Probabilistic Reasoning and Bayesian Belief Networks

Edited by A Gammerman
Royal Holloway, University of London

ALFRED WALLER LIMITED
Henley-on-Thames

in association with UNICOM
Chapter 4

Modelling ignorance in uncertainty theories

Prakash P. Shenoy

4.1 Introduction

How do different uncertainty theories represent ignorance? This paper gives an answer to this question using the axiomatic framework of valuation-based systems (VBS).

The VBS framework was initially proposed by Shenoy [14]. It is able to represent many different uncertainty calculi such as probability theory [18], Dempster-Shafer (D-S) belief-function theory [23], Spohn’s epistemic-belief theory [16, 18], and Zadeh’s possibility theory [20]. The VBS framework is also flexible enough to include propositional logic [15, 24], discrete optimization [17], Bayesian decision analysis [19, 21, 25], and constraint satisfaction [26].
It is commonly believed that the D-S belief-function theory is better able to represent ignorance than, for example, probability theory. We disagree. Ignorance can be represented in probability theory as well as it can be represented in D-S belief-function theory. All uncertainty theories that fit in the VBS framework can represent ignorance equally well.

Complete ignorance is represented in valuation-based systems by identity valuations. In probability theory, complete ignorance is modelled by an equally likely probability distribution. In D-S belief-function theory, complete ignorance is modelled, for example, by a commonality function that is identically one. In Spohn’s epistemic-belief theory, complete ignorance is modelled by a disbelief function that is identically zero. And in Zadeh’s possibility theory, complete ignorance is modelled by a possibility function that is identically one.

In probability theory (and in Spohn’s epistemic-belief theory and Zadeh’s possibility theory), complete ignorance coincides with knowledge that all elements of a frame are equally likely (or equally believed in Spohn’s theory, or equally possible in Zadeh’s theory). In D-S theory, there are many ways to model knowledge that all elements of a frame are equally likely. For example, consider a variable X whose frame has two configurations, say x and ¬x. Then a basic probability assignment (bpa) function m such that m(⟦x⟧) = m(⟦¬x⟧) = p, m(⟦x, ¬x⟧) = 1 − 2p, where 0 ≤ p ≤ 0.5, represents knowledge that all elements of the frame of X are equally likely. However, only one of these bpa functions (with p = 0) represents complete ignorance. This expressiveness of belief-function theory, however, should not be misinterpreted as inability of probability theory (or Spohn’s theory or Zadeh’s theory) to represent complete ignorance.

On a frame with n elements, a probability distribution function (or a Spohnian disbelief function or a possibility function) has n − 1 independent parameters. On the other hand, a D-S belief function (or bpa function or plausibility function or commonality function) has 2^n − 1 independent parameters. Consequently, given a frame of fixed size, a D-S belief function is more expressive than a probability function. Of course, this expressiveness comes at a computational cost. Also,

**Modelling ignorance in uncertainty theories**

Probability theory can achieve the same expressiveness as D-S theory by simply increasing the size of the frame.

Besides complete ignorance, we also define the notion of contextual ignorance. Contextual ignorance for σ is knowledge δ_σ that does not add anything new to the state of knowledge σ, i.e. σ ∩ δ_σ = σ. Complete ignorance is a special case of contextual ignorance for any context σ. These ideas are further explored in this paper.

An outline of the remainder of the paper is as follows. In Section 4.2, we sketch the VBS framework in the abstract. In Section 4.3, we describe four specific instances of VBS, namely probability theory, D-S belief-function theory, Spohn’s epistemic-belief theory, and Zadeh’s possibility theory. In Section 4.4, we discuss how ignorance is modelled in VBS in general and in different uncertainty theories in particular. We also introduce the notion of contextual ignorance. In Section 4.5, we summarize our findings.

### 4.2 The VBS framework

In this section, we describe the axiomatic framework of valuation-based systems (VBSs). Most of the material in this section is taken from [22].

In the VBS framework, we represent knowledge by entities called variables and valuations. We infer conditional independence relations using three operations called combination, marginalization, and removal. We use these operations on valuations.

**Variables.** We assume there is a finite set S whose elements are called variables. Variables are denoted by upper-case Latin alphabets, X, Y, Z, etc. Subsets of S are denoted by lower-case Latin alphabets, r, s, t, etc.

**Valuations.** For each s ⊆ S, there is a set V_s. We call the elements of V_s valuations for s. Let U denote ∪{V_s | s ⊆ S}, the set of all valuations. If σ ∈ U, then we say s is the domain of σ. Valuations are denoted by lower-case Greek alphabets, ρ, σ, τ, etc.
Modelling ignorance in uncertainty theories

Axiom C1 (Domain): If \( \rho \in \mathcal{U}_r \) and \( \sigma \in \mathcal{U}_s \), then \( \rho \oplus \sigma \in \mathcal{U}_{r \cup s} \).

Axiom C2 (Associative): \( \rho \oplus (\sigma \oplus \tau) = (\rho \oplus \sigma) \oplus \tau \).

Axiom C3 (Commutative): \( \rho \oplus \sigma = \sigma \oplus \rho \); and

Axiom C4 (Zero): Suppose zero valuations exist, and suppose \( \sigma \in \mathcal{U}_s \). Then \( \zeta_s \oplus \sigma = \zeta_s \).

If \( \rho \oplus \sigma \), read as \( \rho \) plus \( \sigma \), is a zero valuation, then we say that \( \rho \) and \( \sigma \) are inconsistent. If \( \rho \oplus \sigma \) is a normal valuation, then we say that \( \rho \) and \( \sigma \) are consistent.

Intuitively, combination corresponds to aggregation of knowledge. If \( \rho \) and \( \sigma \) are valuations for \( r \) and \( s \) representing knowledge about variables in \( r \) and \( s \), respectively, then \( \rho \oplus \sigma \) represents the aggregated knowledge about variables in \( r \cup s \).

