Uncertainty, Belief, and Probability*

Ronald Fagin
Joseph Y. Halpern

IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, California 95120-6099
email: fagin@almaden.ibm.com, halpern@almaden.ibm.com

Abstract: We introduce a new probabilistic approach to dealing with uncertainty, based on the observation that probability theory does not require that every event be assigned a probability. For a *nonmeasurable* event (one to which we do not assign a probability), we can talk about only the *inner measure* and *outer measure* of the event. In addition to removing the requirement that every event be assigned a probability, our approach circumvents other criticisms of probability-based approaches to uncertainty. For example, the measure of belief in an event turns out to be represented by an interval (defined by the inner and outer measure), rather than by a single number. Further, this approach allows us to assign a belief (inner measure) to an event E without committing to a belief about its negation $\neg E$ (since the inner measure of an event plus the inner measure of its negation is not necessarily one). Interestingly enough, inner measures induced by probability measures turn out to correspond in a precise sense to Dempster-Shafer belief functions. Hence, in addition to providing promising new conceptual tools for dealing with uncertainty, our approach shows that a key part of the important Dempster-Shafer theory of evidence is firmly rooted in classical probability theory.

*This is an expanded version of a paper that appears in IJCAI 89, pp. 1161–1167. This version is essentially identical to one that appears in *Computational Intelligence* **7**, 1991, pp. 160–173.

1 Introduction

Dealing with uncertainty is a fundamental issue for AI. The most widely-used approach to dealing with uncertainty is undoubtedly the Bayesian approach. It has the advantage of relying on well-understood techniques from probability theory, as well as some philosophical justification on the grounds that a "rational" agent must assign uncertainties to events in a way that satisfies the axioms of probability [Cox46, Sav54]. On the other hand, the Bayesian approach has been widely criticized for requiring an agent to assign a subjective probability to every event. While this can be done in principle by having the agent play a suitable betting game [Jef83],[1] it does have a number of drawbacks. As a practical matter, it may not be possible to provide a reasonable estimate of the probability of some events; and even if it is possible to estimate the probability, the amount of effort required to do so (both in terms of data gathering and computation) may be prohibitive. There is also the issue of whether it is reasonable to describe confidence by a single point rather than a range. While an agent might be prepared to agree that the probability of an event lies within a given range, say between 1/3 and 1/2, he might not be prepared to say that it is precisely .435.

Not surprisingly, there has been a great deal of debate regarding the Bayesian approach (see [Che85] and [Sha76] for some of the arguments). Numerous other approaches to dealing with uncertainty have been proposed, including Dempster-Shafer theory [Dem68, Sha76], Cohen's model of endorsements [Coh85], and various nonstandard, modal, and fuzzy logics (for example, [HR87, Zad75]). A recent overview of the field can be found in [Saf88]. Of particular interest to us here is the Dempster-Shafer approach, which uses *belief functions* and *plausibility functions* to attach numerical lower and upper bounds on the likelihoods of events.

Although the Bayesian approach requires an agent to assign a probability to every event, probability theory does not. The usual reason that mathematicians deal with nonmeasurable events (those that are not assigned a probability) is out of mathematical necessity. For example, it is well known that if the sample space of the probability space consists of all numbers in the real interval $[0,1]$, then we cannot allow every set to be measurable if (like Lebesgue measure) the measure is to be translation-invariant (see [Roy64, page 54]). However, in this paper we allow nonmeasurable events out of choice, rather than out of mathematical necessity. An event E for which an agent has insufficient information to assign a probability is modelled as a nonmeasurable set. The agent is not forced to assign a probability to E in our approach. We can provide meaningful lower and upper bounds on our degree of belief in E by using the standard mathematical notions of *inner measure* and *outer measure* induced by the probability measure [Hal50], which, roughly speaking, are the probability of the largest measurable event contained in E and the smallest measurable event containing E, respectively.

[1] This idea is due to Ramsey [Ram31] and was rediscovered by von Neumann and Morgenstern [NM47]; a clear exposition can be found in [LR57].

Allowing nonmeasurable events has its advantages. It means that the tools of probability can be brought to bear on a problem without having to assign a probability to every set. It also gives us a solid technical framework in which it makes sense to assign an interval, rather than a point-valued probability, to an event as a representation of the uncertainty of the event. In a precise technical sense (discussed below), the inner and outer measure of an event E can be viewed as giving the best bounds on the "true" probability of E; thus, the interval by the inner and outer measure really provides a good measure of our degree of belief in E. (We remark that this point is also made in [Rus87].) Rather than nonmeasurability being a mathematical nuisance, we have turned it here into a desirable feature!

To those familiar with the Dempster-Shafer approach to reasoning about uncertainty, many of the properties of inner and outer measures will seem remimiscent of the properties of the belief and plausibility functions used in that approach. Indeed, the fact that there are connections between inner measures and belief functions has been observed before (we discuss related literature in Section 6 below). What we show here is that the two are, in a precise sense, equivalent. One direction of this equivalence is easy to state and prove: every inner measure is a belief function (i.e., it satisfies the axioms that characterize belief functions). Indeed, this result can be shown to follow from a more general result of Shafer [Sha79, Thm. 5.1(1)].

The converse, that every belief function is an inner measure, is not so obvious. Indeed, the most straightforward way of making this statement precise is false. That is, given a belief function Bel on a space S, it is not in general possible to find a probability function μ on S such that $\mu_* = Bel$. To get the converse implication, we must view both belief functions and probability functions as functions not just on sets, but on *formulas*. The distinction between the two is typically ignored; we talk about the probability of the event "E_1 and E_2" and denote this event $E_1 \wedge E_2$, using a conjunction symbol, rather than $E_1 \cap E_2$. While it is usually safe to ignore the distinction between sets and formulas, there are times when it is useful to make it. We provide straightforward means of viewing probability and belief functions as functions on formulas and show that when viewed as functions on formulas, given a belief function Bel we can always find a probability function μ such that $Bel = \mu_*$. Our technical results thus say that, in a precise sense, we can identify belief functions and inner measures.

The implications of this equivalence are significant. Although some, such as Cheeseman [Che85], consider the theory of belief functions as *ad hoc* and essentially nonprobabilistic (see the discussion by Shafer [Sha86]), our results help show that a key part of the Dempster-Shafer theory of evidence is firmly rooted in classical probability theory. We are, of course, far from the first to show a connection between the Dempster-Shafer theory of evidence and probability theory (see Section 6 for a detailed discussion of other approaches). Nevertheless, we would claim that the the particular relationship we establish is more intuitive, and easier to work with, than others. There is one immediate technical payoff: by combining our results here with those of a companion paper [FHM90], we are able to obtain a sound and complete axiomatization for a rich propositional logic of evidence, and provide a decision procedure for the satisfiability problem, which we show is no harder than that of propositional logic (NP-complete). Our techniques may provide a means for automatically deducing the consequences of a body of evidence; in particular, we can compute when one set of beliefs implies another set of beliefs.

The rest of the paper is organized as follows. In Section 2, we give a brief review of probability theory, describe our approach, give a few examples of its use, and show that in a precise sense, it extends to Nilsson's approach [Nil86]. The remainder of the paper is devoted to examining the relationship between our approach and the Dempster-Shafer approach. In Section 3 we review the Dempster-Shafer approach and show that in a precise sense, belief and plausibility functions are just inner and outer measures induced by a probability measure. In Section 4, we show that Dempster's rule of combination, which provides a technique for combining evidence from two sources, can be captured in our framework by an appropriate rule of combination for probability measures. In Section 5 we show that, by combining the results of Section 3 with those of a companion paper [FHM90], we obtain a complete axiomatization for reasoning about belief functions. In Section 6 we compare our results with those in a number of related papers. Section 7 contains some concluding remarks.

2 Probability theory

To make our discussion precise, it is helpful to recall some basic definitions from probability theory (see [Hal50] for more details). A *probability space* $(S, \mathcal{X}, \mu)$ 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* $\mu: \mathcal{X} \rightarrow [0, 1]$ satisfying the following properties:

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

P2. $\mu(S) = 1$

P3. $\mu(\cup_{i=1}^{\infty} X_i) = \Sigma_{i=1}^{\infty} \mu(X_i)$, 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 $\mathcal{X}$ is a finite set, then we can simplify property P3 above to

P3′. $\mu(X \cup Y) = \mu(X) + \mu(Y)$, if X and Y are disjoint members of $\mathcal{X}$.

This property is called *finite additivity*. Properties P1, P2, and P3′ characterize probability measures in finite spaces. Observe that from P2 and P3′ it follows (taking $Y = \overline{X}$, the complement of X) that $\mu(\overline{X}) = 1 - \mu(X)$. Taking $X = S$, we also get that $\mu(\emptyset) = 0$. We remark for future reference that P3′ is equivalent to the following axiom:

P3″. $\mu(X) = \mu(X \cap Y) + \mu(X \cap \overline{Y})$.

Clearly P3″ is a special case of P3′, since $X \cap Y$ and $X \cap \overline{Y}$ are disjoint and $X = (X \cap Y) \cup (X \cap \overline{Y})$. To see that P3′ follows from P3″, just replace the X in P3″ by $X \cup Y$, and observe that if X and Y are disjoint, then $X = (X \cup Y) \cap \overline{Y}$, while $Y = (X \cup Y) \cap Y$.

A subset $\mathcal{X}'$ of $\mathcal{X}$ is said to be a *basis* (of $\mathcal{X}$) if the members of $\mathcal{X}'$ are nonempty and disjoint, and if $\mathcal{X}$ consists precisely of countable unions of members of $\mathcal{X}'$. 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 in $\mathcal{X}$ none of whose proper nonempty subsets are in $\mathcal{X}$). Note that if $\mathcal{X}$ has a basis, once we know the probability of every set in the basis, we can compute the probability of every measurable set by using countable additivity.

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

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

(where, as usual, sup denotes "least upper bound" and inf denotes "greatest lower bound"). 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.[2] In any case, it is not hard to show by countable additivity that for each set A, there are measurable sets B and C where $B \subseteq A \subseteq C$ such that $\mu(B) = \mu_*(A)$ and $\mu(C) = \mu^*(A)$. Note that if there are no nonempty measurable sets contained in A, then $\mu_*(A) = 0$, and if there are no measurable sets containing A other than the whole space S, then $\mu^*(A) = 1$. The properties of probability spaces guarantee that if X is a measurable set, then $\mu_*(X) = \mu^*(X) = \mu(X)$. In general we have $\mu^*(A) = 1 - \mu_*(\overline{A})$.

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 $(S, \mathcal{X}', \mu')$ is an *extension* of the probability space $(S, \mathcal{X}, \mu)$ if $\mathcal{X}' \supseteq \mathcal{X}$, and $\mu'(A) = \mu(A)$ for all $A \in \mathcal{X}$ (so that μ and μ' agree on $\mathcal{X}$, their common domain).

