Appeared in: Cozman, F. G., R. Nau, T. Seidenfeld (eds.), Proc. of the 4th Int. Symposium on Imprecise Probabilities and Their Applications, 2005, 306--314.

No Double Counting Semantics for Conditional Independence

Prakash P. Shenoy
University of Kansas, School of Business,
1300 Sunnyside Ave., Summerfield Hall,
Lawrence, KS 66045-7585, USA
pshenoy@ku.edu

Abstract

The main goal of this paper is to describe a new semantic for conditional independence in terms of no double counting of uncertain evidence. For ease of exposition, we use probability calculus to state all results. But the results generalize easily to any calculus that fits in the framework of valuation-based systems. Thus, the results described in this paper apply also, for example, to Dempster-Shafer's (D-S) belief function theory, to Spohn's epistemic beliefs theory, and to Zadeh's possibility theory. The concept of independent (or distinct) evidence in D-S belief function theory is analogous to the concept of conditional independence for variables in probability theory.

Keywords. Conditional independence, no double counting semantics, distinct evidence, independent pieces of evidence, graphoid axioms, Dempster-Shafer's belief function theory, Spohn's epistemic beliefs theory, possibility theory, valuation-based systems.

1 Introduction

Conditional independence for probability theory has been traditionally interpreted in terms of irrelevance [Dawid 1979, Spohn 1980, Pearl and Paz 1987]. In Shenoy [1994], I have interpreted it in terms of factorization of the joint probability distribution. In this paper, I describe a new semantic for conditional independence in terms of no double counting of uncertain knowledge.

There are several advantages of these new semantics. First, these semantics provide a new method for building models in domains, such as sensor fusion, where these semantics can be easily applied.

Second, Dempster-Shafer's (D-S) belief function theory [Dempster 1967, Shafer 1976] uses the concept of independent (or distinct) pieces of evidence to qualify when it is proper to combine the corresponding belief functions by Dempster's rule of combination. The concept of independent evidence can be understood as no double counting of uncertain evidence. These semantics of no double counting are paramount in building D-S belief function models. Yet, these semantics are not well understood and remains a mystery for many. The results provided here explain these semantics. This should make D-S belief function theory more appealing to probabilists, and facilitate the integration of these two uncertainty calculi. Thus, e.g., methods for constructing Bayesian network models can be easily adapted to constructing belief function models, and vice-versa.

An outline of the remainder of the paper is as follows. In Section 2, we describe the notation used and basic definitions. In Section 3, we define conditional independence and describe its semantics in terms of irrelevance. In Section 4, we describe factorization semantics of conditional independence. In Section 5, we give a definition of what it means for probability potentials to be independent using the semantics of no double counting of uncertain knowledge, and we provide five sufficient conditions for independence of potentials. In Section 6, we show that our definition of independence of potentials satisfies the graphoid properties. In Section 7, we describe two examples to illustrate our definition and results. Finally, in Section 8, we summarize and conclude.

2 Notation

Upper-case Roman alphabets such as X, Y, Z will be used to denote *variables*. The *state space* of X is denoted by Ω_X. We assume that all variables have finite state spaces. Sets of variables will be denoted by lower-case Roman alphabets such as f, g, h, etc. The state space for a set of variables is the Cartesian product of the state space of variables in the set, $\Omega_f = \times\{\Omega_X \mid X \in f\}$. Lower-case Greek alphabets will be used to denote potentials. A *potential* ϕ for f is a function ϕ: $\Omega_f \to \mathbb{R}^+$, where $\mathbb{R}^+$ denotes the set of non-negative real numbers. Although the domain of ϕ is Ω_f, to simplify terminology, we refer to f as the *domain* of ϕ. In a Bayes net, potentials are also called conditional probability tables. If $\phi(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \Omega_f$, then we say ϕ is a *positive* potential. Potentials can be marginalized and combined. We can also remove a potential from another potential using division (taking care to avoid division by zeroes). The removal operation can be viewed as an inverse of combination. Formal definitions are as follows.

Marginalization. Suppose ϕ is a potential for f, and suppose $X \in f$. The *marginal* of ϕ for $f \setminus \{X\}$, denoted by ϕ^{-X} or by $\phi^{\downarrow f \setminus \{X\}}$, is a potential for $f \setminus \{X\}$ defined as follows.

$$\phi^{-X}(\boldsymbol{y}) = \phi^{\downarrow f \setminus \{X\}}(\boldsymbol{y}) = \Sigma\{\phi(\boldsymbol{y}, x) \mid x \in \Omega_X\} \text{ for all } \boldsymbol{y} \in \Omega_{f \setminus \{X\}}. \quad (2.1)$$

Marginalization is an operation that reduces the domain of potentials by addition over the state space of variables being removed. When we want to emphasize the variable that is being removed, we use the notation ϕ^{-X}, and when we want to emphasize the variables that remain in the domain after marginalization, we use the notation $\phi^{\downarrow f \setminus \{X\}}$.

We can successively remove several variables from the domain of a potential, and the result is independent of the sequence in which the variables are removed. Thus $(\phi^{-X_1})^{-X_2} = (\phi^{-X_2})^{-X_1} = \phi^{-\{X_1, X_2\}}$, or using the alternative notation, $(\phi^{\downarrow f \setminus \{X_1\}})^{\downarrow f \setminus \{X_1, X_2\}} = (\phi^{\downarrow f \setminus \{X_2\}})^{\downarrow f \setminus \{X_1, X_2\}} = \phi^{\downarrow f \setminus \{X_1, X_2\}}$. We can remove all variables from a potential, resulting in a potential with a single value, which we can regard as a potential for the empty set. To keep our notation consistent, we assume that the state space of the empty set has a single element, denoted say by $\blacklozenge$, $\Omega_\varnothing = \{\blacklozenge\}$.

Combination. First we need some notation for projection of states. Suppose $\boldsymbol{y}$ is a state of variables in f, and suppose $g \subseteq f$. Then the projection of $\boldsymbol{y}$ to g, denoted by $\boldsymbol{y}^{\downarrow g}$, is the state of g obtained from $\boldsymbol{y}$ by dropping states of $f \setminus g$. If $g = f$, then $\boldsymbol{y}^{\downarrow g} = \boldsymbol{y}$. If $g = \varnothing$, then $\boldsymbol{y}^{\downarrow g} = \blacklozenge$.