An implication of Axiom C2 is that when we have multiple combinations of valuations, we can write it without using parenthesis. For example, \((\sigma_1 \oplus \sigma_2) \oplus \sigma_3 \oplus \ldots \oplus \sigma_m\) can be written simply as \(\sigma_1 \oplus \ldots \oplus \sigma_m\) without parenthesis. Further, by Axiom C3, we can write \(\sigma_1 \oplus \ldots \oplus \sigma_m\) simply as \(\sigma_1 \oplus \ldots \oplus \sigma_m\), i.e., not only do we not need parenthesis, we need not indicate the order in which the valuations are combined.

An implication of Axioms C1, C2, and C3 is that the set \( \mathcal{N}_s \cup \{\zeta_s\} \) together with the combination operation \( \oplus \) is a commutative semigroup [7]. (If zero valuations do not exist, then \( \mathcal{N}_s \cup \{\zeta_s\} = \mathcal{N}_s \). If zero valuations exist, then Axiom C4 defines the valuation \( \zeta_s \) as the zero of the semigroup \( \mathcal{N}_s \cup \{\zeta_s\} \).

Identity Valuations. We assume the following identity axiom.

Axiom C5 (Identity): For each \( s \subseteq \mathcal{X} \), the commutative semigroup \( \mathcal{N}_s \cup \{\zeta_s\} \) has an identity denoted by \( t_s \).

In other words, Axiom C5 assumes there exists \( t_s \in \mathcal{N}_s \cup \{\zeta_s\} \) such that for each \( \sigma \in \mathcal{N}_s \cup \{\zeta_s\} \), \( \sigma \oplus t_s = \sigma \). Notice that a commutative semigroup may have at most
one identity. From Axiom C4, it follows that \( \delta_s \notin \zeta_s \), therefore \( \delta_s \in \mathcal{N}_s \).

Intuitively, identity valuations represent knowledge that is completely vacuous, i.e., they have no substantive content.

It follows from Axiom C5 that for each \( s \subseteq \mathcal{X} \), and for each \( \sigma \in \mathcal{N}_s \cup \{ \zeta_s \} \), there exists at least one identity for it, i.e., there exists a \( \delta_\sigma \in \mathcal{N}_s \cup \{ \zeta_s \} \) such that \( \sigma \oplus \delta_\sigma = \sigma \). For example, \( \delta_s \) is an identity for each element of \( \mathcal{N}_s \cup \{ \zeta_s \} \). A valuation may have more than one identity. For example, Axiom C4 states that every element of \( \mathcal{N}_s \cup \{ \zeta_s \} \) is an identity for \( \zeta_s \). Notice that if \( \sigma \in \mathcal{N}_s \), then \( \delta_\sigma \in \mathcal{N}_s \). Also, notice that \( \delta_s \) has only one identity, namely itself.

**Positive Normal Valuations.** Let \( \mathcal{U}_s \) denote the subset of \( \mathcal{N}_s \) consisting of all valuations in \( \mathcal{N}_s \) that have unique identities. We call elements of \( \mathcal{U}_s \) positive normal valuations for \( s \). Let \( \mathcal{U} \) denote \( \cup \{ \mathcal{U}_s \mid s \subseteq \mathcal{X} \} \), the set of all positive normal valuations. The concept of positive normal valuations is important because the intersection property of conditional independence holds only for positive normal valuations. Positive normal valuations correspond to strictly positive probability distributions in probability theory. Figure 4.1 shows the relation between different types of valuations.

**Figure 4.1.** The relation between different types of valuations.

---

**Modelling ignorance in uncertainty theories**

We assume the following axiom regarding the normal valuation for the empty set.

**Axiom C6 (Normal Valuations for the Empty Set):** The set \( \mathcal{N}_\emptyset \) consists of exactly one element.

Axiom C6 implies that \( \mathcal{U}_\emptyset = \mathcal{N}_\emptyset = \{ \delta_\emptyset \} \) where \( \delta_\emptyset \) is the identity valuation for the semigroup \( \mathcal{N}_\emptyset \cup \{ \zeta_\emptyset \} \).

**Marginalization.** We assume that for each nonempty \( s \subseteq \mathcal{X} \), and for each \( X \in s \), there is a mapping \( \downarrow(s-\{X\}): \mathcal{U}_s \to \mathcal{U}_{s-\{X\}} \), called marginalization to \( s-\{X\} \), that satisfies the following six axioms:

**Axiom M1 (Order of Deletion):** Suppose \( \sigma \in \mathcal{U}_s \), and suppose \( X_1, X_2 \in s \). Then
\[
(\sigma \downarrow(s-\{X_1\}))(s-\{X_1, X_2\}) = (\sigma \downarrow(s-\{X_1\}))(s-\{X_1\});
\]

**Axiom M2 (Zero):** If zero valuations exist, then \( \zeta_s \downarrow(s-\{X\}) = \zeta_{s-\{X\}} \).

**Axiom M3 (Normal):** \( \sigma \downarrow(s-\{X\}) \in \mathcal{N}_s \) if and only if \( \sigma \in \mathcal{N}_s \).

**Axiom M4 (Positive Normal):** If \( \sigma \in \mathcal{U}_s \), then \( \sigma \downarrow(s-\{X\}) \in \mathcal{U}_s \).

**Axiom CM1 (Combination and Marginalization 1):** Suppose \( \rho \in \mathcal{U}_r \) and \( \sigma \in \mathcal{U}_s \). Suppose \( X \in r \), and \( X \in s \). Then
\[
(\rho \oplus \sigma) \downarrow((r \cup s)-\{X\}) = (\rho \oplus \sigma) \downarrow((r-\{X\})) \quad \text{and}
\]

**Axiom CM2 (Combination and Marginalization 2):** Suppose \( \sigma \in \mathcal{N}_s \), suppose \( r \subseteq s \), and suppose \( \delta_{\sigma \downarrow r} \) is an identity for \( \sigma \downarrow r \) in \( \mathcal{N}_r \). Then \( \delta_{\sigma \downarrow r} \) is an identity for \( \sigma \), i.e., \( \sigma \oplus \delta_{\sigma \downarrow r} = \sigma \).

