

Conditional Independence in Valuation-Based Systems

Prakash P. Shenoy[†]

*School of Business
University of Kansas
Lawrence, Kansas*

ABSTRACT

This paper introduces the concept of conditional independence in valuation-based systems (VBS). VBS is an axiomatic framework capable of representing many different uncertainty calculi. We define conditional independence in terms of factorization of the joint valuation. The definition of conditional independence in VBS generalizes the corresponding definition in probability theory. Besides probability theory, our definition applies also to Dempster-Shafer's belief-function theory, Spohn's epistemic-belief theory, and Zadeh's possibility theory. In fact, it applies to any uncertainty calculi that fit in the VBS framework. We prove that our definition of conditional independence satisfies many of the usual properties associated with it. In particular, it satisfies Pearl and Paz's graphoid axioms.

KEY WORDS: *Conditional independence, valuation-based systems, factorization, graphoid axioms*

1. INTRODUCTION

The concept of conditional independence between two subsets of variables given a third has been extensively studied in probability theory [1, 2, 3, 4, 5, 6, 7]. The concept of conditional independence in probability theory has been interpreted in terms of relevance. If r , s , and t are subsets of variables, then to say that r and s are conditionally independent given t , means that the conditional distribution of r , given any values of s and t , is governed by the value of t alone—further information about the value of s is irrelevant.

The concept of conditional independence for variables has also been studied in Spohn's epistemic-belief theory [8, 9]. However, the concept of conditional independence for *variables*

[†]Address correspondence to Professor Prakash P. Shenoy, School of Business, University of Kansas, Summerfield Hall, Lawrence, KS 66045-7585, USA.

has not been studied in Dempster-Shafer’s theory of belief functions [10, 11] or in Zadeh’s possibility theory [12, 13].¹

An axiomatic framework that unifies various uncertainty calculi is that of valuation-based systems [20, 21, 24]. In valuation-based systems (VBS), knowledge about a set of variables is represented by a valuation for that set of variables. There are three operations in VBS that are used to make inferences. These are called combination, marginalization, and removal. Combination represents aggregation of knowledge. Marginalization represents coarsening of knowledge. And removal represents disaggregation of knowledge.

The VBS framework is able to uniformly represent probability theory, Dempster-Shafer’s belief-function theory, Spohn’s epistemic-belief theory, and Zadeh’s possibility theory. In this paper, we develop the notion of conditional independence for variables in the VBS framework. One advantage of this generality is that all results developed here apply uniformly to all uncertainty calculi that fit in the VBS framework. Thus the results described in this paper apply to, for example, probability theory, Dempster-Shafer’s belief-function theory, Spohn’s epistemic-belief theory, and Zadeh’s possibility theory.

What does it mean for two subsets of variables to be conditionally independent given a third subset? Conditional independence can be described in terms of factorization of the joint valuation. Suppose r , s , and t are disjoint subsets of variables. Suppose τ is a valuation for $r \cup s \cup t$. We say r and s are conditionally independent given t with respect to τ if and only if the valuation τ factors into two valuations, one whose domain involves variables in $r \cup t$, and the other whose domain involves only variables in $s \cup t$.

The conditional independence relation between subsets of variables in probability theory satisfies many different properties. Pearl and Paz [4] have isolated a subset of these properties called the “graphoid axioms.” The graphoid axioms are important because they are also satisfied by many ternary relations besides probabilistic conditional independence. In this paper we show that the definition of conditional independence we propose in the VBS framework satisfies the graphoid axioms.

An outline of this paper is as follows. In section 2, we describe the VBS framework. The VBS framework was described earlier in [20, 21, 24]. In this paper we extend the framework by defining two new sets of valuations called normal, and positive normal. As we will see, the concept of normal valuations is required for the definition of conditional independence, and the concept of positive normal valuations is required to prove the intersection property of conditional independence. Also, we introduce a new operation called removal. The removal operation is required for the definition of conditional valuations. Many of the properties of conditional

¹ Dempster [10], Shafer [11, 14, 15, 16, 17], and Smets [18] have defined independence for belief functions, but not for *variables* on which belief functions are defined. Shafer [11] has defined independence for frames of discernment, a concept further studied by Shafer, Shenoy and Mellouli [19]. Belief functions in belief-function theory are analogs of probability functions in probability theory.

independence are stated using conditional valuations. The exposition in this section is quite abstract. The reader may want to glance ahead at Sections 4–7 for specific examples of each definition in the VBS framework.

In section 3, we define conditional independence for sets of variables. We show that this definition satisfies some well-known properties that have been stated by Dawid [1, 22], Spohn [2], Lauritzen [3], Smith [5], and Pearl and Paz [4] in the context of probability theory. Using Pearl and Paz’s terminology, the conditional independence relation in VBS is a graphoid.

In section 4, we show how probability theory fits in the VBS framework. In particular, we define valuations, zero valuations, proper valuations, normal valuations, positive normal valuations, combination, marginalization, and removal. We also verify that all axioms and assumptions made in Section 2 are satisfied by our definitions.

In section 5, we show how Dempster-Shafer’s theory of belief functions fits in the VBS framework. In section 6, we show how Spohn’s epistemic-belief theory fits in the VBS framework. In section 7, we show how Zadeh’s possibility theory fits in the VBS framework. Finally, in section 8, we make some concluding remarks.

2. THE VALUATION-BASED SYSTEMS FRAMEWORK

In this section, we describe the valuation-based systems (VBS) framework. In VBS, 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.

The VBS framework is described in [20, 21, 24]. The motivation there was to describe a local computational method for computing marginals of the joint valuation. In this paper, we embellish the VBS framework by introducing two new sets of valuations called normal, and positive normal, and by introducing a new operation called removal. Our motivation here is to define conditional independence and describe its properties.

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

Valuations. For each $s \subseteq \mathcal{X}$, there is a set \mathcal{V}_s . We call the elements of \mathcal{V}_s *valuations for s*. Let \mathcal{V} denote $\cup\{\mathcal{V}_s \mid s \subseteq \mathcal{X}\}$, 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 . In probability theory, e.g., a valuation for s is a function from the frame for s to the non-negative real numbers.

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 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, $Z = \emptyset$. We call valuations in $\mathcal{V} - Z$ *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. In probability theory, for example, a zero valuation is a function that is identically zero. The concept of zero valuations is important in the theory of consistent knowledge-based systems [23].

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.

The concept of proper valuations has substance (i.e., \mathcal{P}_s is a proper subset of $\mathcal{V}_s - \{\zeta_s\}$) only in Dempster-Shafer's belief-function theory. In Dempster-Shafer's belief-function theory, a valuation for s is a function from the power set of the frame for s to the non-negative real numbers, and a proper valuation is an unnormalized commonality function. This is explained in detail in Section 5. In probability theory, Spohn's epistemic-belief theory, and Zadeh's possibility theory, $\mathcal{P}_s = \mathcal{V}_s - \{\zeta_s\}$. Proper valuations play no role either in the definitions, 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. In probability theory, e.g., a normal valuation is a function whose values add to 1.

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. For example, in probability theory, a proper normal valuation is a probability distribution function, and in Dempster-Shafer's belief-function theory, a proper normal valuation is a commonality function.

Combination.² We assume there is a mapping $\oplus: \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{N} \cup Z$, called *combination*, that satisfies the following four axioms:

² The definition of combination as a mapping $\oplus: \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{N} \cup Z$ is a slight departure from our earlier definition in [24] as a mapping $\oplus: \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$. The motivation in [24] was to describe a framework for computational purposes. From this perspective, it is wise to postpone normalization to the very end when we have drawn an inference from the computational results. Here, we include normalization in the definition of combination for semantical purposes. Our objective here is not to describe a computational strategy, but to describe a semantically sound theory. Without normalization, we cannot prove the theorems that characterize conditional independence as we do in Section 3.

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 $r \cup s$. In probability theory, e.g., combination corresponds to pointwise multiplication followed by normalization (see Section 4 for a precise definition).

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

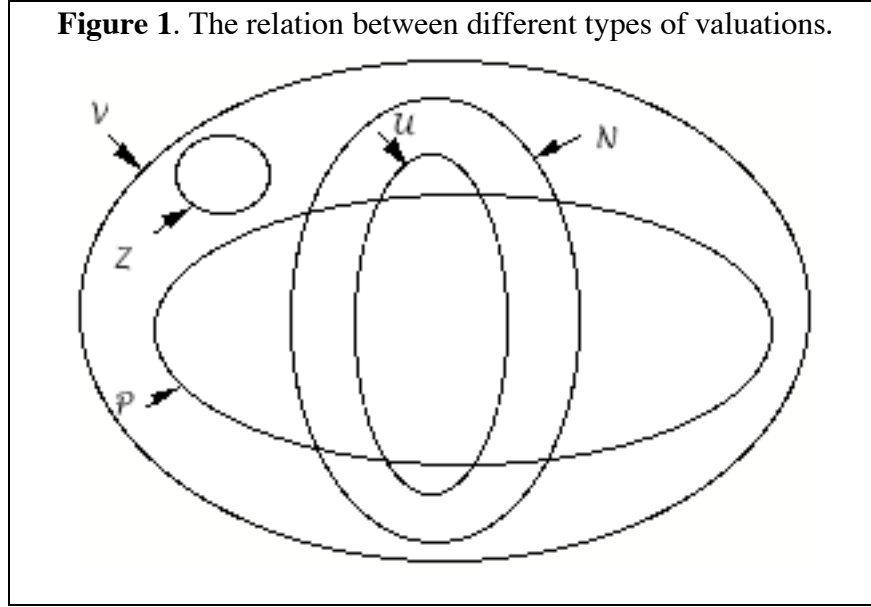
An implication of Axioms C1, C2, and C3 is that the set $N_s \cup \{\zeta_s\}$ together with the combination operation \oplus is a commutative semigroup [25]. (If zero valuations do not exist, then $N_s \cup \{\zeta_s\} = N_s$.) If zero valuations exist, then Axiom C4 defines the valuation ζ_s as the zero of the semigroup $N_s \cup \{\zeta_s\}$.

