the formula in Theorem 2.3 should convince the reader that in general these two definitions of conditional belief will not agree. We give several characterizations of $Bel(\cdot|B)$ and $Bel(\cdot||B)$ that clarify their relationship. We first present an example where the two definitions differ, and show that the DS definition gives counterintuitive answers.

[5] Dempster's definition is usually given as a special application of a more general *rule of combination* for belief functions. It would take us too far afield here to discuss the rule of combination; see the companion paper [HF92] for a discussion of the role of the rule of combination.

Example 2 (The three prisoners revisited): We consider the result of applying the two definitions of conditional belief to analyzing the three prisoners problem. Suppose we are not given a probability that the jailer will say b given that a is pardoned. Thus, using the same notation as in the previous section, we assume that the only measurable sets are those generated by the basis *lives-a*, *lives-b*, and *lives-c*. This means that, for example, we cannot assign a probability to the event that b lives and a is pardoned. Let Pr be the probability function that assigns probability 1/3 to each of these basis sets, and let Bel and Pl be the belief function and plausibility functions, respectively, corresponding to Pr. Using the definition above, it is easy to check that $Bel(\mathit{lives\text{-}a}||\mathit{says\text{-}b}) = Pl(\mathit{lives\text{-}a}||\mathit{says\text{-}b}) = 1/2$. Thus, for the DS notion of conditional probability, the range reduces to the single point 1/2. By way of contrast, recall that we showed in the previous section that $Bel(\mathit{lives\text{-}a}|\mathit{says\text{-}b}) = Pr_*(\mathit{lives\text{-}a}|\mathit{says\text{-}b}) = 0$ while $Pl(\mathit{lives\text{-}a}|\mathit{says\text{-}b}) = Pr^*(\mathit{lives\text{-}a}|\mathit{says\text{-}b}) = 1/2$. ∎

This example shows that the two notions of conditioning can give quite different answers. As we mentioned in the previous section, the range $[0, 1/2]$ determined by $Bel(\mathit{lives\text{-}a}|\mathit{says\text{-}b})$ and $Pl(\mathit{lives\text{-}a}|\mathit{says\text{-}b})$ has a natural probabilistic interpretation: it is determined by taking the probability that the jailer will say b in the one situation that he has a choice between saying b and c, namely, when a is the one pardoned, and letting it range from 0 to 1.

The range $[1/2, 1/2]$ determined by the DS notion seems much more mysterious. The answer 1/2 corresponds to the situation where the jailer says b whenever he can (i.e., whenever a is pardoned or c is pardoned). Why is this a reasonable answer? More importantly, why does it arise? Is there a natural probabilistic interpretation for it?

In order to investigate this question carefully, we first give a characterization of DS conditioning in terms of probability. This characterization shows that we can also interpret the DS conditioning in terms of an inner measures. Using this characterization, we show that the DS definition arises by giving a (somewhat unnatural) twist to a standard probabilistic process for getting our notion conditional belief. This will allow us to give a precise explanation for the answer 1/2 obtained by the DS notion in the three prisoners problem, and, more generally, characterize the conditions under which it will give appropriate answers.

Suppose we start out with a probability space $(S, \mathcal{X}, Pr)$, with a basis $\mathcal{Y}$ for $\mathcal{X}$. (Recall that we are assuming that the probability space is finite, so there is a basis.) Suppose we now observe $B \subseteq S$, where $Pr^*(B) \neq 0$. We construct the conditional probability space $(S, \mathcal{X}_B, Pr_B)$ as follows: $\mathcal{X}_B$ is the space generated by taking as basis the set $\mathcal{Y}_B$ consisting of all the nonempty sets of the form $X \cap B$ and $X \cap \overline{B}$, for $X \in \mathcal{Y}$. If $X \in \mathcal{Y}$, we define $Pr_B(X \cap \overline{B}) = 0$, while

$$Pr_B(X \cap B) = \begin{cases} 0 & \text{if } X \cap B = \emptyset \\ Pr(X)/Pr^*(B) & \text{otherwise.} \end{cases}$$

We then extend Pr_B by additivity. Note that Pr_B is indeed a well-defined probability. In particular, since $\sum_{X \cap B \neq \emptyset, X \in \mathcal{Y}} Pr(X) = Pr^*(B)$, it follows that $Pr_B(S) = 1$. This method of constructing the conditional probability space seems quite natural. Observe that if B is measurable, then $\mathcal{X}_B = \mathcal{X}$ (since every basis set is either a subset of B or disjoint from B) and $Pr_B = Pr(\cdot|B)$. As the following result shows, even if B is not measurable, this process defines the DS notion of conditioning.

Proposition 3.6: *Let $(S, \mathcal{X}, Pr)$ be a probability space, let Bel be the belief function corresponding to Pr, and let $B \subseteq S$ such that $Pr^*(B) > 0$. Then $(Pr_B)_* = Bel(\cdot||B)$ and $(Pr_B)^* = Pl(\cdot||B)$.*

Proof: Given the corresponding relationships between inner and outer measure on the one hand and belief and plausibility on the other, it suffices to show that $(Pr_B)^* = Pl(\cdot||B)$. Let $\mathcal{Y}$ be the basis for $\mathcal{X}$. Recall that the outer measure of a set A is the sum of the measures of all the basis sets that intersect A and that $Pl(A) = Pr^*(A)$. Thus, given an arbitrary set $A \subseteq S$, we have

$$\begin{aligned} (Pr_B)^*(A) &= \textstyle\sum_{X \cap B \cap A \neq \emptyset, X \in \mathcal{Y}} Pr_B(X \cap B) \\ &= \textstyle\sum_{X \cap B \cap A \neq \emptyset, X \in \mathcal{Y}} Pr(X)/Pr^*(B) \\ &= Pr^*(A \cap B)/Pr^*(B) \\ &= Pl(A \cap B)/Pl(B) \\ &= Pl(A||B). \end{aligned}$$

This completes the proof. ∎

In order to explain the difference between $Bel(\cdot|B)$ and $Bel(\cdot||B)$, we now prove an analogous result for $Bel(\cdot|B)$. Given a probability space $(S, \mathcal{X}, Pr)$ with basis $\mathcal{Y}$ and set $B \subseteq S$, let $\mathcal{P}_B$ consist of all probability functions Pr' defined on $\mathcal{X}_B$ such that Pr' is of the form cPr'', where c is a normalizing constant and Pr'' satisfies the following constraints: If $X \in \mathcal{Y}$ then $Pr''(X \cap \overline{B}) = 0$, while

$$Pr''(X \cap B) \begin{cases} = 0 & \text{if } X \cap B = \emptyset \\ = Pr(X) & \text{if } X \subseteq B \\ \in [0, Pr(X)] & \text{if } X \cap B \neq \emptyset,\ X \cap \overline{B} \neq \emptyset. \end{cases}$$

It is easy to see that $Pr_B \in \mathcal{P}_B$ (we just take $Pr''(X \cap B) = Pr(X)$ for all X such that $X \cap B \neq \emptyset$; the appropriate normalizing constant in this case, as we observed above, is $1/Pr^*(B)$). It is the set $\mathcal{P}_B$ of probability functions that characterizes $Bel(\cdot|B)$, as the following proposition shows:

