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 $m(\{x\}) = m(\{\sim x\}) = p$, $m(\{x, \sim x\}) = 1 - 2p$, where $0 \leq p \leq 0.5$, 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' 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, 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. $\sigma \oplus \delta_\sigma = \sigma$. 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 $\mathfrak{X}$ whose elements are called *variables*. Variables are denoted by upper-case Latin alphabets, X, Y, Z, etc. Subsets of $\mathfrak{X}$ are denoted by lower-case Latin alphabets, r, s, t, etc.

Valuations. For each $s \subseteq \mathfrak{X}$, there is a set $\mathcal{V}_s$. We call the elements of $\mathcal{V}_s$ *valuations for s*. Let $\mathcal{V}$ denote $\cup\{\mathcal{V}_s \mid s \subseteq \mathfrak{X}\}$, the set of all *valuations*. If $\sigma \in \mathcal{V}_s$, then we say *s is the domain of* σ. Valuations are denoted by lower-case Greek alphabets, ρ, σ, τ, etc.

Valuations are primitives in our abstract framework and, as such, require no definition. But as we shall see shortly, they are objects that can be combined, marginalized, and removed. Intuitively, a valuation for s represents some knowledge about variables in s.

Zero Valuations. For each $s \subseteq \mathcal{X}$, there is at most one valuation $\zeta_s \in \mathcal{V}_s$ called *the zero valuation for s*. Let $\mathcal{Z}$ denote $\{\zeta_s \mid s \subseteq \mathcal{X}\}$, the set of all *zero valuations*. Notice that we are not assuming zero valuations always exist. If zero valuations do not exist, $\mathcal{Z} = \emptyset$. We call valuations in $\mathcal{V} - \mathcal{Z}$ *nonzero valuations*.

Intuitively, a zero valuation represents knowledge that is internally inconsistent, i.e., knowledge that is a contradiction, or knowledge whose truth value is always false. The concept of zero valuations is important in the theory of consistent knowledge-based systems [Shenoy 1994c].

Proper Valuations. For each $s \subseteq \mathcal{X}$, there is a subset $\mathcal{P}_s$ of $\mathcal{V}_s - \{\zeta_s\}$. We call the elements of $\mathcal{P}_s$ *proper valuations for s*. Let $\mathcal{P}$ denote $\cup\{\mathcal{P}_s \mid s \subseteq \mathcal{X}\}$, the set of all *proper valuations*. Intuitively, a proper valuation represents knowledge that is partially coherent. By coherent knowledge, we mean knowledge that has well-defined semantics. Proper valuations play no role either in the definition, or in the characterizations, or in the properties of conditional independence. The only role of proper valuations is in the semantics of knowledge.

Normal Valuations. For each $s \subseteq \mathcal{X}$, there is another subset $\mathcal{N}_s$ of $\mathcal{V}_s - \{\zeta_s\}$. We call the elements of $\mathcal{N}_s$ *normal valuations for s*. Let $\mathcal{N}$ denote $\cup\{\mathcal{N}_s \mid s \subseteq \mathcal{X}\}$, the set of all *normal valuations*. Intuitively, a normal valuation represents knowledge that is also partially coherent, but in a sense that is different from proper valuations. Normal valuations play an important role in the definition and characterization of conditional independence.

We call the elements of $\mathcal{P} \cap \mathcal{N}$ *proper normal valuations*. Intuitively, a proper normal valuation represents knowledge that is completely coherent, i.e., knowledge that has well-defined semantics.

Combination. We assume there is a mapping $\oplus: \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{N} \cup \mathcal{Z}$, called *combination*, that satisfies the following four axioms:

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

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 $\sigma \in \mathcal{V}_s$. Then $\zeta_r \oplus \sigma = \zeta_{r \cup s}$.

If $\rho \oplus \sigma$, read as ρ plus σ, is a zero valuation, then we say that ρ and σ are *inconsistent*. If $\rho \oplus \sigma$ is a normal valuation, then we say that ρ and σ are *consistent*.

Intuitively, combination corresponds to aggregation of knowledge. If ρ and σ 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, $(\ldots((\sigma_1 \oplus \sigma_2) \oplus \sigma_3) \oplus \ldots \oplus \sigma_m)$ 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 $\oplus\{\sigma_1, \ldots, \sigma_m\}$, 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 ζ_s as the zero of the semigroup $\mathcal{N}_s \cup \{\zeta_s\}$.

Identity Valuations. We assume the following identity axiom.

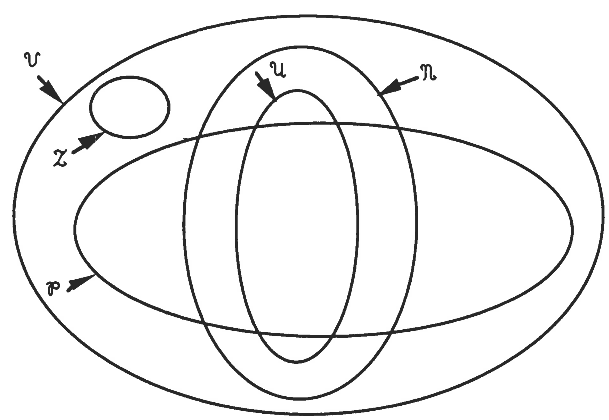
Axiom C5 (Identity): For each $s \subseteq \mathcal{X}$, the commutative semigroup $\mathcal{N}_s \cup \{\zeta_s\}$ has an identity denoted by ι_s.