Identity Valuations. We assume that, for each $s \subseteq \mathcal{X}$, the commutative semigroup $N_s \cup \{\zeta_s\}$ has an identity denoted by ι_s (**Axiom C5**). In other words, there exists $\iota_s \in N_s \cup \{\zeta_s\}$ such that for each $\sigma \in N_s \cup \{\zeta_s\}$, $\sigma \oplus \iota_s = \sigma$. Notice that a commutative semigroup may have at most one identity. From Axiom C4, it follows that $\iota_s \neq \zeta_s$, therefore $\iota_s \in N_s$. Intuitively, identity valuations represent knowledge that is completely vacuous, i.e., they have no substantive content. For example, in probability theory ι_s is the equiprobable probability distribution for s , and in Dempster-Shafer's belief-function theory, ι_s is the vacuous commonality function for s .

It follows from Axiom C5 that for each $s \subseteq \mathcal{X}$, and for each $\sigma \in N_s \cup \{\zeta_s\}$, there exists at least one identity for it in $N_s \cup \{\zeta_s\}$, i.e., there exists a $\delta_\sigma \in N_s \cup \{\zeta_s\}$ such that $\sigma \oplus \delta_\sigma = \sigma$. For example, ι_s is an identity in $N_s \cup \{\zeta_s\}$ for each element of $N_s \cup \{\zeta_s\}$. A valuation in $N_s \cup \{\zeta_s\}$ may have more than one identity in $N_s \cup \{\zeta_s\}$. For example, Axiom C4 states that every element of $N_s \cup \{\zeta_s\}$ is an identity for ζ_s in $N_s \cup \{\zeta_s\}$. Notice that if $\sigma \in N_s$, then $\delta_\sigma \in N_s$ (Proof: If $\delta_\sigma = \zeta_s$, then $\sigma \oplus \delta_\sigma = \sigma \oplus \zeta_s = \zeta_s \neq \sigma$, contradicting the fact that δ_σ is an identity for σ). Also, notice that ι_s has only one identity in N_s , namely itself (Proof: If there exists $\sigma \in N_s \cup \{\zeta_s\}$, $\sigma \neq \iota_s$, such that $\iota_s \oplus \sigma = \iota_s$, then this contradicts the fact that ι_s is an identity for σ).

Positive Normal Valuations. Let \mathcal{U}_s denote the subset of N_s consisting of all valuations in N_s that have unique identities in N_s . 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 only holds for positive normal valuations (as shown in the next section). In probability theory, e.g., positive normal valuations correspond to strictly positive probability distributions. Figure 1 shows the relation between different types of valuations.



Valuations for the Empty Set. We assume that the set N_\emptyset consists of exactly one element (**Axiom C6**).³ This axiom implies that $\mathcal{U}_\emptyset = N_\emptyset = \{\iota_\emptyset\}$ where ι_\emptyset is the identity valuation for the semigroup $N_\emptyset \cup \{\zeta_\emptyset\}$. In probability theory, e.g., ι_\emptyset corresponds to the constant 1.

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 $X \notin r$,

$$\text{and } X \in s. \text{ Then } (\rho \oplus \sigma)^{\downarrow((r \cup s) - \{X\})} = \rho \oplus (\sigma^{\downarrow(s-\{X\})}); \text{ and}$$

³ A similar axiom is stated in [26].

Axiom CM2 (*Combination and Marginalization 2*):⁴ 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 $\sigma \oplus \delta_{\sigma \downarrow r} = \sigma$.

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 . In probability theory, e.g., marginalization σ to $s - \{X\}$ corresponds to summing the values of σ over the frame for X (see Section 4 for a precise definition).

If we regard marginalization as a coarsening of a valuation by deleting variables, then Axiom M1 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 coherence of knowledge. An implication of Axiom M3 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 $(r \cup s) - \{X\}$. The following lemma is an easy consequence of Axiom CM1.

Lemma 2.1. Suppose $\rho \in \mathcal{V}_r$, and $\sigma \in \mathcal{V}_s$. Then $(\rho \oplus \sigma) \downarrow r = \rho \oplus \sigma \downarrow r \cap s$.

Proof of Lemma 2.1: $(\rho \oplus \sigma) \downarrow r = (\rho \oplus \sigma) \downarrow ((r \cup s) - (s - r)) = \rho \oplus \sigma \downarrow (s - (s - r)) = \rho \oplus \sigma \downarrow r \cap s$. ■

Axiom CM2 describes an important property of identity valuations. The following lemma states some implications of Axiom CM2.

Lemma 2.2. Suppose Axioms C1–C6, M1–M4, CM1, and CM2 hold. Then the following statements hold.

- (i). Suppose $\sigma \in \mathcal{V}_s$ and suppose $r \subseteq s$. $\sigma \in \mathcal{N}_s \cup \{\zeta_s\}$ if and only if $\sigma \oplus \iota_r = \sigma$.
- (ii). If $\sigma \in \mathcal{V}_s$ and $r \subseteq s$, then $\sigma \oplus \iota_r = \sigma \oplus \iota_\emptyset$.
- (iii). $\iota_s \oplus \iota_r = \iota_{r \cup s}$.
- (iv). If $r \subseteq s$, then $\iota_s \downarrow r = \iota_r$.

⁴ A similar axiom is stated in [26] and [27].

Proof of Lemma 2.2:

(i). (\Rightarrow): If $\sigma \in \mathcal{N}_s$, then $\sigma \oplus \iota_r = \sigma$ by Axiom CM2 since ι_r is an identity for $\sigma \downarrow^r$. If $\sigma = \zeta_s$, then $\sigma \oplus \iota_\emptyset = \sigma$ by Axiom C4.

(\Leftarrow): If $\sigma \oplus \iota_\emptyset = \sigma$, then, since $\sigma \oplus \iota_\emptyset \in \mathcal{N}_s \cup \{\zeta_s\}$ by definition of the combination operation, $\sigma \in \mathcal{N}_s \cup \{\zeta_s\}$.

(ii). $\sigma \oplus \iota_r = (\sigma \oplus \iota_r) \oplus \iota_\emptyset = (\sigma \oplus \iota_\emptyset) \oplus \iota_r = \sigma \oplus \iota_\emptyset$.

(iii). Suppose $\tau \in \mathcal{N}_{r \cup s}$. Then from Axiom CM2, $\tau \oplus (\iota_s \oplus \iota_r) = (\tau \oplus \iota_s) \oplus \iota_r = \tau \oplus \iota_r = \tau$. If $\tau = \zeta_{r \cup s}$, then from Axiom C4, $\tau \oplus (\iota_s \oplus \iota_r) = \zeta_{r \cup s} \oplus (\iota_r \oplus \iota_s) = \zeta_{r \cup s} = \tau$. Therefore $\iota_s \oplus \iota_r$ must be the identity for $\mathcal{N}_{r \cup s} \cup \{\zeta_{r \cup s}\}$, i.e., $\iota_s \oplus \iota_r = \iota_{r \cup s}$.

(iv). Suppose $\rho \in \mathcal{N}_r$, and suppose $r \subseteq s$. We need to show that $\rho \oplus \iota_s \downarrow^r = \rho$. $\rho \oplus \iota_s \downarrow^r = (\rho \oplus \iota_s) \downarrow^r = (\rho \oplus \iota_r \oplus \iota_{s-r}) \downarrow^r = (\rho \oplus \iota_{s-r}) \downarrow^r = \rho \oplus \iota_{s-r} \downarrow^\emptyset = \rho \oplus \iota_\emptyset = \rho$. ■

Axioms C1, C2, C3, M1, and CM1 make local computation of marginals possible. Suppose $\{\sigma_1, \dots, \sigma_m\}$ is a collection of valuations, and suppose $\sigma_i \in \mathcal{V}_{s_i}$. Suppose $\mathcal{X} = s_1 \cup \dots \cup s_m$, and suppose $X \in \mathcal{X}$. Suppose we wish to compute $(\sigma_1 \oplus \dots \oplus \sigma_m) \downarrow^{\{X\}}$. We can do so by successively deleting all variables but X from the collection of valuation $\{\sigma_1, \dots, \sigma_m\}$. Each time we delete a variable, we do a fusion operation defined as follows. Consider a set of k valuations ρ_1, \dots, ρ_k . Suppose $\rho_i \in \mathcal{V}_{r_i}$. Let $\text{Fus}_Y\{\rho_1, \dots, \rho_k\}$ denote the collection of valuations after fusing the valuations in the set $\{\rho_1, \dots, \rho_k\}$ with respect to variable $Y \in r_1 \cup \dots \cup r_k$. Then

$$\text{Fus}_Y\{\rho_1, \dots, \rho_k\} = \{\rho \downarrow^{(r-\{Y\})}\} \cup \{\rho_i \mid Y \notin r_i\}$$

where $\rho = \bigoplus\{\rho_i \mid Y \in r_i\}$, and $r = \bigcup\{r_i \mid Y \in r_i\}$. After fusion, the set of valuations is changed as follows. All valuations whose domains include Y are combined, and the resulting valuation is marginalized such that Y is eliminated from its domain. The valuations whose domains do not include Y remain unchanged. The following theorem describes the fusion algorithm, a method for computing $(\sigma_1 \oplus \dots \oplus \sigma_m) \downarrow^{\{X\}}$ using only local computations.

Theorem 2.1 [21]. Suppose $\{\sigma_1, \dots, \sigma_m\}$ is a collection of valuations such that $\sigma_i \in \mathcal{V}_{s_i}$. Suppose Axioms C1, C2, C3, M1, and CM1 hold. Let \mathcal{X} denote $s_1 \cup \dots \cup s_m$. Suppose $X \in \mathcal{X}$, and suppose $X_1 X_2 \dots X_{n-1}$ is a sequence covering all variables in $\mathcal{X} - \{X\}$. Then

$$(\sigma_1 \oplus \dots \oplus \sigma_m) \downarrow^{\{X\}} = \bigoplus \left\{ \text{Fus}_{X_{n-1}} \left\{ \dots \text{Fus}_{X_2} \left\{ \text{Fus}_{X_1} \left\{ \sigma_1, \dots, \sigma_m \right\} \right\} \right\} \right\}.$$

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 $\rho \in \mathcal{N}_r \cup \mathcal{Z}_r$. Then $\sigma \ominus \rho \in \mathcal{N}_{r \cup s} \cup \mathcal{Z}_{r \cup s}$;

Axiom R2 (Identity): For each $\rho \in \mathcal{N} \cup \mathcal{Z}$, there exists an identity for ρ in $\mathcal{N} \cup \mathcal{Z}$, denoted by, say, ι_ρ , such that $\rho \ominus \rho = \iota_\rho$; and

Axiom CR (Combination and Removal): 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. In probability theory, e.g., removing corresponds to pointwise division followed by normalization (see Section 4 for a precise definition).