Proposition 3.7: *Let $(S, \mathcal{X}, Pr)$ be a probability space, let Bel be the belief function corresponding to Pr, and let $B \subseteq S$ be such that $Pr_*(B) > 0$. Then for all sets $A \subseteq S$, we have $Bel(A|B) = \min_{Pr' \in \mathcal{P}_B} Pr'(A)$ and $Pl(A|B) = \max_{Pr' \in \mathcal{P}_B} Pr'(A)$.*

Proof: Since $Bel(A|B) = Pr_*(A|B)$, it follows from Theorem 2.2 that $Bel(A|B) = \min_{Pr'} Pr'(A|B)$, where the minimum is taken over all probability functions Pr' extending Pr. It follows easily from the proof of Theorem 2.3 that if we start with a probability space $(S, \mathcal{X}, Pr)$ with a basis, then the minimum is actually taken by a probability function in $\mathcal{P}_B$. The proof in the case of the $Pl(A|B)$ is similar. We leave details to the reader. ∎

Propositions 3.6 and 3.7 show that the difference between our notion of conditional belief and the DS notion of conditional belief comes in the treatment of basis sets X such that both $X \cap B \neq \emptyset$ and $X \cap \overline{B} \neq \emptyset$. For such basis sets, the DS notion gives probability $Pr(X)/Pr^*(B)$, while ours essentially allows the probability to vary from 0 to $cPr(X)$ (where c is the appropriate normalizing constant). In particular, this means that we allow for the possibility that all the probability of X is actually located in B or that none of it is. Note that if B is a measurable set, then there are no basis sets that have a nonempty intersection with both B and $\overline{B}$, so that both notions of conditional belief agree. In general, as an immediate corollary to the observation that $Pr_B \in \mathcal{P}_B$ and Propositions 3.6 and 3.7, we get the following observation, which was already known to Dempster [Dem67; Dem68].

Corollary 3.8: *If Bel is a belief function defined on S and $A, B \subseteq S$, then*

$$Bel(A|B) \leq Bel(A||B) \leq Pl(A||B) \leq Pl(A|B).$$

Thus, in general, $Bel(\cdot||B)$ and $Pl(\cdot||B)$ will define a smaller interval than $Bel(\cdot|B)$ and $Pl(\cdot|B)$. However, as we saw in the discussion of the three prisoners problem, this smaller interval may not always be justified.

We conclude our discussion with a construction that may further explain the difference between $Bel(\cdot|B)$ and $Bel(\cdot||B)$. This construction is a generalization of the "beehive" example in [SK89] (as well as being a formalization of some comments made in [dCLM90]).

Suppose a set S is partitioned into (nonempty) disjoints sets $X_1, \ldots, X_k$. An agent chooses X_i with probability a_i ($a_1 + \cdots + a_k = 1$) and then chooses $x \in X_i$ with some unknown probability. (Equivalently, we can view $(S, \mathcal{X}, Pr_0)$ as a probability space with basis $\mathcal{Y} = \{X_1, \ldots, X_k\}$ where $Pr_0(X_i) = a_i$.) Given subsets A and B of S, we want to know what the probability is that the element x chosen is in A, and the probability that x is in A given that it is in B. If $A = X_i$, then it is clear that the probability that $x \in A$ is a_i. However, if A is not one of the X_i's, then all we can compute are upper and lower bounds on the probability.

Let $\mathcal{P}$ be the set of discrete probability functions on S consistent with this situation; namely, $Pr \in \mathcal{P}$ iff $Pr(X_i) = a_i$, $i = 1, \ldots, k$. Let Bel be the lower envelope of $\mathcal{P}$. By our earlier discussion, we know that $Bel = Pr_*$ and that Bel is indeed a belief function. Moreover, the best upper and lower bounds we can give on the probability that $x \in A$

are $Bel(A)$ and $Pl(A)$. Similarly, the best lower and upper bounds we can give on the probability that $x \in A$ given that $x \in B$ are given by the infimum and supremum of $\{Pr(A|B) : Pr \in \mathcal{P}\}$. These are precisely $Bel(A|B)$ and $Pl(A|B)$.

Now suppose we slightly change the rules of the game. We are told that the probabilistic process that chooses an element in X_i will definitely choose an element in B if possible This does not affect anything if $X_i \subseteq B$ or if $X_i \subseteq \overline{B}$. However, if $X_i \cap B \neq \emptyset$ and $X_i \neq \overline{B} \neq \emptyset$, then, rather than choosing X_i with probability a_i, the probability is now redistributed so that $X_i \cap B$ is chosen with probability a_i, while $X_i \cap \overline{B}$ is chosen with probability 0. The probability that used to be spread over all of X_i is now concentrated on $X_i \cap B$. Now what is the probability that this new process chooses an element of A given that the element chosen is definitely in B? In this case, the bounds are given by $Bel(A||B)$ and $Pl(A||B)$. To see why, first notice that the change in the rules effectively results in a change to the probability function Pr to a probability function Pr' defined on $\mathcal{X}_B$ as follows. We have $Pr'(X) = Pr(X)$ if $X \subseteq B$ or $X \subseteq \overline{B}$. However, if $X \cap B \neq \emptyset$ and $X \cap \overline{B} \neq \emptyset$, then all of the probability of X moves to $X \cap B$; thus we have $Pr'(X \cap B) = Pr(X)$, while $Pr'(X \cap \overline{B}) = 0$. This means that if B is measurable, then Pr' is identical to Pr, since if B is measurable, then every basis set X is contained either in B or in $\overline{B}$. If B is not measurable, then Pr' can be quite different from Pr. Notice that, in fact, Pr' is almost identical to Pr_B, except for the fact that it assigns nonzero probability to subsets $X \subseteq \overline{B}$. However, it is easy to see that $Pr'(\cdot|B) = Pr_B$. The fact that $Bel(A||B)$ and $Pl(A||B)$ provide the lower and upper bounds on the probability that an element of A will be chosen given that the element chosen is definitely in B now follows from Proposition 3.6.

Suppose we now reconsider the three prisoners problem from this point of view. We can now see that $Bel(\textit{lives-a}||\textit{says-b})$ gives the probability that a lives given the extra hypothesis that the jailer says b whenever possible. in particular, this means that the jailer definitely says b if a is the one that is pardoned; i.e., $Pr(\textit{says-b}|\textit{lives-a}) = 1$. Under this revised situation, the probability that a lives given that the jailer says b is indeed exactly 1/2. With this understanding of the DS notion of updating, the result $Bel(\textit{lives-a}||\textit{says-b}) = Pl(\textit{lives-a}||\textit{says-b}) = 1/2$ should come as no surprise.

To summarize, this discussion has shown that $Bel(A||B)$ corresponds to a somewhat unnatural updating process, where before we condition with respect to B, we first try to choose an element in B whenever possible. Although this extra step before updating makes no difference if B is measurable, it will make a difference if B is not measurable. This is the case in the three prisoner problem, and is the cause of the answer 1/2 that we get when we try to apply DS conditioning in this case.