[2] We would have begun this sentence by saying "If there are only countably many measurable sets", except that it turns out that if there are countably many measurable sets, then there are only finitely many! The proof is as follows. Let $\mathcal{Y} \subseteq \mathcal{X}$ be the set of all nonempty minimal measurable sets. If every point in the sample space S is in $\cup\mathcal{Y}$, then $\mathcal{Y}$ is a basis; in this case $\mathcal{X}$ is finite if $\mathcal{Y}$ is finite, and uncountable if $\mathcal{Y}$ is infinite (since the existence of an infinite, pairwise disjoint family of measurable sets implies that there are uncountably many measurable sets). So we can assume that $S \neq \cup\mathcal{Y}$, or else we are done. Let $T_0 = S - \cup\mathcal{Y} \neq \emptyset$. Since T_0 is nonempty, measurable, and not in $\mathcal{Y}$, it follows that T_0 is not minimal. So T_0 has a proper nonempty measurable subset T_1. Similarly, T_1 has a proper nonempty measurable subset T_2, etc. Now the set differences $T_i - T_{i+1}$, for $i = 0, 1, \ldots$ form an infinite, pairwise disjoint family of measurable sets. Again, this implies that there are uncountably many measurable sets.

If $(S, \mathcal{X}', \mu')$ is an extension of $(S, \mathcal{X}, \mu)$, then we say that μ' *extends* μ. The following result seems to be well known. (A proof can be found in [Rus87]; see also [FH91] for further discussion.)

Theorem 2.1: *If $(S, \mathcal{X}', \mu')$ is an extension of $(S, \mathcal{X}, \mu)$ and $A \in \mathcal{X}'$, then $\mu_*(A) \leq \mu'(A) \leq \mu^*(A)$. Moreover, there exist extensions $(S, \mathcal{X}_1, \mu_1)$, $(S, \mathcal{X}_2, \mu_2)$ of $(S, \mathcal{X}, \mu)$ such that $A \in \mathcal{X}_1$, $A \in \mathcal{X}_2$, $\mu_1(A) = \mu_*(A)$, and $\mu_2(A) = \mu^*(A)$.*

Although we shall not need this result in the remainder of the paper, it provides perhaps the best intuitive motivation for using inner and outer measures.

Now suppose we have a situation we want to reason about. Typically we do so by fixing a finite set $\Phi = \{p_1, \ldots, p_n\}$ of primitive propositions, which can be thought of as corresponding to basic events, such as "it is raining now" or "the coin landed heads". The set $\mathcal{L}(\Phi)$ of *(propositional) formulas* is the closure of Φ under the Boolean operations $\wedge$ and $\neg$. For convenience we assume also that there is a special formula *true*. We abbreviate $\neg true$ by *false*. The primitive propositions in Φ do not in general describe mutually exclusive events. To get mutually exclusive events, we can consider all the *atoms*, that is, all the formulas of the form $p_1' \wedge \ldots \wedge p_n'$, where p_i' is either p_i or $\neg p_i$. Let At denote the set of atoms (over Φ).

We have been using the word "event" informally, sometimes meaning "set" and sometimes meaning "formula". As we mentioned in the introduction, we are typically rather loose about this distinction. However, this distinction turns out to be crucial in some of our technical results, so we must treat it with some care here. Formally, a probability measure is a function on sets, not formulas. Fortunately, it is easy to shift focus from sets to formulas.

Using standard propositional reasoning, it is easy to see that any formula can be written as a disjunction of atoms. Thus, a formula φ can be identified with the unique set $\{\delta_1, \ldots, \delta_k\}$ of atoms such that $\varphi \Leftrightarrow \delta_1 \vee \ldots \vee \delta_k$. If we want to assign probabilities to all formulas, we can simply assign probabilities to each of the atoms, and then use the finite additivity property of probability measures to compute the probability of an arbitrary formula. This amounts to taking a probability space of the form $(At, 2^{At}, \mu)$. The states in the probability space are just the atoms, and the measurable subsets are all the sets of atoms (i.e., all formulas). Once we assign a measure to the singleton sets (i.e., to the atoms), we can extend by additivity to any subset. We call such a probability space a *Nilsson structure*, since this is essentially what Nilsson used to give meaning to formulas in his probability logic [Nil86].[3] Given a Nilsson structure $N = (At, 2^{At}, \mu)$ and a formula φ, let $W_N(\varphi)$ denote the *weight* or *probability* of φ in N, which is defined to be $\mu(At(\varphi))$, where $At(\varphi)$ is the set of atoms whose disjunction is equivalent to φ.

A more general approach is to take a *probability structure* to be a tuple $(S, \mathcal{X}, \mu, \pi)$, where $(S, \mathcal{X}, \mu)$ is a probability space, and π associates with each $s \in S$ a truth assignment

[3]Actually, the use of possible worlds in giving semantics to probability formulas goes back to Carnap [Car50].

$\pi(s)\colon \Phi \rightarrow \{\mathbf{true}, \mathbf{false}\}$. We say that p is *true at* s if $\pi(s)(p) = \mathbf{true}$; otherwise, we say that p is *false at* s.

We think of S as consisting of the possible states of the world. We can associate with each state s in S a unique atom describing the truth values of the primitive propositions in s. For example, if $\Phi = \{p_1, p_2\}$, and if $\pi(s)(p_1) = \mathbf{true}$ and $\pi(s)(p_2) = \mathbf{false}$, then we associate with s the atom $p_1 \wedge \neg p_2$. It is perfectly all right for there to be several states associated with the same atom (indeed, there may be an infinite number, since we allow S to be infinite, even though Φ is finite). This situation may occur if a state is not completely characterized by the events that are true there. This is the case, for example, if there are features of worlds that are not captured by the primitive propositions.[4] It may also be the case that there are some atoms not associated with any state.

We can easily extend $\pi(s)$ to a truth assignment on all formulas by taking the usual rules of propositional logic. Then if M is a probability structure, we can associate with every formula φ the set φ^M consisting of all the states in M where φ is true (i.e., the set $\{s \in S \mid \pi(s)(\varphi) = \mathbf{true}\}$). Of course, we assume that π is defined so that $true^M = S$. If p^M is measurable for every primitive proposition $p \in \Phi$, then φ^M is also measurable for every formula φ (since the set $\mathcal{X}$ of measurable sets is closed under complementation and countable union). We say M is a *measurable probability structure* if φ^M is measurable for every formula φ.

It makes sense to talk about the probability of φ in M only if φ^M is measurable; we can then take the probability of φ, which we denote $W_M(\varphi)$, to be $\mu(\varphi^M)$. If φ^M is not measurable, then we cannot talk about its probability. However, we can still talk about its inner measure and outer measure, since these are defined for all subsets. Intuitively, the inner and outer measure provide lower and upper bounds on the probability of φ. In general, if φ^M is not measurable, then we take $W_M(\varphi)$ to be $\mu_*(\varphi^M)$, i.e., the inner measure of φ in M.

We define a probability structure M and a Nilsson structure N to be *equivalent* if $W_M(\varphi) = W_N(\varphi)$ for every formula φ. Intuitively, a probability structure and a Nilsson structure are equivalent if they assign the same probability to every formula. The next theorem shows that there is a natural correspondence between Nilsson structures and measurable probability structures.

Theorem 2.2:

1. *For every Nilsson structure there is an equivalent measurable probability structure.*
2. *For every measurable probability structure there is an equivalent Nilsson structure.*

Proof: Given a Nilsson structure $N = (At, 2^{At}, \mu)$, let M_N be the probability structure $(At, 2^{At}, \mu, \pi)$, where for each $\delta \in At$, we define $\pi(\delta)(p) = \mathbf{true}$ iff δ logically implies

[4] Note that in the possible world semantics for temporal logic [MP81], there are in general many states associated with the same atom. There is a big difference between $p_1 \wedge \neg p_2$ being true today and its being true tomorrow. A similar phenomenon occurs in the multi-agent case of epistemic logic [HM92, RK86].

p (that is, iff p "appears positively" as a conjunct of δ). Clearly, M_N is a measurable probability structure. Further, it is easy to see that N and M_N are equivalent. Conversely, suppose $M = (S, \mathcal{X}, \mu, \pi)$ is a measurable probability structure. Let $N_M = (At, 2^{At}, \mu')$, where $\mu'(\delta) = \mu(\delta^M)$ for each $\delta \in At$. We leave it to the reader to verify that M and N_M are equivalent. Note that this construction does not work if δ^M is not a measurable set for some $\delta \in At$. ∎

Why should we even allow nonmeasurable sets? As the following examples show, using nonmeasurability allows us to avoid assigning probabilities to those events for which we have insufficient information to assign a probability.

Example 2.3: (This is a variant of an example found in [Dem68].) Suppose we want to know the probability of a projectile landing in water. Unfortunately the region where the projectile might land is rather inadequately mapped. We are able to divide the region up into (disjoint) regions, and we know for each of these subregions (1) the probability of the projectile landing in that subregion and (2) whether the subregion is completely covered by water, completely contained on land, or has both land and water in it. Suppose the subregions are $R_1, \ldots, R_n$. We now construct a probability structure $M = (S, \mathcal{X}, \mu, \pi)$ to represent this situation. For every subregion R we have a state (R, w) if R has some water in it and a state (R, ℓ) if it has some land in it. Thus, S has altogether somewhere between n and $2n$ states. Note that to every subregion R there corresponds a set R' consisting of either (R, w) or (R, ℓ) or both. Let $\mathcal{X}$ be the σ-algebra generated by the basis $R'_1, \ldots, R'_n$. Define $\mu(R'_i)$ to be the probability of the projectile landing in R_i. Thus, we have encoded all the information we have about the subregion R into R'; R' tells us whether there is land and water or both in R, and (through μ) the probability of the projectile landing in R. We have two primitive propositions in the language: *land* and *water*. We define π in the obvious way: *land* is true at states of the form (R_i, ℓ) and false at others, while just the opposite is true for *water*. This completes the description of M.

We are interested in the probability of the projectile landing in water. Intuitively, this is the probability of the set $WATER = water^M$. However, this set is not measurable (unless every region is completely contained in either land or water). We do have lower and upper bounds on the probability of the projectile landing in water, given by $\mu_*(WATER)$ and $\mu^*(WATER)$. It is easy to see that $\mu_*(WATER)$ is precisely the probability of landing in a region that is completely covered by water, while $\mu^*(WATER)$ is the probability of landing in a region that has some water in it. ∎

Example 2.4: Ron has two blue suits and two gray suits. He has a very simple method for deciding what color suit to wear on any particular day: he simply tosses a (fair) coin: if it lands heads he wears a blue suit, and if it lands tails he wears a gray suit. Once he's decided what color suit to wear, he just chooses the rightmost suit of that color on the rack. Both of Ron's blue suits are single-breasted, while one of Ron's gray suits is single-breasted and the other is double-breasted. Ron's wife Susan is (fortunately for Ron) a little more fashion-conscious than he is. She also knows how Ron makes his sartorial choices. So, from time to time, she makes sure that the gray suit she considers preferable is to the right (which it is depends on current fashions and perhaps on other whims of Susan).[5] Suppose we don't know about the current fashions (or about Susan's current whims). What can we say about the probability of Ron's wearing a single-breasted suit on Monday?