Suppose f and g are arbitrary subsets of variables, suppose ϕ is a potential for f, and suppose γ is a potential for g. Then the *combination* of ϕ and γ, denoted by $\phi \otimes \gamma$, is a potential for $f \cup g$ defined as follows.

$$(\phi\otimes\gamma)(\boldsymbol{y}) = \begin{cases} K^{-1}\, \phi(\boldsymbol{y}^{\downarrow f})\, \gamma(\boldsymbol{y}^{\downarrow g}) & \text{if } K > 0, \\ 0 & \text{if } K = 0, \end{cases} \quad (2.2)$$

for all $\boldsymbol{y} \in \Omega_{f \cup g}$, where K is a normalization constant defined as follows: $K = \Sigma\{\phi(\boldsymbol{y}^{\downarrow f})\, \gamma(\boldsymbol{y}^{\downarrow g}) \mid \boldsymbol{y} \in \Omega_{f \cup g}\}$. If $K > 0$, we say ϕ and γ are "combinable," and if $K = 0$, we say ϕ and γ are "not combinable." In the latter case, the combination of ϕ and γ results in a potential that is identically zero, which is the potential for $f \cup g$ representing *contradiction*.

The combination operation must be used with care. If ϕ represents our knowledge about f, γ represents our knowledge about g, *and* ϕ and γ are independent, then $\phi\otimes\gamma$ represents the joint knowledge about $f \cup g$. In the succeeding sections, we explain exactly what we mean when we say "ϕ and γ are independent." Notice that there is no restriction on the sets f and g. They may be the same set, they may be disjoint, they may have a non-empty intersection, or they may be subsets of each other. Thus, combination is pointwise multiplication followed by normalization (when normalization is possible).

Some properties of combination are as follows. Combination is commutative and associative: $\phi\otimes\gamma = \gamma\otimes\phi$, and $\phi\otimes(\gamma\otimes\eta) = (\phi\otimes\gamma)\otimes\eta$. Also, marginalization and combination share one important property called transitivity of marginalization over combination. Suppose ϕ is a potential for f, and γ is a potential for g, and suppose $X \in f$ and $X \notin g$. Then

$$(\phi\otimes\gamma)^{-X} = \phi^{-X}\otimes\gamma. \quad (2.3)$$

Finally, a few words on normalization in the definition of combination. When we are combining and marginalizing a lot of potentials, we can skip normalization and do it once at the very end. It makes no difference in the final results, but it is computationally efficient to avoid normalization every time we combine potentials. But, normalization is an important facet of combination in probability. The normalization constant K in (2.2) is a measure of conflict between ϕ and γ. If $K = 0$, the knowledge in ϕ and γ totally conflict with each other, and we should doubt the veracity of either ϕ or γ.

Normal Potentials. Suppose ϕ is a potential for f. We say ϕ is a *normal* potential if $(\phi^{\downarrow\varnothing})(\blacklozenge) = 1$, i.e., the sum of all values of the potential add to 1. Notice that if ϕ is obtained by combining two or more combinable potentials, then ϕ is normal (by definition of combination). Also, if ϕ is a normal potential, then all marginals of ϕ are also normal.

Vacuous Potentials. Let u denote the set of all variables. We assume that u is finite. Consider Φ_f, the set of all potentials with domain f, $f \subseteq u$. Then Φ_f together with the combination operator $\otimes$ forms a commutative semi-group. Let Φ_u denote the set of all potentials (with domains that are subsets of u). Then $(\Phi_u, \otimes)$ also forms a semi-group. Given any two potentials ϕ_1 and ϕ_2, we say $\phi_1 = \phi_2$ if ϕ_1 and ϕ_2 have the same domains and the same values for each state of the domain. Since the combination operator involves normalization, two potentials ϕ_1 and ϕ_2 with the same domain, say f, and whose values differ by a positive constant k, i.e., $\phi_2(\boldsymbol{y}) = k\ \phi_1(\boldsymbol{y})$ for all $\boldsymbol{y} \in \Omega_f$, have essentially the same information since $\phi_1 \otimes \gamma = \phi_2 \otimes \gamma$ for all potentials $\gamma \in \vartheta$. In this case, we say that ϕ_1 and ϕ_2 are *equivalent*, and denote this by $\phi_1 \approx \phi_2$. Notice that $\approx$ is an equivalence relation, and we use the normal potential in this equivalence class as a representative.

Let Λ_f denote the set of normal potentials with domain f. Let ζ_f denote the potential for f that is identically zero, i.e., $\zeta_f(\boldsymbol{y}) = 0$ for all $\boldsymbol{y} \in \Omega_f$. We refer to ζ_f as the *zero* potential for f. Then $(\Lambda_f \cup \{\zeta_f\}, \otimes)$ is a semi-group. This semi-group has a unique identity element ι_f which is the potential for f such that $\iota_f(\boldsymbol{y}) = k$ for all $\boldsymbol{y} \in \Omega_f$ where $k = 1/|\Omega_f|$. ι_f has the property that $\iota_f \otimes \phi = \phi$ for every $\phi \in \Lambda_f \cup \{\zeta_f\}$. In particular, $\iota_f \otimes \iota_f = \iota_f$, i.e., ι_f is idempotent. We call ι_f, and the potentials in its equivalence class, *vacuous* potentials for f.

Observation Potentials. Consider the semi-group $(\Lambda_f \cup \{\zeta_f\}, \otimes)$. Let $\boldsymbol{a} \in \Omega_f$. Consider the potential $\lambda_{\boldsymbol{a}}$ for f defined as follows: $\lambda_{\boldsymbol{a}}(\boldsymbol{a}) = 1$, and $\lambda_{\boldsymbol{a}}(\boldsymbol{y}) = 0$ for all $\boldsymbol{y} \in \Omega_f$, $\boldsymbol{y} \neq \boldsymbol{a}$. Clearly $\lambda_{\boldsymbol{a}} \in \Lambda_f$. $\lambda_{\boldsymbol{a}}$ has the property that $\lambda_{\boldsymbol{a}} \otimes \phi = \lambda_{\boldsymbol{a}}$ for all $\phi \in \Lambda_f$ such that ϕ is combinable with $\lambda_{\boldsymbol{a}}$, and $\lambda_{\boldsymbol{a}} \otimes \phi = \zeta_f$ for all $\phi \in \Lambda_f$ such that ϕ is not combinable with $\lambda_{\boldsymbol{a}}$. In particular, $\lambda_{\boldsymbol{a}} \otimes \lambda_{\boldsymbol{a}} = \lambda_{\boldsymbol{a}}$, i.e., $\lambda_{\boldsymbol{a}}$ is idempotent. We call such potentials *observation* potentials.