In other words, Axiom C5 assumes there exists $\iota_s \in \mathcal{N}_s \cup \{\zeta_s\}$ such that for each $\sigma \in \mathcal{N}_s \cup \{\zeta_s\}$, $\sigma \oplus \iota_s = \sigma$. Notice that a commutative semigroup may have at most one identity. From Axiom C4, it follows that $\iota_s \neq \zeta_s$, therefore $\iota_s \in \mathcal{N}_s$. 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, ι_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 ζ_s. Notice that if $\sigma \in \mathcal{N}_s$, then $\delta_\sigma \in \mathcal{N}_s$. Also, notice that ι_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.



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

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

Axiom C6 implies that $\mathcal{U}_\varnothing = \mathcal{N}_\varnothing = \{\iota_\varnothing\}$ where $\iota_\varnothing$ is the identity valuation for the semigroup $\mathcal{N}_\varnothing \cup \{\zeta_\varnothing\}$.

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

> **Axiom M1** (*Order of Deletion*): Suppose $\sigma \in \mathcal{V}_s$, and suppose $X_1, X_2 \in s$. Then $(\sigma^{\downarrow(s-\{X_1\})})^{\downarrow(s-\{X_1, X_2\})} = (\sigma^{\downarrow(s-\{X_2\})})^{\downarrow(s-\{X_1, X_2\})}$;
>
> **Axiom M2** (*Zero*): If zero valuations exist, then $\zeta_s^{\downarrow(s-\{X\})} = \zeta_{s-\{X\}}$;
>
> **Axiom M3** (*Normal*): $\sigma^{\downarrow(s-\{X\})} \in \mathcal{N}$ if and only if $\sigma \in \mathcal{N}$;
>
> **Axiom M4** (*Positive Normal*): If $\sigma \in \mathcal{U}$, then $\sigma^{\downarrow(s-\{X\})} \in \mathcal{U}$;
>
> **Axiom CM1** (*Combination and Marginalization 1*): Suppose $\rho \in \mathcal{V}_r$ and $\sigma \in \mathcal{V}_s$. Suppose $X \notin r$, and $X \in s$. Then $(\rho \oplus \sigma)^{\downarrow((r\cup s)-\{X\})} = \rho \oplus (\sigma^{\downarrow(s-\{X\})})$; 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 σ, i.e., $\sigma \oplus \delta_{\sigma^{\downarrow r}} = \sigma$.

We call $\sigma^{\downarrow(s-\{X\})}$ the *marginal of σ for s–{X}*.

Intuitively, marginalization corresponds to coarsening of knowledge. If σ is a valuation for s representing some knowledge about variables in s, and $X \in s$, then

$\sigma^{\downarrow(s-\{X\})}$ represents the knowledge about variables in s–{X} implied by σ if we disregard variable X.

If we regard marginalization as a coarsening of a valuation by deleting variables, then Axiom M2 says that the order in which the variables are deleted does not matter. One implication of this axiom is that $(\sigma^{\downarrow(s-\{X_1\})})^{\downarrow(s-\{X_1, X_2\})}$ can be written simply as $\sigma^{\downarrow(s-\{X_1, X_2\})}$, i.e., we need not indicate the order in which the variables are deleted.

Axioms M2, M3 and M4 state that marginalization preserves the coherence of knowledge. An implication of Axiom M4 is that a valuation σ for s is normal if and only if $\sigma^{\downarrow\emptyset} = \iota_\emptyset$.

Axiom CM1 states that the computation of $(\rho\oplus\sigma)^{\downarrow((r\cup s)-\{X\})}$ can be accomplished without having to compute $\rho\oplus\sigma$. The combination $\rho\oplus\sigma$ is a valuation for $r\cup s$ whereas the combination $\rho\oplus(\sigma^{\downarrow(s-\{X\})})$ is a valuation for $(r\cup s)-\{X\}$.

Axiom CM2 states an important property of identity valuations. It follows from Axiom CM2 that $\iota_s\oplus\iota_r = \iota_{r\cup s}$. Also, if $r\subseteq s$, then $\iota_s^{\downarrow r} = \iota_r$ [22].

Next, we define another binary operation called removal. The removal operation is an inverse of the combination operation.

Removal. We assume there is a mapping $\ominus: \mathcal{V}\times(\mathcal{N}\cup\mathcal{Z})\rightarrow(\mathcal{N}\cup\mathcal{Z})$, called *removal*, that satisfies the following three axioms:

Axiom R1 (*Domain*): Suppose $\sigma\in\mathcal{V}_s$, and $\rho\in\mathcal{N}_r\cup\{\zeta_r\}$. Then $\sigma\ominus\rho\in\mathcal{N}_{r\cup s}\cup\{\zeta_{r\cup s}\}$;

Axiom CR1 (*Combination and Removal 1*): For each $\rho\in\mathcal{N}\cup\mathcal{Z}$, there exists an identity for ρ, denoted by, say, ι_ρ, such that $\rho\ominus\rho = \iota_\rho$; and

Axiom CR2 (*Combination and Removal 2*): Suppose $\pi, \theta\in\mathcal{V}$, and $\rho\in\mathcal{N}\cup\mathcal{Z}$. Then, $(\pi\oplus\theta)\ominus\rho = \pi\oplus(\theta\ominus\rho)$.