We call \( \sigma \downarrow(s-\{X\}) \) the marginal of \( \sigma \) for \( s-\{X\} \).

Intuitively, marginalization corresponds to coarsening of knowledge. If \( \sigma \) is a valuation for \( s \) representing some knowledge about variables in \( s \), and \( X \in s \), then
Modelling ignorance in uncertainty theories

We call \( \sigma \ominus \rho \), read as \( \sigma \) minus \( \rho \), the valuation resulting after removing \( \rho \) from \( \sigma \). Intuitively, \( \sigma \ominus \rho \) can be interpreted as follows. If \( \sigma \) and \( \rho \) represent some knowledge, and if we remove the knowledge represented by \( \rho \) from \( \sigma \), then \( \sigma \ominus \rho \) describes the knowledge that remains.

Axioms CR1 and CR2 define the removal operation as an "inverse" of the combination operation in the sense that arithmetic division is inverse of arithmetic multiplication, and in the sense that arithmetic subtraction is inverse of arithmetic addition.

**Conditionals.** Suppose \( \sigma \in \mathcal{N}_s \), and suppose \( a \) and \( b \) are disjoint subsets of \( s \). The valuation \( \sigma^{(a \cup b)} \otimes \sigma^{(a)} \), for \( a \cup b \), plays an important role in uncertainty theories. Borrowing terminology from probability theory, we call \( \sigma^{(a \cup b)} \otimes \sigma^{(a)} \) the conditional for \( b \) given \( a \) with respect to \( \sigma \). Let \( \sigma(b|a) \) denote \( \sigma^{(a \cup b)} \otimes \sigma^{(a)} \).

We call \( b \) the head of the domain of \( \sigma(b|a) \), and we call \( a \) the tail of the domain of \( \sigma(b|a) \).

**Theorem 4.1** [22]. Suppose \( \sigma \in \mathcal{N}_s \), and suppose \( a \), \( b \), and \( c \) are disjoint subsets of \( s \).

(i). \( \sigma(a) = \sigma^{(a)} \).

(ii). \( \sigma(a) \otimes \sigma(b|a) = \sigma(a \cup b) \).

(iii). \( \sigma(b|a) \otimes \sigma(c|b) = \sigma(b \cup c) \).

(iv). Suppose \( b \subseteq b \). Then \( \sigma(b|a) \otimes \sigma(c|b) = \sigma(b|a) \).

(v). \( \sigma(b|a) \otimes \sigma(c|b) = \sigma(c|a) \).

(vi). \( \sigma(b|a) = \sigma(a) \).

(vii). \( \sigma(b|a) \in \mathcal{N}_{a \cup b} \).

4.3 Instances of VBS

In this section, we describe four specific instances of valuation-based systems, namely probability theory, D-S belief-function theory, Spohn's epistemic-belief theory, and Zadeh's possibility theory. First we start with the notation.
Frames and Configurations. We use the symbol $\mathcal{W}_X$ for the set of possible values of a variable $X$, and we call $\mathcal{W}_X$ the frame for $X$. We assume that one and only one of the elements of $\mathcal{W}_X$ is the true value of $X$. We assume that all the variables in $\mathcal{X}$ have finite frames.

Given a nonempty set $s$ of variables, let $\mathcal{W}_s$ denote the Cartesian product of $\mathcal{W}_X$ for $X$ in $s$; $\mathcal{W}_s = \times \{ \mathcal{W}_X \mid X \in s \}$. We call $\mathcal{W}_s$ the frame for $s$. We call the elements of $\mathcal{W}_s$ configurations of $s$. We use this terminology even when $s$ is a singleton subset. Thus elements of $\mathcal{W}_s$ are configurations of $X$. We use lower-case, bold-faced letters such as $x$, $y$, etc., to denote configurations.

It is convenient to extend this terminology to the case where the set of variables $s$ is empty. We adopt the convention that the frame for the empty set $\emptyset$ consists of a single configuration, and we use the symbol $\circ$ to name that configuration; $\mathcal{W}_\emptyset = \{ \circ \}$.

Projection of Configurations. Projection simply means dropping extra coordinates; for example, if $(w, x, y, z)$ is a configuration of $\{W, X, Y, Z\}$, then the projection of $(w, x, y, z)$ to $\{W, Y\}$ is simply $(w, y)$, which is a configuration of $\{W, Y\}$. If $r$ and $s$ are sets of variables, $r \subseteq s$, and $x$ is a configuration of $s$, then $x_{1:r}$ denotes the projection of $x$ to $r$. If $r = \emptyset$, of course, $x_{1:r} = \circ$.

If $x$ is a configuration of $r$, $y$ is a configuration of $s$, and $r \cap s = \emptyset$, then there is a unique configuration $z$ of $r \cup s$ such that $x_{1:r} = x$, and $z_{1:s} = y$. Let $(x, y)$ or $(y, x)$ denote $z$. As per this notation, $(x, \circ) = (\circ, x) = x$.

Let $2^\mathcal{W}$ denote the set of all nonempty subsets of $\mathcal{W}_s$. Elements of $2^\mathcal{W}$ will be denoted by $a$, $b$, $c$, etc. Let $\mathbb{R}^+$ denote the set of all non-negative real numbers.

4.3.1 Probability theory

In this subsection, we show how probability theory fits in the VBS framework. More precisely, we define valuations, zero valuations, proper valuations, normal valuations, combinaation, marginalization, and removal.

