Fundamental Concepts of Qualitative Probabilistic Networks*

Michael P. Wellman
AI Technology Office, WRDC/TXI, Wright-Patterson AFB, OH 45433, USA

ABSTRACT

Graphical representations for probabilistic relationships have recently received considerable attention in AI. Qualitative probabilistic networks abstract from the usual numeric representations by encoding only qualitative relationships, which are inequality constraints on the joint probability distribution over the variables. Although these constraints are insufficient to determine probabilities uniquely, they are designed to justify the deduction of a class of relative likelihood conclusions that imply useful decision-making properties.

Two types of qualitative relationship are defined, each a probabilistic form of monotonicity constraint over a group of variables. Qualitative influences describe the direction of the relationship between two variables. Qualitative synergies describe interactions among influences.

The probabilistic definitions chosen justify sound and efficient inference procedures based on graphical manipulations of the network. These procedures answer queries about qualitative relationships among variables separated in the network and determine structural properties of optimal assignments to decision variables.

1. Introduction

Many knowledge representation schemes, including the various flavors of "causal networks" [36, 44, 57], qualitative physical models [58], and belief networks [39], model the world as a collection of states, events, or other ontological primitives connected by links that describe their interrelationships. The representations differ widely in the nature of the fundamental objects and in the precision and expressiveness of the relationship links.

Qualitative probabilistic networks (QPNs) occupy a region in representation space where the objects are arbitrary variables, and the relationships are qualitative constraints on the joint probability distribution among them. This area is important for AI research because the relation among variables is often uncertain due to incomplete knowledge or modeling, and because strictly numeric representations are inappropriately precise for many applications.

* Portions of this paper previously appeared in the Proceedings of AAAI-87, Seattle, WA [60].

Artificial Intelligence **44** (1990) 257–303
0004-3702/90/$03.50

Excess precision leads to knowledge bases applicable in only narrow domains and to diminished modularity because interactions increasingly arise at finer levels of detail [20, 64].

The qualitative relationships expressible in the QPN formalism are designed to afford robustness yet permit a reasoner to deduce useful properties about optimal assignments to the specially designated decision variables in the network. These "useful properties" are facts that enable a planner to reduce the search space of possible courses of action. The nature of these decision properties and the qualitative relationships leading to them are developed in the body of this paper.

1.1. Motivation

Expected benefits of the analysis of qualitative probabilistic networks fall into three primary categories.

(1) *Probabilistic semantics for a common knowledge base construct.* Relations similar in intent to those expressible in QPNs have been applied widely in AI knowledge bases without serious attempts at formalization, probabilistic or otherwise. The analysis below suggests how such constructs might be interpreted and in some cases dictates how they *must* be interpreted to justify inferences drawn by associated reasoners.

(2) *Qualitative reasoning methods for domains where signs of associations are not guaranteed, and functional relations are not deterministically fixed.* Many applications of qualitative reasoning are not faithful to the underlying assumptions behind a "qualitative differential equations" interpretation. Taking an explicit probabilistic approach reveals the possible pitfalls of such violations. This issue is discussed further in Section 8.4 below.

(3) *Efficient inference techniques to support tasks in planning under uncertainty.* As mentioned above, the qualitative relationships are designed to assist a planner by determining some facts about the admissible plans. Indeed, this representation was originally developed within a planning context [61, 62]. In the examples and discussion below we will see how inferences derived from QPNs can constrain the *structure* of strategies that need to be considered by a planner.

1.2. Preview of the paper

Section 2 formally introduces qualitative probabilistic networks, relates them to numeric graphical probabilistic representations, and presents an example from the domain of digitalis therapy. The digitalis example illustrates the use of *qualitative influences*, qualitative relations describing the sign (direction) of the relationship between a pair of variables.

The next four sections elaborate the semantics, properties, and application of qualitative influences. A formal probabilistic definition for them is motivated and developed in Section 3. Section 4 describes inference mechanisms that are sound with respect to this definition and presents an efficient algorithm for answering queries about the qualitative influences holding among arbitrary variables in the network. Section 5 considers alternative probabilistic semantics and shows that the definition of Section 3 is the weakest satisfying the inference mechanisms of Section 4. Application of these techniques to the digitalis example is the subject of Section 6.

Qualitative synergies, which describe the qualitative interaction among influences, are defined, defended, and analyzed in Section 7. This section also presents graphical algorithms for reasoning about synergies in QPNs similar to those for qualitative influences. Analysis of the digitalis model enhanced with synergy assertions demonstrates that useful properties of the preferred therapy plan follow from purely qualitative assertions.

Section 8 contrasts the qualitative probabilistic network representation with related work in AI, decision theory, and statistics. The relevance of these results to previous qualitative reasoning applications is also discussed. A perspective on the significance of this work is offered in the final section.

2. Qualitative Probabilistic Networks

2.1. Network models

A network model is a graph-like structure with nodes that represent variables and edges and hyper-edges that describe relationships among them. In a probabilistic model, the values of variables as well as their interrelationships are uncertain, defined by a probability distribution over the joint value space. Probabilistic network models have attracted much recent attention in AI, for example in Pearl's work on belief networks [39] and related formalisms [5, 29, 53]. The network formalism developed here is accurately viewed as a qualitative abstraction of *influence diagrams* [22], which are belief networks with additional constructs to support decision making. In particular, all properties holding of a belief network or influence diagram by virtue of structure alone are also true of the corresponding QPN. Some QPN terminology, notation, and even solution concepts (by analogy) are borrowed from Shachter's work on influence diagram evaluation [47, 48].

Formally, a qualitative probabilistic network is a pair $G = (V, Q)$. V is the set of *variables*, or vertices of the graph. Q is a set of *qualitative relationships* among the variables. The qualitative influences and synergies in Q correspond to directed edges and hyper-edges, respectively, in the graph G. To be a valid QPN, G must be acyclic with respect to influence edges.