In terms of possible worlds, it is clear that there are four possible worlds, one corresponding to each of the suits that Ron could choose. For definiteness, suppose states s_1 and s_2 correspond to the two blue suits, s_3 corresponds to the single-breasted gray suit, and s_4 corresponds to the double-breasted gray suit. Let $S = \{s_1, s_2, s_3, s_4\}$. There are two features of interest about a suit: its color and whether it is single-breasted or double-breasted. Let the primitive proposition g denote "the suit is gray" and let db denote "the suit is double-breasted", and define the truth assignment π in the obvious way. Note that the atom $\neg g \wedge \neg db$ is associated with both states s_1 and s_2. Since the two blue suits are both single-breasted, these two states cannot be distinguished by the formulas in our language.

What are the measurable events? Besides S itself and the empty set, the only other candidates are $\{s_1, s_2\}$ ("Ron chooses a blue suit") and $\{s_3, s_4\}$ ("Ron chooses a gray suit"). However, $SB = \{s_1, s_2, s_3\}$ ("Ron chooses a single-breasted suit") is nonmeasurable. The reason is that we do not have a probability on the event "Ron chooses a single-breasted suit, given that Ron chooses a gray suit", since this in turn depends on the probability that Susan put the single-breasted suit to the right of the other gray suit, which we do not know. Susan's choice might be characterizable by a probability distribution; it might also be deterministic, based on some complex algorithm which even she might not be able to describe; or it might be completely nondeterministic, in which case it is not technically meaningful to talk about the "probability" of Susan's actions! Our ignorance here is captured by nonmeasurability. Informally, we can say that the probability of Ron choosing a single-breasted suit lies somewhere in the interval $[1/2, 1]$, since it is bounded below by the probability of Ron choosing a blue suit. This is an informal statement because formally it does not make sense to talk about the probability of a nonmeasurable event. The formal analogue is simply that the inner measure of SB is $1/2$, while its outer measure is 1. ∎

3 The Dempster-Shafer theory of evidence

The Dempster-Shafer theory of evidence [Sha76] provides another approach to attaching likelihoods to events. This theory starts out with a *belief function* (sometimes called a *support 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

[5] Any similarity between the characters in this example and the first author of this paper and his wife Susan is not totally accidental.

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. Just as we defined probability structures, we can define a *DS structure* (where, of course, "DS" stands for Dempster-Shafer) to be a tuple (S, Bel, π), where S and π are as before, and where $Bel: 2^S \rightarrow [0,1]$ is a function satisfying:

B1. $Bel(\emptyset) = 0$

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

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

A belief function is typically defined on a *frame of discernment*, consisting of mutually exclusive and exhaustive propositions describing the domain of interest. We think of the set S of states in a DS structure as being this frame of discernment. We could always choose S to be some subset of At, the set of atoms, so that its elements are in fact propositions in the language. In general, given a DS structure $D = (S, Bel, \pi)$ and formula φ, we define the weight $W_D(\varphi)$ to be $Bel(\varphi^D)$, where φ^D is the set of states where φ is true. Thus we can talk about the degree of belief in φ in the DS structure D, described by $W_D(\varphi)$, by identifying φ with the set φ^D and considering the belief in φ^D. As before, we define a probability structure M (resp., a Nilsson structure N, a DS structure D') and a DS structure D to be *equivalent* if $W_M(\varphi) = W_D(\varphi)$ (resp., $W_N(\varphi) = W_D(\varphi)$, $W_{D'}(\varphi) = W_D(\varphi)$) for every formula φ.

Property B3 may seem unmotivated. Perhaps the best way to understand it is as an analogue to the usual inclusion-exclusion rule for probabilities [Fel57, p. 89], which is obtained by replacing the inequality by equality (and the belief function Bel by a probability measure μ). In particular, B3 holds for probability measures (we prove a more general result, namely that it holds for all inner measures induced by probability measures, in Proposition 3.1 below). Hence, if $(S, \mathcal{X}, \mu)$ is a probability space and $\mathcal{X} = 2^S$ (making every subset of S measurable), then μ is a belief function. (This fact has been observed frequently before; see, for example, [Sha76].) It follows that every Nilsson structure is a DS structure.

It is easy to see that the converse does not hold. For example, suppose there is only one primitive proposition, say p, in the language, so that $At = \{p, \neg p\}$, and let $D_0 = (At, Bel, \pi)$ be such that $Bel(\{p\}) = 1/2$, $Bel(\{\neg p\}) = 0$, and π is defined in the obvious way. Intuitively, there is weight of evidence $1/2$ for p, and no evidence for $\neg p$. Thus $W_{D_0}(p) = 1/2$ and $W_{D_0}(\neg p) = 0$. D_0 is not equivalent to any Nilsson structure, since if N is a Nilsson structure such that $W_N(p) = 1/2$, then we must have $W_N(\neg p) = 1/2$.

These observations tell us that in some sense belief functions are more general than probability measures, provided we restrict attention to probability spaces where all sets are measurable. This fact is well known. Indeed, in [Sha76], Shafer makes explicit use of the greater generality of belief functions. While he does consider events E such that $Bel(\neg E) = 1 - Bel(E)$ (he calls such events *probabilistic*), he also wants to allow non-probabilistic events. He gives examples of events where the fact that we would like to assign weight .8 to our belief in event E does not mean that we want to assign weight .2 to our belief in $\neg E$. In our framework, where we allow nonmeasurable sets, we can view probabilistic events as corresponding to measurable sets, while non-probabilistic events do not. We can push this analogy much further. Not only do nonmeasurable sets correspond to non-probabilistic events, but the inner measures induced by probability measures correspond to belief functions.

The following result is the first step to proving the correspondence between inner measures and belief functions. It says that every inner measure is a belief function (and thus generalizes the statement that every probability measure is a belief function). As we mentioned in the introduction, the result actually follows from a more general result in [Sha79]. We give a proof here to keep this paper self-contained. (Later, we give an easier proof for the case where the probability space is finite.)

Proposition 3.1: *If $(S, \mathcal{X}, \mu)$ is a probability space, then μ_* is a belief function on S.*

Proof: Clearly μ_* satisfies B1 and B2, so it suffices to show that it satisfies B3. Given sets $A_1, \ldots, A_k$, let $B_1, \ldots, B_k$ be measurable sets such that $B_i \subseteq A_i$ and $\mu_*(A_i) = \mu(B_i)$, $i = 1, \ldots, k$. We now show that for any subset I of $\{1, \ldots, k\}$, we have $\mu_*(\bigcap_{i \in I} A_i) = \mu(\bigcap_{i \in I} B_i)$. Since $\bigcap_{i \in I} B_i \subseteq \bigcap_{i \in I} A_i$, we must have $\mu_*(\bigcap_{i \in I} A_i) \geq \mu(\bigcap_{i \in I} B_i)$. To see that equality holds, let C be a measurable set such that $C \subseteq \bigcap_{i \in I} A_i$ and $\mu(C) = \mu_*(\bigcap_{i \in I} A_i)$ (as we noted earlier, there is such a set C). We can assume that $(\bigcap_{i \in I} B_i) \subseteq C$ (by replacing C by $C \cup (\bigcap_{i \in I} B_i)$ if necessary). Let $D = C \cap \overline{\bigcap_{i \in I} B_i}$. Since D is a Boolean combination of measurable sets, it is measurable. We now show that $\mu(D) = 0$. Let i be a fixed member of I. We know that $D \subseteq C \subseteq A_i$. If $\mu(D \cap \overline{B_i}) \neq 0$, then $B_i \cup D$ would be a subset of A_i with bigger measure than B_i, which would contradict our choice of B_i. So $\mu(D \cap \overline{B_i}) = 0$, for each $i \in I$. Since $D \subseteq \overline{\bigcap_{i \in I} B_i} = \bigcup_{i \in I} \overline{B_i}$, it follows that $\mu(D) = \mu(D \cap \bigcup_{i \in I} \overline{B_i}) = \mu(\bigcup_{i \in I}(D \cap \overline{B_i})) \leq \sum_{i \in I} \mu(D \cap \overline{B_i}) = 0$, as desired. Since $(\bigcap_{i \in I} B_i) \subseteq C$ and $\mu(C \cap \overline{\bigcap_{i \in I} B_i}) = \mu(D) = 0$, it follows that $\mu(C) = \mu(\bigcap_{i \in I} B_i)$. So $\mu_*(\bigcap_{i \in I} A_i) = \mu(\bigcap_{i \in I} B_i)$, as desired. Since $\bigcup_{i \in I} A_i \supseteq \bigcup_{i \in I} B_i$, we also get $\mu_*(\bigcup_{i \in I} A_i) \geq \mu(\bigcup_{i \in I} B_i)$ (although in this case we do not in general get equality).

Now applying the inclusion-exclusion rule for probabilities, we get:

$$
\begin{array}{rl}
 & \mu_*(A_1 \cup \ldots \cup A_m) \\
\geq & \mu(B_1 \cup \ldots \cup B_m) \\
= & \sum_{I \subseteq \{1, \ldots, m\}, I \neq \emptyset} (-1)^{|I|+1} \mu(\bigcap_{i \in I} B_i) \\
= & \sum_{I \subseteq \{1, \ldots, m\}, I \neq \emptyset} (-1)^{|I|+1} \mu_*(\bigcap_{i \in I} A_i)
\end{array}
$$

This completes the inductive proof. ∎

Corollary 3.2: *For every probability structure, there is an equivalent DS structure.*

Proof: Given a probability structure $M = (S, \mathcal{X}, \mu, \pi)$, consider the DS structure $D = (S, \mu_*, \pi)$, where μ_* is the inner measure on S induced by μ. By Proposition 3.1, μ_* is an inner measure. Clearly $W_M(\varphi) = \mu_*(\varphi^M) = \mu_*(\varphi^D) = W_D(\varphi)$ for all formulas φ, so D and M are equivalent. ∎

Proposition 3.1 says that every inner measure is a belief function. The converse does not quite hold. For example, consider the DS structure D_0 defined above. There is no probability measure μ that we can define on $\{p, \neg p\}$ such that $\mu_* = Bel$. However, it is easy to define a probability structure M_0 such that $\mu_*(p^{M_0}) = 1/2$ and $\mu_*(\neg p^{M_0}) = 0$. For example, we can take M_0 to consist of three states s_1, s_2, s_3, such that p is true at s_1 and s_2, and p is false at s_3. Further suppose that the only nontrivial measurable sets are $\{s_1\}$ and $\{s_2, s_3\}$, each of which has probability $1/2$. It is now easy to check that M_0 is equivalent to D_0. Thus, although in this case we could not find an inner measure equal to Bel when viewed as a function on *sets*, we can find an inner measure equivalent to Bel when viewed as a function on *formulas*.