Modelling ignorance in uncertainty theories

In probability theory, the basic representational unit is called a probability function.

Probability Function. A probability function $\sigma$ for $s$ is a function $\sigma: 2^\mathcal{W} \rightarrow \mathbb{R}^+$ such that:

(P1). $\Sigma(\{\sigma(x) \mid x \in \mathcal{W}_s\}) = 1$; and

(P2). $\sigma(a) = \Sigma(\{\sigma(x) \mid x \in a\})$ for all $a \subseteq 2^\mathcal{W}$.

Notice that although a probability function is defined for the set of all nonempty subsets of $\mathcal{W}_s$, it is clear from condition (P2) that it is completely specified by its values for all singleton subsets of $\mathcal{W}_s$.

In probability theory, a valuation for $s$ is a function $\sigma: \mathcal{W}_s \rightarrow \mathbb{R}^+$. Zero valuations exist—a valuation $\zeta_s$ for $s$ is zero if and only if all values of $\zeta_s$ are zeros, i.e., $\zeta_s(x) = 0$ for all $x \in \mathcal{W}_s$. Suppose $\sigma$ is a valuation for $s$. We call $\sigma$ proper if and only if $\sigma \neq \zeta_s$, i.e., all nonzero valuations are proper. Suppose $\sigma$ is a valuation for $s$. We call $\sigma$ normal if and only if $\Sigma(\{\sigma(x) \mid x \in \mathcal{W}_s\}) = 1$. A normal valuation can be regarded as a probability function defined only for singleton subsets.

Combination. In probability theory, combination is pointwise multiplication followed by normalization (if normalization is possible). Suppose $\rho \in \mathcal{V}_s$, and $\sigma \in \mathcal{V}_s$. Let $K = \Sigma(\rho(x)\sigma(x) \mid x \in \mathcal{W}_s \cap \mathcal{G})$. The combination of $\rho$ and $\sigma$, denoted by $\rho \Theta \sigma$, is the valuation for $r \cap s$ given by

$$
(\rho \Theta \sigma)(x) = \begin{cases} 
K^{-1}\rho(x)\sigma(x) & \text{if } K > 0 \\
0 & \text{if } K = 0 
\end{cases}
$$

for all $x \in \mathcal{W}_s \cap \mathcal{G}$. If $K = 0$, $\rho \Theta \sigma = \zeta_{s \cap \mathcal{G}}$. If $K > 0$, then $K$ is a normalization constant that ensures $\rho \Theta \sigma$ is a normal valuation.

It is easy to see that Axioms C1–C6 are satisfied by the definition of combination in (4.1). The identity $I_s$ for $\mathcal{G}_s \cup \{\zeta_s\}$ is given by $I_s(x) = 1/|\mathcal{W}_s|$ for all $x \in \mathcal{W}_s$. Suppose $\sigma \in \mathcal{G}_s$. An identity $\delta_\sigma$ for $\sigma$ in $\mathcal{G}_s$ is a normal valuation for $s$ such that $\delta_\sigma(x) = K^{-1}$ if $\sigma(x) > 0$, and $\delta_\sigma(x) = K^{-1}r$ if $\sigma(x) = 0$, where $r$ is any
non-negative real number, and $K$ is the normalization constant. Suppose $\sigma \in \mathcal{N}_s$. Notice that $\sigma$ is positive normal if and only if $\sigma(x) > 0$ for all $x \in \mathcal{W}_s$.

**Marginalization.** For valuations in probability theory, marginalization is addition. Suppose $\sigma \in \mathcal{U}_s$, and $X \in s$. The marginal of $\sigma$ for $s\setminus X$, denoted by $\sigma^s_{s\setminus X}$, is the valuation for $s\setminus \{\mathcal{X}\}$ defined as follows:

$$
\sigma^s_{s\setminus X}(y) = \sum \{ \sigma(y,x) \mid x \in \mathcal{W}_X \}
$$

for all $y \in \mathcal{W}_{s\setminus \{X\}}$.

The above definition of marginalization follows from condition (P2) in the definition of a probability function since a proposition $\{y\}$ about variables in $s\setminus \{X\}$ is the same as proposition $\{y\} \times \mathcal{W}_X$ about variables in $s$.

It is easy to see that the definition of marginalization in (4.2) satisfies Axioms M1–M4. It can be easily shown that Axioms CM1 and CM2 hold [27].

**Removal.** In probability theory, removal is division followed by normalization (if normalization is possible). Division by zero can be defined arbitrarily. For the sake of simplicity of exposition, we define division of any real number by zero as resulting in zero. Suppose $\sigma \in \mathcal{V}_s$, and $\rho \in \mathcal{N}_r \cup \mathcal{Z}_r$. Let $K = \sum \{ \sigma(x^+\rho(x^-)) \mid x \in \mathcal{W}_{r\cup s} \text{ s.t. } \rho(x^-) > 0 \}$. Then the valuation resulting from the removal of $\rho$ from $\sigma$, denoted by $\sigma \oplus \rho$, is the valuation for $r \cup s$ given by

$$
(\sigma \oplus \rho)(x) = \begin{cases} 
K^{-1} \sigma(x^+\rho(x^-)/\rho(x^-)) & \text{if } K > 0 \text{ and } \rho(x^-) > 0 \\
0 & \text{if } K = 0 \text{ or } \rho(x^-) = 0
\end{cases}
$$

for all $x \in \mathcal{W}_{r\cup s}$.

If $K > 0$, $K$ is the normalization constant that ensures $\sigma \oplus \rho$ is a normal valuation. It is easy to see that Axioms R1, R2, and CR hold. Suppose $\rho \in \mathcal{N}_r \cup \mathcal{Z}_r$. The identity $t_\rho$ for $\rho$ defined in Axiom R2 is the normal valuation for $r$ such that $t_\rho(x) = K^{-1}$ if $\rho(x) > 0$, and $t_\rho(x) = 0$ if $\rho(x) = 0$, where $K$ is the normalization constant.