Axioms R2 and CR 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. The following lemma describes some implications of Axioms R1, R2, and CR.

Lemma 2.3. Suppose $\pi, \theta \in \mathcal{V}$, and suppose $\rho \in \mathcal{N} \cup \mathcal{Z}$.

- (i). $(\pi \oplus \theta) \ominus \rho = (\pi \ominus \rho) \oplus \theta$.
- (ii). If $\sigma \in \mathcal{V}_s$, and $r \subseteq s$, then $\sigma \ominus \iota_r = \sigma \oplus \iota_\emptyset$.
- (iii). $[(\pi \oplus \rho) \ominus \rho] \oplus \rho = \pi \oplus \rho$.

Proof of Lemma 2.3:

- (i). $(\pi \oplus \theta) \ominus \rho = (\theta \oplus \pi) \ominus \rho = \theta \oplus (\pi \ominus \rho) = (\pi \ominus \rho) \oplus \theta$.
- (ii). Let $\sigma \in \mathcal{V}_s$, and $r \subseteq s$. Then, $\sigma \ominus \iota_r = (\sigma \ominus \iota_r) \oplus \iota_r = \sigma \oplus (\iota_r \ominus \iota_r) = \sigma \oplus \iota_\emptyset = \sigma \oplus \iota_r = \sigma \oplus \iota_\emptyset$.
- (iii). $[(\pi \oplus \rho) \ominus \rho] \oplus \rho = [\pi \oplus (\rho \ominus \rho)] \oplus \rho = (\pi \oplus \iota_\emptyset) \oplus \rho = \pi \oplus (\iota_\emptyset \oplus \rho) = \pi \oplus \rho$. ■

Suppose $\rho \in \mathcal{N} \cup \mathcal{Z}$. Define $\rho^{-1} = \iota_\emptyset \ominus \rho$. We call ρ^{-1} *the inverse of ρ* . If $\rho \in \mathcal{N}_r \cup \mathcal{Z}_r$, then $\rho^{-1} \in \mathcal{N}_r \cup \mathcal{Z}_r$. The following lemma describes two properties of inverses.

Lemma 2.4. Suppose $\sigma \in \mathcal{V}$, $\rho \in \mathcal{N} \cup \mathcal{Z}$.

(i). $\rho^{-1} \oplus \rho = \rho \oplus \rho^{-1} = \iota_\rho$.⁵

(ii). $\sigma \ominus \rho = \sigma \oplus \rho^{-1}$.

Proof of Lemma 2.4:

(i). $\rho \oplus \rho^{-1} = \rho \oplus (\iota_\emptyset \ominus \rho) = (\rho \ominus \rho) \oplus \iota_\emptyset = \rho \ominus \rho = \iota_\rho$.

(ii). $\sigma \ominus \rho = (\sigma \ominus \rho) \oplus \iota_\emptyset = (\sigma \oplus \iota_\emptyset) \ominus \rho = \sigma \oplus (\iota_\emptyset \ominus \rho) = \sigma \oplus \rho^{-1}$. ■

The following lemma states an important consequence of Axioms R1, R2, CR, and CM1.

Lemma 2.5. Suppose $\sigma \in \mathcal{V}_s$, $\rho \in \mathcal{N}_r \cup \mathcal{Z}_r$, $X \in s$, and $X \notin r$. Then

$$(\sigma \ominus \rho) \downarrow^{((r \cup s) - \{X\})} = \sigma \downarrow^{(s - \{X\})} \ominus \rho.$$

Proof of Lemma 2.5: Suppose $\sigma \in \mathcal{V}_s$, $\rho \in \mathcal{V}_r$, $X \in s$, and $X \notin r$. Then

$$(\sigma \ominus \rho) \downarrow^{((r \cup s) - \{X\})} = (\sigma \oplus \rho^{-1}) \downarrow^{((r \cup s) - \{X\})} = \sigma \downarrow^{(s - \{X\})} \oplus \rho^{-1} = \sigma \downarrow^{(s - \{X\})} \ominus \rho. \quad \blacksquare$$

Conditional Valuations. 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 the theory of conditional independence. 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 $\sigma \downarrow^{(a \cup b)} \ominus \sigma \downarrow^a$. Also, if $a = \emptyset$, let $\sigma(b)$ denote $\sigma(b|\emptyset)$.

The following lemma states some important properties of conditional valuations.

Lemma 2.6. 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). $\sigma(b|a) \in \mathcal{N}_{a \cup b}$.

Proof of Lemma 2.6:

⁵ If we assume that u is closed under combination, then statement (i) of Lemma 2.4 implies that (u, \oplus) is an Abelian (commutative) group [28]. One implication of this is that if $\rho, \sigma \in u$, then $(\rho \oplus \sigma)^{-1} = \rho^{-1} \oplus \sigma^{-1}$ and $(\rho^{-1})^{-1} = \rho$. Also, if $\pi \in v$, and $\theta, \rho \in u$, then $\pi \ominus (\theta \oplus \rho) = (\pi \ominus \theta) \ominus \rho$, and $\pi \ominus (\theta \ominus \rho) = (\pi \ominus \theta) \oplus \rho$.

- (i). $\sigma(a) = \sigma^{\downarrow a} \ominus \sigma^{\downarrow \emptyset} = \sigma^{\downarrow a} \ominus \iota_{\emptyset} = \sigma^{\downarrow a} \oplus \iota_{\emptyset} = \sigma^{\downarrow a}$.
- (ii). $\sigma(a) \oplus \sigma(b|a) = \sigma^{\downarrow a} \oplus (\sigma^{\downarrow(a \cup b)} \ominus \sigma^{\downarrow a}) = (\sigma^{\downarrow a} \ominus \sigma^{\downarrow a}) \oplus \sigma^{\downarrow(a \cup b)} = \iota_{\sigma^{\downarrow a}} \oplus \sigma^{\downarrow(a \cup b)} = \sigma^{\downarrow(a \cup b)} = \sigma(a \cup b)$.
- (iii). $\sigma(b|a) \oplus \sigma(c|a \cup b) = (\sigma^{\downarrow(a \cup b)} \ominus \sigma^{\downarrow a}) \oplus (\sigma^{\downarrow(a \cup b \cup c)} \ominus \sigma^{\downarrow(a \cup b)}) = \sigma^{\downarrow(a \cup b \cup c)} \oplus [\sigma^{\downarrow(a \cup b)} \ominus \sigma^{\downarrow(a \cup b)}] \ominus \sigma^{\downarrow a} = \sigma^{\downarrow(a \cup b \cup c)} \oplus \iota_{\sigma^{\downarrow(a \cup b)}} \ominus \sigma^{\downarrow a} = \sigma^{\downarrow(a \cup b \cup c)} \ominus \sigma^{\downarrow a} = \sigma(b \cup c|a)$.
- (iv). $\sigma(b|a)^{\downarrow(a \cup b')} = (\sigma^{\downarrow(a \cup b)} \ominus \sigma^{\downarrow a})^{\downarrow(a \cup b')} = (\sigma^{\downarrow(a \cup b)})^{\downarrow(a \cup b')} \ominus \sigma^{\downarrow a} = \sigma^{\downarrow(a \cup b')} \ominus \sigma^{\downarrow a} = \sigma(b'|a)$.
- (v). This follows directly from (iii) and (iv).
- (vi). $\sigma(b|a)^{\downarrow a} = (\sigma^{\downarrow(a \cup b)} \ominus \sigma^{\downarrow a})^{\downarrow a} = (\sigma^{\downarrow(a \cup b)})^{\downarrow a} \ominus \sigma^{\downarrow a} = \sigma^{\downarrow a} \ominus \sigma^{\downarrow a} = \sigma(a) \ominus \sigma(a) = \iota_{\sigma(a)}$.
- (vii). $\sigma(b|a)$ is either normal or zero. If $\sigma(b|a)$ is zero, then $\sigma(b|a) \oplus \sigma(a) = \zeta_{a \cup b} \neq \sigma(a \cup b)$ contradicting statement (ii). Therefore $\sigma(b|a)$ is normal. ■

3. CONDITIONAL INDEPENDENCE

In this section, we define conditional independence in terms of factorization of the joint valuation. Also, we show that this definition implies the well-known properties of conditional independence in probability theory [1, 2, 3] and in other domains [4, 5, 29].

The essence of conditional independence is as follows. Suppose r , s , and v are disjoint subsets. We say r and s are conditionally independent given v with respect to a valuation τ if and only if $\tau^{\downarrow(r \cup s \cup v)}$ factors into two valuations $\alpha_{r \cup v} \in \mathcal{V}_{r \cup v}$, and $\alpha_{s \cup v} \in \mathcal{V}_{s \cup v}$.

The definition of conditional independence is either objective or subjective depending on whether we have an objective or subjective measure of knowledge represented by valuation τ . In probability theory for example, in some cases, we start with an objective specification of a joint probability distribution of all variables. This joint probability distribution then serves as an objective measure of knowledge, and all statements of conditional independence are objective with respect to this state of knowledge. In other cases, however, we do not start always with a joint probability distribution. In such cases, the first task is to specify a joint probability distribution. To make this specification task simpler, we make assertions of conditional independence that are necessarily subjective. However, once we have a specification of a joint probability distribution (obtained either objectively or subjectively), all further statements of conditional independence are necessarily objective with respect to the joint probability distribution.

If τ is normal, statement (i) of Lemma 2.6 tells us that $\tau(a) = \tau^{\downarrow a}$. In this case, we will use the simpler and more intuitive conditional notation to denote the marginals, i.e., we will use, for example, $\tau(a)$ in place of $\tau^{\downarrow a}$.

Definition 3.1 (*Conditional Independence*). Suppose $\tau \in \mathcal{N}_w$, and suppose r , s , and v are disjoint subsets of w . We say r and s are conditionally independent given v with respect to τ , written as $r \perp_{\tau} s \mid v$, if and only if there exist $\alpha_{r \cup v} \in \mathcal{V}_{r \cup v}$, and $\alpha_{s \cup v} \in \mathcal{V}_{s \cup v}$ such that $\tau(r \cup s \cup v) = \alpha_{r \cup v} \oplus \alpha_{s \cup v}$.

When it is clear that all conditional independence statements are with respect to τ , we simply say ‘ r and s are conditionally independent given v ’ instead of ‘ r and s are conditionally independent given v with respect to τ ,’ and use the simpler notation $r \perp s \mid v$ instead of $r \perp_{\tau} s \mid v$. Also, if $v = \emptyset$, we say ‘ r and s are independent’ instead of ‘ r and s are conditionally independent given \emptyset ’ and use the simpler notation $r \perp s$ instead of $r \perp s \mid \emptyset$.