4 Other updating rules

Taking conditional probability with respect to B only makes sense if we have definite evidence that B actually occurred. More typically, the best we can say is that B occurred with some probability. *Jeffrey's rule* [Jef83] is designed to deal with a generalization of this situation. Suppose our initial situation is described by a probability distribution Pr. Suppose that $B_1, \ldots, B_k$ are mutually exclusive events, measurable with respect to Pr, with $Pr(B_i) > 0$, and we make an observation Ob which tells us that B_i holds with probability a_i, $i = 1, \ldots, k$, where $a_1 + \cdots + a_k = 1$. Then, according to Jeffrey's rule, the probability that we should assign to event A (which is again assumed to be measurable with respect to Pr) given Ob, which we denote by $Pr(A|Ob)$, is $a_1 Pr(A|B_1) + \cdots + a_k Pr(A|B_k)$. Notice that conditional probability with respect to B is a special case of this rule (where we assume that we observe that B occurs with probability 1). We call an observation that places probability 1 on some event B a *simple observation.*

Now what should we do if some $A, B_1, \ldots, B_k$ are not measurable? By analogy with our previous definitions, we take $Pr_*(A|Ob)$ and $Pr^*(A|Ob)$ to be the appropriate lower and upper envelopes. Thus, we define $Pr_*(A|Ob)$ to be

$$\inf\{Pr'(A|Ob) : Pr' \text{ extends } Pr \text{ and } A, B_1, \ldots B_k \text{ are measurable with respect to } Pr'\}.$$

The definition of $Pr^*(A|Ob)$ is analogous. We conjecture that $Pr_*(\cdot|Ob)$ defined in this way is a belief function, just as in the case of simple observations. We leave it to the reader to check that it is indeed the case that $Pr^*(\overline{A}|Ob) = 1 - Pr_*(\overline{A}|Ob)$, so that if $Pr_*(\cdot|Ob)$ is a belief function, then $Pr^*(\cdot|Ob)$ is the corresponding plausibility function.

Unfortunately, unlike the case of simple observations, we do not believe that there is any closed-form expression for $Pr_*(A|Ob)$ in the general case. A natural conjecture is that $Pr_*(A|Ob) = a_1 Pr_*(A|B_1) + \cdots + a_k Pr_*(A|B_k)$. Unfortunately, this is incorrect, as the following counterexample shows.

Example 4.1: Let S consist of four distinct points a, b, c, d, let $A = \{a, b\}$, and let $B = \{a, c\}$. Assume that the the basis consists of the four sets $\{a, b\}$, $\{c\}$, and $\{d\}$, and that $Pr(\{a, b\}) = 1/2$, and $Pr(\{c\}) = Pr(\{d\}) = 1/4$. Let Ob be an observation that tells us that B holds with probability 2/3, and $\overline{B}$ holds with probability 1/3. When we consider probability distributions Pr' that extend Pr, the only issue is what measure should be assigned to $\{a\}$. If $Pr'(\{a\}) = \alpha$ and $Pr'(\{b\}) = \beta$ (where $\alpha + \beta = 1/2$), then

$$Pr'(A|Ob) = \frac{2}{3}\frac{\alpha}{\alpha + \frac{1}{4}} + \frac{1}{3}\frac{\beta}{\beta + \frac{1}{4}} \tag{1}$$

By definition, $Pr(A|Ob)$ is the infimum of the right-hand side of (1), subject to $\alpha + \beta = 1/2$, along with $\alpha \geq 0$ and $\beta \geq 0$. A straightforward calculation shows that the infimum is 2/9, which is attained when $\alpha = 0$. However, the proposed estimate given by the conjecture we are disproving is

$$\frac{2}{3} Pr_*(A|B) + \frac{1}{3} Pr_*(A|\overline{B}),$$

which is 0, since $Pr_*(A|B) = 0$ and $Pr_*(A|\overline{B}) = 0$. ∎

5 Belief functions and lower envelopes

Theorem 3.3 says that each belief function is the lower envelope of a set of probability functions, and each plausibility function an upper envelope. Unfortunately, as we mentioned above, the lower envelope of an arbitrary set of probability functions is not in general a belief function, nor is the upper envelope of an arbitrary set of probability functions in general a plausibility function. A very nice example of this situation is provided in [Pea89] (where it is credited to N. Dalkey). Suppose a space is partitioned into three disjoint sets E_1, E_2, E_3; all we know about these sets is that the probability of each is at most 1/2. Thus, we are considering the class of all probability functions Pr such that $Pr(E_1) + Pr(E_2) + Pr(E_3) = 1$ and $Pr(E_i) \leq 1/2$, $i = 1, 2, 3$. If we now minimize and maximize over this collection of probability functions, we get a function Bel such $Bel(E_i) = 0$ and $Pl(E_i) = 1/2$, $i = 1, 2, 3$ (since the probability of each set E_i can be as low as 0 and as high as 1/2). However, Bel is not a belief function. To see this, first observe that if it were, we would also have $Bel(\overline{E_i}) = 1 - Pl(E_i) = 1/2$. Since $\overline{E_i} \cap \overline{E_j} = E_k$ if i, j, and k are all distinct, it follows that $Bel(\overline{E_i} \cap \overline{E_j}) = 0$. However, it can now be shown that Bel violates axiom B3 of belief functions, since $1 = Bel(\overline{E_1} \cup \overline{E_2} \cup \overline{E_3}) < Bel(\overline{E_1}) + Bel(\overline{E_2}) + Bel(\overline{E_3})$.

Nevertheless, our results show that there are natural sets of probability functions that do induce belief and plausibility functions. For example, Theorem 2.1 and the fact that every inner measure is a belief function shows that the set of all probability functions extending a particular probability function Pr is one such example. Theorem 3.5 shows that a similar situation holds for conditional probability functions. Although a general characterization is lacking, the discussion in Section 4 and further examples in [HF92] suggest that these may not be isolated examples. It seems that there are many situations where naturally defined sets of probability functions do induce belief and plausibility functions.

Even if a set $\mathcal{P}$ of probability functions does induce a belief and plausibility function, say Bel and Pl, it is reasonable to ask whether we *should* represent $\mathcal{P}$ by Bel and Pl. Clearly the answer depends very much on the intended application. However, it is worth noting that this representation of $\mathcal{P}$ might result in a loss of valuable information. Let $\mathcal{P}$ consist of all probability functions on $\{a, b, c\}$ with the following three properties: (1) $1/4 \leq Pr(\{a\}) \leq 1/2$, (2) $1/4 \leq Pr(\{b\}) \leq 1/2$, and (3) $Pr(\{a\}) = Pr(\{b\})$. It is not hard to show that the lower envelope of $\mathcal{P}$ is a belief function Bel whose mass function m satisfies $m(a) = m(b) = m(\{a, c\}) = m(\{b, c\}) = 1/4$. Note that $Bel(\{a\}) = Bel(\{b\}) = 1/4$ and $Pl(\{a\}) = Pl(\{b\}) = 1/2$. Thus, we retain the information that the probability of a and b both range between 1/4 and 1/2. However, we have lost the information that the probabilities of a and b are the same in all the probability functions in $\mathcal{P}$.