### 4.3.2 D-S belief-function theory

In this subsection, we show how D-S belief-function theory [1, 9, 28] fits in the VBS framework. More precisely, we define valuations, zero valuations, proper valuations, normal valuations, combination, marginalization, and removal.

In D-S belief-function theory, proper normal valuations correspond to either basic probability assignment functions, belief functions, plausibility functions, or commonality functions. For simplicity of exposition, we describe D-S belief-function theory in terms of commonality functions. We define commonality functions in terms of basic probability assignment functions. Remember that $2^\mathcal{W}$ denotes the set of all nonempty subsets of $\mathcal{W}_s$.

**Basic Probability Assignment Function.** A basic probability assignment (bpa) function for $s$ is a function $\mu : 2^\mathcal{W}_s \rightarrow \mathbb{R}$ such that

- (B1). $\mu(a) \geq 0$ for all $a \in 2^\mathcal{W}_s$,
- (B2). $\sum \{ \mu(a) \mid a \in 2^\mathcal{W}_s \} = 1$.

**Commonality Function.** A function $\theta : 2^\mathcal{W}_s \rightarrow \mathbb{R}$ is a commonality function for $s$ if there exists a bpa function $\mu$ for $s$ such that

$$
\theta(a) = \sum \{ \mu(c) \mid c \supseteq a \}.
$$

for all $a \in 2^\mathcal{W}_s$.

It is evident from (B1), (B2), and (4.4) that $0 \leq \theta(a) \leq 1$, and that $\theta(a) \geq \theta(b)$ whenever $a \subseteq b$.

The following two lemmas from [13] will help us understand the mathematical properties of commonality functions.

**Lemma 4.1.** Suppose $\mu$ and $\theta$ are real-valued functions defined on $2^\mathcal{W}_s$. Then (4.4) holds for every $a \in 2^\mathcal{W}_s$ if and only if

$$
\mu(a) = \sum \{ (-1)^{|c\cap a|} \theta(c) \mid c \supseteq a \}
$$

holds for all $a \in 2^\mathcal{W}_s$. 

---

**Modelling ignorance in uncertainty theories**

---

83
Lemma 4.2. Suppose $\mu$ and $\theta$ are real-valued functions defined on $2^\mathbb{W}$, and suppose (4.4) holds for every $a \in 2^\mathbb{W}$. Then
\[
\Sigma \{ \mu(a) | a \in 2^\mathbb{W} \} = \Sigma \{ (-1)^{|a|} \theta(a) | a \in 2^\mathbb{W} \}.
\]

These lemmas can be proven by the methods used in the appendix of Ch. 2 of [9].

From Lemma 4.1, we see that a basic probability assignment is completely determined by the commonality function. From Lemmas 4.1 and 4.2, and conditions (B1) and (B2), we see that a function $\theta : 2^\mathbb{W} \rightarrow \mathbb{R}$ is a commonality function if and only if two conditions are satisfied:
\[
\Sigma \{ (-1)^{|c|} \theta(c) | c \supseteq a \} \geq 0 \quad (4.5)
\]
for every $a \in 2^\mathbb{W}$, and
\[
\Sigma \{ (-1)^{|a|} \theta(a) | a \in 2^\mathbb{W} \} = 1. \quad (4.6)
\]
Condition (4.5) follows from condition (B1) and Lemma 4.1, and condition (4.6) follows from condition (B2) and Lemma 4.2.

In belief-function theory, a valuation for $s$ is a function $\sigma : 2^\mathbb{W} \rightarrow \mathbb{R}^+$. Zero valuations exist — a valuation $\sigma_s$ for $s$ is zero if and only if all values of $\sigma_s$ are zeros, i.e., $\sigma_s(a) = 0$ for all $a \in 2^\mathbb{W}$. Suppose $\sigma$ is a nonzero valuation for $s$. We call $\sigma$ proper if and only if $\Sigma \{ (-1)^{|c|} \theta(c) | c \supseteq a \} \geq 0$ for all $a \in 2^\mathbb{W}$.

Suppose $\sigma$ is a nonzero valuation for $s$. We say $\sigma$ is normal if and only if
\[
\Sigma \{ (-1)^{|a|} \theta(a) | a \in 2^\mathbb{W} \} = 1. \quad (4.6)
\]

Proper normal valuations are commonality functions.

In belief-function theory, combination is pointwise multiplication of commonality functions followed by normalization [9]. Before we can give a formal definition of combination, we need the definition of projection of subsets of configurations.

Projection of Subsets of Configurations. If $r$ and $s$ are sets of variables, $r \subseteq s$, and $a \in 2^\mathbb{W}$, then the projection of $a$ to $r$, denoted by $a^r$, is the element of $2^\mathbb{W}$ given by $a^r = \{ x^r | x \in a \}$.

Modelling ignorance in uncertainty theories

Combination. Suppose $p \in \mathcal{V}_r$ and $\sigma \in \mathcal{V}_s$. Let $K = \Sigma \{ (-1)^{|a^r|} \rho(a^r) \sigma(a^s) | a \in 2^\mathbb{W} \}$. The combination of $\rho$ and $\sigma$, denoted by $\rho \oplus \sigma$, is the valuation for $r \cup s$ given by
\[
(p \oplus \sigma)(a) = \begin{cases} 
K^{-1} \rho(a^r) \sigma(a^s) & \text{if } K \neq 0 \\
0 & \text{if } K = 0
\end{cases} \quad (4.7)
\]
for all $a \in 2^\mathbb{W}$. If $K = 0$, then $\rho \oplus \sigma = \zeta_{r \cup s}$. If $K \neq 0$, then $K$ is the normalization constant that ensures $\rho \oplus \sigma$ is a normal valuation. It is shown by Shafer [9, p. 61] that if $\rho$ and $\sigma$ are commonality functions (proper normal valuations), and $K \neq 0$, then $\rho \oplus \sigma$ is a commonality function.