Division. Suppose $f \subseteq g$, suppose γ is a potential for g, and suppose ϕ is a positive potential for f. Then we define $\gamma \oslash \phi$, read as γ divided by ϕ, as a potential for g defined by

$$(\gamma \oslash \phi)(\boldsymbol{x}) = \gamma(\boldsymbol{x})/\phi(\boldsymbol{x}^{\downarrow f}) \text{ for all } \boldsymbol{x} \in \Omega_g. \tag{2.4}$$

If ϕ is a positive potential for f, then $\phi \oslash \phi$ is a vacuous potential for f. If γ is a potential for g, then $(\gamma \otimes \phi) \oslash \phi = \gamma \otimes (\phi \oslash \phi) = \gamma \otimes \iota_f$ where ι_f is the normal vacuous potential for f.

Conditionals. In probability theory, we often build large probability models using conditioning. Conditioning is important since it allows us to factor a joint distribution into several potentials. Suppose f and g are disjoint subsets of variables, and suppose γ is a potential for $f \cup g$ such that γ^{-g} is a vacuous potential for f. Then γ is said to be a *conditional* for g given f.

A rationale for this definition is as follows. Suppose ϕ is a normal positive potential for f. Then consider the probability distribution $P = \phi \otimes \gamma$ for $f \cup g$ where γ is a conditional for g given f. Then $P^{\downarrow f} = (\phi \otimes \gamma)^{\downarrow f} = (\phi \otimes \gamma)^{-g} = \phi \otimes \gamma^{-g} = \phi$ (since γ^{-g} is vacuous for f). Thus ϕ represents the marginal distribution of f. Now consider the conditional distribution of g given f. Suppose $\boldsymbol{x} \in \Omega_g$ and $\boldsymbol{y} \in \Omega_f$. Then $P(\boldsymbol{x} \mid \boldsymbol{y}) = P(\boldsymbol{x}, \boldsymbol{y})/P(\boldsymbol{y}) = (\phi \otimes \gamma)(\boldsymbol{x}, \boldsymbol{y})/\phi(\boldsymbol{y}) = (K^{-1}\ \phi(\boldsymbol{y})\ \gamma(\boldsymbol{x}, \boldsymbol{y}))/\phi(\boldsymbol{y}) = K^{-1}\ \gamma(\boldsymbol{x}, \boldsymbol{y})$. Thus, the values of γ can be regarded as conditional probabilities (up to a normalization constant).

3 CI as Irrelevance

Suppose we are constructing a probability model for a collection of three variables as follows. A flight departs from Los Angeles, stops in Denver and arrives in Kansas City. Let D_1 denote departure from LA with states 'o_{LA}' (for on-time departure from LA) and 'l_{LA},' (for late departure from LA), let D_2 denote departure of the same flight from Denver with the similar two states, o_D and l_D, and let A denote arrival in Kansas City with two states, o_{KC} and l_{KC}.

One method of constructing the joint distribution for D_1, D_2, and A is to assess the prior distribution of D_1 denoted by, say δ_1, the conditional distribution of D_2 given D_1, denoted by δ_2, and the conditional distribution of A given D_1 and D_2, denoted by α. Thus, $\delta_1(x) = P(x)$ for all $x \in \Omega_{D_1}$, $\delta_2(x, y) = P(y \mid x)$ for all $x \in \Omega_{D_1}$, and $y \in \Omega_{D_2}$, and $\alpha(x, y, z) = P(z \mid x, y)$ for all $x \in \Omega_{D_1}$, $y \in \Omega_{D_2}$, and $z \in \Omega_A$. Probability theory tells us that we can construct the joint distribution for D_1, D_2, and A, denoted by π, by combining these three potentials, i.e.,

$$\pi = \delta_1 \otimes \delta_2 \otimes \alpha, \tag{3.1}$$

i.e., $\pi(x, y, z) = P(x)\ P(y \mid x)\ P(z \mid x, y)$ for $x \in \Omega_{D_1}$, $y \in \Omega_{D_2}$, and $z \in \Omega_A$. In this model, it is reasonable to assume that once we know the value of D_2, further knowledge of D_1 is irrelevant to A. Thus, given a value of D_2 (departure from Denver), D_1 (departure from LA) is irrelevant for assessing the distribution of A (arrival in Kansas City), i.e.,

$$P(A = l_{KC} \mid D_1 = l_{LA}, D_2 = l_D) = P(A = l_{KC} \mid D_1 = o_{LA}, D_2 = l_D), \text{ and}$$

$$P(A = l_{KC} \mid D_1 = l_{LA}, D_2 = o_D) = P(A = l_{KC} \mid D_1 = o_{LA}, D_2 = o_D). \tag{3.2}$$

A formal way of saying this is A is conditionally independent of D_1 given D_2, or using symbols, $A \perp D_1 \mid D_2$. Thus, we define conditional independence as a tri-nary relation between disjoint subsets of variables as follows.

Suppose P is a joint probability distribution for variables in $\boldsymbol{X}$, and suppose $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\boldsymbol{X}_3$ are three mutually disjoint subsets of $\boldsymbol{X}$. We say $\boldsymbol{X}_1$ is *conditionally independent* of $\boldsymbol{X}_2$ *given* $\boldsymbol{X}_3$ with respect to P, written as $\boldsymbol{X}_1 \perp \boldsymbol{X}_2 \mid \boldsymbol{X}_3$, if and only if

$$P(\boldsymbol{x}_1 \mid \boldsymbol{x}_2, \boldsymbol{x}_3) = P(\boldsymbol{x}_1 \mid \boldsymbol{x}_3) \quad (3.3)$$

for all states $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$ of $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\boldsymbol{X}_3$, respectively, such that $P(\boldsymbol{x}_2, \boldsymbol{x}_3) > 0$. If $\boldsymbol{X}_3 = \varnothing$, then we say $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are *marginally independent*.

4 CI as Factorization

Consider the example discussed earlier about departures and arrival, in which $A \perp D_1 \mid D_2$. We shall show that this means the joint distribution for D_1, D_2, and A factorizes into two potentials whose domains are $\{A, D_2\}$ and $\{D_1, D_2\}$. It is easy to see that if (3.2) holds, then

$$\begin{aligned} &P(A = l_{KC} \mid D_1 = l_{LA}, D_2 = l_D) \\ &= P(A = l_{KC} \mid D_1 = o_{LA}, D_2 = l_D) \\ &= P(A = l_{KC} \mid D_2 = l_D), \end{aligned}$$

and

$$\begin{aligned} &P(A = l_{KC} \mid D_1 = l_{LA}, D_2 = o_D) \\ &= P(A = l_{KC} \mid D_1 = o_{LA}, D_2 = o_D) \\ &= P(A = l_{KC} \mid D_2 = o_D). \end{aligned}$$

This means that we can replace $P(A \mid D_1, D_2)$ in (1) by $P(A \mid D_2)$. Thus the joint distribution for D_1, D_2, and A factorizes into two potentials, a potential $P(D_1) \otimes P(D_2 \mid D_1)$ whose domain is $\{D_1, D_2\}$, and a potential $P(A \mid D_2)$ whose domain is $\{A, D_2\}$.