We make four observations about our definition of conditional independence. First, notice that $\alpha_{r \cup v}$ and $\alpha_{s \cup v}$ are arbitrary valuations, they need not be normal. τ is necessarily normal. Second, notice that we do not use the removal operation in the definition of conditional independence. If there is a removal operation as defined in the previous section, then we can characterize conditional independence in terms of conditionals. This is done in Lemma 3.1 below. Third, if $\tau \in \mathcal{N}_w$, and r and v are disjoint subsets of w , then $r \perp_{\tau} \emptyset \mid v$. This property is called *trivial conditional independence* by Geiger and Pearl [29]. Fourth, if $\tau \in \mathcal{N}_x$, $w \subseteq x$, and r and s are disjoint subsets of w , then $r \perp_{\tau} s$ if and only if $r \perp_{\tau(w)} s$.

The following lemma gives seven characterizations of conditional independence.

Lemma 3.1.⁶ Suppose $\tau \in \mathcal{N}_w$, and suppose r , s , and v are disjoint subsets of w . The following statements are equivalent.

- (i). $r \perp s \mid v$.
- (ii). $\tau(r \cup s \mid v) = \beta_{r \cup v} \oplus \beta_{s \cup v}$, where $\beta_{r \cup v} \in \mathcal{V}_{r \cup v}$, and $\beta_{s \cup v} \in \mathcal{V}_{s \cup v}$.
- (iii). $\tau(r \cup s \cup v) = \tau(v) \oplus \tau(r \mid v) \oplus \tau(s \mid v)$.
- (iv). $\tau(r \cup s \mid v) = \tau(r \mid v) \oplus \tau(s \mid v)$.
- (v). $\tau(r \cup s \cup v) \oplus \tau(v) = \tau(r \cup v) \oplus \tau(s \cup v)$.
- (vi). $\tau(r \cup s \cup v) = \tau(r \mid v) \oplus \tau(s \cup v)$.
- (vii). $\tau(r \mid s \cup v) = \tau(r \mid v) \oplus \iota_{\tau(s \cup v)}$.
- (viii). $\tau(r \mid s \cup v) = \alpha_{r \cup v} \oplus \iota_{\tau(s \cup v)}$, where $\alpha_{r \cup v} \in \mathcal{V}_{r \cup v}$.

⁶ The statements of Lemma 3.1 are analogs of corresponding statements in probability theory [1]. Our contribution here is in showing that these statements hold in the more general framework of VBS.

Proof of Lemma 3.1: We will prove that (i) implies (ii), (ii) implies (iii), ..., (vii) implies (i).

To prove (i) implies (ii), suppose $\tau(rUsUv) = \alpha_{rUv} \oplus \alpha_{sUv}$, where $\alpha_{rUv} \in \mathcal{V}_{rUv}$, and $\alpha_{sUv} \in \mathcal{V}_{sUv}$. Removing $\tau(v)$ from both sides of the preceding equality, we get $\tau(rUsUv) \ominus \tau(v) = \alpha_{rUv} \oplus \alpha_{sUv} \ominus \tau(v)$. Since $\tau(rUsUv) \ominus \tau(v) = \tau(rUs|v)$, and $[\alpha_{sUv} \ominus \tau(v)] \in \mathcal{V}_{sUv}$, we have proved (ii).

To prove (ii) implies (iii), suppose $\tau(rUs|v) = \beta_{rUv} \oplus \beta_{sUv}$, where $\beta_{rUv} \in \mathcal{V}_{rUv}$, and $\beta_{sUv} \in \mathcal{V}_{sUv}$. Adding $\tau(v)$ to both sides, we get $\tau(rUsUv) = \beta_{rUv} \oplus \beta_{sUv} \oplus \tau(v)$. Deleting variables in s from both sides, we get $\tau(rUv) = \beta_{rUv} \oplus \beta_{sUv}^{\downarrow v} \oplus \tau(v)$. Removing $\tau(v)$ from both sides, $\tau(r|v) = \beta_{rUv} \oplus \beta_{sUv}^{\downarrow v} \oplus \iota_{\tau(v)}$. Similarly, we can show that $\tau(s|v) = \beta_{rUv}^{\downarrow v} \oplus \beta_{sUv} \oplus \iota_{\tau(v)}$. Deleting variables in rUs from both sides of $\tau(rUsUv) = \beta_{rUv} \oplus \beta_{sUv} \oplus \tau(v)$, we get $\tau(v) = \beta_{rUv}^{\downarrow v} \oplus \beta_{sUv}^{\downarrow v} \oplus \tau(v)$. Now, $\tau(v) \oplus \tau(r|v) \oplus \tau(s|v) = \tau(v) \oplus [\beta_{rUv} \oplus \beta_{sUv}^{\downarrow v} \oplus \iota_{\tau(v)}] \oplus [\beta_{rUv}^{\downarrow v} \oplus \beta_{sUv} \oplus \iota_{\tau(v)}] = [\beta_{rUv} \oplus \beta_{sUv}] \oplus [\beta_{rUv}^{\downarrow v} \oplus \beta_{sUv}^{\downarrow v} \oplus \tau(v) \oplus \iota_{\tau(v)} \oplus \iota_{\tau(v)}] = [\beta_{rUv} \oplus \beta_{sUv}] \oplus [\beta_{rUv}^{\downarrow v} \oplus \beta_{sUv}^{\downarrow v} \oplus \tau(v)] = \tau(rUs|v) \oplus \tau(v) = \tau(rUsUv)$.

To prove (iii) implies (iv), suppose $\tau(rUsUv) = \tau(v) \oplus \tau(r|v) \oplus \tau(s|v)$. Removing $\tau(v)$ from both sides of the preceding equality, $\tau(rUsUv) \ominus \tau(v) = \tau(r|v) \oplus \tau(s|v) \oplus \iota_{\tau(v)}$, i.e., $\tau(rUs|v) = \tau(r|v) \oplus \tau(s|v)$.

To prove (iv) implies (v), suppose $\tau(rUs|v) = \tau(r|v) \oplus \tau(s|v)$. Adding $\tau(v) \oplus \tau(v)$ to both sides, we get $\tau(rUs|v) \oplus \tau(v) \oplus \tau(v) = \tau(r|v) \oplus \tau(s|v) \oplus \tau(v) \oplus \tau(v)$, i.e., $\tau(rUsUv) \oplus \tau(v) = \tau(rUv) \oplus \tau(sUv)$.

To prove (v) implies (vi), suppose $\tau(rUsUv) \oplus \tau(v) = \tau(rUv) \oplus \tau(sUv)$. Removing $\tau(v)$ from both sides, we get $\tau(rUsUv) = \tau(rUv) \oplus \tau(sUv) \ominus \tau(v) = \tau(r|v) \oplus \tau(sUv)$.

To prove that (vi) implies (vii), suppose $\tau(rUsUv) = \tau(r|v) \oplus \tau(sUv)$. Removing $\tau(sUv)$ from both sides of the equality, we get $\tau(rUsUv) \ominus \tau(sUv) = (\tau(r|v) \oplus \tau(sUv)) \ominus \tau(sUv)$, i.e., $\tau(r|sUv) = \tau(r|v) \oplus \iota_{\tau(sUv)}$.

To prove that (vii) implies (viii), notice that $\tau(r|v) \in \mathcal{V}_{rUv}$.

To prove that (viii) implies (i), suppose $\tau(r|sUv) = \alpha_{rUv} \oplus \iota_{\tau(sUv)}$, where $\alpha_{rUv} \in \mathcal{V}_{rUv}$. Adding $\tau(sUv)$ to both sides of the equality, we get $\tau(rUsUv) = \alpha_{rUv} \oplus \iota_{\tau(sUv)} \oplus \tau(sUv) = \alpha_{rUv} \oplus \tau(sUv)$. The result now follows from Definition 3.1. ■

The following corollary to Lemma 3.1 gives three characterizations of the independence relation.

Corollary 3.1. Suppose $\tau \in \mathcal{N}_w$, and suppose r and s are disjoint subsets of w . The following statements are equivalent.

- (i). $r \perp s$.
- (ii). $\tau(rUs) = \tau(r) \oplus \tau(s)$.
- (iii). $\tau(r|s) = \tau(r) \oplus \iota_{\tau(s)}$.

(iv). $\tau(r|s) = \rho \oplus_{\tau(s)}$, where $\rho \in \mathcal{V}_r$.

Proof of Corollary 3.1: Statement (ii) follows from statement (iii) (or (iv) or (v) or (vi)) of Lemma 3.1, statement (iii) follows from statement (vii) of Lemma 3.1, and statement (iv) follows from statement (viii) of Lemma 3.1. ■

Theorem 3.1 below states the symmetry property of conditional independence.

Theorem 3.1 (Symmetry). Suppose $\tau \in \mathcal{N}_w$, and suppose r, s , and v are disjoint subsets of w . $r \perp s | v$ if and only if $s \perp r | v$.

Proof of Theorem 3.1: The proof follows immediately from the definition of conditional independence and Axiom C3 (commutativity of combination). ■

Definition 3.2 generalizes Definition 3.1 for any number of subsets of variables.

Definition 3.2 (Joint Conditional Independence). Suppose $\tau \in \mathcal{N}_w$, and suppose r_1, \dots, r_n, v are disjoint subsets of w . We say r_1, \dots, r_n are *conditionally independent given v with respect to τ* , written as $\perp_{\tau} \{r_1, \dots, r_n\} | v$, if and only if there exist $\alpha_{r_i \cup v} \in \mathcal{V}_{r_i \cup v}$ for $i = 1, \dots, n$, such that $\tau(r_1 \cup \dots \cup r_n \cup v) = \alpha_{r_1 \cup v} \oplus \dots \oplus \alpha_{r_n \cup v}$.

Definition 3.2 is a generalization of Definition 3.1. Notice that $r \perp s | v$ if and only if $\perp \{r, s\} | v$. We know from probability theory that functions of independent random variables are independent. If X_1 and X_2 are independent random variables, then $f(X_1)$ and $g(X_2)$ are also independent random variables. More generally, if X_1, \dots, X_n are conditionally independent given X , $\{N_1, \dots, N_k\}$ is a partition of the set $\{X_1, \dots, X_n\}$, and Y_j is a function of the X_i in N_j , then Y_1, \dots, Y_k are conditionally independent given X . The following lemma makes an analogous statement.