We call $\sigma\ominus\rho$, read as σ minus ρ, the *valuation resulting after removing ρ from σ*. Intuitively, $\sigma\ominus\rho$ can be interpreted as follows. If σ and ρ represent some knowledge, and if we remove the knowledge represented by ρ from σ, 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 $\sigma^{\downarrow(a\cup b)}\ominus\sigma^{\downarrow a}$ for $a\cup b$ plays an important role in uncertainty theories. Borrowing terminology from probability theory, we call $\sigma^{\downarrow(a\cup b)}\ominus\sigma^{\downarrow a}$ the *conditional for b given a with respect to σ*. Let $\sigma(b|a)$ denote $\sigma^{\downarrow(a\cup b)}\ominus\sigma^{\downarrow a}$ We call *b* the *head* of the domain of $\sigma(b|a)$, and we call *a* the *tail* of the domain of $\sigma(b|a)$. Also, if $a = \emptyset$, let $\sigma(b)$ denote $\sigma(b|\emptyset)$. The following theorem states some important properties of conditionals.

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

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

(ii). $\sigma(a)\oplus\sigma(b|a) = \sigma(a\cup b)$.

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

(iv). Suppose $b'\subseteq b$. Then $\sigma(b|a)^{\downarrow(a\cup b')} = \sigma(b'|a)$.

(v). $(\sigma(b|a)\oplus\sigma(c|a\cup b))^{\downarrow(a\cup c)} = \sigma(c|a)$

(vi). $\sigma(b|a)^{\downarrow a} = \iota_{\sigma(a)}$.

(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 $\mathcal{W}_X$ are called 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 $\blacklozenge$ to name that configuration; $\mathcal{W}_\emptyset = \{\blacklozenge\}$.

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 $\mathbf{x}^{\downarrow r}$ denotes the projection of **x** to r. If $r = \emptyset$, then of course, $\mathbf{x}^{\downarrow r} = \blacklozenge$.

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 $\mathbf{z}^{\downarrow r} = \mathbf{x}$, and $\mathbf{z}^{\downarrow s} = \mathbf{y}$. Let (**x**,**y**) or (**y**,**x**) denote **z**. As per this notation, $(\mathbf{x},\blacklozenge) = (\blacklozenge,\mathbf{x}) = \mathbf{x}$.

Let $2^{\mathcal{W}_s}$ denote the set of all nonempty subsets of $\mathcal{W}_s$. Elements of $2^{\mathcal{W}_s}$ 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, combination, marginalization, and removal.

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

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

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

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

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 ζ_s for s is zero if and only if all values of ζ_s are zeros, i.e., $\zeta_s(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{W}_s$. Suppose σ is a valuation for s. We call σ *proper* if and only if $\sigma \neq \zeta_s$, i.e., all nonzero valuations are proper. Suppose σ is a valuation for s. We call σ *normal* if and only if $\Sigma\{\sigma(\mathbf{x}) \mid \mathbf{x} \in \mathcal{W}_s\} = 1$. 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 $\rho \in \mathcal{V}_r$, and $\sigma \in \mathcal{V}_s$. Let $K = \Sigma\{\rho(\mathbf{x}^{\downarrow r})\sigma(\mathbf{x}^{\downarrow s}) \mid \mathbf{x} \in \mathcal{W}_{r \cup s}\}$. The *combination of ρ and σ*, denoted by $\rho \oplus \sigma$, is the valuation for $r \cup s$ given by

$$(\rho \oplus \sigma)(\mathbf{x}) = \begin{cases} K^{-1}\rho(\mathbf{x}^{\downarrow r})\sigma(\mathbf{x}^{\downarrow s}) & \text{if } K > 0 \\ 0 & \text{if } K = 0 \end{cases} \tag{4.1}$$

for all $\mathbf{x} \in \mathcal{W}_{r \cup s}$. If $K = 0$, $\rho \oplus \sigma = \zeta_{r \cup s}$. If $K > 0$, then K is a normalization constant that ensures $\rho \oplus \sigma$ is a normal valuation.

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

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

$$\sigma^{\downarrow(s-\{X\})}(\mathbf{y}) = \Sigma\{\sigma(\mathbf{y},\mathbf{x}) \mid \mathbf{x} \in \mathcal{W}_X\} \tag{4.2}$$

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

The above definition of marginalization follows from condition (P2) in the definition of a probability function since a proposition $\{\mathbf{y}\}$ about variables in $s-\{X\}$ is the same as proposition $\{\mathbf{y}\}\times\mathcal{W}_X$ 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 $K = \Sigma\{\sigma(\mathbf{x}^{\downarrow s})/\rho(\mathbf{x}^{\downarrow r}) \mid \mathbf{x} \in \mathcal{W}_{r\cup s} \text{ s.t. } \rho(\mathbf{x}^{\downarrow r}) > 0\}$. Then the valuation resulting from the removal of ρ from σ, denoted by $\sigma \ominus \rho$, is the valuation for $r \cup s$ given by

$$(\sigma\ominus\rho)(\mathbf{x}) = \begin{cases} K^{-1}\sigma(\mathbf{x}^{\downarrow s})/\rho(\mathbf{x}^{\downarrow r}) & \text{if } K > 0 \text{ and } \rho(\mathbf{x}^{\downarrow r}) > 0 \\ 0 & \text{if } K = 0 \text{ or } \rho(\mathbf{x}^{\downarrow r}) = 0 \end{cases} \tag{4.3}$$

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

If $K > 0$, K is the normalization constant that ensures $\sigma\ominus\rho$ 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 ι_ρ for ρ defined in Axiom R2 is the normal valuation for r such that $\iota_\rho(\mathbf{x}) = K^{-1}$ if $\rho(\mathbf{x}) > 0$, and $\iota_\rho(\mathbf{x}) = 0$ if $\rho(\mathbf{x}) = 0$, 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 $2^{\mathcal{W}_s}$ 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 $\mu: 2^{\mathcal{W}_s} \to \mathbb{R}$ such that

(B1). $\mu(\mathbf{a}) \geq 0$ for all $\mathbf{a} \in 2^{\mathcal{W}_s}$

(B2). $\Sigma\{\mu(\mathbf{a}) \mid \mathbf{a} \in 2^{\mathcal{W}_s}\} = 1$.