The converse is also true, i.e., if we assume that the joint distribution π factorizes into two factors, ϕ with domain $\{D_1, D_2\}$, and γ with domain $\{D_2, A\}$, then $A \perp D_1 \mid D_2$. Let a, d_1, d_2 denote any states of A, D_1, and D_2, respectively. Then

$$\begin{aligned} P(a \mid d_1, d_2) &= P(a, d_1, d_2) \,/\, P(d_1, d_2) \\ &= (\phi(d_1, d_2)\, \gamma(d_2, a)) \,/\, (\phi(d_1, d_2)\, \gamma^{\downarrow D_2}(d_2)) \\ &= \gamma(d_2, a) \,/\, \gamma^{\downarrow D_2}(d_2). \end{aligned}$$

Also

$$\begin{aligned} P(a \mid d_2) &= P(a, d_2) \,/\, P(d_2) \\ &= (\phi^{\downarrow D_2}(d_2)\, \gamma(d_2, a)) \,/\, (\phi^{\downarrow D_2}(d_2)\, \gamma^{\downarrow D_2}(d_2)) \\ &= \gamma(d_2, a) \,/\, \gamma^{\downarrow D_2}(d_2). \end{aligned}$$

Therefore, it follows that $P(a \mid d_1, d_2) = P(a \mid d_2)$, proving our claim.

5 CI as No Double Counting

In the previous cases, we used the language of independence of subset of variables with respect to a joint probability distribution. Here we will shift our language and talk about independence of knowledge with respect to a joint probability distribution. By knowledge, we mean probability potentials. In this section, we provide a formal definition of independence of potentials, which is motivated by the notion of no double counting of uncertain knowledge.

In constructing a joint probability distribution for several variables, we break our knowledge into smaller pieces and then combine these pieces to get the joint distribution. In the process of combining various pieces of knowledge, we need to ensure that we are not double-counting some uncertain knowledge. To take a small example, suppose our prior distribution of D_1 is vacuous, and based on some evidence we get likelihoods as follows: $\delta_1(l_{LA}) = 0.1$, $\delta_1(o_{LA}) = 0.9$. If we include this piece of information twice, e.g., we get the potential $\delta_1 \otimes \delta_1$ which has values $(\delta_1 \otimes \delta_1)(l_{LA}) \approx 0.01$, $(\delta_1 \otimes \delta_1)(o_{LA}) \approx 0.99$. Thus, $\delta_1 \otimes \delta_1 \neq \delta_1$. Double counting of uncertain knowledge overestimates more probable states and underestimates low probable states. In uncertain reasoning, one must be cautious not to double-count uncertain information.

There are exceptions, of course. Some probability potentials are idempotent. Examples are deterministic (or categorical) knowledge. If we observe that $D_1 = l_{LA}$ and express this knowledge by the likelihood potential $\gamma(l_{LA}) = 1$, $\gamma(o_{LA}) = 0$, then $\gamma \otimes \gamma = \gamma$. In propositional logic, all knowledge is expressed without uncertainty and double counting is not an issue.

There is another case of idempotent knowledge. Suppose based on some evidence for D_1, we get equally likely likelihoods, i.e., $\delta_1(l_{LA}) = \delta_1(o_{LA}) = 1/2$. This is a vacuous potential for D_1 and it is also idempotent, and double counting of such knowledge is not a problem.

Definition. We formally define independence of potentials as follows. Suppose ϕ is a potential for f, and γ is a potential for g, and π is a normal potential for $f \cup g$. We say ϕ and γ are *independent* with respect to π if and only if

$$\phi \otimes \gamma = \pi \quad (5.1)$$

Next, we shall state several interesting sufficient conditions for independence of potentials.

Trivial Independence I (Vacuous Likelihoods). Suppose π is a normal potential for g, and suppose ϕ is a vacuous potential for f, where $f \subseteq g$. Since ϕ is vacuous, $\pi \otimes \phi = \pi$. Thus there is no harm in multiplying a normal potential with a vacuous potential. We shall state this as our first lemma.

Lemma 1. Suppose π is a normal potential for g, and suppose ϕ is a vacuous potential for f, where $f \subseteq g$. Then π and ϕ are independent with respect to π.

Trivial Independence II (Observations). There is another case of trivial independence. Suppose we have a joint distribution π of some set of variables g. Now suppose we observe the values $\boldsymbol{a} \in \Omega_f$ of variables in f, where $f \subseteq g$, and $\pi^{\downarrow f}(\boldsymbol{a}) > 0$, and this is represented by observation potential $\lambda_{\boldsymbol{a}}$. The posterior distribution of variables in g is $\pi \otimes \lambda_{\boldsymbol{a}}$. Suppose $\pi^{\downarrow f}$ is a positive potential for f. Since $\lambda_{\boldsymbol{a}}$ is an observation potential, there is no effect by double counting information about f since $\pi \otimes \lambda_{\boldsymbol{a}} = \pi^{\downarrow f} \otimes (\pi \oslash \pi^{\downarrow f}) \otimes \lambda_{\boldsymbol{a}} = (\pi \oslash \pi^{\downarrow f}) \otimes \lambda_{\boldsymbol{a}}$.

Lemma 2. Suppose π is a normal potential for g, and suppose $\lambda_{\boldsymbol{a}}$ is an observation potential for f where $f \subseteq g$, and $\pi^{\downarrow f}(\boldsymbol{a}) > 0$. Then π and $\lambda_{\boldsymbol{a}}$ are independent with respect to $\pi \otimes \lambda_{\boldsymbol{a}}$.

Conditioning. Suppose we wish to construct a joint distribution π of two variables, say D_1 and D_2. Probability theory tells us that one way to do this is to factor π into two pieces, δ_1 with domain D_1 (representing the prior distribution of D_1), and δ_2 with domain $\{D_1, D_2\}$, such that δ_2 is a conditional for D_2 given D_1, i.e., $\delta_2^{-D_2}$ is vacuous. δ_1 contains some information about D_1, and δ_2 contains information of how D_2 is related to D_1. Potentially, there could be double counting of information about D_1. But since δ_2 is a conditional for D_2 given D_1, δ_2 tells us nothing about D_1. So there is no double counting in combining δ_1 and δ_2. We state this as our third lemma.

Lemma 3. Suppose f and g are disjoint subsets of variables, and suppose π is a normal potential for $f \cup g$ such that $\pi^{\downarrow f}$ is positive. Then $\pi^{\downarrow f}$ and $\pi \oslash \pi^{\downarrow f}$ are independent with respect to π.