This observation generalizes to give us the converse of Corollary 3.2.

Theorem 3.3: *For every DS structure there is an equivalent probability structure.*

In order to prove Theorem 3.3, we first need to consider finite DS structures (those with only finitely many states). In much of the work on belief functions (for example, [Sha76]), the set of states in a DS structure is assumed to be finite. As we now show, as far as equivalence is concerned, restricting to a finite set of states does not result in any loss of generality.

Proposition 3.4: *Every DS structure is equivalent to a finite DS structure.*

Proof: Fix a DS structure $D = (S, Bel, \pi)$. Let At be the set of atoms. Since we have assumed that there are only finitely many primitive propositions in the language, At is finite. For every subset $A = \{\delta_1, \ldots, \delta_k\} \subseteq At$, define $\psi_A = \delta_1 \vee \ldots \vee \delta_k$; we identify $\psi_\emptyset$ with the formula *false*. Let $D' = (At, Bel', \pi')$, where $\pi'(\delta)(p) = \mathbf{true}$ iff p is one of the conjuncts of the atom δ, and $Bel'(A) = Bel((\psi_A)^D)$ for $A \subseteq At$. We need to check that Bel' is indeed a belief function. It clearly satisfies B1 and B2. For B3, observe that

$$\begin{array}{rl} & Bel'(A_1 \cup \ldots \cup A_k) \\ = & Bel((\psi_{A_1 \cup \ldots \cup A_k})^D) \\ = & Bel(\psi_{A_1}^D \cup \ldots \cup \psi_{A_k}^D) \\ \geq & \sum_{I \subseteq \{1,\ldots,k\}, I \neq \emptyset} (-1)^{|I|+1} Bel(\bigcap_{i \in I} \psi_{A_i}^D) \\ = & \sum_{I \subseteq \{1,\ldots,k\}, I \neq \emptyset} (-1)^{|I|+1} Bel((\bigwedge_{i \in I} \psi_{A_i})^D) \\ = & \sum_{I \subseteq \{1,\ldots,k\}, I \neq \emptyset} (-1)^{|I|+1} Bel((\psi_{\cap_{i \in I} A_i})^D) \\ = & \sum_{I \subseteq \{1,\ldots,k\}, I \neq \emptyset} (-1)^{|I|+1} Bel'(\cap_{i \in I} A_i). \end{array}$$

Thus, Bel' is a belief function. Moreover, for every formula φ we have $W_{D'}(\varphi) = Bel'(\varphi^{D'}) = Bel'(At(\varphi)) = Bel((\psi_{At(\varphi)})^D) = Bel(\varphi^D) = W_D(\varphi)$, so D and D' are equivalent. ∎

We remark that this proof depends on the fact that Φ is finite, so that At is finite. This is the one place we use the finiteness of Φ. However, even if Φ is infinite, for any finite subset $\Phi' \subseteq \Phi$, we can construct a finite DS D' which is equivalent to D for all formulas in $\mathcal{L}(\Phi')$. As we shall see, this suffices for the results we are interested in even if Φ is infinite.

In the case of finite DS structures, we can characterize belief functions in terms of *mass functions*. A mass function on S is simply a function $m\colon 2^S \to [0,1]$ 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 3.5: ([Sha76, page 39])

1. *If m is a mass function on S, then the function $Bel\colon 2^S \to [0,1]$ defined by $Bel(A) = \sum_{B\subseteq A} m(B)$ is a belief function on S.*
2. *If Bel is a belief function on S and S is finite, then there is a unique mass function m on S such that $Bel(A) = \sum_{B\subseteq A} m(B)$ for every subset A of S.*

We remark that the assumption that S is finite in part (2) of Proposition 3.5 is needed.

We remark that if we restrict to finite structures, then Proposition 3.5 gives us an easy way of proving the result of Proposition 3.1, that every inner measure is a belief function. Suppose $(S, \mathcal{X}, \mu)$ is a probability space, and S is finite. Let $\mathcal{Y}$ be a basis of $\mathcal{X}$ (since S is finite, $\mathcal{X}$ must have a basis). Define $m(A) = \mu(A)$ for $A \in \mathcal{Y}$ and $m(A) = 0$ otherwise. Clearly m is a mass function on S. Moreover,

$$\mu_*(B) = \sum_{A\subseteq B, A\in\mathcal{Y}} \mu(A) = \sum_{A\subseteq B, A\in\mathcal{Y}} m(A).$$

By Propostion 3.5, μ_* is the belief function corresponding to m.

We now have the machinery we need to prove Theorem 3.3:

Proof of Theorem 3.3: Let $D = (S, Bel, \pi)$ be a DS structure. By Theorem 3.4, we can assume without loss of generality that D is a finite structure. Let m be the mass function corresponding to the belief function Bel as in part (2) of Proposition 3.5. Let

$M = (S', \mathcal{X}, \mu, \pi')$ be defined as follows. Define $S' = \{(A, s) \mid A \subseteq S,\ s \in A\}$. For $A \subseteq S$, let $A^* = \{(A, s) \mid s \in A\}$. Note that if $A \neq B$ then A^* and B^* are disjoint; moreover, $\cup_{A \subseteq S} A^* = S'$. Since S is finite, so is S'. We take $\{A^* \mid A \subseteq S\}$ to be the basis for $\mathcal{X}$; thus every element of $\mathcal{X}$ is a union of sets of the form A^*. Define $\mu(A^*) = m(A)$; we then extend μ to all of $\mathcal{X}$ by finite additivity. Finally, let $\pi'(A, s) = \pi(s)$.

We now want to show that M is equivalent to D. For each formula φ, we have $A^* \subseteq \varphi^M$ iff $(A, s) \in \varphi^M$ for all $s \in A$ iff $\pi(s)(\varphi) = \mathbf{true}$ for all $s \in A$ iff $A \subseteq \varphi^D$. Thus, the largest measurable set contained in φ^M is $\cup_{A \subseteq \varphi^D} A^*$. It now follows that

$$\begin{array}{rcl} W_M(\varphi) & = & \mu_*(\varphi^M) \\ & = & \mu(\cup_{A \subseteq \varphi^D} A^*) \\ & = & \sum_{A \subseteq \varphi^D} \mu(A^*) \\ & = & \sum_{A \subseteq \varphi^D} m(A) \\ & = & Bel(\varphi^D) \\ & = & W_D(\varphi) \end{array}$$

Thus, D and M are equivalent, as desired. ∎

Intuitively, Corollary 3.2 and Theorem 3.3 show that belief functions and inner measures induced by probability measures are precisely the same if their domains are considered to be formulas rather than sets. Recall the DS structure D_0 introduced above, with $w_{D_0}(p) = 1/2$ and $w_{D_0}(\neg p) = 0$. Although we saw that there is no probability measure μ that we can define on $\{p, \neg p\}$ such that $\mu_* = Bel$, we also showed that there is a probability structure M_0 such that $\mu_*(p^{M_0}) = 1/2$ and $\mu_*(\neg p^{M_0}) = 0$. This is a special case of Theorem 3.3, which says intuitively that every belief function is an inner measure, provided we consider the domains to be formulas rather than sets. Of course, we can still relate belief functions and inner measures if we look at sets, although the relationship is not quite as elegant. One connection is provided by the following corollary to the proof of Theorem 3.3, which can be viewed as a partial converse to Proposition 3.1. In the language of [Sha76, Chap. 6], it says that every belief function is the *restriction* of some probability function (although we do not restrict attention—as Shafer does in [Sha76]—to probability functions with respect to which every subset of the sample space is measurable).

Corollary 3.6: *Given a belief function Bel defined on a finite set S, there is a probability space $(S', \mathcal{X}, \mu)$ and a surjection $f : S' \to S$ such that for each $A \subseteq S$, we have $Bel(A) = \mu_*(f^{-1}(A))$.*

Proof: Given S and Bel, construct the probability space $(S', \mathcal{X}, \mu)$ as in the proof of Theorem 3.3. Define $f((A, s)) = s$, for $s \in A \subseteq S$. Clearly f is a surjection from S' to S. Moreover, it is easy to see that $f^{-1}(A) = \{(B, s) | s \in A \cap B\}$. It follows that $B^* \subseteq f^{-1}(A)$ iff $B \subseteq A$, where $B^* = \{(B, s) | s \in B\}$, as defined in the proof of Theorem 3.3. From this observation, we get

$$\begin{array}{rcl} \mu_*(f^{-1}(A)) & = & \sum_{B \subseteq A} \mu(B^*) \\ & = & \sum_{B \subseteq A} m(B) \\ & = & Bel(A). \quad \blacksquare \end{array}$$

Thus, we have proved our claim that inner measures and belief functions are identical, providing we view them both as being functions on formulas rather than on sets. Note that all that is required for this equivalence is that the formulas be allowed to correspond to nonmeasurable sets. As our earlier examples have shown, this possibility arises in a natural way in many examples. As we shall see in Section 5, one consequence of the equivalence of inner measures and belief functions is that the same axioms characterize both. But first we show how our approach can capture the spirit of Dempster's rule of combination.

4 Combining evidence

An important issue for belief functions, each of which can be viewed as representing a distinct body of evidence, is how to combine them to obtain a new belief function that somehow reflects the combined evidence. A way of doing so is provided by Dempster's *rule of combination*, which was introduced by Dempster [Dem68] and was further developed and studied in an elegant and rather complete manner by Shafer [Sha76]. Since the definition of rule of combination is usually given in terms of mass functions and finite Dempster-Shafer structures, we do so here as well. (We remark that a more general version of the rule of combination for infinite Dempster-Shafer structures is given in [Sha]. All our results can be extended to the infinite case as well.)

If m_1 and m_2 are mass functions with the same domain 2^S, let $m_1 \oplus m_2$ be the mass function m where $m(A) = c\sum_{\{B_1,B_2 \,|\, B_1 \cap B_2 = A\}} m_1(B_1)m_2(B_2)$ for each nonempty $A \subseteq S$, and where c is a normalizing constant chosen so that the sum of all of the $m(A)$'s is 1. It is easy to check that $c = (\sum_{\{B_1,B_2 \,|\, B_1 \cap B_2 \neq \emptyset\}} m_1(B_1)m_2(B_2))^{-1}$. Note that if there is no pair B_1, B_2 where $B_1 \cap B_2 \neq \emptyset$ and $m_1(B_1)m_2(B_2) > 0$, then we cannot find such a normalizing constant c. In this case $m_1 \oplus m_2$ is undefined. If Bel_1 and Bel_2 are belief functions with mass functions m_1 and m_2 respectively, then the belief function (denoted $Bel_1 \oplus Bel_2$) that is the result of combining Bel_1 and Bel_2 is the belief function with mass function $m_1 \oplus m_2$ (and is undefined if $m_1 \oplus m_2$ is undefined). If $D_1 = (S, Bel_1, \pi)$ and $D_2 = (S, Bel_2, \pi)$ are DS structures with the same sample space S and the same truth assignment function π, then $D_1 \oplus D_2$ is the DS structure $(S, Bel_1 \oplus Bel_2, \pi)$ (and is undefined if $Bel_1 \oplus Bel_2$ is undefined).