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

$$\theta(\mathbf{a}) = \Sigma\{\mu(\mathbf{c}) \mid \mathbf{c} \supseteq \mathbf{a}\}. \tag{4.4}$$

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

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

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

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

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

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

Lemma 4.2. Suppose μ and θ are real-valued functions defined on $2^{\mathcal{W}_s}$, and suppose (4.4) holds for every $\mathbf{a} \in 2^{\mathcal{W}_s}$. Then

$$\Sigma\{\mu(\mathbf{a}) \mid \mathbf{a} \in 2^{\mathcal{W}_s}\} = \Sigma\{(-1)^{|\mathbf{a}|+1}\theta(\mathbf{a}) \mid \mathbf{a} \in 2^{\mathcal{W}_s}\}.$$

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 θ: $2^{\mathcal{W}_s} \to \mathbb{R}$ is a commonality function if and only if two conditions are satisfied:

$$\Sigma\{(-1)^{|\mathbf{c}-\mathbf{a}|}\theta(\mathbf{c}) \mid \mathbf{c} \supseteq \mathbf{a}\} \geq 0 \tag{4.5}$$

for every $\mathbf{a} \in 2^{\mathcal{W}_s}$, and

$$\Sigma\{(-1)^{|\mathbf{a}|+1}\theta(\mathbf{a}) \mid \mathbf{a} \in 2^{\mathcal{W}_s}\} = 1. \tag{4.6}$$

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 σ: $2^{\mathcal{W}_s} \to \mathbb{R}^+$. Zero valuations exist — a valuation ζ_s for s is *zero* if and only if all values of ζ_s are zeros, i.e., $\zeta_s(\mathbf{a}) = 0$ for all $\mathbf{a} \in 2^{\mathcal{W}_s}$. Suppose σ is a nonzero valuation for s. We call σ *proper* if and only if $\Sigma\{(-1)^{|\mathbf{c}-\mathbf{a}|}\theta(\mathbf{c}) \mid \mathbf{c} \supseteq \mathbf{a}\} \geq 0$ for all $\mathbf{a} \in 2^{\mathcal{W}_s}$. Suppose σ is a nonzero valuation for s. We say σ is *normal* if and only if $\Sigma\{(-1)^{|\mathbf{a}|+1}\theta(\mathbf{a}) \mid \mathbf{a} \in 2^{\mathcal{W}_s}\} = 1$. 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 $\mathbf{a} \in 2^{\mathcal{W}_s}$, then the *projection of a to r*, denoted by $\mathbf{a}^{\downarrow r}$, is the element of $2^{\mathcal{W}_r}$ given by $\mathbf{a}^{\downarrow r} = \{\mathbf{x}^{\downarrow r} \mid \mathbf{x} \in \mathbf{a}\}$.

Combination. Suppose $\rho \in \mathcal{V}_r$ and $\sigma \in \mathcal{V}_s$. Let $K = \Sigma\{(-1)^{|\mathbf{a}|+1}\rho(\mathbf{a}^{\downarrow r})\sigma(\mathbf{a}^{\downarrow s}) \mid \mathbf{a} \in 2^{\mathcal{W}_{r\cup s}}\}$. The *combination of ρ and σ*, denoted by $\rho\oplus\sigma$, is the valuation for $r\cup s$ given by

$$(\rho\oplus\sigma)(\mathbf{a}) = \begin{cases} K^{-1}\rho(\mathbf{a}^{\downarrow r})\sigma(\mathbf{a}^{\downarrow s}) & \text{if } K \neq 0 \\ 0 & \text{if } K = 0 \end{cases} \tag{4.7}$$

for all $\mathbf{a} \in 2^{\mathcal{W}_{r\cup s}}$. 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 ρ and σ 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 ι_s for $\mathcal{N}_s \cup \{\zeta_s\}$ is given by $\iota_s(\mathbf{a}) = 1$ for all $\mathbf{a} \in 2^{\mathcal{W}_s}$. Suppose $\sigma \in \mathcal{N}_s$. An identity δ_σ for σ in $\mathcal{N}_s$ is a normal valuation for s such that $\delta_\sigma(\mathbf{a}) = K^{-1}$ if $\sigma(\mathbf{a}) > 0$, and $\delta_\sigma(\mathbf{a}) = K^{-1}r$ if $\sigma(\mathbf{a}) = 0$, where r is any non-negative real number, and K is the normalization constant. Suppose $\sigma \in \mathcal{N}_s$. Notice that σ is *positive normal* if and only if $\sigma(\mathbf{a}) > 0$ for all $\mathbf{a} \in 2^{\mathcal{W}_s}$.

Marginalization. Suppose $\sigma \in \mathcal{V}_s$, and suppose $X \in s$. The *marginal of σ for s–{X}*, denoted by $\sigma^{\downarrow(s-\{X\})}$, is the valuation for s–{X} defined as follows:

$$\sigma^{\downarrow(s-\{X\})}(\mathbf{a}) = \Sigma\{(-1)^{|\mathbf{b}-\mathbf{c}|}\sigma(\mathbf{b}) \mid \mathbf{b}, \mathbf{c} \in 2^{\mathcal{W}_s} \text{ s.t. } \mathbf{c}^{\downarrow(s-\{X\})} \supseteq \mathbf{a}, \text{ and } \mathbf{b} \supseteq \mathbf{c}\} \tag{4.8}$$

for all $\mathbf{a} \in 2^{\mathcal{W}_{s-\{X\}}}$.