Marginal Independence. Next, consider the following case of marginal independence in probability theory. Suppose f and g are disjoint subsets of variables that are independent with respect to the joint distribution π with domain $f \cup g$. This means that $\pi(\boldsymbol{x}, \boldsymbol{y}) = \pi^{\downarrow f}(\boldsymbol{x})\ \pi^{\downarrow g}(\boldsymbol{y})$ for all $\boldsymbol{x} \in \Omega_f$, and $\boldsymbol{y} \in \Omega_g$. Thus, we can state our fourth lemma as follows.

Lemma 4. Suppose f and g are disjoint subsets of variables that are marginally independent with respect to the joint distribution π for $f \cup g$. Then the potentials $\pi^{\downarrow f}$ with domain f and $\pi^{\downarrow g}$ with domain g are independent with respect to π.

The condition that f and g are marginally independent with respect to the joint distribution π is crucial in the above lemma. Suppose this condition is not true. Assume $\pi^{\downarrow f}$ is positive. Then we can decompose π into $\phi = \pi^{\downarrow f}$ with domain f and $\gamma = \pi \oslash \pi^{\downarrow f}$ with domain $f \cup g$. This means $\pi = \phi \otimes \gamma$, and $\pi^{\downarrow g} = (\phi \otimes \gamma)^{\downarrow g}$. Thus $\pi^{\downarrow f}$ and $\pi^{\downarrow g}$ are not independent since $\pi^{\downarrow f}$ is being counted twice—it also appears in $\pi^{\downarrow g}$. If indeed f and g are marginally independent, then $\gamma = \pi \oslash \pi^{\downarrow f} = (\pi^{\downarrow f} \otimes \pi^{\downarrow g}) \oslash \pi^{\downarrow f} = \pi^{\downarrow g} \otimes \iota_f$, where ι_f is a vacuous valuation for f, and thus there is no double counting of $\pi^{\downarrow f}$. By a similar reasoning, we can establish that $\pi^{\downarrow f}$ and $\pi^{\downarrow g}$ are not independent since $\pi^{\downarrow g}$ is being double counted (in the same sense as $\pi^{\downarrow f}$).

In the above analysis, double counting of $\pi^{\downarrow f}$ does not literally mean that in combining $\pi^{\downarrow f}$ and $\pi^{\downarrow g}$, we are multiplying $\pi^{\downarrow f}$ twice. Instead, $\pi^{\downarrow f} \otimes \pi^{\downarrow g} = \pi^{\downarrow f} \otimes (\pi^{\downarrow f} \otimes \gamma)^{\downarrow g}$. Thus even if $\pi^{\downarrow f}$ were idempotent by virtue of being equally likely, we would still not have independence between $\pi^{\downarrow f}$ and $\pi^{\downarrow g}$. Even if we had a probability model in which $\pi^{\downarrow f}$ and $\pi^{\downarrow g}$ were both vacuous, we still wouldn't have independence of $\pi^{\downarrow f}$ and $\pi^{\downarrow g}$ with respect to π. A simple numerical example is shown in Table 1.

Consider the following case. Initially, we express our knowledge about variables in $f \cup g$ using potentials $\pi^{\downarrow f}$ with domain f and potential γ with domain $f \cup g$ such that $\pi^{\downarrow f} \otimes \gamma = \pi$. Suppose that f and g are not independent with respect to π. Therefore, $\pi^{\downarrow f}$ and $(\pi^{\downarrow f} \otimes \gamma)^{\downarrow g}$ are not independent with respect to π. Now we get an additional piece of knowledge, say an observation $\boldsymbol{a}$ of the variables in f. Let us denote this knowledge by observation potential $\lambda_{\boldsymbol{a}}$ with domain f. Since $\lambda_{\boldsymbol{a}}$ represents an observation, $\lambda_{\boldsymbol{a}}(\boldsymbol{a}) = 1$ for $\boldsymbol{a} \in \Omega_f$, and $\lambda_{\boldsymbol{a}}(\boldsymbol{y}) = 0$ for all $\boldsymbol{y} \in \Omega_f$, $\boldsymbol{y} \neq \boldsymbol{a}$. Including this knowledge, the joint knowledge is $\lambda_{\boldsymbol{a}} \otimes \pi^{\downarrow f} \otimes \gamma = \pi'$, say. What can we say about independence of variables with respect to the revised distribution π'? First, notice that $\lambda_{\boldsymbol{a}} \otimes \pi^{\downarrow f} = \lambda_{\boldsymbol{a}}$. Second, $\pi'^{\downarrow f} = \lambda_{\boldsymbol{a}}$ since γ is a conditional for g given f. Third, $\pi'^{\downarrow g} = (\lambda_{\boldsymbol{a}} \otimes \gamma)^{\downarrow g}$. Fourth, $\pi'^{\downarrow f} \otimes \pi'^{\downarrow g} = \lambda_{\boldsymbol{a}} \otimes (\lambda_{\boldsymbol{a}} \otimes \gamma)^{\downarrow g} = \pi'$ since $\pi'(\boldsymbol{a}, \boldsymbol{y}) = \lambda_{\boldsymbol{a}}(\boldsymbol{a})\ \gamma(\boldsymbol{a}, \boldsymbol{y}) = \gamma(\boldsymbol{a}, \boldsymbol{y})$, and $(\lambda_{\boldsymbol{a}} \otimes (\lambda_{\boldsymbol{a}} \otimes \gamma)^{\downarrow g})(\boldsymbol{a}, \boldsymbol{y}) = \lambda_{\boldsymbol{a}}(\boldsymbol{a})\ (\lambda_{\boldsymbol{a}} \otimes \gamma)^{\downarrow g}(\boldsymbol{y}) = (\lambda_{\boldsymbol{a}} \otimes \gamma)^{\downarrow g}(\boldsymbol{y}) = \lambda_{\boldsymbol{a}}(\boldsymbol{a})\ \gamma(\boldsymbol{a}, \boldsymbol{y}) + \Sigma\{\lambda_{\boldsymbol{a}}(\boldsymbol{x})\ \gamma(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \neq \boldsymbol{a}\} = \gamma(\boldsymbol{a}, \boldsymbol{y})$, and for $\boldsymbol{x} \neq \boldsymbol{a}$, $\pi'(\boldsymbol{x}, \boldsymbol{y}) = (\lambda_{\boldsymbol{a}} \otimes (\lambda_{\boldsymbol{a}} \otimes \gamma)^{\downarrow g})(\boldsymbol{x}, \boldsymbol{y}) = 0$. Therefore, f and

g are independent with respect to π'. And consequently, $\lambda_{\boldsymbol{a}}$ and $(\lambda_{\boldsymbol{a}}\otimes\gamma)^{\downarrow g}$ are independent with respect to $\lambda_{\boldsymbol{a}}\otimes\gamma$.