This loss of information has some serious repercussions. One consequence is that updates do not commute. To make this precise, let $S = \{a, b, c, d\}$ and suppose we have a basis of S consisting of the three measurable sets $\{a\}$, $\{b\}$, and $\{c, d\}$. Suppose that the probability function Pr is such that $Pr(\{a\}) = 1/4$, $Pr(\{b\}) = 1/4$, and $Pr(\{c, d\}) =$

1/2, and let Bel be the belief function corresponding to Pr. Finally, let $A = \{a\}$, $B = \{a, b\}$, and $C = \{a, b, c\}$. It is easy to see that $Bel(A|B) = Pl(A|B) = 1/2$, since A and B are measurable sets, $A \subseteq B$, and the probability of A is half that of B. It is also easy to check that $Bel(A|C) = 1/4$, $Pl(A|C) = 1/2$, $Bel(B|C) = 1/2$, and $Pl(B|C) = 1$. Let $\mathcal{P}$ consist of all probability functions extending Pr, let $\mathcal{P}^B$ consist of all probability functions $Pr'(\cdot|B)$ such that $Pr' \in \mathcal{P}$, and let $\mathcal{P}^C$ consist of all probability functions $Pr'(\cdot|C)$ such that $Pr' \in \mathcal{P}$. By our earlier results, we know that $Bel(\cdot|C)$ and $Pl(\cdot|C)$ are the lower and upper envelope, respectively, of the probability functions in $\mathcal{P}^C$. It is also easy to check that for all $Pr' \in \mathcal{P}^C$, we have $Pr'(B|C) = 2Pr'(A|C)$. Just as in our earlier example, for each function $Pr' \in \mathcal{P}$, we have $Pr'(\{a\}|C) = Pr'(\{b\}|C)$, but this information is lost when we take the lower envelope. We can easily construct a probability function Pr'' consistent with $Bel(\cdot|C)$ and $Pl(\cdot|C)$ such that $Pr''(A|C) = 1/4$ and $Pr''(B|C) = 1$.

Now suppose we start with the set $\mathcal{P}$ of probability functions and then observe B. The result of this observation is the set $\mathcal{P}^B$. Since $B \subseteq C$, if we next observe C, this does not change anything. The set $\mathcal{P}^B$ describes this set of probability functions. Changing the order of observations still results in the same final set. This is true in general for probability functions; the order of updating is irrelevant as long as we do our updating by conditioning and have measurable sets at every step of the way. (See [HF92] for a proof of this well-known fact.) Unfortunately, this is not the case for belief functions. For example, if we start with Bel, observe C, and then observe B, we get the belief function $Bel_C(\cdot|B)$, where $Bel_C = Bel(\cdot|C)$. It is easy to check that $Bel_C(A|B) = 1/3$ and $Pl_C(A|B) = 2/3$. On the other hand, if we first observe B then observe C, this is equivalent to just observing B, so it is easy to check that $Bel_B(A|C) = Bel_B(A) = 1/2$ and $Pl_B(A|C) = Pl_B(A) = 1/2$, where $Bel_B = Bel(\cdot|B)$. Notice also that updating by $B \wedge C$ is equivalent to updating by B, so $Bel(\cdot|B \wedge C) = Bel(\cdot|B)$. Thus, updating by C and then updating by B is not the same as updating by $B \wedge C$ or updating by B then updating by C. The key point here is that information is lost if we represent $\mathcal{P}^C$ by $Bel(\cdot|C)$ and $Pl(\cdot|C)$, namely, that the probability of B is twice that of A. (By way of contrast, the DS rule of conditioning is commutative. Conditioning with respect to C and then with respect to B is equivalent to conditioning with respect to B and then with respect to C, since both are equivalent to conditioning with respect to $B \wedge C$. However, as we have pointed out, the DS rule of conditioning has other problems when viewed as a technique for updating beliefs.)

These observations suggest to us that the question of the "best" representation of evidence does not have a unique answer. It may be easier to compute with a pair of belief and plausibility functions than to have to carry around a whole set of probability functions. Nevertheless, since information may be lost in this process, this ease of computation comes at a cost. (See [Pea89] for further examples of this phenomenon.)

6 Conclusions

In [FH91] we advocated the use of nonmeasurable sets as a way of representing uncertainty, and showed that the Dempster-Shafer notions of belief and plausibility could be understood in terms of the classical notions of inner and outer measure. Here we have shown that this viewpoint can be extended in a natural way to deal with the process of conditioning. In particular, this extension allows us to define a new notion of conditional belief, distinct from the DS notion, that leads to more intuitive results.

There have been many other attempts at relating the Dempster-Shafer approach to probability theory. Perhaps the most prominent among these include the original papers by Dempster [Dem67; Dem68], and later works by Shafer [Sha79], Ruspini [Rus87], Kyburg [Kyb87], and Pearl [Pea88]. A detailed comparison of our approach to these others can be found in [FH91], so will not be repeated here. Of course, Dempster's viewpoint leads to a notion of conditioning, namely, the DS notion. It is not clear to what extent the techniques used in these other papers can be extended to deal with conditioning. The fact that we can define such a natural notion of conditioning lends support to the usefulness of thinking in terms of inner and outer measures.

Our notion also allows us to avoid some paradoxes associated with the DS notion. For example, we would expect that if both an agent's belief in a proposition p given q and his belief in p given $\neg q$ are at least α, then his belief in p should be at least α, whether or not he learns anything about q. This is essentially what Savage [Sav54] has called the *sure thing principle*. It is easy to see that conditional probability satisfies the sure thing principle, but the DS conditioning rule does not. An example is provided by the three prisoner problem. Recall that we showed that $Bel(lives\text{-}a || says\text{-}b) = 1/2$. Now if the jailer does not say b, then he must say c, and by symmetry we have $Bel(lives\text{-}a || \neg says\text{-}b) = Bel(lives\text{-}a || says\text{-}c) = 1/2$. However, $Bel(lives\text{-}a) = 1/3$. On the other hand, it is easy to see that our notion of conditioning does satisfy the sure thing principle. (This is also observed in [Pea89].) For suppose we have an arbitrary belief function Bel such that $Bel(p|q) \geq \alpha$ and $Bel(p|\neg q) \geq \alpha$. Choose an arbitrary probability function Pr compatible with Bel. By our definition of conditional belief as an infimum, we see that $Pr(p|q) \geq \alpha$ and $Pr(p|\neg q) \geq \alpha$. So $Pr(p) \geq \alpha$. Thus, $Pr(p) \geq \alpha$ for all probability functions Pr compatible with Bel. So, from Theorem 3.3, it follows that $Bel(p) \geq \alpha$.

Although our results show that belief functions can play a useful role even when one wants to think probabilistically, the observations of the previous section do show that information can be lost if we pass to belief functions. This suggests they should be used with care.