It is easy to see that axioms C1–C6 are satisfied by the definition of combination in (4.7). The identity $\iota_x$ for $\mathcal{V}_s \cup (\zeta_s)$ is given by $\iota_x(a) = 1$ for all $a \in 2^\mathbb{W}$. Suppose $\sigma \in \mathcal{V}_s$. An identity $\delta_x$ for $\sigma$ in $\mathcal{V}_s$ is a normal valuation for $s$ such that $\delta_x(a) = K^{-1}$ if $\sigma(a) > 0$, and $\delta_x(a) = K^{-1}$ if $\sigma(a) = 0$, where $r$ is any non-negative real number, and $K$ is the normalization constant. Suppose $\sigma \in \mathcal{V}_s$. Notice that $\sigma$ is positive normal if and only if $\sigma(a) > 0$ for all $a \in 2^\mathbb{W}$.

Marginalization. Suppose $\sigma \in \mathcal{V}_s$, and suppose $X \in s$. The marginal of $\sigma$ for $s \setminus X$, denoted by $\sigma^{s \setminus (X)}$, is the valuation for $s \setminus X$ defined as follows:
\[
\sigma^{s \setminus (X)}(a) = \Sigma \{ (-1)^{|b|} \sigma(b) | b, c \in 2^\mathbb{W}, \text{s.t.} c^{(s \setminus X)} \supseteq a, \text{and } b \supseteq c \} \quad (4.8)
\]
for all $a \in 2^\mathbb{W}$. (4.8)

It is easy to see that the definition of marginalization in (4.8) satisfies Axioms M1–M4. It can be easily shown that Axioms CM1 and CM2 hold. Formal proofs that Axioms M1 and CM2 hold can be found in [23].

Removal. We define removal as pointwise division followed by normalization (if normalization is possible). Division by zero can be defined arbitrarily. For the sake of simplicity of exposition, we define division of any real number by zero as resulting in zero. Suppose $\sigma \in \mathcal{V}_s$, and $p \in \mathcal{V}_r \cup Z_r$. Let $K = \Sigma \{ (-1)^{|a^r|} \sigma(a^s) \rho(a^r) | a \in 2^\mathbb{W} \}$ s.t. $\rho(a^r) > 0$. Then the valuation
resulting from the removal of \( \rho \) from \( \sigma \), denoted by \( \sigma \oplus \rho \), is the valuation for \( r \subseteq s \) given by
\[
(\sigma \oplus \rho)(a) = \begin{cases} 
K^{-1}\sigma(a^{1r})/\rho(a^{1r}) & \text{if } K > 0 \text{ and } \rho(a^{1r}) > 0 \\
0 & \text{if } K = 0 \text{ or } \rho(a^{1r}) = 0 
\end{cases}
\] (4.9)
for all \( a \in 2^{W_{r \subseteq s}} \).

If \( K > 0 \), \( K \) is the normalization constant that ensures \( \sigma \oplus \rho \) is a normal valuation. It can be easily shown that Axioms R1, R2, and CR hold. Suppose \( \rho \in \mathcal{N}_r \cap P_r \). The identity \( \iota_{ag} \) for \( \rho \) defined in Axiom R2 is the normal valuation for \( r \) such that \( \iota_{ag}(a) = K^{-1} \) if \( \rho(a) > 0 \), and \( \iota_{ag}(a) = 0 \) if \( \rho(a) = 0 \), where \( K \) is the normalization constant.

Notice that if \( \sigma \) and \( \rho \) are commonality functions, it is possible that \( \sigma \oplus \rho \) may not be a commonality function because condition (4.5) may not be satisfied by \( \sigma \oplus \rho \). In fact, if \( \sigma \) is a commonality function for \( s \), and \( r \subseteq s \), then even \( \sigma \oplus \sigma^{1r} \) may fail to be a commonality function. This fact is the reason why we need the concept of proper valuations as distinct from nonzero and normal valuations in the general VBS framework.

### 4.3.3 Spohn’s epistemic-belief theory

In this section, we show how Spohn’s epistemic-belief theory [30, 31, 16] fits in the VBS framework. More precisely, we define valuations, proper valuations, normal valuations, combination, marginalization, and removal.

In Spohn’s theory, a basic representational unit is called a disbelief function. Let \( \mathbb{N} \) denote the set of all natural numbers.

**Disbelief Function.** A disbelief function for \( s \) is a function \( \sigma: 2^{W_s} \rightarrow \mathbb{N} \) such that

- (D1). there exists a configuration \( x \in W_s \) such that \( \sigma(\{x\}) = 0 \); and
- (D2). \( \sigma(a) = \text{MIN}\{ \sigma(\{x\}) \mid x \in a \} \) for all \( a \in 2^{W_s} \).

Notice that from condition (D2) in the definition of a disbelief function, a disbelief function is completely determined by its values for singleton subes.

Intuitively, \( \sigma(a) \) represents the degree of disbelief in proposition \( a \) (the proposition that the true configuration of \( s \) is in \( a \)). The degree of belief in proposition \( a \) is given by \( \sigma(\neg a) \), where \( \neg a = W_s \setminus a \). Thus \( \sigma \) represents an epistemic state in which \( a \) is believed if and only if \( \sigma(a) > 0 \), \( a \) is disbelieved if and only if \( \sigma(a) < 0 \), and \( a \) is neither believed nor disbelieved if \( \sigma(a) = 0 \). Also, in epistemic state \( \sigma \), \( a \) is more believed than \( b \) if \( \sigma(a) > \sigma(b) > 0 \), and \( a \) is more disbelieved than \( b \) if \( \sigma(a) < \sigma(b) < 0 \).