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 $\rho \in \mathcal{N}_r \cup \mathcal{Z}_r$. Let $K = \Sigma\{(-1)^{|\mathbf{a}|+1}\sigma(\mathbf{a}^{\downarrow s})/\rho(\mathbf{a}^{\downarrow r}) \mid \mathbf{a} \in 2^{\mathcal{W}_{r\cup s}} \text{ s.t. } \rho(\mathbf{a}^{\downarrow r}) > 0\}$. Then the valuation resulting from the removal of ρ from σ, denoted by $\sigma\ominus\rho$, is the valuation for $r\cup s$ given by

$$(\sigma\ominus\rho)(\mathbf{a}) = \begin{cases} K^{-1}\sigma(\mathbf{a}^{\downarrow s})/\rho(\mathbf{a}^{\downarrow r}) & \text{if } K>0 \text{ and } \rho(\mathbf{a}^{\downarrow r})>0 \\ 0 & \text{if } K=0 \text{ or } \rho(\mathbf{a}^{\downarrow r})=0 \end{cases} \tag{4.9}$$

for all $\mathbf{a}\in 2^{\mathcal{W}_{r\cup s}}$.

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

Notice that if σ and ρ are commonality functions, it is possible that $\sigma\ominus\rho$ may not be a commonality function because condition (4.5) may not be satisfied by $\sigma\ominus\rho$. In fact, if σ is a commonality function for s, and $r\subseteq s$, then even $\sigma\ominus\sigma^{\downarrow r}$ 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 $\sigma: 2^{\mathcal{W}_s}\to\mathbb{N}$ such that

(D1). there exists a configuration $\mathbf{x}\in\mathcal{W}_s$ such that $\sigma(\{\mathbf{x}\}) = 0$; and

(D2). $\sigma(\mathbf{a}) = \text{MIN}\{\sigma(\{\mathbf{x}\}) \mid \mathbf{x}\in\mathbf{a}\}$ for all $\mathbf{a}\in 2^{\mathcal{W}_s}$.

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

Intuitively, $\sigma(\mathbf{a})$ 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 $\sigma(\sim\mathbf{a})$, where $\sim\mathbf{a} = \mathcal{W}_s - \mathbf{a}$. Thus σ represents an epistemic state in which **a** is believed if and only if $\sigma(\sim\mathbf{a})>0$, **a** is disbelieved if and only if $\sigma(\mathbf{a})>0$, and **a** is neither believed nor disbelieved if $\sigma(\mathbf{a}) = \sigma(\sim\mathbf{a}) = 0$. Also, in epistemic state σ, **a** is more believed than **b** if $\sigma(\sim\mathbf{a}) > \sigma(\sim\mathbf{b}) > 0$, and **a** is more disbelieved than **b** if $\sigma(\mathbf{a})>\sigma(\mathbf{b})>0$.

In Spohn's epistemic-belief theory, a valuation for s is a function $\sigma:\mathcal{W}_s\to\mathbb{N}$. Zero valuations do not exist, i.e., all valuations are nonzero. Also, all valuations are proper.

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

Combination. In Spohn's theory, combination is simply pointwise addition followed by normalization [30, 16]. If $\rho\in\mathcal{V}_r$, and $\sigma\in\mathcal{V}_s$, then their *combination*, denoted by $\rho\oplus\sigma$, is the valuation for $r\cup s$ given by

$$(\rho\oplus\sigma)(\mathbf{x}) = \rho(\mathbf{x}^{\downarrow r}) + \sigma(\mathbf{x}^{\downarrow s}) - K \tag{4.10}$$

for all $\mathbf{x}\in\mathcal{W}_{r\cup s}$, where K is a constant defined as follows:

$$K = \text{MIN}\{\rho(\mathbf{x}^{\downarrow r}) + \sigma(\mathbf{x}^{\downarrow s}) \mid \mathbf{x}\in\mathcal{W}_{r\cup s}\}.$$

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 ι_s for $\mathcal{N}_s\cup\{\zeta_s\}$ is given by $\iota_s(\mathbf{x}) = 0$ for all $\mathbf{x}\in\mathcal{W}_s$. 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 $\sigma\in\mathcal{V}_s$, and suppose $X\in s$. The *marginal of σ for s–{X}*, denoted by $\sigma^{\downarrow(s-\{X\})}$, is the valuation for $s-\{X\}$ defined as follows:

$$\sigma^{\downarrow(s-\{X\})}(\mathbf{y}) = \text{MIN}\{\sigma(\mathbf{y},\mathbf{x}) \mid \mathbf{x}\in\mathcal{W}_X\} \tag{4.11}$$

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

The above definition of marginalization follows from condition (D2) in the definition of a disbelief function since a proposition $\{y\}$ about variables in s–{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 (4.11) 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 [16].