One thing we have not really discussed in this paper is what is considered perhaps the key component of the Dempster-Shafer approach, namely, the *rule of combination*. This rule is a way of combining two belief functions to obtain a third one. The reason we have not discussed it is that we feel that the rule of combination does not fit in well with the viewpoint of belief functions as a generalization of probability functions that is discussed in this paper. This point came up indirectly when we pointed out how Shafer's definition of conditional belief, which is defined in terms of the rule of combination, gives unintuitive answers in certain cases. However, there is another way of viewing belief functions, which is as representations of evidence. This is in fact the view taken in [Sha76]. When belief is viewed as a representation of evidence, then the rule of combination becomes more appropriate. These issues are discussed in more detail in a companion paper [HF92].

Acknowledgements: The second author would like to thank Judea Pearl for a series of net conversations that inspired the development of the definitions of $Pr_*(\cdot|B)$ and $Pr^*(\cdot|B)$. We also gratefully acknowledge Nati Linial for his help in the proof of Theorem 3.5. Comments by Hector Levesque, Philippe Smets, and Moshe Vardi inspired a number of useful changes. Finally, we thank Judea Pearl for sending us [dCLM90], Tom Strat for sending us [SK89], Sue Andrews for sending us [Wal81], and Enrique Ruspini for pointing out that our expressions for inner and outer conditional probability actually appear in [Dem67].

Appendix: Proof of Theorem 3.5

Recall that we want to show $Bel(\cdot|B)$ is a belief function, and $Pl(\cdot|B)$ is the corresponding plausibility function. We are working under the assumption that S is finite. We remark that using techniques of [FH91] we could drop this assumption.

It is easy to see, using the formulas in Theorem 3.4, that

$$\begin{aligned} Pl(\overline{A}|B) &= \frac{Pl(\overline{A}\cap B)}{Pl(\overline{A}\cap B)+Bel(A\cap B)} \\ &= 1-\frac{Bel(A\cap B)}{Pl(\overline{A}\cap B)+Bel(A\cap B)} \\ &= 1-Bel(A|B). \end{aligned}$$

Thus, once we show that $Bel(\cdot|B)$ is a belief function, it will immediately follow that $Pl(\cdot|B)$ is the corresponding plausibility function.

Suppose Bel is a belief function defined on the space S. Let m be the underlying mass function of Bel, as guaranteed by Proposition 3.1. In order to show that $Bel(\cdot|B)$ is a belief function, assume that $Bel(B)>0$. Let Bel' be the function defined on 2^B such that for each subset A of B,

$$Bel'(A)=Bel(A)/(Bel(A)+Pl(\overline{A}\cap B)).$$

It clearly suffices to show that Bel' is a belief function, since for all subsets C of S, we have $Bel(C|B)=Bel'(C\cap B)$. Once we show that Bel' satisfies axioms B0–B3, it immediately follows that $Bel(\cdot|B)$ does.

It is clear that Bel' satisfies B0–B2. All that remains is to show that Bel' satisfies B3. Thus we must show that the following inequality holds:

$$Bel'(A_1 \cup \ldots \cup A_k) \geq \sum_{I \subseteq \{1,\ldots,k\},\, I \neq \emptyset} (-1)^{|I|+1} Bel'(\bigcap_{i \in I} A_i).$$

Let $B_1, \ldots, B_t$ be the distinct sets with positive mass contained in B. Let $A'_1, \ldots, A'_n$ be the distinct sets with positive mass that intersect B but are not subsets of B, and let $A_i = A'_i \cap B$, for $1 \leq i \leq n$. Since $Bel(B) > 0$, we know that there is some B_i (but there may be no A_i). Let $\alpha'_i = m(A'_i)$, and $\beta'_i = m(B_i)$, for each i. Let $N = \sum_{i=1}^{n} \alpha'_i + \sum_{i=1}^{t} \beta'_i$. Note that $N > 0$, since there is some B_i. Let $\alpha_i = \alpha'_i / N$, and $\beta_i = \beta'_i / N$, for each i. Thus, the α_i's and β_i's are normalized versions of the α'_i's and β'_i's.

We want to define a mass function m' corresponding to Bel'. We first need to do a small detour. If $s_1 \ldots s_k$ is a string, and if $1 \leq i_1 < \cdots < i_p \leq k$, then we call $s_{i_1} \ldots s_{i_p}$ a *substring of* $s_1 \ldots s_k$, which we write as $s_{i_1} \ldots s_{i_p} \preceq s_1 \ldots s_k$. For example, $s_1 s_3 s_4$ is a substring of $s_1 s_2 s_3 s_4 s_5$. The substring is *proper* if it does not equal the full string $s_1 \ldots s_k$; we then write $s_{i_1} \ldots s_{i_p} \prec s_1 \ldots s_k$. We now define a function m'', whose domain is $\{A_1, \ldots, A_n, B_1, \ldots, B_t\}^*$, the set of finite strings over the alphabet consisting of the names of the sets with positive mass that intersect B. (We shall usually not bother to distinguish between a set and the name of a set, but, as we shall see, it is convenient to consider explicitly strings of names of sets.) First, we let $m''(B_i) = \beta_i$, for $1 \leq i \leq t$. Assume now that we have defined $m''(B_i A_{j_1} \cdots A_{j_s})$ whenever $s < r$ and $j_1 < \cdots < j_s$. Assume that $j_1 < \cdots < j_r$. Let

$$m''(B_i A_{j_1} \cdots A_{j_r}) = \frac{\beta_i}{1 - \alpha_{j_1} - \cdots - \alpha_{j_r}} - \sum_{Y \prec A_{j_1} \cdots A_{j_r}} m''(B_i Y). \qquad (2)$$

If A is not of the form $B_i A_{j_1} \cdots A_{j_r}$ with $j_1 < \cdots < j_r$, then $m''(A) = 0$.

We are now ready to define the alleged mass function m'. If X is the string $B_i A_{j_1} \cdots A_{j_r}$, where $j_1 < \cdots < j_r$, then we say that X *represents* the set given by $B_i \cup A_{j_1} \cup \cdots \cup A_{j_r}$. We would like to let m' be simply m'' (that is, by letting m' applied to a set be equal to m'' applied to a string that represents the set, and let $m'(A) = 0$ for sets not of the form $B_i A_{j_1} \cdots A_{j_r}$). The problem is that several distinct strings may represent the same set; for example, it is quite possible that, say, the sets $B_1 \cup A_1$ and $B_2 \cup A_4 \cup A_5$ are the same. We define $m'(A)$ to be $\sum_{X \text{ represents } A} m''(X)$. For example, if the set A equals both $B_1 \cup A_1$ and $B_2 \cup A_4 \cup A_5$, but if A is not of the form $B_i \cup A_{j_1} \cdots \cup A_{j_r}$ for any other choices of $B_i, A_{j_1}, \ldots, A_{j_r}$ with $j_1 < \cdots < j_r$, then $m'(A) = m''(B_1 A_1) + m''(B_2 A_4 A_5)$. We shall prove that m' is a mass function, and that $Bel'(A) = \sum_{C \subseteq A} m'(C)$. This will show that Bel' is a belief function.