In Spohn’s epistemic-belief theory, a valuation for \( s \) is a function \( \sigma: W_s \rightarrow \mathbb{N} \). Zero valuations do not exist, i.e., all valuations are nonzero. Also, all valuations are proper.

Suppose \( \sigma \in V_s \). We say \( \sigma \) is normal if and only if \( \text{MIN}\{ \sigma(x) \mid x \in W_s \} = 0 \). A normal valuation for \( s \) can be regarded as a disbelief function for \( s \) defined only for singleton subsets of \( 2^{W_s} \).

**Combination.** In Spohn’s theory, combination is simply pointwise addition followed by normalization [30, 16]. If \( \rho \in V_r \), and \( \sigma \in V_s \), then their combination, denoted by \( \rho \oplus \sigma \), is the valuation for \( r \subseteq s \) given by
\[
(\rho \oplus \sigma)(x) = \rho(x^{1r}) + \sigma(x^{1s}) - K
\] (4.10)
for all \( x \in W_{r \subseteq s} \), where \( K \) is a constant defined as follows:
\[
K = \text{MIN}\{ \rho(x^{1r}) + \sigma(x^{1s}) \mid x \in W_{r \subseteq s} \}.
\]

\( K \) is the normalization constant that ensures that \( \rho \oplus \sigma \) is a normal valuation.

It is easy to see that axioms C1–C6 are satisfied by the definition of combination in (4.10). The identity \( \iota_s \) for \( \mathcal{N}_s \cap (a) \) is given by \( \iota_s(x) = 0 \) for all \( x \in W_s \). Every normal valuation in \( \mathcal{N}_s \) has a unique identity in \( \mathcal{N}_s \), therefore a normal valuation is also positive normal.

**Marginalization.** Suppose \( \sigma \in V_s \), and suppose \( X \subseteq s \). The marginal of \( \sigma \) for \( s \setminus X \), denoted by \( \sigma^{(s\setminus X)} \), is the valuation for \( s \setminus X \) defined as follows:
\[
\sigma^{(s\setminus X)}(y) = \text{MIN}\{ \sigma(y,x) \mid x \in W_{s \setminus X} \}
\] (4.11)
for all \( y \in W_{s \setminus X} \).
Modelling ignorance in uncertainty theories

Notice that from condition (S2) in the definition of a possibility function, a possibility function is completely determined by its values for singleton subsets.

A possibility function is a complete representation of a consistent possibilistic state [20]. A is possible in state π if and only if π(a) = 1, and a is not possible in state π if and only if π(a) < 1. A possibility function consists of more than a representation of a consistent possibilistic state. It also includes degrees to which proposition are possible and degrees to which propositions are not possible. π(a) can be interpreted as the degree to which proposition a is possible, and 1 − π(a) can be interpreted as the degree to which proposition a is not possible, i.e., a is more possible than b if π(a) > π(b) and conversely, a is more impossible than b if π(a) < π(b) < 1.

In Zadeh’s possibility theory, a valuation σ for s is a function σ: \( \mathcal{W}_s \rightarrow \mathbb{R}^+ \). Zero valuations exist — a valuation \( \zeta_s \) for s is zero if and only if all values of \( \zeta_s \) are zeros, i.e., \( \zeta_s(x) = 0 \) for all \( x \in \mathcal{W}_s \).

Suppose \( \sigma \) is a valuation for s. We say \( \sigma \) is proper if and only if \( \sigma \neq \zeta_s \), i.e., all nonzero valuations are proper.

Suppose \( \sigma \) is a valuation for s. We say \( \sigma \) is normal if and only if \( \text{MAX}\{\sigma(x) | x \in \mathcal{W}_s\} = 1 \). A normal valuation can be regarded as a possibility function defined only for singleton subsets.

Combination.\(^1\) We define combination as multiplication followed by normalization (if normalization is possible). Suppose \( \rho \in \mathcal{V}_r \), and suppose \( \sigma \in \mathcal{U}_s \). Suppose \( K = \text{MAX}\{\rho(x) \sigma(x) | x \in \mathcal{W}_{r,s}\} \). The combination of \( \rho \) and \( \sigma \), denoted by \( \rho \Theta \sigma \), is the valuation for \( r \cup s \) given by

\[^1\] There are several definitions of combination in possibility theory. Zadeh [32] has defined combination as pointwise minimization (with no normalization). However, several alternative definitions of combination have been suggested in the fuzzy set literature [see, e.g., 2, pp.78–85]. Any triangular norm can be regarded as a definition of combination. In the VBS framework, combination has to be associative, and the combination of two valuations has to be either normal or zero. These two requirements restrict the definition of combination to pointwise multiplication (since pointwise minimization followed by normalization, for example, fails to be associative).
for all \( x \in \mathcal{W}_{r,s} \). If \( K = 0 \), then \( K = \frac{1}{\rho(x)} \). If \( K > 0 \), then \( K \) is the normalization constant that ensures \( \rho \odot \sigma \) is a normal valuation.

It is easy to see that axioms C1–C6 are satisfied by the definition of combination in (4.13). The identity \( t \) is determined by \( \tau = \{ t \} \) for all \( x \in \mathcal{W}_s \). Suppose \( \sigma \in \mathcal{N}_s \). An identity \( \delta \) is a normal valuation for \( s \) such that \( \delta \equiv 1 \) if \( \sigma(x) > 0 \), and \( \delta \equiv 0 \) if \( \sigma(x) = 0 \), where \( r \) is any real number in the interval \([0, 1]\). Suppose \( \delta \in \mathcal{N}_s \). Notice that \( \delta \) is positive normal if and only if \( \delta(x) > 0 \) for all \( x \in \mathcal{W}_s \).