We now give a natural way (in the spirit of Dempster's rule) to define the combination of two probability spaces $(S, \mathcal{X}_1, \mu_1)$ and $(S, \mathcal{X}_2, \mu_2)$ with the same finite sample space S. Suppose $\mathcal{X}_i$ has basis $\mathcal{X}_i'$, $i = 1, 2$. (We restrict S to be finite in order to ensure that $\mathcal{X}_i$ has a basis.) Let $\mathcal{X}_1 \oplus \mathcal{X}_2$ be the σ-algebra generated by the basis consisting of the nonempty sets of the form $X_1 \cap X_2$, $X_i \in \mathcal{X}_i'$, $i = 1, 2$. Define $\mu_1 \oplus \mu_2$ to be the probability measure on $\mathcal{X}_1 \oplus \mathcal{X}_2$ such that $(\mu_1 \oplus \mu_2)(X_1 \cap X_2) = c\mu_1(X_1)\mu_2(X_2)$ for $X_1 \in \mathcal{X}_1'$ and $X_2 \in \mathcal{X}_2'$ where $c = (\sum_{\{X_1 \in \mathcal{X}_1', X_2 \in \mathcal{X}_2' \,|\, X_1 \cap X_2 \neq \emptyset\}} \mu_1(X_1)\mu_2(X_2))^{-1}$ is a normalizing constant chosen so that the measure of the whole space is 1. (If there is no pair X_1, X_2 where $\mu_1(X_1)\mu_2(X_2) > 0$ then, as before, $\mu_1 \oplus \mu_2$ is undefined.) Finally, if $M_1 = (S, \mathcal{X}_1, \mu_1, \pi)$ and $M_2 = (S, \mathcal{X}_2, \mu_2, \pi)$ are probability structures with the same finite sample space S and same truth assignment function π, then we define $M_1 \oplus M_2$ to be the probability structure $(S, \mathcal{X}_1 \oplus X_2, \mu_1 \oplus \mu_2, \pi)$ (as before, $M_1 \oplus M_2$ is undefined if $\mu_1 \oplus \mu_2$ is undefined).

Providing a detailed discussion of the motivation of this way of combining probabilities is beyond the scope of this paper. The intuition behind it is very similar to that behind the rule of combination. Suppose we have two tests, T_1 and T_2. Further suppose that, according to test T_i, an element $s \in S$ is in $X_i \in \mathcal{X}_i'$ with probability $\mu_i(X_i)$, $i = 1, 2$. In that case, if we combine the results of both tests and we assume that they are independent, we might say that the probability of a randomly $s \in S$ being in $X_1 \cap X_2$ is $c\mu_1(X_1)\mu_2(X_2)$, where c is the appropriate normalizing constant.

The next theorem shows how the spirit of Dempster's rule of combination can be captured within our framework.

Theorem 4.1: *Let D_1 and D_2 be DS structures where $D_1 \oplus D_2$ is defined. There are probability structures M_1 and M_2 such that (a) D_1 is equivalent to M_1, (b) D_2 is equivalent to M_2, and (c) $D_1 \oplus D_2$ is equivalent to $M_1 \oplus M_2$.*

Proof: We give a construction which is a variation of that in the proof of Theorem 3.3. Let D_i be (S, Bel_i, π) and let m_i be the mass function corresponding to Bel_i, $i = 1, 2$. We now define $M_i = (S', X_i, \mu_i, \pi)$, $i = 1, 2$, as follows. Let $S' = \{(A, B, s) \,|\, A, B \subseteq S \text{ and } s \in A \cap B\}$. For $A \subseteq S$, define $A^1 = \{(A, B, s) \,|\, B \subseteq S \text{ and } s \in A \cap B\}$ and $A^2 = \{(B, A, s) \,|\, B \subseteq S \text{ and } s \in A \cap B\}$. Recall we have restricted to finite DS structures for the purposes of this section. It follows that there are only finitely many sets of the form A^i, $i = 1, 2$. Let $\mathcal{X}_i$ be the σ-algebra generated by the basis sets A^i and let $\mu_i(A^i) = m_i(A)$, $i = 1, 2$. Extend μ_i to all of $\mathcal{X}_i$ by finite additivity. Finally, define $\pi'(A, B, s) = \pi(s)$. The same arguments as those used in Theorem 3.3 can now be used to show that D_i is equivalent to M_i, $i = 1, 2$; we leave details to the reader.

We now want to show that $D_1 \oplus D_2$ is equivalent to $M_1 \oplus M_2$. Let us denote $D_1 \oplus D_2$ by D, $M_1 \oplus M_2$ by M, $m_1 \oplus m_2$ by m, and $\mu_1 \oplus \mu_2$ by μ. Note that $A^1 \cap B^2 = \{(A, B, s) \,|\, s \in A \cap B\}$, so that $A^1 \cap B^2$ is nonempty iff $A \cap B$ is nonempty. Since $m_1 \oplus m_2$ is defined by assumption, it follows that $m_1(A)m_2(B) > 0$ for some $A, B \subseteq S$, and hence that $\mu_1(A^1)\mu_2(B^2) > 0$, so that $\mu_1 \oplus \mu_2$ is defined. Moreover, we have $\sum_{\{A, B \,|\, A \cap B \neq \emptyset\}} m_1(A)m_2(B) = \sum_{\{A^1, B^2 \,|\, A^1 \cap B^2 \neq \emptyset\}} \mu_1(A^1)\mu_2(B^2)$, so that the normalizing constant for $m_1 \oplus m_2$ is the same as that for $\mu_1 \oplus \mu_2$; let us call them both c.

Just as in the proof of Theorem 3.3, we can show that for any formula φ, we have

$A^1 \cap B^2 \subseteq \varphi^M$ iff $A \cap B \subseteq \varphi^D$. Thus we have

$$\begin{array}{rcl} Bel(\rho^D) & = & \sum_{\{C \mid C \subseteq \varphi^D\}} m(C) \\ & = & \sum_{\{C \mid C \subseteq \varphi^D, C \neq \emptyset\}} m(C) \\ & = & \sum_{\{C \mid C \subseteq \varphi^D, C \neq \emptyset\}} c(\sum_{\{A,B \mid A \cap B = C\}} m_1(A) m_2(B)) \\ & = & c \sum_{\{A,B \mid A \cap B \subseteq \varphi^D, A \cap B \neq \emptyset\}} m_1(A) m_2(B) \\ & = & c \sum_{\{A^1,B^2 \mid A^1 \cap B^2 \subseteq \varphi^M, A^1 \cap B^2 \neq \emptyset\}} \mu_1(A^1) \mu_2(B^2) \\ & = & \mu_*(\varphi^M). \end{array}$$

Thus D and M are equivalent, as desired. ∎

We have shown that given DS structures D_1 and D_2, there exist probability structures M_1 and M_2 such that D_1 is equivalent to M_1, D_2 is equivalent to M_2, and $D_1 \oplus D_2$ is equivalent to $M_1 \oplus M_2$. The reader might wonder if it is the case that for *any* probability structures M_1 and M_2 such that D_i is equivalent to M_i, $i = 1, 2$, we have that $D_1 \oplus D_2$ is equivalent to $M_1 \oplus M_2$. This is not the case, as the following example shows. Let $D_1 = D_2 = D_0$, where D_0 is the DS structure defined in the previous section. The mass function m associated with both D_1 and D_2 has $m(\{p\}) = 1/2$, $m(\{p, \neg p\}) = 1/2$, and $m(A) = 0$ for all other subsets $A \subseteq \{p, \neg p\}$. Now let $M_1 = M_2 = (\{a, b, c, d\}, \mathcal{X}, \mu, \pi')$, where $\pi(a)(p) = \pi(b)(p) = \pi(c)(p) = \mathbf{true}$, $\pi(d)(p) = \mathbf{false}$, the sets $\{a\}$, $\{b\}$, and $\{c, d\}$ form a basis for $\mathcal{X}$, and $\mu(\{a\}) = \mu(\{b\}) = 1/4$, $\mu\{c, d\} = 1/2$. It is easy to see that M_i is equivalent to D_i, for $i = 1, 2$. However, it is also easy to check that $\mu \oplus \mu$ assigns probability 1/6 to each of $\{a\}$ and $\{b\}$, and probability 2/3 to $\{c, d\}$, while $m \oplus m = m$. It now follows that $D_1 \oplus D_2$ is not equivalent to $M_1 \oplus M_2$.

While Theorem 4.1 shows that there is a sense in which we can define a rule of combination on probability spaces that simulates Dempster's rule of combination, it does not provide an explanation of the rule of combination in terms of inner measures. (This is in contrast to our other results that show how we can view belief functions in terms of inner measures.) There is good reason for this: we feel that Dempster's rule of combination does not fit into the framework of viewing belief functions in terms of inner measures. A discussion of this point is beyond the scope of this paper; for more details, the reader is encouraged to consult [FH91, HF92].

5 Reasoning about belief and probability

We are often interested in the inferences we can make about probabilities or beliefs given some information. In order to do this, we need a language for doing such reasoning. In [FHM90], two languages are introduced for reasoning about probability, and results regarding complete axiomatizations and decision procedures are proved. We review these results here, and show that by combining them with Corollary 3.2 and Theorem 3.3, we obtain analogous results for reasoning about belief functions. We consider the simpler language first. A *term* in this language is an expression of the form $a_1 w(\varphi_1) + \cdots +$

$a_k w(\varphi_k)$, where $a_1, \ldots, a_k$ are integers and $\varphi_1, \ldots, \varphi_k$ are propositional formulas. A *basic weight formula* is one of the form $t \geq b$, where t is a term and b is an integer. A *weight formula* is a Boolean combination of basic weight formulas. We sometimes use obvious abbreviations, such as $w(\varphi) \geq w(\psi)$ for $w(\varphi) - w(\psi) \geq 0$, $w(\varphi) \leq b$ for $-w(\varphi) \geq -b$, $w(\varphi) > b$ for $\neg(w(\varphi) \leq b)$, and $w(\varphi) = b$ for $(w(\varphi) \geq b) \wedge (w(\varphi) \leq b)$. A formula such as $w(\varphi) \geq 1/3$ can be viewed as an abbreviation for $3w(\varphi) \geq 1$; we can always allow rational numbers in our formulas as abbreviations for the formula that would be obtained by clearing the denominator.