Thus, we must show that

A. $m'(\emptyset) = 0$.

B. $m'(A) \geq 0$, for each $A \subseteq B$.

C. $\sum_{A \subseteq B} m'(A) = 1$.

D. $Bel'(A) = \sum_{C \subseteq A} m'(C)$.

By definition of m'' and m', we know that (A) holds. We now prove (D). Let $A_{k_1}, \ldots, A_{k_q}$ (where $k_1 < \cdots < k_q$) be the A_i's contained in A, and let $B_{i_1}, \ldots, B_{i_s}$ be the B_i's contained in A. What is $Bel'(A)$? As before, let $N = \sum_{i=1}^{n} \alpha'_i + \sum_{i=1}^{t} \beta'_i$. It is easy to see that $Bel(A) = \beta'_{i_1} + \cdots + \beta'_{i_s}$, and $Pl(\overline{A} \cap B) = N - (\alpha'_{k_1} + \cdots + \alpha'_{k_q} + \beta'_{i_1} + \cdots + \beta'_{i_s})$. Hence,

$$Bel'(A) = Bel(A)/(Bel(A) + Pl(\overline{A} \cap B)) = (\beta'_{i_1} + \cdots + \beta'_{i_s})/(N - \alpha'_{k_1} - \cdots - \alpha'_{k_q}).$$

When we divide numerator and denominator by N, we see that

$$Bel'(A) = (\beta_{i_1} + \cdots + \beta_{i_s})/(1 - \alpha_{k_1} - \cdots - \alpha_{k_q}). \tag{3}$$

To prove (D), we must show that $\sum_{C \subseteq A} m'(C)$ equals the right-hand side of (3). Let us call an expression $m''(B_i A_{j_1} \cdots A_{j_r})$, where i is a member of $\{i_1, \ldots, i_s\}$, and where $j_1 < \ldots < j_r$ are members of $\{k_1, \ldots, k_q\}$, a *good term*. Note that if $m''(B_i A_{j_1} \cdots A_{j_r})$ is a good term, then $B_i \cup A_{j_1} \cup \cdots \cup A_{j_r} \subseteq A$. Now $\sum_{C \subseteq A} m'(C)$ equals the sum of all good terms. This is because (a) each good term is a part of the sum defining $m'(C)$ for exactly one $C \subseteq A$, and (b) if $C \subseteq A$, then $m'(C)$ is defined as the sum of certain good terms. So we must show that the sum of all of the good terms equals the right-hand side of (3). Now let i be a fixed member of $\{i_1, \ldots, i_s\}$. The sum of all good terms of the form $m''(B_i A_{j_1} \cdots A_{j_r})$ *except* for the good term $m''(B_i A_{k_1} \cdots A_{k_q})$ is simply $\sum_{Y \prec A_{k_1} \cdots A_{k_q}} m''(B_i Y)$. Since

$$m''(B_i A_{k_1} \cdots A_{k_q}) = \frac{\beta_i}{1 - \alpha_{k_1} - \cdots - \alpha_{k_q}} - \sum_{Y \prec A_{k_1} \cdots A_{k_q}} m''(B_i Y),$$

it follows that the sum of all good terms of the form $m''(B_i A_{j_1} \cdots A_{j_r})$ equals $\beta_i/(1 - \alpha_{k_1} - \cdots - \alpha_{k_q})$. So the sum of all good terms is $(\beta_{i_1} + \cdots + \beta_{i_s})/(1 - \alpha_{k_1} - \cdots - \alpha_{k_q})$, as desired. This proves (D).

Now (C) follows from (D), since it is easy to see that $Bel'(B) = 1$. So we need only prove (B). Since $m'(A) = \sum_{X \text{ represents } A} m''(X)$, we need only show that m'' is nonnegative. Thus, we must show that each $m''(B_i A_{j_1} \cdots A_{j_r})$ is nonnegative. For ease in notation, we replace i by 1, and $j_1, \ldots, j_r$ by $1, \ldots, r$, and show that $m''(B_1 A_1 \cdots A_r)$ is nonnegative.

We now show that

$$m''(B_1 A_1 \cdots A_r) = \sum_{I \subseteq \{1, \ldots, r\}} (-1)^{r-|I|} \frac{\beta_1}{1 - \sum_{j \in I} \alpha_j}. \tag{4}$$

If $r = 0$, then (4) says that $m''(B_1) = \beta_1$, as desired (there is one summand, where $I = \emptyset$). If $r = 1$, then (4) says

$$m''(B_1 A_1) = \frac{\beta_1}{1-\alpha_1} - \beta_1, \tag{5}$$

and if $r = 2$, then (4) says

$$m''(B_1 A_1 A_2) = \frac{\beta_1}{1-\alpha_1-\alpha_2} - \frac{\beta_1}{1-\alpha_1} - \frac{\beta_1}{1-\alpha_2} + \beta_1. \tag{6}$$

We prove (4) by induction on r. As we noted, the base case ($r = 0$) simply says that $m''(B_1) = \beta_1$. As for the induction step, we replace each $m''(B_1 Y)$ on the right-hand side of (2) by the corresponding formula as given by the right-hand side of (4); this is all right by induction hypothesis, since Y is a proper substring of $A_1 \cdots A_r$. Let us denote $\beta_1/(1 - \sum_{j \in I} \alpha_j)$ by E_I. We now compute the coefficient of E_I in the expression resulting from our substitution we just described. If $I = \{1, \ldots, r\}$, then it is clear that E_I appears only once (and positively). This is $(-1)^{r-|I|}\beta_1/(1 - \sum_{j \in I} \alpha_j)$, as desired. Let $|I| = s$, and assume that $s < r$. The number of subsets of $\{1, \ldots, r\}$ that contain I where $|I| = t$ is 0, if $t < s$, and $\binom{r-s}{t-s}$ otherwise. So the total contribution of the expression E_I by proper substrings Y of $A_1 \cdots A_r$ of length t is $(-1)^{t-s+1} \binom{r-s}{t-s} E_I$. Hence, the grand total contribution is

$$E_I \sum_{t=s}^{r-1} (-1)^{t-s+1} \binom{r-s}{t-s}.$$

Note that the largest index of the sum is $t = r - 1$ rather than $t = r$, since in (2) we consider only *proper* substrings Y. By changing index, this value is

$$E_I \sum_{u=0}^{r-s-1} (-1)^{u+1} \binom{r-s}{u}. \tag{7}$$

Now $\sum_{u=0}^{r-s} (-1)^{u+1} \binom{r-s}{u} = 0$, since it is -1 times the binomial expansion of $(1-1)^{r-s}$. So (7) equals $(-1)^{r-s} E_I$, that is, $(-1)^{r-|I|} E_I$, as desired.