Lemma 3.2.⁷ Suppose $\tau \in \mathcal{N}_w$, and suppose r_1, \dots, r_n, v are disjoint subsets of w . Suppose $\perp \{r_1, \dots, r_n\} | v$. Suppose $\{N_1, \dots, N_k\}$ is a partition of $\{1, \dots, n\}$, i.e., $N_i \cap N_j = \emptyset$ if $i \neq j$, and $N_1 \cup \dots \cup N_k = \{1, \dots, n\}$. Suppose $s_j \subseteq (\cup \{r_i | i \in N_j\})$, for $j = 1, \dots, k$. Then $\perp \{s_1, \dots, s_k\} | v$.

Proof of Lemma 3.2: From Definition 3.2, $\tau(r_1 \cup \dots \cup r_n \cup v) = \alpha_{r_1 \cup v} \oplus \dots \oplus \alpha_{r_n \cup v}$. Let $\beta_j = \oplus \{\alpha_{r_i \cup v} | i \in N_j\}$, $j = 1, \dots, k$. Since $\{N_1, \dots, N_k\}$ is a partition of $\{1, \dots, n\}$, $\tau(r_1 \cup \dots \cup r_n \cup v) = \beta_1 \oplus \dots \oplus \beta_k$. If we delete variables in $((\cup \{r_i | i \in N_1\}) - s_1) \cup \dots \cup ((\cup \{r_i | i \in N_k\}) - s_k)$ from both sides

⁷ An analogous statement is stated and proved in [19] in the context of qualitative conditional independence.

of the preceding equality, using Axiom CM1 we get $\tau(s_1 \cup \dots \cup s_k \cup v) = \beta_1^{\downarrow(s_1 \cup v)} \oplus \dots \oplus \beta_k^{\downarrow(s_k \cup v)}$. Therefore, $\perp\{s_1, \dots, s_k\} \mid v$. ■

The statement in the following theorem is called decomposition by Pearl [30]. It is a special case of Lemma 3.2.

Theorem 3.2 (Decomposition). Suppose $\tau \in \mathcal{N}_w$, suppose r, s, t , and v are disjoint subsets of w . If $r \perp s \cup t \mid v$, then $r \perp s \mid v$.

Proof of Theorem 3.2: The result follows directly from Lemma 3.2. ■

The following lemma gives three alternative characterizations of joint conditional independence in terms of binary conditional independence.

Lemma 3.4.⁸ Suppose $\tau \in \mathcal{N}_w$, and suppose r_1, \dots, r_n , and v are disjoint subsets of w . The following statements are equivalent.

- (i). $\perp\{r_1, \dots, r_n\} \mid v$.
- (ii). $\perp\{r_1, \dots, r_{n-1}\} \mid v$ and $(r_1 \cup \dots \cup r_{n-1}) \perp r_n \mid v$.
- (iii). $r_i \perp \cup\{r_j \mid j = 1, \dots, n, j \neq i\} \mid v$, for $i = 1, \dots, n$.
- (iv). $r_j \perp (r_1 \cup \dots \cup r_{j-1}) \mid v$ for $j = 2, \dots, n$.

Proof of Lemma 3.4: We will prove that (i) implies (ii), (ii) implies (iii), ..., (iv) implies (i).

That (i) implies (ii) follows directly from Lemma 3.2.

To prove (ii) implies (iii), we will prove (ii) implies (i), and (i) implies (iii). Suppose $\perp_{\tau}\{r_1, \dots, r_{n-1}\} \mid v$, and $(r_1 \cup \dots \cup r_{n-1}) \perp r_n \mid v$. Since $(r_1 \cup \dots \cup r_{n-1}) \perp r_n \mid v$, $\tau(r_1 \cup \dots \cup r_n \mid v) = \tau(r_1 \cup \dots \cup r_{n-1} \mid v) \oplus \tau(r_n \mid v)$. Since $\perp_{\tau}\{r_1, \dots, r_{n-1}\} \mid v$, $\tau(r_1 \cup \dots \cup r_{n-1} \cup v) = \alpha_{r_1 \cup v} \oplus \dots \oplus \alpha_{r_{n-1} \cup v}$, where $\alpha_{r_i \cup v} \in \mathcal{V}_{r_i \cup v}$, $i = 1, \dots, n-1$. Therefore, $\tau(r_1 \cup \dots \cup r_n \cup v) = \tau(r_1 \cup \dots \cup r_n \mid v) \oplus \tau(v) = \tau(r_1 \cup \dots \cup r_{n-1} \mid v) \oplus \tau(r_n \mid v) \oplus \tau(v) = \tau(r_1 \cup \dots \cup r_{n-1} \cup v) \oplus \tau(r_n \mid v) = \alpha_{r_1 \cup v} \oplus \dots \oplus \alpha_{r_{n-1} \cup v} \oplus \tau(r_n \mid v)$. By Definition 3.2, $\perp\{r_1, \dots, r_n\} \mid v$. That (i) implies (iii) follows directly from Lemma 3.2.

That (iii) implies (iv) follows directly from Lemma 3.2.

To show (iv) implies (i), suppose $r_i \perp (r_1 \cup \dots \cup r_{i-1}) \mid v$ for $i = 2, \dots, n$. We are given $\perp\{r_1, r_2\} \mid v$. It suffices to show that if $\perp\{r_1, \dots, r_{j-1}\} \mid v$, then $\perp\{r_1, \dots, r_j\} \mid v$. The proof of this latter assertion is similar to the proof given above to show (ii) implies (i). ■

⁸ The statements in Lemma 3.2 are analogs of corresponding statements in [19] in the context of qualitative conditional independence.

Theorem 3.3 states another property of conditional independence. This property is called weak union by Pearl [30].

Theorem 3.3 (*Weak Union*). Suppose $\tau \in \mathcal{N}_w$, and suppose $r, s, t,$ and v are disjoint subsets of w . If $r \perp s \cup t \mid v$, then $r \perp s \mid t \cup v$.

Proof of Theorem 3.3: Suppose $r \perp s \cup t \mid v$. Then by Definition 3.1, $\tau(r \cup s \cup t \cup v) = \alpha_{r \cup v} \oplus \alpha_{s \cup t \cup v}$, where $\alpha_{r \cup v} \in \mathcal{V}_{r \cup v}$, and $\alpha_{s \cup t \cup v} \in \mathcal{V}_{s \cup t \cup v}$. Therefore $\tau(r \cup s \cup t \cup v) = \alpha_{r \cup v} \oplus \alpha_{s \cup t \cup v} = (\alpha_{r \cup v} \oplus \alpha_{s \cup t \cup v}) \oplus \iota_t = (\alpha_{r \cup v} \oplus \iota_t) \oplus \alpha_{s \cup t \cup v}$. Since $\alpha_{r \cup v} \oplus \iota_t \in \mathcal{V}_{r \cup t \cup v}$, and $\alpha_{s \cup t \cup v} \in \mathcal{V}_{s \cup t \cup v}$, the result follows. ■

Theorem 3.4 states another property of conditional independence. This property is called contraction by Pearl [30].

Theorem 3.4 (*Contraction*). Suppose $\tau \in \mathcal{N}_w$, and suppose $r, s, t,$ and v are disjoint subsets of w . If $r \perp s \mid v$, and $r \perp t \mid s \cup v$, then $r \perp s \cup t \mid v$.

Proof of Theorem 3.4: Suppose $r \perp s \mid v$, and $r \perp t \mid s \cup v$. Therefore, $\tau(r \cup s \cup t \cup v) = \tau(r \cup s \cup v) \oplus \tau(t \mid r \cup s \cup v) = [\tau(r \mid v) \oplus \tau(s \cup v)] \oplus \tau(t \mid s \cup v) = \tau(r \mid v) \oplus [\tau(s \cup v) \oplus \tau(t \mid s \cup v)] = \tau(r \mid v) \oplus \tau(s \cup t \cup v)$. Therefore, by Definition 3.1, $r \perp s \cup t \mid v$. ■

The next theorem states a property of conditional independence that holds only if the joint valuation τ is positive normal. This property is called intersection by Pearl [30].

Theorem 3.5 (*Intersection*). Suppose $\tau \in \mathcal{U}_w$, and suppose $r, s, t,$ and v are disjoint subsets of w . If $r \perp s \mid t \cup v$, and $r \perp t \mid s \cup v$, then $r \perp s \cup t \mid v$.

Proof of Theorem 3.5: Suppose $r \perp s \mid t \cup v$, and $r \perp t \mid s \cup v$. Since $r \perp s \mid t \cup v$, by statement (vii) of Lemma 3.1, $\tau(r \mid s \cup t \cup v) = \tau(r \mid t \cup v) \oplus \iota_{\tau(s \cup t \cup v)}$. Since τ is positive normal, $\tau(s \cup t \cup v)$ is positive normal. Therefore $\iota_{\tau(s \cup t \cup v)} = \iota_{s \cup t \cup v} = \iota_s \oplus \iota_{t \cup v}$. Therefore, $\tau(r \mid s \cup t \cup v) = \tau(r \mid t \cup v) \oplus \iota_s \oplus \iota_{t \cup v} = \tau(r \mid t \cup v) \oplus \iota_s$. Similarly, since $r \perp t \mid s \cup v$, $\tau(r \mid s \cup t \cup v) = \tau(r \mid s \cup v) \oplus \iota_t$. Since the left hand sides of the preceding two equalities are the same, the right hand sides must be equal, i.e., $\tau(r \mid t \cup v) \oplus \iota_s = \tau(r \mid s \cup v) \oplus \iota_t$. Adding $\tau(t \cup v) \oplus \tau(s \cup v)$ to both sides of the preceding equality, we get $\tau(r \mid t \cup v) \oplus \iota_s \oplus \tau(t \cup v) \oplus \tau(s \cup v) = \tau(r \mid s \cup v) \oplus \iota_t \oplus \tau(t \cup v) \oplus \tau(s \cup v)$, i.e., $\tau(r \cup t \cup v) \oplus \tau(s \cup v) = \tau(r \cup s \cup v) \oplus \tau(t \cup v)$. Deleting s from both sides of the preceding equality, we get $\tau(r \cup t \cup v) \oplus \tau(v) = \tau(r \cup v) \oplus \tau(t \cup v)$, i.e., $r \perp t \mid v$. Earlier we had $\tau(r \mid s \cup t \cup v) = \tau(r \mid t \cup v) \oplus \iota_s = \tau(r \mid v) \oplus \iota_t \oplus \iota_s$. Therefore, $\tau(r \mid s \cup t \cup v) \oplus \tau(s \cup t \cup v) = [\tau(r \mid v) \oplus \iota_t \oplus \iota_s] \oplus \tau(s \cup t \cup v)$, i.e., $\tau(r \cup s \cup t \cup v) = \tau(r \mid v) \oplus \tau(s \cup t \cup v)$. Therefore, $r \perp s \cup t \mid v$. ■

Pearl and Paz [4] call a conditional independence relation that satisfies symmetry, decomposition, weak union, contraction, and intersection a *graphoid*. From Theorems 3.1–3.5, it follows that the definition of conditional independence in Definition 3.1 is a graphoid.