Removal. In Spohn's theory, removal is subtraction followed by normalization [16]. Suppose $\sigma \in \mathcal{V}_s$, and $\rho \in \mathcal{N}_r \cup \mathcal{Z}_r$. Then the normal valuation resulting from the removal of ρ from σ, denoted by $\sigma \ominus \rho$, is given by

$$(\sigma \ominus \rho)(\mathbf{x}) = \sigma(\mathbf{x}^{\downarrow s}) - \rho(\mathbf{x}^{\downarrow r}) - K \tag{4.12}$$

for all $\mathbf{x} \in \mathcal{W}_{r \cup s}$, where K is a constant given by

$$K = \mathrm{MIN}\{\sigma(\mathbf{x}^{\downarrow s}) - \rho(\mathbf{x}^{\downarrow r}) \mid \mathbf{x} \in \mathcal{W}_{r \cup s}\}.$$

K is the normalization constant that ensures $\sigma \ominus \rho$ is a normal valuation. It can be easily shown that Axioms R1, R2, and CR hold. Suppose $\rho \in \mathcal{N}_r \cup \mathcal{Z}_r$. Since every normal valuation is positive normal, $\iota_\rho = \iota_r$.

4.3.4 Zadeh's possibility theory

In this section, we describe how Zadeh's possibility theory [32, 2] fits in the framework of valuation-based systems. More precisely, we define valuations, normal valuations, proper valuations, combination, marginalization, and removal.

The basic representational unit in Zadeh's possibility theory is called a possibility function.

Possibility Function. A *possibility function π for s* is a function π: $2^{\mathcal{W}_s} \to \mathbb{R}^+$ such that

(S1). there exists a configuration $\mathbf{x} \in \mathcal{W}_s$ such that $\pi(\{\mathbf{x}\}) = 1$; and

(S2). $\pi(\mathbf{a}) = \mathrm{MAX}\{\pi(\{\mathbf{x}\}) \mid \mathbf{x} \in \mathbf{a}\}$ for all $\mathbf{a} \in 2^{\mathcal{W}_s}$.

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 $\pi(\mathbf{a}) = 1$, and **a** is not possible in state π if and only if $\pi(\mathbf{a}) < 1$. 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. $\pi(\mathbf{a})$ can be interpreted as the degree to which proposition **a** is possible, and $1-\pi(\mathbf{a})$ can be interpreted as the degree to which proposition **a** is not possible, i.e., **a** is more possible than **b** if $\pi(\mathbf{a}) > \pi(\mathbf{b})$ and conversely, **a** is more impossible than **b** if $\pi(\mathbf{a}) < \pi(\mathbf{b}) < 1$.

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

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

Suppose σ is a valuation for s. We say σ is *normal* if and only if $\mathrm{MAX}\{\sigma(\mathbf{x}) \mid \mathbf{x} \in \mathcal{W}_s\} = 1$. 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 $\sigma \in \mathcal{V}_s$. Suppose $K = \mathrm{MAX}\left\{\rho(\mathbf{x}^{\downarrow r})\sigma(\mathbf{x}^{\downarrow s}) \mid \mathbf{x} \in \mathcal{W}_{r \cup s}\right\}$. The *combination of ρ and σ*, denoted by $\rho \oplus \sigma$, 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).

$$(\rho\oplus\sigma)(\mathbf{x}) = \begin{cases} K^{-1}\,\rho(\mathbf{x}^{\downarrow r})\sigma(\mathbf{x}^{\downarrow s}) & \text{if } K > 0 \\ 0 & \text{if } K = 0 \end{cases} \tag{4.13}$$

for all $\mathbf{x} \in \mathcal{W}_{r\cup s}$. If K = 0, $\rho\oplus\sigma = \zeta_{r\cup s}$. If K > 0, then 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 (4.13). The identity ι_s for $\mathcal{N}_s\cup\{\zeta_s\}$ is given by $\iota_s(\mathbf{x}) = 1$ for all $\mathbf{x} \in \mathcal{W}_s$. Suppose $\sigma \in \mathcal{N}_s$. An identity δ_σ for σ in $\mathcal{N}_s$ is a normal valuation for s such that $\delta_\sigma(\mathbf{x}) = 1$ if $\sigma(\mathbf{x}) > 0$, and $\delta_\sigma(\mathbf{x}) \in r$ if $\sigma(\mathbf{x}) = 0$, where r is any real number in the interval [0, 1]. Suppose $\sigma \in \mathcal{N}_s$. Notice that σ is *positive normal* if and only if $\sigma(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{W}_s$.

Marginalization. Suppose $\sigma \in \mathcal{V}_s$, and $X \in s$. The *marginal of σ for s–{X}*, denoted by $\sigma^{\downarrow(s-\{X\})}$, is the valuation for s–{X} defined as follows:

$$\sigma^{\downarrow(s-\{X\})}(\mathbf{y}) = \text{MAX}\{\sigma(\mathbf{y},\mathbf{x}) \mid \mathbf{x} \in \mathcal{W}_X\} \tag{4.14}$$

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

The above definition of marginalization follows from condition (S2) in the definition of a possibility function since a proposition $\{\mathbf{y}\}$ about variables in s–{X} is the same as proposition $\{\mathbf{y}\}\times\mathcal{W}_X$ 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 $\sigma \in \mathcal{V}_s$, $\rho \in \mathcal{N}_r\cup\mathcal{Z}_r$. Suppose K = $\text{MAX}\left\{\sigma(\mathbf{x}^{\downarrow s})/\rho(\mathbf{x}^{\downarrow r}) \mid \mathbf{x} \in \mathcal{W}_{r\cup s} \text{ such that } \rho(\mathbf{x}^{\downarrow r}) > 0\right\}$. Then the valuation resulting from the removal of ρ from σ, denoted by $\sigma\ominus\rho$, is given by,

$$(\sigma\ominus\rho)(\mathbf{x}) = \begin{cases} K^{-1}\,\sigma(\mathbf{x}^{\downarrow s})/\rho(\mathbf{x}^{\downarrow r}) & \text{if } K > 0, \text{ and } \rho(\mathbf{x}^{\downarrow r}) > 0 \\ 0 & \text{if } K = 0 \text{ or } \rho(\mathbf{x}^{\downarrow r}) = 0 \end{cases} \tag{4.15}$$