To prove that m'' is nonnegative, we now need only show that the right-hand side of (4) is nonnegative. Dividing through by β_1, we see that we must show that

$$\sum_{I \subseteq \{1,\ldots,r\}} (-1)^{r-|I|} \frac{1}{1 - \sum_{j \in I} \alpha_j} \geq 0 \tag{8}$$

Note that each α_j is positive, and $\sum_{j=1}^{r} \alpha_j < 1$ (since $\beta_1 > 0$, and the sum of the α_j's and β_i's is at most 1).

Since $0 \le \sum_{j\in I} \alpha_j < 1$, it follows that $1/(1 - \sum_{j\in I} \alpha_j) = \sum_{k=0}^{\infty} (\sum_{j\in I} \alpha_j)^k$, where we interpret $(\sum_{j\in I} \alpha_j)^0$ to be 1, even when $\sum_{j\in I} \alpha_j = 0$ (which occurs when $I = \emptyset$). So we need only show that for each r and k,

$$\sum_{I\subseteq\{1,\ldots,r\}} (-1)^{r-|I|} (\sum_{j\in I} \alpha_j)^k \ge 0 \tag{9}$$

Define $f_{r,k}$ to be a function with r arguments, with domain those tuples $(\alpha_1, \ldots, \alpha_r)$ where each α_i is nonnegative, and

$$f_{r,k}(\alpha_1, \ldots, \alpha_r) = \sum_{I\subseteq\{1,\ldots,r\}} (-1)^{r-|I|} (\sum_{j\in I} \alpha_j)^k$$

(where we make the convention that if $r = 0$, then $f_{r,k}$ is a constant function that is equal to 0, with no arguments). To prove (9), we need only show that each $f_{r,k}$ is nonnegative. Define $g_{r,k}$ to be a function with the same domain as $f_{r,k}$, and

$$g_{r,k}(\alpha_1, \ldots, \alpha_r) = \sum_{I\subseteq\{1,\ldots,r\}, r\in I} (-1)^{r-|I|} (\sum_{j\in I} \alpha_j)^k$$

if $r > 0$ (and as before, 0 if $r = 0$).

We shall show, by induction on r, that for every k, the functions $f_{r,k}$ and $g_{r,k}$ are nonnegative. The base case ($r = 0$) is immediate. Assume inductively that $r \ge 1$, and that $f_{r-1,k}$ and $g_{r-1,k}$ are nonnegative for every k. We now show, by a second induction on k, that $f_{r,k}$ and $g_{r,k}$ are nonnegative. For the base case ($k = 0$), note that $f_{r,0}(\alpha_1, \ldots, \alpha_r) = \sum_{I\subseteq\{1,\ldots,r\}} (-1)^{r-|I|}$, which equals 0, since it is the same as the binomial expansion of $(1-1)^r$. Now $g_{r,0}(\alpha_1, \ldots, \alpha_r) = \sum_{I\subseteq\{1,\ldots,r\}, r\in I} (-1)^{r-|I|} = \sum_{I'\subseteq\{1,\ldots,r-1\}} (-1)^{r-1-|I'|} = f_{r-1,0}(\alpha_1, \ldots, \alpha_{r-1})$, so also $g_{r,0}$ is nonnegative.

For the inductive step, assume that $f_{r,k-1}$ and $g_{r,k-1}$ are nonnegative. To show that $f_{r,k}$ and $g_{r,k}$ are nonnegative, we shall show that $f_{r,k}(\alpha_1, \ldots, \alpha_r)$ and $g_{r,k}(\alpha_1, \ldots, \alpha_r)$ are nonnegative when $\alpha_r = 0$, and that the first derivatives with respect to α_r are nonnegative.

If $\alpha_r = 0$, then

$$\begin{aligned}
f_{r,k}(\alpha_1, \ldots, \alpha_r) &= \sum_{I\subseteq\{1,\ldots,r\}} (-1)^{r-|I|} (\sum_{j\in I} \alpha_j)^k \\
&= \sum_{I\subseteq\{1,\ldots,r\}, r\in I} (-1)^{r-|I|} (\sum_{j\in I} \alpha_j)^k + \sum_{I\subseteq\{1,\ldots,r\}, r\notin I} (-1)^{r-|I|} (\sum_{j\in I} \alpha_j)^k \\
&= \sum_{I'\subseteq\{1,\ldots,r-1\}} (-1)^{r-1-|I'|} (\sum_{j\in I'} \alpha_j)^k + \sum_{I\subseteq\{1,\ldots,r-1\}} (-1)^{r-|I|} (\sum_{j\in I} \alpha_j)^k \\
&= f_{r-1,k}(\alpha_1, \ldots, \alpha_{r-1}) - f_{r-1,k}(\alpha_1, \ldots, \alpha_{r-1}) \\
&= 0.
\end{aligned}$$

If $\alpha_r = 0$, then $g_{r,k}(\alpha_1, \ldots, \alpha_r) = f_{r-1,k}(\alpha_1, \ldots, \alpha_{r-1})$, which by inductive assumption on r is nonnegative.

As for the derivatives: The derivative of $f_{r,k}(\alpha_1, \ldots, \alpha_r)$ with respect to α_r is

$$k \sum_{I \subseteq \{1,\ldots,r\}, r \in I} (-1)^{r-|I|} (\sum_{j \in I} \alpha_j)^{k-1} = k g_{r,k-1}(\alpha_1, \ldots, \alpha_r),$$

which is nonnegative by inductive assumption on k. The derivative of $g_{r,k}(\alpha_1, \ldots, \alpha_r)$ with respect to α_r is $k g_{r,k-1}(\alpha_1, \ldots, \alpha_r)$, which again is nonnegative by inductive assumption on k. ∎

We conclude with a few remarks on some technical details from the proof of our result. By Proposition 3.1, a belief function on a finite sample space uniquely determines the mass function. Therefore, the mass function constructed in the previous proof is uniquely determined. It is interesting to consider which sets have positive mass; these are what Shafer calls the *focal elements*. In the construction, we saw that the only sets that were possibly assigned positive mass are sets of the form $B_i \cup A_{j_1} \cup \cdots \cup A_{j_s}$, where the B_i's and A_i's are as in the previous proof. Do all of these sets indeed have positive mass? As we now show, the answer is yes: that is, $m'(A) > 0$ iff A is of the form $B_i \cup A_{j_1} \cup \cdots \cup A_{j_s}$.

If A is B_i, then $m'(A) = \beta_i > 0$, so we can assume that A is of the form $B_i \cup A_{j_1} \cup \cdots \cup A_{j_s}$ where $s > 0$. We need only show that if r and $\alpha_1, \ldots, \alpha_r$ are each positive, then $f_{r,k}(\alpha_1, \ldots, \alpha_r) > 0$ for some k, where $f_{r,k}$ is as in the previous proof. In fact, we shall show that under these assumptions, $f_{r,r}(\alpha_1, \ldots, \alpha_r) > 0$.