**Marginalization.** Suppose \( \sigma \in \mathcal{U}_s \), and \( X \in s \). The marginal of \( \sigma \) for \( s \setminus \{ X \} \), denoted by \( \sigma^{s \setminus \{ X \}} \), is the valuation for \( s \setminus \{ X \} \) defined as follows:

\[
\sigma^{s \setminus \{ X \}}(y) = \operatorname{MAX} \{ \sigma(y, x) \mid x \in \mathcal{W}_X \}
\]

for all \( y \in \mathcal{W}_{s \setminus \{ X \}} \).

The above definition of marginalization follows from condition (S2) in the definition of a possibility function since a proposition \( y \) about variables in \( s \setminus \{ X \} \) is the same as \( y \times \mathcal{W}_X \) about variables in \( s \).

It is easy to see that the definition of marginalization in (4.14) satisfies Axioms M1–M4. It can be easily shown that Axioms CM1 and CM2 hold. Formal proofs that Axioms M1 and CM1 hold can be found in [20].

**Removal.** In possibility theory, removal is division followed by normalization (if normalization is possible). Division by zero can be defined arbitrarily. For the sake of simplicity of exposition, we define division of any real number by zero as resulting in zero. Suppose \( \sigma \in \mathcal{U}_s \), \( \rho \in \mathcal{N}_r \setminus \mathcal{Z}_r \). Suppose \( K = \operatorname{MAX} \{ \sigma(x) / \rho(x) \mid x \in \mathcal{W}_{r,s} \) such that \( \rho(x) > 0 \} \). Then the valuation resulting from the removal of \( \rho \) from \( \sigma \), denoted by \( \sigma \odot \rho \), is given by

\[
(\sigma \odot \rho)(x) = \begin{cases} 
K^{-1} \sigma(x) / \rho(x) & \text{if } K > 0 \text{ and } \rho(x) > 0 \\
0 & \text{if } K = 0 \text{ or } \rho(x) = 0
\end{cases}
\]

for all \( x \in \mathcal{W}_{r,s} \).

If \( K > 0 \), then \( K \) is the normalization constant that ensures \( \sigma \odot \rho \) is a normal valuation. It can be easily shown that Axioms R1, R2, and CR hold. Suppose \( \rho \in \mathcal{N}_r \setminus \mathcal{Z}_r \). The identity \( t \) is valid for \( \rho \) defined in Axiom R2 is the normal valuation for \( r \) such that \( t_r(x) = \rho(x) \) for all \( \rho(x) > 0 \), and \( t_r(x) = 0 \) for all \( \rho(x) = 0 \).

### 4.4 Ignorance in VBS

In this section, we describe how ignorance is modelled in VBS and its instances.

First, let us describe the setting. Suppose \( s \) is an arbitrary subset of variables. We will define complete ignorance about variables in \( s \) as a proper normal valuation \( t \) such that \( t_s \equiv \sigma \), for all \( \sigma \in \mathcal{P}_{r \setminus \mathcal{N}_t} \). (The requirement that \( t_s \) is proper normal is simply so that \( t_s \) represents coherent knowledge.) In words, regardless of our current knowledge \( \sigma \), \( t_s \) does not add anything new. Notice that in the VBS terminology, this simply means that \( t_s \) is the unique identity for the semigroup \( \mathcal{N}_s \). As we saw in Section 4.2, complete ignorance has certain properties. First, \( t_s \odot t_s = t_s \), i.e., the combination of complete ignorance about variables in \( r \) results in complete ignorance about variables in \( r \). Second, if \( r \subseteq s \), then \( t_r \equiv t_s \). In words, if we are completely ignorant about variables in \( s \), and we disregard variables in \( r \), then we are completely ignorant about variables in \( r \).

As we saw in Section 4.3, complete ignorance about variables in \( s \) can be represented in probability theory by the probability distribution \( t_s(x) = K^{-1} \) for all \( x \in \mathcal{W}_s \), where \( K = \mathcal{W}_s \), i.e., \( t_s \) is the equally likely probability distribution for \( s \). In D-S theory of belief functions, complete ignorance is represented by the commonality function \( t_s(\mathcal{A}) = 1 \) for all \( \mathcal{A} \subseteq \mathcal{W}_s \). This commonality function
Modelling ignorance in uncertainty theories

be represented in D-S belief-function theory. All uncertainty theories that fit in the
VBS framework can represent ignorance equally well.

It is true that in D-S belief-function theory, representation of ignorance is
distinct from representation of an equally likely distribution, whereas in
probability theory, the representation of ignorance is the same as the
representation of an equally likely distribution. What this means is that if we have
a probabilistic encoding of knowledge, it is not possible to automatically translate
a probabilistic representation to a D-S belief-function representation without
knowing what knowledge the probabilistic representation represents [5].

The axiomatic VBS framework reveals that there are a lot of similarities
between the different uncertainty theories. Of course, there are also many
differences. The question then is which uncertainty theory should one use in a
particular application. The answer to this question depends on the nature of
knowledge available in the particular application. Each uncertainty theory is based
on some semantics. For example, since probability theory is based on frequency
semantics, if we have frequency information in the application under
consideration, then probability theory is the appropriate uncertainty calculus that
should be used. Semantics for D-S belief-function theory can be found in Shafer
[10, 11, 12] and Smets [29]. Semantics for Spohn's epistemic-belief theory can
be found in Pearl [6], and Goldszmidt and Pearl [3, 4]. Semantics for possibility
theory can be found in [8] and [20].

Acknowledgments

This work is based upon work supported in part by the National Science
Foundation under Grant No. SES-9213558, and in part by the General Research
Fund of the University of Kansas under Grant No. 3605-XX-0038. Any
opinions, findings, and conclusions or recommendations expressed in this paper
are those of the author and do not necessarily reflect the views of the National
Science Foundation or the University of Kansas.
Modelling ignorance in uncertainty theories