	$\pi^{\downarrow A}$		γ	π
a_1	½	b_1	1	½
		b_2	0	0
a_2	½	b_1	0	0
		b_2	1	½

	$\pi^{\downarrow A}$
a_1	½
a_2	½

	$\pi^{\downarrow B}$
b_1	½
b_2	½

	$\pi^{\downarrow A}$	$\pi^{\downarrow B}$	$\pi^{\downarrow A}\otimes\pi^{\downarrow B}$	π
a_1, b_1	½	½	¼	½
a_1, b_2	½	½	¼	0
a_2, b_1	½	½	¼	0
a_2, b_2	½	½	¼	½

Table 1. A probability model for two variables in which $\pi^{\downarrow A}$ and $\pi^{\downarrow B}$ are vacuous and yet $\pi^{\downarrow A}$ and $\pi^{\downarrow B}$ are not independent.

Conditional Independence. Next, we discuss conditional independence. Suppose f, g, and h are mutually disjoint subsets of variables such that f and g are conditionally independent given h with respect to π. This means that $\pi = \pi^{\downarrow h\cup f}\otimes\lambda_g$, where λ_g has domain $g\cup h$. If $\pi^{\downarrow h}$ is positive, then we can write $\lambda_g = \pi^{\downarrow g\cup h}\oslash\pi^{\downarrow h}$, i.e., λ_g is the conditional for g given h. We can now state our next lemma.

> **Lemma 5**. Suppose f, g, and h are mutually disjoint subsets of variables such that f and g are conditionally independent given h with respect to π, and suppose $\pi^{\downarrow h}$ is strictly positive. Then $\pi^{\downarrow f\cup h}$, and $\pi^{\downarrow g\cup h}\oslash\pi^{\downarrow h}$ are independent with respect to π.

Notice that under the assumptions of Lemma 5, $\pi^{\downarrow f\cup h}$ and $\pi^{\downarrow g\cup h}$ may not be independent with respect to π since $\pi^{\downarrow h}$ is double-counted. However, if $\pi^{\downarrow h}$ is idempotent, then $\pi^{\downarrow f\cup h}$ and $\pi^{\downarrow g\cup h}$ are in fact independent with respect to π. This is true even if $\pi^{\downarrow h}$ is not strictly positive. Finally, suppose f and g are conditionally independent given h with respect to π, and suppose we observe the values of variables in h. Let $\lambda_{\boldsymbol{a}}$ denote the observation likelihood potential. Then $(\pi\otimes\lambda_{\boldsymbol{a}})^{\downarrow f}$ and $(\pi\otimes\lambda_{\boldsymbol{a}})^{\downarrow g}$ are independent with respect to $(\pi\otimes\lambda_{\boldsymbol{a}})^{\downarrow(f\cup g)}$.

6 Independence of Potentials and Graphoid Properties

In this section, we show that the definition of independence of potentials stated in (5.1) satisfies the graphoid properties of conditional independence (Dawid [1979], Spohn [1980], Pearl and Paz [1987]).

> **Proposition 1** (*Symmetry*). If ϕ and γ are independent with respect to π, then γ and ϕ are independent with respect to π.

The symmetry property follows trivially from the commutative property of combination. Next, we examine the decomposition property.

> **Proposition 2** (*Decomposition*). Suppose ϕ is a potential for f, and γ is a potential for g such that ϕ and γ are independent with respect to π. Suppose $X\in g$, and $X\notin f$. Then ϕ and γ^{-X} are independent with respect to π^{-X}.

The decomposition property follows from the transitivity of marginalization over combination, a key property that is essential for local computation. Next, we examine the property called weak union.

> **Proposition 3** (*Weak Union*). Suppose ϕ is a potential for f, and γ is a potential for g such that ϕ and γ are independent with respect to π. Suppose $X\in g$, and $X\notin f$, and suppose we observe value a of X. Then ϕ and $(\gamma\otimes\lambda_a)^{-X}$ are independent with respect to $(\pi\otimes\lambda_a)^{-X}$.

Proof. Since ϕ and γ are independent with respect to π, it follows from Lemma 2 that ϕ and $\gamma\otimes\lambda_a$ are independent with respect to $\pi\otimes\lambda_a$. From proposition 2, it follows that ϕ and $(\gamma\otimes\lambda_a)^{-X}$ are independent with respect to $(\pi\otimes\lambda_a)^{-X}$. ■

Next, we look at the contraction property.

> **Proposition 4** (*Contraction*). Suppose f, g, and h are disjoint sets of variables such that $\pi^{\downarrow h\cup f}$, and $\pi^{\downarrow h\cup g}\oslash\pi^{\downarrow h}$ are independent with respect to π. Also suppose that $\pi^{\downarrow h}$ and $\pi^{\downarrow f}$ are independent with respect to $\pi^{\downarrow h\cup f}$. Then $\pi^{\downarrow f}$ and $\pi^{\downarrow h\cup g}$ are independent with respect to π.

Proof. It follows from the first hypothesis that $\pi^{\downarrow h\cup f}\otimes(\pi^{\downarrow h\cup g}\oslash\pi^{\downarrow h}) = \pi$. It follows from the second hypothesis that $\pi^{\downarrow h\cup f} = \pi^{\downarrow h}\otimes\pi^{\downarrow f}$. Substituting the second equality in the first, we get $(\pi^{\downarrow h}\otimes\pi^{\downarrow f})\otimes(\pi^{\downarrow h\cup g}\oslash\pi^{\downarrow h}) = \pi$, i.e., $\pi^{\downarrow f}\otimes\pi^{\downarrow h\cup g} = \pi$, thus proving the proposition. ■

Finally we examine the intersection property.

Proposition 5 (*Intersection*). Suppose π is a positive normal potential for $f \cup g \cup h$ where f, g, and h are disjoint sets of variables. Suppose that $\pi^{\downarrow h \cup f}$, and $\pi^{\downarrow h \cup g} \oslash \pi^{\downarrow h}$ are independent with respect to π, and suppose that $\pi^{\downarrow g \cup f}$, and $\pi^{\downarrow h \cup g} \oslash \pi^{\downarrow g}$ are independent with respect to π. Then $\pi^{\downarrow f}$ and $\pi^{\downarrow g \cup h}$ are independent with respect to π.