We give semantics to the formulas in our language with respect to all the structures we have been considering. Let K be either a Nilsson structure, a probability structure, or a DS structure, and let f be a weight formula. We now define what it means for K to *satisfy* f, written $K \models f$. For a basic weight formula,

$$K \models a_1 w(\varphi_1) + \cdots + a_k w(\varphi_k) \geq b \text{ iff } a_1 W_K(\varphi_1) + \cdots + a_k W_K(\varphi_k) \geq b.$$

We then extend $\models$ in the obvious way to conjunctions and negations. The interpretation of $w(\varphi)$ is either "the probability of φ" (if we are dealing with Nilsson structures or measurable probability structures), "the inner measure of φ" (if we are dealing with general probability structures), or "the belief in φ" (if we are dealing with DS structures). Consider, for example, the formula $w(\varphi_1) \geq 2w(\varphi_2)$. In a Nilsson structure N, we would interpret this as "φ_1 is twice as probable as φ_2". In a DS structure D, the formula would look structurally identical, but the interpretation would be that our belief in φ_1 is twice as great as our belief in φ_2. Notice how allowing linear combinations of weights adds to the expressive power of our language.

Let $\mathcal{K}$ be a class of structures (in the cases of interest to us, $\mathcal{K}$ is either probability structures, measurable probability structures, Nilsson structures, or DS structures). As usual, we define a weight formula f to be *satisfiable with respect to* $\mathcal{K}$ if $K \models f$ for some $K \in \mathcal{K}$. Similarly, f is *valid with respect to* $\mathcal{K}$ if $K \models f$ for all $K \in \mathcal{K}$.

We now turn our attention to complete axiomatizations. Consider the following axiom system AX_{MEAS} for reasoning about measurable probability structures, taken from [FHM90]. The system divides nicely into three parts, which deal respectively with propositional reasoning, reasoning about linear inequalities, and reasoning about probabilities.

Propositional reasoning:

Taut. All instances of propositional tautologies[6]

MP. From f and $f \Rightarrow g$ infer g (modus ponens)

Reasoning about inequalities:

[6] We remark we could replace Taut by a simpler collection of axioms that characterize propositional tautologies (see, for example [Men64]). We have not done so here because we want to focus on the other axioms.

I1. $(a_1w(\varphi_1) + \cdots + a_kw(\varphi_k) \geq b) \Leftrightarrow (a_1w(\varphi_1) + \cdots + a_kw(\varphi_k) + 0w(\varphi_{k+1}) \geq b)$ (adding and deleting 0 terms)

I2. $(a_1w(\varphi_1) + \cdots + a_kw(\varphi_k) \geq b) \Rightarrow (a_{j_1}w(\varphi_{j_1}) + \cdots + a_{j_k}w(\varphi_{j_k}) \geq b)$, if $j_1, \ldots, j_k$ is a permutation of $1, \ldots, k$ (permutation)

I3. $(a_1w(\varphi_1) + \cdots + a_kw(\varphi_k) \geq b) \wedge (a'_1w(\varphi_1) + \cdots + a'_kw(\varphi_k) \geq b') \Rightarrow (a_1 + a'_1)w(\varphi_1) + \cdots + (a_k + a'_k)w(\varphi_k) \geq (b + b')$ (addition of coefficients)

I4. $(a_1w(\varphi_1) + \cdots + a_kw(\varphi_k) \geq b) \Leftrightarrow (ca_1w(\varphi_1) + \cdots + ca_kw(\varphi_k) \geq cb)$ if $c > 0$ (multiplication and division of nonzero coefficients)

I5. $(t \geq b) \vee (t \leq b)$ if t is a term (dichotomy)

I6. $(t \geq b) \Rightarrow (t > b')$ if t is a term and $b > b'$ (monotonicity)

Reasoning about probabilities:

W1. $w(\varphi) \geq 0$ (nonnegativity)

W2. $w(true) = 1$ (the probability of the event $true$ is 1)

W3. $w(\varphi \wedge \psi) + w(\varphi \wedge \neg\psi) = w(\varphi)$ (additivity)

W4. $w(\varphi) = w(\psi)$ if $\varphi \Leftrightarrow \psi$ is a propositional tautology (distributivity)[7]

Note that axioms W1, W2, and W3 correspond precisely to P1, P2, and P3$''$, the axioms that characterize probability measures in finite spaces. We could replace axioms I1–I6 by a single axiom that represents all instances of valid formulas about Boolean combinations of linear inequalities, analogously to what we did with the axioms Taut and W4. Axioms I1–I6, along with the axiom $w(\varphi) \geq w(\varphi)$, the axiom Taut, and the rule MP, was shown in [FHM90] to be a sound and complete axiomatization for Boolean combinations of linear inequalities (where $w(\varphi_i)$ is treated like a variable x_i). The axiom $w(\varphi) \geq w(\varphi)$ is redundant here, because of the axiom W4.

As is shown in [FHM90], AX_{MEAS} characterizes the valid formulas for measurable probability structures. Every formula that is valid with respect to measurable probability structures is provable from AX_{MEAS}, and only these formulas are provable.

Theorem 5.1: ([FHM90]) *AX_{MEAS} is a sound and complete axiomatization for weight formulas with respect to measurable probability structures.*

This result, together with Theorem 2.2, immediately gives us

[7]Just as in the case of Taut, we could make use of a complete axiomatization for propositional equivalences to create a collection of elementary axioms that could replace W4. Again, we have not done so here because we want to focus on the other axioms.

Corollary 5.2: *AX_{MEAS} is a sound and complete axiomatization for weight formulas with respect to Nilsson structures.*

Of course, AX_{MEAS} is not sound with respect to arbitrary probability structures, where $w(\varphi)$ is interpreted as the inner measure of φ^M. In particular, axiom W3 no longer holds: inner measures are not finitely additive. Let AX be obtained from AX_{MEAS} by replacing W3 by the following two axioms, which are obtained from conditions B1 and B3 for belief functions in an obvious way:

W5. $w(false) = 0$

W6. $w(\varphi_1 \vee \ldots \vee \varphi_k) \geq \sum_{I \subseteq \{1,\ldots,k\}, I \neq \emptyset} (-1)^{|I|+1} w(\bigwedge_{i \in I} \varphi_i)$

(We remark that $w(false) = 0$ no longer follows from the other axioms as it did in the system AX_{MEAS}, so we explicitly include it in AX.)

Theorem 5.3: ([FHM90]) *AX is a sound and complete axiomatization for weight formulas with respect to probability structures.*

Applying Corollary 3.2 and Theorem 3.3, we immediately get

Corollary 5.4: *AX is a sound and complete axiomatization for weight formulas with respect to DS structures.*

Thus, using AX, we can derive all consequences of a collection of beliefs. This result holds even if we allow Φ, the set of primitive propositions, to be infinite. Given a weight formula φ with primitive propositions in Φ, let Φ' consist of the primitive propositions that actually appear in φ. Clearly Φ' is finite. Since for every DS structure D there is a probability structure that is equivalent to D with respect to all formulas whose primitive propositions are contained in Φ', it follows that φ is valid with respect to DS structures iff φ is valid with respect to probability structures.

Combining the preceding results with results of [FHM90], we can also characterize the complexity of reasoning about probability and belief.

Theorem 5.5: *The complexity of deciding whether a weight formula is satisfiable with respect to probability structures (respectively, measurable probability structures, Nilsson structures, DS structures) is NP-complete.*

(This result in the case of Nilsson structures was obtained independently in [GKP88].) Note that Theorem 5.5 says that reasoning about probability and belief is, in a precise sense, exactly as difficult as propositional reasoning. This is the best we could expect, since it is easy to see that reasoning about probability and belief is at least as hard as propositional reasoning (the propositional formula φ is satisfiable iff the weight formula $w(\varphi) > 0$ is satisfiable).

The key to obtaining these results is to reduce the problem of reasoning about weight formulas to a linear programming problem, and then to apply well known techniques from linear programming. The details can all be found in [FHM90]. We can use linear programming exactly because weight formulas allow only linear combinations of terms such as $w(\varphi)$. However, this restriction prevents us from doing general reasoning about conditional probabilities.

To see why, suppose we interpret the formula $w(p_1 \,|\, p_2) \geq 1/2$ to say "the probability of p_1 given p_2 is at least 1/2". We can express this in the language described above by rewriting $w(p_1 \,|\, p_2)$ as $w(p_1 \wedge p_2)/w(p_2)$ and then clearing the denominator to get $w(p_1 \wedge p_2) - 2w(p_2) \geq 0$. However, we cannot express more complicated expressions such as $w(p_2 \,|\, p_1) + w(p_1 \,|\, p_2) \geq 1/2$ in our language, because clearing the denominator in this case leaves us with a nonlinear combination of terms. We can deal with conditional probabilities by extending the language to allow products of terms, such as $2w(p_1 \wedge p_2)w(p_2) + 2w(p_1 \wedge p_2)w(p_1) \geq w(p_1)w(p_2)$ (this is what we get when we clear the denominator in the conditional expression above). In [FHM90], the question of decision procedures and complete axiomatizations for this extended language is addressed. We briefly review the results here and discuss how they relate to reasoning about beliefs.

Although we can no longer reduce the question of the validity of a formula to a linear programming problem as we did before, it turns out we can reduce it to the validity of a quantifier-free formula in the theory of real closed fields [Sho67]. By a recent result of Canny [Can88], it follows that we can get a polynomial space decision procedure for validity (and satisfiability) with respect to all classes of structures in which we are interested. We also consider the effect of further extending our language to allow quantification over probabilities (thus allowing such formulas as $\exists y(w(\varphi) \geq y)$. We exploit the fact that the quantified theory of real closed fields has an elegant complete axiomatization [Tar51, Sho67]. If we do extend our language to allow quantification over probabilities, we can get a complete axiomatization with respect to measurable probability structures and Nilsson structures by combining axioms W1–W4 and the complete axiomatization for real closed fields. If we replace W3 by W5 and W6, we get a complete axiomatization with respect to probability structures and DS structures. Finally, by using results of [BKR86] on the complexity of the decision problem for real closed fields, we can get an exponential space decision procedure for the validity problem in all cases. See [FHM90] for further details.

To summarize this discussion, by combining the results of this paper with those of [FHM90], we are able to provide elegant complete axiomatizations for reasoning about belief and probability, as well as giving a decision procedure for the validity problem. The key point is that reasoning about belief functions is identical to reasoning about inner measures induced by probability measures.

6 Related work

Although we believe we are the first to propose using inner and outer measures as a way of dealing with uncertainty, there are a number of other works with similar themes. We briefly discuss them here.

A number of authors have argued that we should think in terms of an interval in which the probability lies, rather than a unique numerical probability (see, for example, [Kyb61, Kyb88]). Good [Goo62], Koopman [Koo40a, Koo40b], and Smith [Smi61] try to derive reasonable properties for the intuitive notions of *lower* and *upper* probability, which are somehow meant to capture lower and upper bounds on an agent's belief in a proposition. Good observes that "The analogy [between lower and upper probability and] inner and outer measure is obvious. But the axioms for upper and lower probability do not all follow from the theory of inner and outer measure."