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

If K > 0, then K is the normalization constant that ensures $\sigma\ominus\rho$ is a normal valuation. It can be easily shown that Axioms R1, R2, and CR hold. Suppose $\rho \in \mathcal{N}_r\cup\mathcal{Z}_r$. The identity ι_ρ for ρ defined in Axiom R2 is the normal valuation for r such that $\iota_\rho(\mathbf{x}) = 1$ if $\rho(\mathbf{x}) > 0$, and $\iota_\rho(\mathbf{x}) = 0$ if $\rho(\mathbf{x}) = 0$.

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 ι_s such that $\sigma\oplus\iota_s = \sigma$ for all $\sigma \in \mathcal{P}_s\cap\mathcal{N}_s$. (The requirement that ι_s is proper normal is simply so that ι_s represents coherent knowledge.) In words, regardless of our current knowledge σ, ι_s does not add any thing new. Notice that in the VBS terminology, this simply means that ι_s is the unique identity for the semigroup $\mathcal{N}_s\cup\mathcal{Z}_s$.

As we saw in Section 4.2, complete ignorance has certain properties. First, $\iota_r\oplus\iota_s = \iota_{r\cup s}$, i.e., the combination of complete ignorance about variables in r and complete ignorance about variables in s results in complete ignorance about variables in $r\cup s$. Second, if $r\subseteq s$, then $\iota_s^{\downarrow r} = \iota_r$. In words, if we are completely ignorant about variables in s, and we disregard variables in s – 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 $\iota_s(\mathbf{x}) = K^{-1}$ for all $\mathbf{x} \in \mathcal{W}_s$, where $K = |\mathcal{W}_s|$, i.e., ι_s is the equally likely probability distribution for s. In D-S theory of belief functions, complete ignorance is represented by the commonality function $\iota_s(\mathbf{a}) = 1$ for all $\mathbf{a} \subseteq \mathcal{W}_s$. This commonality function corresponds to the bpa function that assigns probability 1 to the frame $\mathcal{W}_s$, and consequently, probabilities 0 to all subsets of $\mathcal{W}_s$. In Spohn's epistemic-belief theory, complete ignorance about variables in s is the disbelief function ι_s such that $\iota_s(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{W}_s$. Finally, in Zadeh's possibility theory, complete ignorance about variables in s is the possibility function ι_s given by $\iota_s(\mathbf{x}) = 1$ for all $\mathbf{x} \in \mathcal{W}_s$.

Contextual Ignorance. Suppose σ is a proper normal valuation for s. Suppose σ represents our current state of knowledge about variables in s. We will define *ignorance in context of* σ as a proper normal valuation δ_σ such that $\sigma \oplus \delta_\sigma = \sigma$. Unlike complete ignorance ι_s, ignorance in context of σ, δ_σ may not be unique. If σ is positive proper normal, then δ_σ is unique and same as ι_s. If σ is not positive proper normal, then δ_σ is not unique. For example, in probability theory, consider a probability distribution σ for {X, Y} such that $\sigma(x, y) = 0$, $\sigma(x, \sim y) = 0$, $\sigma(\sim x, y) = 0.2$, $\sigma(\sim x, \sim y) = 0.8$. Notice that σ is not positive proper normal. Then δ_σ is any probability distribution for {X, Y} such that $\delta_\sigma(x, y) = p$, $\delta_\sigma(x, \sim y) = q$, $\delta_\sigma(\sim x, y) = r$, $\delta_\sigma(\sim x, \sim y) = r$, where $p \geq 0$, $q \geq 0$, $r > 0$, and $p + q + 2r = 1$.

Contextual ignorance arises when we marginalize the heads of conditionals. From statement (vi) of Theorem 4.1, we know that if *a* and *b* are disjoint subsets, $\sigma(b|a)$ is a conditional for *b* given *a*, and we marginalize variables in *b* out of $\sigma(b|a)$, then the result $\sigma(b|a)^{\downarrow a}$ is ignorance in context of $\sigma(a)$. What this means is that the conditional $\sigma(b|a)$ has no additional information regarding variables in *a* given that we have information $\sigma(a)$.

4.5 Conclusion

It is commonly believed that the D-S belief-function theory is better able to represent ignorance than, for example, probability theory. As we have argued, this is not true. 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.

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.

References

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

[2] Dubois, D., Prade, H. (1988), *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, Plenum Press, New York, NY.

[3] Goldszmidt, M., Pearl, J. (1992), Rank-based systems: A simple approach to belief revision, belief update, and reasoning about evidence and actions, *Proceedings of the Third International Conference on Knowledge Representation and Reasoning,* Cambridge, MA.

[4] Goldszmidt, M., J. Pearl (1992), Stratified rankings for causal modeling, *Proceedings of the Fourth International Workshop on Nonmonotonic Reasoning,* 99–110, Plymouth, VT.

[5] Parsons, S., Saffiotti, A. (1993), Integrating uncertainty handling formalisms in distributed artificial intelligence, in Clarke, M., Kruse, R., Moral, S. (eds.), *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, 304–309, Springer-Verlag, Berlin.

[6] Pearl, J. (1991), Probabilistic semantics for nonmonotonic reasoning: A survey, in Cummins, R., Pollock, J. (eds.), *Philosophy and Artificial Intelligence: Essays at the Interface*, 157–187, Bradford Press/MIT Press, Cambridge, MA.