4. PROBABILITY THEORY

In this section, 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. Also, we show that all axioms made in Section 2 hold. First we start with 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 \mathbf{x} , \mathbf{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 \mathbf{x} is a configuration of s , then $\mathbf{x}^{\downarrow r}$ denotes the projection of \mathbf{x} to r . If $r = \emptyset$, then of course, $\mathbf{x}^{\downarrow r} = \blacklozenge$.

If \mathbf{x} is a configuration of r , \mathbf{y} is a configuration of s , and $r \cap s = \emptyset$, then there is a unique configuration \mathbf{z} of $r \cup s$ such that $\mathbf{z}^{\downarrow r} = \mathbf{x}$, and $\mathbf{z}^{\downarrow s} = \mathbf{y}$. Let (\mathbf{x}, \mathbf{y}) or (\mathbf{y}, \mathbf{x}) denote \mathbf{z} . As per this notation, $(\mathbf{x}, \blacklozenge) = (\blacklozenge, \mathbf{x}) = \mathbf{x}$.

In probability theory, the basic representational unit is called a probability function. 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 \mathbf{a} , \mathbf{b} , \mathbf{c} , etc. Let $[0, 1]$ denote the unit interval.

Probability Function. A *probability function* σ for s is a function $\sigma: 2^{\mathcal{W}_s} \rightarrow [0, 1]$ such that

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

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

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 $\sigma: \mathcal{W}_s \rightarrow [0, 1]$. 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 $\sum\{\sigma(\mathbf{x}) \mid \mathbf{x} \in \mathcal{W}_s\} = 1$. 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 $K = \sum\{\rho(\mathbf{x} \downarrow^r) \sigma(\mathbf{x} \downarrow^s) \mid \mathbf{x} \in \mathcal{W}_{r \cup s}\}$. The *combination of ρ and σ* , denoted by $\rho \oplus \sigma$, is the valuation for $r \cup s$ given by

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

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:

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

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.

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 $K = \sum\{\sigma(\mathbf{x} \downarrow^s) / \rho(\mathbf{x} \downarrow^r) \mid \mathbf{x} \in \mathcal{W}_{r \cup s} \text{ s.t. } \rho(\mathbf{x} \downarrow^r) > 0\}$. 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

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

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.

Does the VBS framework capture all aspects of probability theory? The answer is no. The VBS framework only captures the important features of probability theory. For example, Studeny [31] has proved the following property of conditional independence in probability theory: If r, s, t , and u are disjoint subsets of variables, then $t \perp u \mid r \cup s$, $t \perp u$, $r \perp s \mid t$, and $r \perp s \mid u$ if and only if $r \perp s \mid t \cup u$, $r \perp s$, $t \perp u \mid r$, and $t \perp u \mid s$. However, Spohn [32] has shown that this property does not hold in his theory. Therefore, it is clear that the above property of conditional independence does not hold in the VBS framework.

5. DEMPSTER-SHAFER'S BELIEF-FUNCTION THEORY

In this section, we show how Dempster-Shafer's belief-function theory fits in the VBS framework. More precisely, we define valuations, zero valuations, proper valuations, normal valuations, combination, marginalization, and removal. Also, we show that all axioms and assumptions made in Section 2 hold.

In Dempster-Shafer'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 Dempster-Shafer'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 $\mu: 2^{\mathcal{W}_s} \rightarrow [0, 1]$ such that

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

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

Commonality Function. A function $\theta: 2^{\mathcal{W}_s} \rightarrow [0, 1]$ is a *commonality function* for s if there exists a bpa function μ for s such that

$$\theta(\mathbf{a}) = \sum \{\mu(\mathbf{c}) \mid \mathbf{c} \supseteq \mathbf{a}\}. \quad (5.1)$$

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

It is evident from (B1), (B2), and (5.1) 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 [33] will help us understand the mathematical properties of commonality functions.

Lemma 5.1. Suppose μ and θ are real-valued functions defined on $2^{\mathcal{W}_s}$. Then

(5.1) holds for every $\mathbf{a} \in 2^{\mathcal{W}_s}$ if and only if

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

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

Lemma 5.2. Suppose μ and θ are real-valued functions defined on $2^{\mathcal{W}_s}$, and

suppose (5.1) holds for every $\mathbf{a} \in 2^{\mathcal{W}_s}$. Then

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

These lemmas can be proven by the methods used in the appendix of Ch. 2 of [11].

From Lemma 5.1, we see that a basic probability assignment is completely determined by the commonality function. From Lemmas 5.1 and 5.2, and conditions (B1) and (B2), we see that a function $\theta: 2^{\mathcal{W}_s} \rightarrow [0, 1]$ is a commonality function if and only if two conditions are satisfied:

$$\sum\{(-1)^{|\mathbf{c}-\mathbf{a}|} \theta(\mathbf{c}) \mid \mathbf{c} \supseteq \mathbf{a}\} \geq 0 \quad (5.2)$$

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

$$\sum\{(-1)^{|\mathbf{a}|+1} \theta(\mathbf{a}) \mid \mathbf{a} \in 2^{\mathcal{W}_s}\} = 1. \quad (5.3)$$

Condition (5.2) follows from condition (B1) and Lemma 5.1, and condition (5.3) follows from condition (B2) and Lemma 5.2.

In belief-function theory, a *valuation for s* is a function $\sigma: 2^{\mathcal{W}_s} \rightarrow [0, 1]$. 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 $\sum\{(-1)^{|\mathbf{c}-\mathbf{a}|} \sigma(\mathbf{c}) \mid \mathbf{c} \supseteq \mathbf{a}\} \geq 0$ 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 $\sum\{(-1)^{|\mathbf{a}|+1} \sigma(\mathbf{a}) \mid \mathbf{a} \in 2^{\mathcal{W}_s}\} = 1$. It is clear from (5.2) and (5.3) that proper normal valuations are commonality functions.

In belief-function theory, combination is pointwise multiplication of commonality functions followed by normalization [11]. 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 $K = \sum\{(-1)^{|\mathbf{a}|+1} \rho(\mathbf{a} \downarrow r) \sigma(\mathbf{a} \downarrow s) \mid \mathbf{a} \in 2^{\mathcal{W}_{r \cup s}}\}$. The *combination of ρ and σ* , denoted by $\rho \oplus \sigma$, is the valuation for $r \cup s$ given by

$$(\rho \oplus \sigma)(\mathbf{a}) = \begin{cases} K^{-1} \rho(\mathbf{a} \downarrow r) \sigma(\mathbf{a} \downarrow s) & \text{if } K \neq 0 \\ 0 & \text{if } K = 0 \end{cases} \quad (5.4)$$

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 [11, 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 (5.4). 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:

$$\sigma \downarrow (s - \{X\})(\mathbf{a}) = \sum \{ (-1)^{|\mathbf{b} - \mathbf{c}|} \sigma(\mathbf{b}) \mid \mathbf{b}, \mathbf{c} \in 2^{\mathcal{W}_s} \text{ such that } \mathbf{c} \downarrow (s - \{X\}) \supseteq \mathbf{a}, \text{ and } \mathbf{b} \supseteq \mathbf{c} \} \quad (5.5)$$

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

It is easy to see that the definition of marginalization in (5.5) 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 [34].

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 $K = \sum \{ (-1)^{|\mathbf{a}|+1} \sigma(\mathbf{a} \downarrow s) / \rho(\mathbf{a} \downarrow r) \mid \mathbf{a} \in 2^{\mathcal{W}_{r \cup s}} \text{ s.t. } \rho(\mathbf{a} \downarrow r) > 0 \}$. 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

$$(\sigma \ominus \rho)(\mathbf{a}) = \begin{cases} K^{-1} \sigma(\mathbf{a} \downarrow s) / \rho(\mathbf{a} \downarrow r) & \text{if } K > 0 \text{ and } \rho(\mathbf{a} \downarrow r) > 0 \\ 0 & \text{if } K = 0 \text{ or } \rho(\mathbf{a} \downarrow r) = 0 \end{cases} \quad (5.6)$$

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 (5.2) 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. An implication of this fact is that

conditionals may lack semantical coherence in the Dempster-Shafer's theory. This is the primary reason why conditionals are neither natural nor widely studied in the Dempster-Shafer's belief-function theory.

6. SPOHN'S EPISTEMIC-BELIEF THEORY

In this section, we show how Spohn's epistemic-belief theory [8, 35, 36] fits in the VBS framework. More precisely, we define valuations, proper valuations, normal valuations, combination, marginalization, and removal. Also, we show that all axioms and assumptions made in Section 2 hold.

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

Disbelief Function. A *disbelief function* for s is a function $\sigma: 2^{\mathcal{W}_s} \rightarrow \mathcal{N}$ 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 \mathbf{a} (the proposition that the true configuration of s is in \mathbf{a}). The degree of belief in proposition \mathbf{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 \mathbf{a} is believed if and only if $\sigma(\sim\mathbf{a}) > 0$, \mathbf{a} is disbelieved if and only if $\sigma(\mathbf{a}) > 0$, and \mathbf{a} is neither believed nor disbelieved if $\sigma(\mathbf{a}) = \sigma(\sim\mathbf{a}) = 0$. Also, in epistemic state σ , \mathbf{a} is more believed than \mathbf{b} if $\sigma(\sim\mathbf{a}) > \sigma(\sim\mathbf{b}) > 0$, and \mathbf{a} is more disbelieved than \mathbf{b} if $\sigma(\mathbf{a}) > \sigma(\mathbf{b}) > 0$.