In some papers (e.g., [Smi61, Wal81, WF82]), the phrase "lower probability" is used to denote the inf of a family of probability functions. That is, given a set $\mathcal{P}$ of probability functions defined on a σ-algebra $\mathcal{X}$, the lower probability of $\mathcal{P}$ is taken to be the function f such that for each $A \in \mathcal{X}$, we have $f(A) = \inf\{\mu(A) : \mu \in \mathcal{P}\}$. The upper probability is then taken to be the corresponding sup. We use the phrases *lower envelope* and *upper envelope* to denote these notions, in order to distinguish them from Dempster's definition of lower and upper probability [Dem67, Dem68], which we now discuss.

Dempster starts with a tuple $(S, \mathcal{X}, \mu, T, \Gamma)$, where $(S, \mathcal{X}, \mu)$ is a probability space, T is another set, and $\Gamma : S \rightarrow 2^T$ is a function which Dempster calls a "multivalued mapping from S to T" (since $\Gamma(s)$ is a subset of T for each $s \in S$). We call such a structure a *Dempster structure*. Given $A \subseteq T$, we define subsets A_* and A^* of S as follows:

$$A_* = \{s \in S \mid \Gamma(s) \neq \emptyset,\ \Gamma(s) \subseteq A\}$$
$$A^* = \{s \in S \mid \Gamma(s) \cap A \neq \emptyset\}.$$

It is easy to check that T_* and T^* both equal $\{s \in S \mid \Gamma(S) \neq \emptyset\}$, and so $T_* = T^*$. Provided $T^* \in \mathcal{X}$ and $\mu(T^*) \neq 0$, Dempster defines the lower and upper probabilities of A for all sets A such that A_* and A^* are in $\mathcal{X}$, written $P_*(A)$ and $P^*(A)$ respectively, as follows:

$$P_*(A) = \mu(A_*)/\mu(T^*)$$
$$P^*(A) = \mu(A^*)/\mu(T^*).$$

(Notice that dividing by $\mu(T^*)$ has the effect of normalizing so that $P_*(T) = P^*(T) = 1$.)

It is well known that there is a straightforward connection between Dempster's lower and upper probabilities and belief and plausibility functions. The connection is summarized in the following two propositions. The first essentially says that every lower probability is a belief function. Since it is easy to check that $P^*(A) = 1 - P_*(\overline{A})$, it follows that the upper probability is the corresponding plausibility function.

Proposition 6.1: *Let $(S, \mathcal{X}, \mu, T, \Gamma)$ be a Dempster structure such that $A_* \in \mathcal{X}$ for all $A \subseteq T$. Then the lower probability P_* is a belief function on T.*

Proof: It is easy to see that P_* satisfies B1 and B2: $P_*(\emptyset) = 0$ and $P_*(T) = 1$. To see that it satisfies B3, first observe that $(C \cap D)_* = C_* \cap D_*$ and $(C \cup D)_* \supseteq C_* \cup D_*$ for all $C, D \subseteq T$. Using these observations and the standard inclusion-exclusion rule for probabilities, we get

$$\begin{array}{rl} & P_*(A_1 \cup \ldots \cup A_m) \\ = & \mu((A_1 \cup \ldots \cup A_m)_*)/\mu(T^*) \\ \geq & \mu((A_1)_* \cup \ldots \cup (A_m)_*)/\mu(T^*) \\ = & \sum_{I \subseteq \{1,\ldots,m\},\, I \neq \emptyset} (-1)^{|I|+1} \mu(\bigcap_{i \in I} (A_i)_*)/\mu(T^*) \\ = & \sum_{I \subseteq \{1,\ldots,m\},\, I \neq \emptyset} (-1)^{|I|+1} \mu((\bigcap_{i \in I} A_i)_*)/\mu(T^*) \\ = & \sum_{I \subseteq \{1,\ldots,m\},\, I \neq \emptyset} (-1)^{|I|+1} P_*(\bigcap_{i \in I} A_i) \end{array}$$

Thus P_* satisfies B3, and so is a belief function. ∎

We can get a more direct proof that P_* is a belief function if T is finite. In fact, if we define the function m on T by taking $m(A) = \mu(\{s : \Gamma(s) = A\})/\mu(T^*)$ for $A \neq \emptyset$, then m is easily seen to be a mass function, and P_* is the belief function corresponding to m. (We remark that if we assume that the set A_* is measurable for every set A, then by induction on the cardinality of A, it can be shown that the set $\{s : \Gamma(s) = A\}$ is also measurable.) However, our original proof of Theorem 6.1, besides holding even when T is not finite, has an additional advantage. Suppose we extend the definition of P_* to sets A such that $A_* \notin \mathcal{X}$ by taking $P_*(A) = \mu_*(A_*)$. Then a slight modification of our proof (using ideas in the proof of Proposition 3.1) shows that this extension still makes P_* a belief function. This observation was first made by Kampé de Fériet [dF82].

The converse to Proposition 6.1 essentially holds as well, and seems to be somewhat of a folk theorem in the community. A proof can be found (using quite different notions) in [Ngu78]. We also provide a proof here, since the result is so straightforward.

Proposition 6.2: *Let Bel be a belief function on a finite space T. There exists a Dempster structure $(S, \mathcal{X}, \mu, T, \Gamma)$ with lower probability P_* such that $Bel = P_*$.*

Proof: Let m be the mass function corresponding to Bel, and let $(2^T, 2^{2^T}, \mu, T, \Gamma)$ be the Dempster structure where we define $\mu(\{A\}) = m(A)$ for $A \subseteq T$, and then extend μ by additivity to arbitrary subsets of 2^T, and define $\Gamma(\{A\}) = A$ and $\Gamma(B) = \emptyset$ if B is not a singleton subset of 2^T. Then easy to see that $A_* = \{\{B\} : B \subseteq A\}$, from which it follows that

$$P_*(A) = \mu(A_*) = \sum_{B \subseteq A} m(B) = Bel(B). \quad ∎$$

As might be expected, there is also a close connection between Dempster's lower probability and the notion of a lower envelope. In fact, it is well known every lower probability (and hence, every belief function) is a lower envelope. We briefly sketch why here. (The observation is essentially due to [Dem67]; see [Rus87] or [FH91] for more details.) Given a belief function Bel on a finite set S with corresponding plausibility function Pl, we say that a probability function μ defined on 2^S is *consistent with Bel* if $Bel(A) \leq \mu(A) \leq Pl(A)$ for all $A \subseteq S$.[8] Let $\mathcal{P}_{Bel}$ consist of all probability functions consistent with Bel. It is then not hard to show that Bel is the lower envelope of $\mathcal{P}_{Bel}$.

Theorem 6.3: *If Bel is a belief function on S, then for all $A \subseteq S$, we have*

$$Bel(A) = \inf_{\mu \in \mathcal{P}_{Bel}} \mu(A).$$

Although this result shows that every belief function is a lower envelope, the converse does not hold. Again, this remark can already be found in [Dem67]; a counterexample with further discussion appears in [Kyb87]. Intuitively, the fact that every belief function is a lower envelope means that we can view the belief function to be the result of collecting a group of experts, and taking the belief in an event E to be the minimum of the probabilities that the experts assign to E. However, since not every lower envelope is a belief function, we cannot characterize belief functions in this way. Further discussion on the relationship between lower envelopes and belief functions can also be found in [FH91, HF92].

We next turn our attention to the connection between Dempster's lower probabilities and inner measures. Since, as we have shown, lower probabilities are equivalent to belief functions, which in turn are essentially equivalent to inner measures, we know that they are closely related. In fact, the relationship is quite direct. Let us reconsider Example 2.3, where we estimate the probability of the projectile landing in water. In this example, we constructed a probability structure $M = (S, \mathcal{X}, \mu, \pi)$. Consider the Dempster structure (S', μ', T, Γ) defined as follows. S' consists of the sets R_i' that form a basis for $\mathcal{X}$. We define $\mu'(\{R_i'\}) = \mu(R_i)$, and extend to arbitrary subsets of S' by additivity. We take T to consist of all the propositional formulas in the language (where the only primitive propositions are *land* and *water*). Finally, $\Gamma(R')$ consists of all the propositional formulas that are true at some point in the set R'. It is easy to check that $\{water\}_*$ consists of all R' such that R is completely contained in water, while $\{water\}^*$ consists of all R' such that R has some water in it. It immediately follows that $P_*(\{water\}) = \mu_*(water^M)$ and $P^*(\{water\}) = \mu^*(water^M)$.

This close relationship between lower and upper probabilities and inner and outer measures induced by a probability measure holds in general. Given a probability structure $M = (S, \mathcal{X}, \mu, \pi)$ where S is finite, let $(\mathcal{X}', \mu', T, \Gamma)$ be the Dempster structure where (1) $\mathcal{X}'$ is a basis for $\mathcal{X}$, (2) μ' is a probability measure defined on $2^{\mathcal{X}'}$ by taking $\mu'(\{A\}) = \mu(A)$ for $A \in \mathcal{X}'$ and then extending to all subsets of $\mathcal{X}'$ by finite additivity, (3) T consists of all propositional formulas, and (4) for $A \in \mathcal{X}'$, we define $\Gamma(A)$ to consist of all formulas φ such that φ is true at some point in A (in the structure M). Thus Γ is a multivalued mapping from $\mathcal{X}'$ to T. It is easy to check that $P_*(\{\varphi\}) = \mu_*(\varphi^M)$ and $P^*(\{\varphi\}) = \mu^*(\varphi^M)$ for all formulas φ. This shows that every inner measure is a lower probability, and thus

[8] We remark that it suffices to require $Bel(A) \leq \mu(A)$ for all $A \subseteq S$. It then follows that $Pl(A) = 1 - Bel(\overline{A}) \geq 1 - \mu(\overline{A}) = \mu(A)$.

corresponds to Proposition 3.1 (or, more accurately, the proof of Proposition 3.1 in the case of finite sample spaces given after the proof of Proposition 3.5). It also follows from our results that every lower probability is equivalent to an inner measure (when viewed as a function on formulas, rather than sets); the proof is an analogue to that of Theorem 3.3.