To prove this, we show by induction on $r \geq 1$ that if $\alpha_1, \ldots, \alpha_r$ are each positive, then $g_{r,r-1}(\alpha_1, \ldots, \alpha_r) > 0$ and $f_{r,r}(\alpha_1, \ldots, \alpha_r) > 0$. As for the base case $r = 1$, we have that $g_{1,0}(\alpha_1) = 1$ and $f_{1,1}(\alpha_1) = \alpha_1$, so both are positive if $\alpha_1 > 0$. Assume inductively that $r \geq 2$, and that $g_{r-1,r-2}(\alpha'_1, \ldots, \alpha'_{r-1}) > 0$ and $f_{r-1,r-1}(\alpha'_1, \ldots, \alpha'_{r-1}) > 0$ whenever $\alpha'_1, \ldots, \alpha'_r$ are each positive.

If $\alpha_r = 0$, then $g_{r,r-1}(\alpha_1, \ldots, \alpha_r) = f_{r-1,r-1}(\alpha_1, \ldots, \alpha_{r-1})$, which by inductive assumption is positive if $\alpha_1, \ldots, \alpha_{r-1}$ are each positive. Since the derivative of $g_{r,r-1}(\alpha_1, \ldots, \alpha_r)$ is, as we saw, nonnegative, it follows that $g_{r,r-1}(\alpha_1, \ldots, \alpha_r) > 0$ whenever $\alpha_1, \ldots, \alpha_{r-1}$ are each positive, and in particular whenever $\alpha_1, \ldots, \alpha_r$ are each positive.

The derivative of $f_{r,r}(\alpha_1, \ldots, \alpha_r)$ with respect to α_r is, as we saw, $r g_{r,r-1}(\alpha_1, \ldots, \alpha_r)$, which, as we just showed, is positive if $\alpha_1, \ldots, \alpha_r$ are each positive. Since, as we saw, $f_{r,r}(\alpha_1, \ldots, \alpha_r)$ is nonnegative when $\alpha_r = 0$, it then follows that $f_{r,r}(\alpha_1, \ldots, \alpha_r)$ is positive when $\alpha_1, \ldots, \alpha_r$ are each positive, as we now show. If instead, $f_{r,r}(\alpha_1, \ldots, \alpha_r) = 0$, then since the derivative at α'_r when $0 \leq \alpha'_r \leq \alpha_r$ is nonnegative, it would follow that $f_{r,r}(\alpha_1, \ldots, \alpha_{r-1}, \alpha'_r) = 0$ when $0 \leq \alpha'_r \leq \alpha_r$. But then the derivative at α'_r when $0 < \alpha'_r < \alpha_r$ would be 0, whereas we showed that the derivative is positive when $\alpha_1, \ldots, \alpha_{r-1}, \alpha'_r$ are each positive.

References

[Abe88] S. Abel. The sum-and-lattice-points method based on an evidential-reasoning system applied to the real-time vehicle guidance problem. In *Uncertainty in Artificial Intelligence 2*, pages 365–370, 1988.

[Ait68] J. Aitchison. Discussion on Professor Dempster's paper. *Journal of the Royal Statistical Society, Series B*, 30:234–237, 1968.

[Bla87] P. Black. Is Shafer general Bayes? In *Proc. Third AAAI Uncertainty in Artificial Intelligence Workshop*, pages 2–9, 1987.

[dCLM90] L. M. de Campos, M. T. Lamata, and S. Moral. The concept of conditional fuzzy measure, to appear. *International Journal of Intelligent Systems*, 1990.

[Dem67] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[Dem68] A. P. Dempster. A generalization of Bayesian inference. *Journal of the Royal Statistical Society, Series B*, 30:205–247, 1968.

[Dia78] P. Diaconis. Review of "A Mathematical Theory of Evidence". *Journal of the American Statistical Society*, 73(363):677–678, 1978.

[DP88] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4:244–264, 1988.

[DZ86] P. Diaconis and S. L. Zabell. Some alternatives to Bayes's rule. In B. Grofman and G. Owen, editors, *Proc. Second University of California, Irvine, Conference on Political Economy*, pages 25–38, 1986.

[Fal88] B. Falkenhainer. Towards a general-purpose belief maintenance system. In *Uncertainty in Artificial Intelligence 2*, pages 125–131, 1988.

[Fel57] W. Feller. *An Introduction to Probability Theory and its Applications*, volume 1. John Wiley & Sons, New York, 2nd edition, 1957.

[FH91] R. Fagin and J. Y. Halpern. Uncertainty, belief, and probability. *Computational Intelligence*, 7(3):160–173, 1991.

[Gar61] M. Gardner. *Second Scientific American Book of Mathematical Puzzles and Diversions*. Simon & Schuster, 1961.

[Hal50] P. Halmos. *Measure Theory*. Van Nostrand, 1950.

[HF92] J. Y. Halpern and R. Fagin. Two views of belief: belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54:275–317, 1992.

[Hun87] D. Hunter. Dempster-Shafer vs. probabilistic logic. In *Proc. Third AAAI Uncertainty in Artificial Intelligence Workshop*, pages 22–29, 1987.

[Jaf90] J.-Y. Jaffray. Bayesian updating and belief functions. Manuscript, 1990.

[Jef83] R. C. Jeffrey. *The Logic of Decision.* University of Chicago Press, Chicago, 1983.

[Kyb87] H. E. Kyburg, Jr. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, 31:271–293, 1987.

[Lem86] J. F. Lemmer. Confidence factors, empiricism, and the Dempster-Shafer theory of evidence. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 167–196. North-Holland, Amsterdam, 1986.

[LG83] J. D. Lowrance and T. D. Garvey. Evidential reasoning: an implementation for multisensor integration. Technical Report TN 307, SRI International, Menlo Park, Calif., 1983.

[LU88] Z. Li and L. Uhr. Evidential reasoning in a computer vision system. In *Uncertainty in Artificial Intelligence 2*, pages 403–412, 1988.

[Pea88] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, San Francisco, Calif., 1988.

[Pea89] J. Pearl. Reasoning with belief functions: a critical assessment. Technical Report R-136, UCLA, 1989.

[Rus87] E. H. Ruspini. The logical foundations of evidential reasoning. Research Note 408, revised version, SRI International, Menlo Park, Calif., 1987.

[Sav54] L. J. Savage. *Foundations of Statistics.* John Wiley & Sons, New York, 1954.

[Sha76] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, N.J., 1976.

[Sha79] G. Shafer. Allocations of probability. *Annals of Probability*, 7(5):827–839, 1979.

[SK89] P. Smets and R. Kennes. The transferable belief model: comparison with Bayesian models, manuscript, IRIDIA, Université Libre de Bruxelles, 1050 Brussels, Belgium. 1989.

[SW91] C. Sundberg and C. Wagner. Generalized finite differences and Bayesian conditioning of Choquet capacities. Manuscript, 1991.

[Wal81] P. Walley. Coherent lower (and upper) probabilities. Manuscript, Dept. of Statistics, University of Warwick, 1981.

[Zad84] L. A. Zadeh. A mathematical theory of evidence (book review). *AI Magazine*, 5(3):81–83, 1984.

[Zha89] L. Zhang. A new proof to Theorem 3.2 of Fagin and Halpern's paper. Memorandum, 1989.