In Spohn's epistemic-belief theory, a valuation for s is a function $\sigma: \mathcal{W}_s \rightarrow \mathcal{N}$. 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 [8, 36]. 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

$$(\rho \oplus \sigma)(\mathbf{x}) = \rho(\mathbf{x} \downarrow^r) + \sigma(\mathbf{x} \downarrow^s) - K \quad (6.1)$$

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 (6.1). The identity ι_s for $N_s \cup \{\zeta_s\}$ is given by $\iota_s(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{W}_s$. Every normal valuation in N_s has a unique identity in 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:

$$\sigma^{\downarrow(s - \{X\})}(\mathbf{y}) = \text{MIN}\{\sigma(\mathbf{y}, \mathbf{x}) \mid \mathbf{x} \in \mathcal{W}_X\} \quad (6.2)$$

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 $\{\mathbf{y}\}$ 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 (6.2) 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 [36].

Removal. In Spohn's theory, removal is subtraction followed by normalization [36]. Suppose $\sigma \in \mathcal{V}_s$, and $\rho \in N_r \cup Z_r$. Then the normal valuation resulting from the removal of ρ from σ , denoted by $\sigma \ominus \rho$, is given by

$$(\sigma \ominus \rho)(\mathbf{x}) = \sigma(\mathbf{x}^{\downarrow s}) - \rho(\mathbf{x}^{\downarrow r}) - K \quad (6.3)$$

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

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

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 $\rho \in N_r \cup Z_r$. Since every normal valuation is positive normal, $\iota_\rho = \iota_r$.

7. ZADEH'S POSSIBILITY THEORY

In this section, we describe how Zadeh's possibility theory [12, 13] fits in the VBS framework. More precisely, we define valuations, normal valuations, proper valuations, combination, marginalization, and removal. Also, we show that all axioms and assumptions made in Section 2 hold.

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 $\pi: 2^{\mathcal{W}_s} \rightarrow [0, 1]$ such that

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

(S2). $\pi(\mathbf{a}) = \text{MAX}\{\pi(\{\mathbf{x}\}) \mid \mathbf{x} \in \mathbf{a}\}$ 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 [37]. \mathbf{a} is possible in state π if and only if $\pi(\mathbf{a}) = 1$, and \mathbf{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 \mathbf{a} is possible, and $[1 - \pi(\mathbf{a})]$ can be interpreted as the degree to which proposition \mathbf{a} is not possible, i.e., \mathbf{a} is more possible than \mathbf{b} if $\pi(\mathbf{a}) > \pi(\mathbf{b})$ and conversely, \mathbf{a} is more impossible than \mathbf{b} if $\pi(\mathbf{a}) < \pi(\mathbf{b}) < 1$.

In Zadeh's possibility theory, a *valuation* σ for s is a function $\sigma: \mathcal{W}_s \rightarrow [0, 1]$. 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 $\text{MAX}\{\sigma(\mathbf{x}) \mid \mathbf{x} \in \mathcal{W}_s\} = 1$. A normal valuation can be regarded as a possibility function defined only for singleton subsets.

Combination.⁹ 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 $K = \text{MAX}\{\rho(\mathbf{x}^{\downarrow r})\sigma(\mathbf{x}^{\downarrow s}) \mid \mathbf{x} \in \mathcal{W}_{r \cup s}\}$. The *combination of ρ and σ* , denoted by $\rho \oplus \sigma$, is the valuation for $r \cup s$ given by

$$(\rho \oplus \sigma)(\mathbf{x}) = \begin{cases} K^{-1} \rho(\mathbf{x}^{\downarrow r})\sigma(\mathbf{x}^{\downarrow s}) & \text{if } K > 0 \\ 0 & \text{if } K = 0 \end{cases} \quad (7.1)$$

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 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 (7.1). 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:

$$\sigma^{\downarrow(s - \{X\})}(\mathbf{y}) = \text{MAX}\{\sigma(\mathbf{y}, \mathbf{x}) \mid \mathbf{x} \in \mathcal{W}_X\} \quad (7.2)$$

⁹ There are several definitions of combination in possibility theory. Zadeh [12] 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., [13], 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 followed by normalization (since pointwise minimization followed by normalization, for example, fails to be associative).

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 (7.2) 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 [37].

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,

$$(\sigma \ominus \rho)(\mathbf{x}) = \begin{cases} K^{-1} \sigma(\mathbf{x}^{\downarrow s}) / \rho(\mathbf{x}^{\downarrow r}) & \text{if } K > 0, \text{ and } \rho(\mathbf{x}^{\downarrow r}) > 0 \\ 0 & \text{if } K = 0 \text{ or } \rho(\mathbf{x}^{\downarrow r}) = 0 \end{cases} \quad (7.3)$$

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 $\rho(\mathbf{x}) = 0$.

Most of the literature on Zadeh's possibility theory defines combination as pointwise minimization with no normalization. With this definition, combination is always idempotent, i.e., $\pi \oplus \pi = \pi$, and consequently, conditional independence always holds for any disjoint subsets of variables. Therefore, conditional independence has not been widely studied in the possibility theory literature. A problem with the definition of combination as pointwise minimization with no normalization is that it is semantically inadequate. If we define combination as pointwise minimization with normalization, then combination is not associative. This poses other problems because now we have to worry about the sequence in which we combine possibility valuations and what the sequence represents. Our definition of combination as pointwise multiplication followed by normalization makes possibility theory more similar to probability theory and Spohn's epistemic-belief theory. In our version of possibility theory, possibility valuations are no longer idempotent and therefore the conditional independence theory is no longer trivial. We believe this version of possibility theory merits more study than it has received in the literature.

8. CONCLUSION

The main objective of this paper was to define conditional independence in the VBS framework. Although this concept has been defined and extensively studied in probability theory, it has not been extensively studied in non-probabilistic uncertainty theories.

Drawing upon the literature on conditional independence in probability theory [1, 2, 3, 4, 5, 6, 7], we define conditional independence in VBS. The VBS framework was defined earlier by Shenoy [20, 21]. However, the VBS framework defined there is inadequate for the purposes of studying properties of conditional independence. In this paper, we embellish the framework by including three new classes of valuations called proper, normal, and positive normal, and by including a new operation called removal. The new definitions are stated in the form of axioms. These axioms are general enough to include probability theory, Dempster-Shafer's belief-function theory, Spohn's epistemic-belief theory, and Zadeh's possibility theory.

The VBS framework enables us to define conditional independence, and to prove some major properties of conditional independence that have been derived in probability theory. Conditional independence is defined in terms of factorization of the joint valuation. Thus, not only do we have a deeper understanding of conditional independence in probability theory, we also understand what conditional independence means in various non-probabilistic uncertainty theories. This should deflect some criticism that non-probabilistic uncertainty theories are not as well developed as probability theory. Also, the VBS framework allows us to translate results from probability theory to non-probabilistic uncertainty theories, and vice-versa.

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. I am grateful to Satya Mandal, Serafin Moral, Pierre Ndilikilikisha, Glenn Shafer, Philippe Smets, Leen-Kiat Soh, Wolfgang Spohn, Milan Studeny, Hong Xu, and three anonymous referees for comments and discussions.

REFERENCES

1. Dawid, A. P., Conditional independence in statistical theory (with discussion), *J. Roy. Stat. Soc., Ser. B*, **41**(1), 1–31, 1979.
2. Spohn, W., Stochastic independence, causal independence, and shieldability, *J. Phil. Logic*, **9**, 73–99, 1980.

3. Lauritzen, S. L., Lectures on contingency tables, 3rd edition, Technical Report No. R-89-24, Institute for Electronic Systems, University of Aalborg, Denmark, 1989.
4. Pearl, J. and A. Paz, Graphoids: Graph-based logic for reasoning about relevance relations, in *Advances in Artificial Intelligence - II* (B. D. Boulay, D. Hogg and L. Steele, Eds.), North-Holland, Amsterdam, 357–363, 1987.
5. Smith, J. Q., Influence diagrams for statistical modelling, *Ann. of Stat.*, **17**(2), 654–672, 1989.
6. Studeny, M., Multiinformation and the problem of characterization of conditional independence relations, *Problems of Control and Information Theory*, **18**(1), 3–16, 1989.
7. Geiger, D., Graphoids: A qualitative framework for probabilistic inference, Ph.D. dissertation, Department of Computer Science, University of California at Los Angeles, CA, 1990.
8. Spohn, W., Ordinal conditional functions: A dynamic theory of epistemic states, in *Causation in Decision, Belief Change, and Statistics* (W. L. Harper and B. Skyrms, Eds.), D. Reidel, Dordrecht, Holland, **2**, 105-134, 1988.
9. Hunter, D., Graphoids and natural conditional functions, *Int. J. of Approx. Reas.*, **5**(6), 489–504, 1991.
10. Dempster, A. P., Upper and lower probabilities induced by a multivalued mapping,” *Annals of Mathematical Statistics*, **38**, 325-339, 1967.
11. Shafer, G., *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ., 1976.
12. Zadeh, L. A., A theory of approximate reasoning, in *Machine Intelligence* (J. E. Ayes, D. Mitchie and L. I. Mikulich, Eds.), Ellis Horwood, Chichester, U.K., **9**, 149–194, 1979.
13. Dubois, D. and H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, Plenum Press, New York, NY, 1988.
14. Shafer, G., Belief functions and parametric models (with discussion), *J. Roy. Stat. Soc., Ser. B*, **44**(3), 322-352, 1982.
15. Shafer, G., The problem of dependent evidence, Working Paper No. 164, School of Business, University of Kansas, Lawrence, KS, 1984.
16. Shafer, G., Belief functions and possibility measures, in *The Analysis of Fuzzy Information* (J. Bezdek, Ed.), CRC Press, Boca Raton, FL, **1**, 51-84, 1987.
17. Shafer, G., Perspectives on the theory and practice of belief functions, *Int. J. Approx. Reas.*, **4**(5–6), 323-362, 1990.
18. Smets, P., Combining non distinct evidences, *Proc. of the North American Fuzzy Information Processing Society Conference* (NAFIPS’86), New Orleans, LA, 544–548, 1986.
19. Shafer, G., P. P. Shenoy, and K. Mellouli, Propagating belief functions in qualitative Markov trees, *Int. J. Approx. Reas.*, **1**(4), 349–400, 1987.
20. Shenoy, P. P., A valuation-based language for expert systems, *Int. J. Approx. Reas.*, **3**(5), 383-411, 1989.