Proof. Since $\pi = \pi^{\downarrow h \cup f} \otimes (\pi^{\downarrow h \cup g} \oslash \pi^{\downarrow h})$, it follows that $\pi \oslash \pi^{\downarrow h \cup g} = (\pi^{\downarrow h \cup f} \oslash \pi^{\downarrow h}) \otimes \iota_{h \cup g} = (\pi^{\downarrow h \cup f} \oslash \pi^{\downarrow h}) \otimes \iota_g$. Also, since $\pi = \pi^{\downarrow g \cup f} \otimes (\pi^{\downarrow h \cup g} \oslash \pi^{\downarrow g})$, it follows that $\pi \oslash \pi^{\downarrow h \cup g} = (\pi^{\downarrow g \cup f} \oslash \pi^{\downarrow g}) \otimes \iota_h$. Since the LHS of the last two expressions are equal, their RHS must be equal, i.e., $(\pi^{\downarrow h \cup f} \oslash \pi^{\downarrow h}) \otimes \iota_g = (\pi^{\downarrow g \cup f} \oslash \pi^{\downarrow g}) \otimes \iota_h$. Multiplying both sides by $\pi^{\downarrow h} \otimes \pi^{\downarrow g}$, we get $(\pi^{\downarrow h \cup f} \oslash \pi^{\downarrow h}) \otimes \iota_g \otimes \pi^{\downarrow h} \otimes \pi^{\downarrow g} = (\pi^{\downarrow g \cup f} \oslash \pi^{\downarrow g}) \otimes \iota_h \otimes \pi^{\downarrow h} \otimes \pi^{\downarrow g}$, i.e., $\pi^{\downarrow h \cup f} \otimes \pi^{\downarrow g} = \pi^{\downarrow g \cup f} \otimes \pi^{\downarrow h}$. Marginalizing h from both sides of the equality, we get $\pi^{\downarrow f} \otimes \pi^{\downarrow g} = \pi^{\downarrow g \cup f}$, i.e., $\pi^{\downarrow f}$ and $\pi^{\downarrow g}$ are independent with respect to $\pi^{\downarrow g \cup f}$, and marginalizing g from both sides of the equality, we get $\pi^{\downarrow h \cup f} = \pi^{\downarrow f} \otimes \pi^{\downarrow h}$, i.e., $\pi^{\downarrow f}$ and $\pi^{\downarrow h}$ are independent with respect to $\pi^{\downarrow h \cup f}$. Substituting this last assertion in the first assumption of the proposition, we get $\pi = \pi^{\downarrow h \cup f} \otimes (\pi^{\downarrow h \cup g} \oslash \pi^{\downarrow h}) = \pi^{\downarrow h} \otimes \pi^{\downarrow f} \otimes (\pi^{\downarrow h \cup g} \oslash \pi^{\downarrow h}) = \pi^{\downarrow f} \otimes \pi^{\downarrow h \cup g}$, i.e., $\pi^{\downarrow f}$ and $\pi^{\downarrow g \cup h}$ are independent with respect to π. ■

7 Two Examples

In this section, we describe two examples to illustrate the definition and results in the previous sections.

Example 1. Consider a graphical probability model for three variables, H (hypothesis), E_1 (evidence 1), and E_2 (evidence 2), as shown in Figure 1. According to this graphical model, $E_1 \perp E_2 \mid H$.

Assume all three variables are binary with states h and nh for H, e_1 and ne_1 for E_1, and e_2 and ne_2 for E_2. Let η denote the prior distribution for H, ε_1 the conditional distributions for E_1 given H, and ε_2 the conditional distributions for E_2 given H. By Lemmas 3 and 5, η, ε_1, and ε_2 are independent (with respect to π, the prior joint distribution of H, E_1, and E_2). Thus, the prior joint distribution of H, E_1, and E_2 is given by $\pi = \eta \otimes \varepsilon_1 \otimes \varepsilon_2$. Suppose we have two pieces of evidence: we observe states e_1 of E_1, and e_2 of E_2, such that $\pi^{\downarrow E_1}(e_1) > 0$ and $\pi^{\downarrow E_2}(e_2) > 0$. We can represent these pieces of evidence by observation potentials λ_{e_1} and λ_{e_2}. By Lemma 2, the posterior joint distribution of H, E_1, and E_2 is given by $\eta \otimes \varepsilon_1 \otimes \varepsilon_2 \otimes \lambda_{e_1} \otimes \lambda_{e_2}$. The marginal posterior distribution of H is given by $(\eta \otimes \varepsilon_1 \otimes \varepsilon_2 \otimes \lambda_{e_1} \otimes \lambda_{e_2})^{-\{E_1, E_2\}} = \eta \otimes (\varepsilon_1 \otimes \lambda_{e_1})^{-E_1} \otimes (\varepsilon_2 \otimes \lambda_{e_2})^{-E_2}$ (using Proposition 3). Thus we conclude that η, $(\varepsilon_1 \otimes \lambda_{e_1})^{-E_1}$, and $(\varepsilon_2 \otimes \lambda_{e_2})^{-E_2}$ are independent potentials for H and combining these gives the correct posterior marginal distribution for H based on the prior and the two independent pieces of evidence. Using symbols,

$$\begin{pmatrix} P(h \mid e_1, e_2) \\ P(nh \mid e_1, e_2) \end{pmatrix} = \begin{pmatrix} P(h) \\ P(nh) \end{pmatrix} \otimes \begin{pmatrix} K_1 P(e_1 \mid h) \\ K_1 P(e_1 \mid nh) \end{pmatrix} \otimes \begin{pmatrix} K_2 P(e_2 \mid h) \\ K_2 P(e_2 \mid nh) \end{pmatrix}$$
$$= \begin{pmatrix} K_3 P(h) P(e_1 \mid h) P(e_2 \mid h) \\ K_3 P(nh) P(e_1 \mid nh) P(e_2 \mid nh) \end{pmatrix},$$

where K_1, K_2, K_3 are normalization constants that do not depend on the states of H. Notice that independence of evidence potentials $(\varepsilon_1 \otimes \lambda_{e_1})^{-E_1}$ and $(\varepsilon_2 \otimes \lambda_{e_2})^{-E_2}$ follows from the Markov assumption of the graphical model that $E_1 \perp E_2 \mid H$. ■

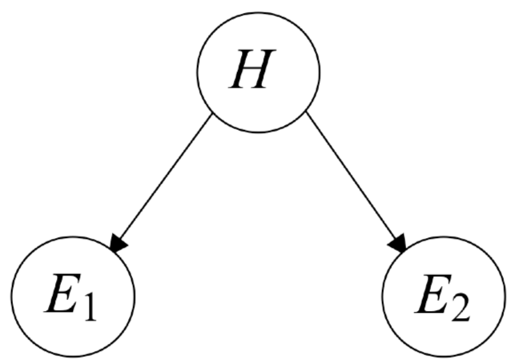

Figure 1. A graphical model for H, E_1, and E_2.