[7] Petrich, M. (1973), *Introduction to Semigroups*, Charles E. Merrill Publishing Co., Columbus, OH.

[8] Ruspini, E. H. (1991), On the semantics of fuzzy logic, *International Journal of Approximate Reasoning*, vol. 5 (1), 45–88.

[9] Shafer, G. (1976), *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ.

[10] Shafer, G. (1982), Belief functions and parametric models (with discussion), *Journal of the Royal Statistical Society, Series B*, vol. 44 (3), 322–352.

[11] Shafer, G. (1987), Belief functions and possibility measures, in Bezdek, J. (ed.), *The Analysis of Fuzzy Information*, vol. 1, 51–84, CRC Press, Boca Raton, FL.

[12] Shafer, G. (1990), Perspectives on the theory and practice of belief functions, *International Journal of Approximate Reasoning*, vol. 4 (5–6), 323–362.

[13] Shafer, G., Shenoy, P. P. (1988), Bayesian and belief-function propagation, *Working Paper No. 192*, School of Business, University of Kansas, Lawrence, KS.

[14] Shenoy, P. P. (1989), A valuation-based language for expert systems, *International Journal of Approximate Reasoning*, vol. 3 (5), 383–411.

[15] Shenoy, P. P. (1990), Valuation-based systems for propositional logic, in Ras, Z. W., Zemankova, M., Emrich, M. L. (eds.), *Methodologies for Intelligent Systems*, vol. 5, 305–312, North-Holland, Amsterdam.

[16] Shenoy, P. P. (1991), On Spohn's rule for revision of beliefs, *International Journal of Approximate Reasoning*, vol. 5 (2), 149–181.

[17] Shenoy, P. P. (1991), Valuation-based systems for discrete optimization, in Bonissone, P. P., Henrion, M., Kanal, L. N., Lemmer, J. F. (eds.), *Uncertainty in Artificial Intelligence, 6*, 385–400, North-Holland, Amsterdam.

[18] Shenoy, P. P. (1992), Valuation-based systems: A framework for managing uncertainty in expert systems, in Zadeh, L. A., Kacprzyk, J. (eds.), *Fuzzy Logic for the Management of Uncertainty*, 83–104, John Wiley & Sons, New York, NY.

[19] Shenoy, P. P. (1992), Valuation-based systems for Bayesian decision analysis, *Operations Research*, vol. 40 (3), 463–484.

[20] Shenoy, P. P. (1992), Using possibility theory in expert systems, *Fuzzy Sets and Systems*, vol. 52 (2), 129–142.

[21] Shenoy, P. P. (1993), A new method for representing and solving Bayesian decision problems, in Hand, D. J., (ed.), *Artificial Intelligence Frontiers in Statistics: AI and Statistics III*, 119–138, Chapman & Hall, London.

[22] Shenoy, P. P. (1994), Conditional independence in valuation-based systems, *International Journal of Approximate Reasoning*, vol. 10 (3), 203–234.

[23] Shenoy, P. P. (1994), Using Dempster-Shafer's belief-function theory in expert systems, in Yager, R. R., Fedrizzi, M., Kacprzyk, J. (eds.), *Advances in the Dempster-Shafer Theory of Evidence*, 395–414, John Wiley and Sons, New York, NY.

[24] Shenoy, P. P. (1994), Consistency in valuation-based systems, *ORSA Journal on Computing*, vol. 6 (3), in press.

[25] Shenoy, P. P. (1994), A comparison of graphical techniques for decision analysis, *European Journal of Operational Research*, vol. 78 (1), 1–21.

[26] Shenoy, P. P., G. Shafer (1988), Constraint propagation, *Working Paper No. 208*, School of Business, University of Kansas, Lawrence, KS.

[27] Shenoy, P. P., Shafer, G. (1990), Axioms for probability and belief-function propagation, in Shachter, R. D., Levitt, T. S., Lemmer, J. F., Kanal, L. N. (eds.), *Uncertainty in Artificial Intelligence, 4*, 169–198, North-Holland, Amsterdam. Reprinted in: Shafer, G., Pearl, J. (eds.), *Readings in Uncertain Reasoning*, 575–610, 1990, Morgan Kaufmann, San Mateo, CA.

[28] Smets, P. (1988), Belief functions, in Smets, P., Mamdani, E. H., Dubois, D., Prade, H. (eds.), *Non-Standard Logics for Automated Reasoning*, 253–286, Academic Press, New York, NY.

[29] Smets, P. (1991), What is Dempster-Shafer's model?, *Technical Report No. 91-20*, IRIDIA, University Libre de Bruxelles, Belgium

[30] Spohn, W. (1988), Ordinal conditional functions: A dynamic theory of epistemic states, in Harper, W. L., Skyrms, B. (eds.), *Causation in Decision, Belief Change, and Statistics*, vol. 2, 105–134, D. Reidel, Dordrecht, Holland.

[31] Spohn, W. (1990), A general non-probabilistic theory of inductive reasoning, in Shachter, R. D., Levitt, T. S., Lemmer, J. F., Kanal, L. N. (eds.), *Uncertainty in Artificial Intelligence, 4*, 149–158, North-Holland, Amsterdam.

[32] Zadeh, L. A. (1979), A theory of approximate reasoning, in Ayes, J. E., Mitchie, D., Mikulich, L. I. (eds.), *Machine Intelligence*, vol. 9, 149–194, Ellis Horwood, Chichester, U.K.