21. Shenoy, P. P., Valuation-based systems: A framework for managing uncertainty in expert systems, in *Fuzzy Logic for the Management of Uncertainty* (L. A. Zadeh and J. Kacprzyk, Eds.), John Wiley & Sons, New York, NY, 83–104, 1992.
22. Dawid, A. P., Conditional independence for statistical operations, *Ann. of Stat.*, **8**(3), 598–617, 1980.
23. Shenoy, P. P., Consistency in valuation-based systems, *ORSA J. on Comp.*, 1993, to appear.
24. Shenoy, P. P. and G. Shafer, Axioms for probability and belief-function propagation, in *Uncertainty in Artificial Intelligence* (R. D. Shachter, T. S. Levitt, J. F. Lemmer and L. N. Kanal, Eds.), North-Holland, Amsterdam, **4**, 169–198, 1990.
25. Petrich, M., *Introduction to Semigroups*, Charles E. Merrill Publishing Co., Columbus, OH, 1973.
26. Cano, J. E., M. Delgado, and S. Moral, An axiomatic framework for propagating uncertainty in directed acyclic networks, *Int. J. Approx. Reas.*, **8**(4), 253–280, 1993.
27. Shafer, G., An axiomatic study of computation in hypertrees, Working Paper No. 232, School of Business, University of Kansas, Lawrence, KS, 1991.
28. Herstein, I. N., *Topics in Algebra*, 2nd edition, Xerox College Publishing, Lexington, MA, 1975.
29. Geiger, D. and J. Pearl, Logical and algorithmic properties of conditional independence and graphical models, CIS Report No. 9108, Technion - Israel Institute of Technology, Computer Science Department, Haifa, Israel, 1991.
30. Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, CA, 1988.
31. Studeny, M., Attempts at axiomatic description of conditional independence, *Kybernetika*, **25**(3), 72–79, 1989.
32. Spohn, W., On the properties of conditional independence, in *Patrick Suppes—Mathematical Philosopher*, (P. Humphreys, Ed.), Synthese, 1993, to appear.
33. Shafer, G. and P. P. Shenoy, Bayesian and belief-function propagation, Working Paper No. 192, School of Business, University of Kansas, Lawrence, KS, 1988.
34. Shenoy, P. P., Using Dempster-Shafer's belief-function theory in expert systems, in *Advances in the Dempster-Shafer Theory of Evidence* (M. Federizzi, J. Kacprzyk and L. A. Zadeh, Eds.), John Wiley & Sons, New York, 1993, to appear.
35. Spohn, W., A general non-probabilistic theory of inductive reasoning, in *Uncertainty in Artificial Intelligence* (R. D. Shachter, T. S. Levitt, J. F. Lemmer and L. N. Kanal, Eds.), North-Holland, Amsterdam, **4**, 149–158, 1990.
36. Shenoy, P. P., On Spohn's rule for revision of beliefs, *Int. J. Approx. Reas.*, **5**(2), 149–181, 1991.
37. Shenoy, P. P., Using possibility theory in expert systems, *Fuzzy Sets and Systems*, **52**(2), 129–142, 1992.

SELECTED WORKING PAPERS

Unpublished working papers are available from the respective authors at:

School of Business
University of Kansas
Summerfield Hall
Lawrence, KS 66045-2003 USA

- No. 184. "Propagating Belief Functions with Local Computations," Prakash P. Shenoy and Glenn Shafer, February 1986. Appeared in *IEEE Expert*, **1**(3), 1986, 43-52.
- No. 190. "Propagating Belief Functions in Qualitative Markov Trees," Glenn Shafer, Prakash P. Shenoy, and Khaled Mellouli, June 1987. Appeared in *International Journal of Approximate Reasoning*, **1**(4), 1987, 349-400.
- No. 197. "AUDITOR'S ASSISTANT: A Knowledge Engineering Tool for Audit Decisions," Glenn Shafer, Prakash P. Shenoy, and Rajendra Srivastava, April 1988. Appeared in *Auditing Symposium IX: Proceedings of 1988 Touche Ross/University of Kansas Symposium on Auditing Problems*, 61-84, School of Business, University of Kansas, Lawrence, KS.
- No. 198. "Studies on Finding Hypertree Covers of Hypergraphs," Lianwen Zhang, April 1988.
- No. 199. "An Axiomatic Framework for Bayesian and Belief-Function Propagation," Prakash P. Shenoy and Glenn Shafer, July 1988. Appeared in *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, MN, 307-314.
- No. 200. "Probability Propagation," Glenn Shafer and Prakash P. Shenoy, August 1988. Appeared in *Annals of Mathematics and Artificial Intelligence*, **2**(1-4), 1990, 327-352.
- No. 201. "Local Computation in Hypertrees," Glenn Shafer and Prakash P. Shenoy, August 1988.
- No. 203. "A Valuation-Based Language for Expert Systems," Prakash P. Shenoy, August 1988. Appeared in *International Journal of Approximate Reasoning*, **3**(5), 1989, 383-411.
- No. 208. "Constraint Propagation," Prakash P. Shenoy and Glenn Shafer, November 1988.
- No. 209. "Axioms for Probability and Belief-Function Propagation," Prakash P. Shenoy and Glenn Shafer, November 1988. Appeared in Shachter, R. D., M. Henrion, L. N. Kanal, and J. F. Lemmer, eds., *Uncertainty in Artificial Intelligence*, **4**, 1990, 169-198. Reprinted in Shafer, G. and J. Pearl (eds.), *Readings in Uncertain Reasoning*, 1990, 575-610, Morgan Kaufmann, San Mateo, CA.
- No. 211. "MacEvidence: A Visual Evidential Language for Knowledge-Based Systems," Yen-Teh Hsia and Prakash P. Shenoy, March 1989. A condensed version of this paper appeared as: Hsia, Y-T. and P. P. Shenoy, "An evidential language for expert systems," in Ras, Z. W., ed., *Methodologies for Intelligent Systems*, **4**, 1989, 9-16.
- No. 213. "On Spohn's Rule for Revision of Beliefs," Prakash P. Shenoy, July 1989. Appeared in *International Journal of Approximate Reasoning*, **5**(2), 1991, 149-181.
- No. 216. "Consistency in Valuation-Based Systems," Prakash P. Shenoy, February 1990, revised May 1991. To appear in *ORSA Journal on Computing*.

- No. 218. “Perspectives on the Theory and Applications of Belief Functions,” Glenn Shafer, April 1990. Appeared in *International Journal of Approximate Reasoning*, **4**(5–6), 1990, 323–362.
- No. 220. “Valuation-Based Systems for Bayesian Decision Analysis,” Prakash P. Shenoy, April 1990, revised May 1991. Appeared in *Operations Research*, **40**(3), 1992, 463–484.
- No. 221. “Valuation-Based Systems for Discrete Optimization,” Prakash P. Shenoy, June 1990. Appeared in Bonissone, P. P., M. Henrion, L. N. Kanal, and J. F. Lemmer, eds., *Uncertainty in Artificial Intelligence*, **6**, 1991, 385–400, North-Holland, Amsterdam.
- No. 223. “A New Method for Representing and Solving Bayesian Decision Problems,” Prakash P. Shenoy, September 1990, revised February 1991. To appear in 1993 in Hand, D. J., ed., *Artificial Intelligence Frontiers in Statistics: AI and Statistics III*, 119–138, Chapman & Hall, London, England. A 9-page summary of this paper appeared as “A Fusion Algorithm for Solving Bayesian Decision Problems” in D’Ambrosio, B., P. Smets, and P. P. Bonissone (eds.), *Uncertainty in Artificial Intelligence*, 1991, 361–369, Morgan Kaufmann, San Mateo, CA.
- No. 225. “Why Should Statisticians be Interested in Artificial Intelligence?” Glenn Shafer, November, 1990. Appeared in *Proceedings of the Fifth Annual Conference on Making Statistics More Effective in Schools of Business*, 1990, 16–58, Lawrence, KS.
- No. 226. “Valuation-Based Systems: A Framework for Managing Uncertainty in Expert Systems,” Prakash P. Shenoy, March, 1991. Revised July 1991. Appeared in Zadeh, L. A. and J. Kacprzyk, eds., *Fuzzy Logic for the Management of Uncertainty*, 1992, 83–104, John Wiley and Sons, New York, NY.
- No. 227. “Valuation Networks, Decision Trees, and Influence Diagrams: A Comparison,” Prakash P. Shenoy, June 1991. Revised June 1992.
- No. 228. “The Early Development of Mathematical Probability,” Glenn Shafer, August 1991. To appear in Grattan-Guinness, I., ed., *Encyclopedia of the History and Philosophy of the Mathematical Sciences*, Routledge, London.
- No. 229. “What is Probability?,” Glenn Shafer, August 1991. To appear in Hoaglin, D. C. and D. S. Moore, eds., *Perspectives on Contemporary Statistics*, Mathematical Association of America.
- No. 230. “Can the Various Meanings of Probability be Reconciled?,” Glenn Shafer, August 1991. To appear in Keren, G. and C. Lewis, eds., *Methodological and Quantitative Issues in the Analysis of Psychological Data*, Lawrence Erlbaum, Hillsdale, NJ.
- No. 231. “Rejoinders to Comments on “Perspectives on the Theory and Practice of Belief Functions”,” Glenn Shafer, August 1991. Appeared in *International Journal of Approximate Reasoning*, **6**(3), 1992, 445–480.
- No. 232. “An Axiomatic Study of Computation in Hypertrees,” Glenn Shafer, August 1991.
- No. 233. “Using Possibility Theory in Expert Systems,” Prakash P. Shenoy, September 1991. To appear in *Fuzzy Sets and Systems*, **51**(3), 1992.
- No. 234. “Using Dempster-Shafer’s Belief-Function Theory in Expert Systems,” Prakash P. Shenoy, September 1991. To appear in 1992 in *Advances in the Dempster-Shafer Theory of Evidence*, Federizzi, M., J. Kacprzyk, and L. A. Zadeh, eds., John Wiley & Sons, New York, NY.
- No. 235. “Potential Influence Diagrams,” Pierre Ndilikilikisha, September 1991.

- No. 236. “Conditional Independence in Valuation-Based Systems,” Prakash P. Shenoy, September 1991. Revised November 1991. Revised September 1992. An 8-page summary of the November 1991 version appeared as “Conditional Independence in Uncertainty Theories” in Dubois, D., M. P. Wellman, B. D’Ambrosio, and P. Smets (eds.), *Uncertainty in Artificial Intelligence*, 1992, 284–291, Morgan Kaufmann, San Mateo, CA.