Ruspini [Rus87] also considers giving semantics to probability formulas by using possible worlds, but he includes epistemic notions in the picture. Briefly, his approach can be described as follows (where have taken the liberty of converting some of his notation to ours, to make the ideas easier to compare). Fix a set $\{p_1, \ldots, p_n\}$ of primitive propositions. Instead of considering just propositional formulas, Ruspini allows epistemic formulas; he obtains his language by closing off under the propositional connectives $\wedge$, $\vee$, $\Rightarrow$, and $\neg$, as well as the epistemic operator K. Thus, a typical formula in his language would be $K(p_1 \Rightarrow K(p_2 \wedge p_3))$. (A formula such as $K\varphi$ should be read "the agent knows φ.") Rather than considering arbitrary sample spaces as we have done here, where at each point in the sample space some subset of primitive propositions is true, Ruspini considers one fixed sample space S (which he calls a *sentence space*) whose points consist of all the possible truth assignments to these formulas consistent with the axioms of the modal logic S5. (See, for example, [HM92] for an introduction to S5. We remark that it can be shown that there are less than $2^n 2^{2^n}$ consistent truth assignments, so that S is finite.) We can define an equivalence relation $\sim$ on S by taking $s \sim t$ if s and t agree on the truth values of all formulas of the form $K\varphi$. The equivalence classes form a basis for a σ-algebra of measurable subsets of S. Let $\mathcal{X}$ be this σ-algebra. For any formula φ, let φ^S consist of all the truth assignments in S that make φ true. It is easy to check that $(K\varphi)^S$, the set of truth assignments that make $K\varphi$ true, is the union of equivalence classes, and hence is measurable. Let μ be any probability measure defined on $\mathcal{X}$. Given μ, we can consider the probability structure $(S, \mathcal{X}, \mu, \pi)$, where we take $\pi(s)(p) = s(p)$. (Since s is a truth assignment, this is well defined.) The axioms of S5 guarantee us that $(K\varphi)^S$ is the largest measurable subset contained in φ^M; thus, $\mu_*(\varphi^M) = \mu((K\varphi)^S)$.

Ruspini then considers the DS structure (At, Bel, π'), where π' is defined in the obvious way on the atoms in At, and $Bel(\varphi^D) = \mu((K\varphi)^S)(= \mu_*(\varphi^M))$.[9] Ruspini shows that Bel defined in this way is indeed a belief function. (Since $Bel(\varphi^D) = \mu_*(\varphi^M)$, the result follows using exactly the same techniques as those in the proof of Proposition 3.1.) Thus, Ruspini shows a close connection between probabilities, inner measures, and belief functions in the particular structures that he considers. He does not show a general relationship between inner measures and belief functions; in particular, he does not show that DS structures are equivalent to probability structures, as we do in Theorem 3.3. Nevertheless, Ruspini's viewpoint is very similar in spirit to ours. He states and proves Theorem 2.1, and stresses the idea that the inner and outer measure (and hence the belief and plausibility function) can be viewed as the best approximation to the "true" probability, given our lack of information. Ruspini also considers ways of combining

[9]Ruspini actually defines the belief function directly on formulas; i.e., he defines $Bel(\varphi)$. In our notation, what he is doing is defining a weight function W_D.

two probability structures defined on sentence spaces and shows that he can capture Dempster's rule of combination in this way. Although his results are not the same as Theorem 4.1, again, they are similar in spirit.

We have characterized belief functions as being essentially inner measures induced by probability measures. Another characterization of belief functions in terms of probability theory is discussed by Shafer in [Sha79]. He shows that it follows directly from the integral representation of Choquet [Cho53] that under natural assumptions, every belief function is of the form $\mu \circ r$, where μ is a probability measure and r is an $\cap$-homomorphism (that is, r maps the empty set onto the empty set and the whole space onto the whole space, and $r(A \cap B) = r(A) \cap r(B)$). Moreover, every function of the form $\mu \circ r$, is a belief function. Thus, belief functions can be characterized as the result of composing probability measures and $\cap$-homomorphisms.

Finally, Pearl [Pea88] informally characterizes belief functions in terms of "probability of provability". Although the details are not completely spelled out, it appears that this characterization is equivalent to that of Proposition 3.5, which shows that a belief function can be characterized in terms of a mass function; the mass of a formula can be associated with its probability of provability. We can slightly reformulate Pearl's ideas as follows: we are given a collection of theories (sets of formulas) $T_1, \ldots, T_n$, each with a probability, such that the probabilities sum to 1. The belief in a formula φ is the sum of the probabilities of the theories from which φ follows as a logical consequence. Note that both a formula and its negation might have belief 0, since neither might follow from any of the theories. This approach can be shown to be closely related to that of Ruspini [Rus87], and, just as Ruspini's, can be put into our framework as well.

7 Conclusions

We have introduced a new way of dealing with uncertainty, where nonmeasurability of certain events turns out to be a crucial feature, rather than a mathematical nuisance. This approach seems to correspond to our intuitions in a natural way in many examples, and gets around some of the objections to the Bayesian approach, while still retaining many of the attractive features of using probability theory. Surprisingly, our approach helps point out a tight connection between the Dempster-Shafer approach and classical probability theory. In particular, we are able to characterize belief functions as being essentially inner measures induced by probability measures. We hope that this characterization will give added insight into belief functions, and lead to better tools for reasoning about uncertainty. It has already enabled us to provide a complete axiomatization and decision procedure for reasoning about belief functions. More recently, it has led us to define new approaches to updating belief functions, different from those defined using Dempster's rule of combination (see [Sha76] for details). The idea is to first consider what it means to take a conditional probability with respect to a nonmeasurable set, by defining notions of inner and outer conditional probabilities and then proving a result analogous to Theorem 2.1.

Given the tight connection between inner measures and belief functions described here, this quickly leads us to notions of conditional belief and conditional plausibility. Our definitions seem to avoid many of the problems that arise when using the more standard definition (see, for example, [Ait68, Bla87, Dia78, DZ86, Hun87, Lem86, Pea89]). These results are reported in [FH91].

While our approach seems natural, and works well in a number of examples we have considered, we do not feel it is necessarily the right approach to take in all cases. More experience is required with real-world examples in order to understand when it is appropriate. We feel that our ideas and approach will also lead to a deeper understanding of when belief functions can be used. We report some preliminary results along these lines in [HF92].

Acknowledgments: The authors are grateful to Glenn Shafer, Moshe Vardi, and Henry Kyburg for helpful comments and suggestions, and to an anonymous referee for pointing out that Theorem 3.1 followed from a result in [Sha79].

References

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

[BKR86] M. Ben-Or, D. Kozen, and J. H. Reif. The complexity of elementary algebra and geometry. *Journal of Computer and System Sciences*, 32(1):251–264, 1986.

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

[Can88] J. F. Canny. Some algebraic and geometric computations in PSPACE. In *Proc. 20th ACM Symp. on Theory of Computing*, pages 460–467, 1988.

[Car50] R. Carnap. *Logical Foundations of Probability.* University of Chicago Press, Chicago, 1950.

[Che85] P. Cheeseman. In defense of probability. In *Proc. Ninth International Joint Conference on Artificial Intelligence (IJCAI '85)*, pages 1002–1009, 1985.

[Cho53] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier (Grenoble)*, 5:131–295, 1953.

[Coh85] P. R. Cohen. *Heuristic Reasoning about Uncertainty: An Artificial Intelligence Approach.* Pitman, 1985.

[Cox46] R. Cox. Probability, frequency, and reasonable expectation. *American Journal of Physics*, 14(1):1–13, 1946.

[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.

[dF82] J. Kampé de Fériet. Interpretations of membership functions of fuzzy sets in terms of plausibility and belief. In M. M. Gupta and E. Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 93–98. North-Holland, Amsterdam, 1982.

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

[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.

[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. A new approach to updating beliefs. In P.P. Bonissone, M. Henrion, L.N. Kanal, and J. Lemmer, editors, *Uncertainty in Artificial Intelligence: Volume VI*, pages 347–374. Elsevier Science Publishers, Amsterdam, 1991.

[FHM90] R. Fagin, J. Y. Halpern, and N. Megiddo. A logic for reasoning about probabilities. *Information and Computation*, 87(1/2):78–128, 1990.

[GKP88] G. Georgakopoulos, D. Kavvadias, and C. H. Papadimitriou. Probabilistic satisfiability. *Journal of Complexity*, 4(1):1–11, 1988.

[Goo62] I. J. Good. The measure of a non-measurable set. In E. Nagel, P. Suppes, and A. Tarski, editors, *Logic, Methodology, and Philosophy of Science*, pages 319–329. Stanford University Press, Stanford, Calif., 1962.

[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.

[HM92] J. Y. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992.

[HR87] J. Y. Halpern and M. O. Rabin. A logic to reason about likelihood. *Artificial Intelligence*, 32(3):379–405, 1987.

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

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

[Koo40a] B. O. Koopman. The axioms and algebra of intuitive probability. *Annals of Mathematics*, 41:269–292, 1940.

[Koo40b] B. O. Koopman. The bases of probability. *Bulletin of the American Mathematical Society*, 46:763–774, 1940.

[Kyb61] H. E. Kyburg, Jr. *Probability and the Logic of Rational Belief.* Wesleyan University Press, Middletown, Conn., 1961.

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

[Kyb88] H. E. Kyburg, Jr. Higher order probabilities and intervals. *International Journal of Approximate Reasoning*, 2:195–209, 1988.

[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.

[LR57] R. D. Luce and H. Raiffa. *Games and Decisions.* Wiley, New York, 1957.

[Men64] E. Mendelson. *Introduction to Mathematical Logic.* Van Nostrand, New York, 1964.

[MP81] Z. Manna and A. Pnueli. Verification of temporal programs: the temporal framework. In R. S. Boyer and J. S. Moore, editors, *The Correctness Problem in Computer Science.* Academic Press, New York, 1981.

[Ngu78] H. Nguyen. On random sets and belief functions. *Journal of Mathematical Analysis and Applications*, 65:531–542, 1978.

[Nil86] N. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28:71–87, 1986.

[NM47] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, Princeton, N.J., 2nd edition, 1947.

[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.

[Ram31] F. P. Ramsey. Truth and probability. In R. B. Braithwaite, editor, *The Foundations of Probability and other Logical Essays*. Harcourt Brace, New York, 1931.

[RK86] S. J. Rosenschein and L. P. Kaelbling. The synthesis of digital machines with provable epistemic properties. In J. Y. Halpern, editor, *Theoretical Aspects of Reasoning about Knowledge: Proc. 1986 Conference*, pages 83–97. Morgan Kaufmann, San Francisco, Calif., 1986.

[Roy64] H. L. Royden. *Real Analysis*. Macmillan, 1964.

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

[Saf88] A. Saffiotti. An AI view of the treatment of uncertainty. *The Knowledge Engineering Review*, 2(2), 1988.

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

[Sha] G. Shafer. Dempster's rule of combination. Manuscript, circa 1980.

[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.

[Sha86] G. Shafer. The combination of evidence. *International Journal of Intelligent Systems*, 1:155–179, 1986.

[Sho67] J. R. Shoenfield. *Mathematical Logic*. Addison-Wesley, Reading, Mass., 1967.

[Smi61] C. A. B. Smith. Consistency in statistical inference and decision. *Journal of the Royal Statistical Society, Series B*, 23:1–25, 1961.

[Tar51] A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. Univ. of California Press, 2nd edition, 1951.

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

[WF82] P. Walley and T. L. Fine. Towards a frequentist theory of upper and lower probability. *Annals of Statistics*, 10(3):741–761, 1982.

[Zad75] L. A. Zadeh. Fuzzy logics and approximate reasoning. *Synthese*, 30:407–428, 1975.