Example 2. Consider a graphical probability model for three variables, H_1 (hypothesis 1), H_2 (hypothesis 2), and E (evidence), as shown in Figure 2. According to this graphical model, $H_1 \perp H_2$.

Assume all three variables are binary with states h_1 and nh_1 for H_1, h_2 and nh_2 for H_2, and e and ne for E. Let η_1 denote the prior distribution for H_1, η_2 the prior distribution for H_2, and ε the conditional distribution for E given H_1 and H_2. By Lemmas 4 and 3, the potentials η_1, η_2, and ε are independent (with respect to π, the prior joint distribution of H_1, H_2, and E). Thus, the prior joint distribution of H_1, H_2, and E is given by $\pi = \eta_1 \otimes \eta_2 \otimes \varepsilon$.

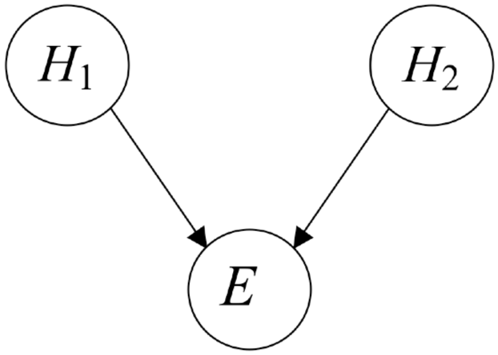

Figure 2. A graphical model for H_1, H_2, and E.

Suppose we have one piece of evidence: we observe state e of E such that $\pi^{\downarrow E}(e) > 0$. We can represent this piece of evidence by observation potential λ_e for E. By Lemma 2, the posterior joint distribution of $\{H_1, H_2, E\}$ is given by $\eta_1 \otimes \eta_2 \otimes \varepsilon \otimes \lambda_e$. The posterior marginal distribution of H_1, H_2 is given by $(\eta_1 \otimes \eta_2 \otimes \varepsilon \otimes \lambda_e)^{-E}$ = $(\eta_1 \otimes \eta_2 \otimes (\varepsilon \otimes \lambda_e)^{-E}$ (using Proposition 3). Thus we conclude that η_1, η_2, and $(\varepsilon \otimes \lambda_2)^{-E}$ are independent potentials (for H_1, H_2, and $\{H_1, H_2\}$, respectively) and combining these gives the correct posterior distribution for $\{H_1, H_2\}$ based on the priors and the evidence. Using symbols,

$$\begin{pmatrix} P(h_1, h_2 \mid e) \\ P(h_1, nh_2 \mid e) \\ P(nh_1, h_2 \mid e) \\ P(nh_1, nh_2 \mid e) \end{pmatrix} = \begin{pmatrix} P(h_1) \\ P(nh_1) \end{pmatrix} \otimes \begin{pmatrix} P(h_2) \\ P(nh_2) \end{pmatrix} \otimes \begin{pmatrix} K_1 P(e \mid h_1, h_2) \\ K_1 P(e \mid h_1, nh_2) \\ K_1 P(e \mid nh_1, h_2) \\ K_1 P(e \mid nh_1, nh_2) \end{pmatrix}$$

$$= \begin{pmatrix} K_2 P(h_1) P(h_2) P(e \mid h_1, h_2) \\ K_2 P(h_1) P(nh_2) P(e \mid h_1, nh_2) \\ K_2 P(nh_1) P(h_2) P(e \mid nh_1, h_2) \\ K_2 P(nh_1) P(nh_2) P(e \mid nh_1, nh_2) \end{pmatrix},$$

where K_1 and K_2 are normalization constants that do not depend on the states of $\{H_1, H_2\}$. Notice that although H_1 and H_2 are marginally independent, they are not independent after the evidence if the evidence is not vacuous. ■

8 Summary and Conclusions

We have defined conditional independence in terms of no double counting of uncertain knowledge. We state five sufficient conditions for independence of potentials, and we have showed that our definition of independence of potentials satisfies the graphoid properties of conditional independence.

Although we have stated all results so far for probability theory, the results generalize easily to any uncertainty calculi that fit in the framework of valuation-based systems. Shenoy [1994] shows that probability theory, Dempster-Shafer's belief function theory, Spohn's [1988] theory of epistemic beliefs, and Zadeh's [1979] possibility theory (with combination defined as pointwise multiplication followed by normalization when normalization is possible) are some examples of uncertainty calculi that are captured by the axiomatic framework of valuation-based systems. Smets's [1998] transferable belief model, and Kohlas-Monney's [1995] theory of hints are mathematically the same as D-S theory, and consequently, the results stated here apply to these theories as well. Qualitative possibility theory, in which the combination rule is pointwise minimization does not fit the axiomatic framework of valuation-based system, and consequently, the results stated here do not apply to it.

The results stated here also apply to the D-S belief function theory. In this theory, Dempster's rule of combination is only supposed to be used for "independent" belief functions. Using the results in this paper, and the framework of valuation-based systems [Shenoy 1994], it is easy to see that independence of belief functions is mathematically equivalent to the conditional independence theory for variables in probability theory.

Acknowledgements

This article was written during my visit to the Decision Support Systems group in the Department of Computer Science, Aalborg University, in Spring 2004. I am grateful to Finn V. Jensen, Thomas D. Nielsen and other members of this group for many valuable comments and for providing an intellectually stimulating environment for this research. I am also grateful to Philippe Smets, Arthur P. Dempster, Liping Liu, and three anonymous reviewers of ISIPTA for comments on an earlier draft.

References

[1] Dawid, A. P. (1979). Conditional independence in statistical theory (with discussion). *Journal of the Royal Statistical Society,* Series B **41**(1): 1–31.

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

[3] Kohlas, J. and P.-A. Monney (1995). *A Mathematical Theory of Hints: An Approach to Dempster-Shafer Theory of Evidence*. Lecture Notes in Economics and Mathematical Systems, **425**: Springer-Verlag.

[4] Pearl, J. and A. Paz (1987). Graphoids: A graph based logic for reasoning about relevance relations. *Advances in Artificial Intelligence*. B. D. Boulay *et al*. Amsterdam, North-Holland. **2:** 357–363.

[5] Smets, P. (1998). The transferable belief model for quantified belief representation. *Handbook of Defeasible Reasoning and Uncertainty Management Systems*. D.M. Gabbay and P. Smets. Dordrecht, Kluwer Academic Publishers. **1**: 267–301.

[6] Spohn, W. (1980). Stochastic independence, causal independence, and shieldability. *Journal of Philosophical Logic*, **9**: 73–99.

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

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

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

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