Variables, named by lower-case symbols, are associated with a set of possible values, for example, boolean for propositional event variables, or real intervals for continuous parameters. Unlike most numeric schemes, there is no practical requirement to reformulate the value spaces into discrete, finite sets. Let $X(a)$

denote the domain of variable a. The domain of a tuple of variables is the product space of the individual domains, for example, $X(\langle a, b\rangle) = X(a) \times X(b)$. The tuple is written as a set when the ordering is insignificant. Subscripted symbols denote values in the domain of a variable.

The variable set V may contain one special variable v, called the *value node*.[1] Relationships involving v express preferences over the other variables.

It is also useful to distinguish a set $D \subseteq V - \{v\}$ of *decision variables*. A decision-making program takes variables in D to be under its control and therefore focuses on deriving the implications on v of choosing different values for them. The remainder of the variables in the network are random variables not under direct control of the decision maker.

Qualitative relationships express constraints on the joint probability distribution over the variables. Unlike the numeric conditional probabilities specified in belief networks and influence diagrams, they are not generally sufficient to determine the exact distribution. In fact, in a purely qualitative network the absolute likelihood of any joint event is completely unconstrained! Nevertheless, the qualitative relationships are carefully designed to justify the deduction of a class of *relative* likelihood conclusions that in turn imply useful decision-making properties. Note that nothing prevents us from building hybrid models combining qualitative relationships with those more precise, although the present work does not pursue that possibility.

There are two types of qualitative relationships in QPNs. *Qualitative influences* describe the direction of the relationships between two variables. *Qualitative synergies* describe interactions among influences. These concepts form the basis of the QPN formalism and are developed in detail below.

2.2. Example: The Digitalis Therapy Advisor

The development of QPN concepts is illustrated with a simple causal model taken from Swartout's programs for digitalis therapy [55]. The model, shown in Fig. 1, is a fragment of the knowledge base that Swartout used to re-implement the Digitalis Therapy Advisor [16] via an automatic programmer.

In the figure the circular nodes represent random variables. The rectangular node is a decision variable, in this case denoting the dosage of digitalis (*dig*) administered to the patient. The value node v is drawn within a hexagon and represents the utility of the outcome of the patient. Qualitative influences among the variables are indicated by dependence links, annotated with a sign denoting the direction of the relationship. The link asserts that the variables are related monotonically, in a precise probabilistic sense elucidated below.[2]

[1] This name is unfortunate because v actually represents a *utility function*, often distinguished in decision theory from the *value function*. Nevertheless, the term is retained because it is well entrenched in the vocabulary of influence diagrams.

[2] Discussion of qualitative synergies holding in this example is deferred to Section 7.

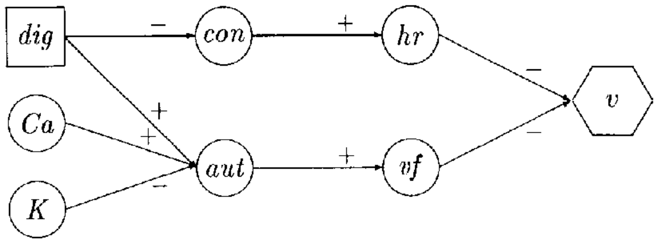

Fig. 1. Part of the causal model for digitalis therapy. The sign on a link form node a to node b indicates the effect of an increase in a on b.

According to the model, digitalis negatively influences conduction (*con*) and positively influences automaticity (*aut*). The former is the desired effect of the drug, because a decrease in conduction *decreases* the heart rate (*hr*), which is considered beneficial for patients with tachycardia, the population of interest here. The desirability of lower heart rates is represented by the negative influence on the value node, asserting that lower rates increase expected utility. The negative influence is obviously valid only to a point; a universal objective of therapy, after all, is to keep heart rates significantly above zero. In interpreting conclusions from these models it is important to heed the qualifying assumption that variables remain within the monotonic range of their relationships.

The increase in automaticity is an undesired side-effect of digitalis because this variable is positively related to the probability of ventricular fibrillation (*vf*), a life-threatening cardiac state. Calcium (*Ca*) and potassium (*K*) levels also influence the level of automaticity.

There are no links into the decision variable because the digitalis dosage is considered by the model to be under direct control.

A qualitative encoding of this model is appropriate for the knowledge base of a general digitalis therapy program because a numeric description would require additional context information or be inaccurate. While the exact probabilistic relationships among these variables vary from patient to patient, the *directions* of the relations are reliably taken as constant. Conclusions drawn from this model are therefore valid for a broad class of patients.

The conclusions we would like our programs to derive from the digitalis model are those taken for granted in the description above. For example, we unthinkingly assumed that the effects of digitalis on conduction and of conduction on heart rate would combine to imply that digitalis reduces the heart rate. Further, because lower heart rates are desirable, digitalis is *therapeutic* along the upper path. Conversely, it is *toxic* along its lower path to the value node. The tradeoff between therapy and toxicity cannot be resolved by mere qualitative influences.

The immediate task of this paper is to develop a semantics for these qualitative influences that justifies the kind of inferences we require while providing the maximum robustness. In the sections below, I provide such a semantics in terms of a probabilistic definition for qualitative influences. In Section 5 we will see that this definition is the weakest in a reasonable class that justifies the conclusions mentioned above.

3. Qualitative Influences Defined

3.1. Influence notation

The qualitative links in the digitalis model above can be represented formally as edges in the graph annotated by sign. Let $S^{\delta}(a, b, G)$ denote the assertion that a qualitative influence of a on b in direction (that is, sign) δ holds in graph $G = (V, Q)$.

Definition 3.1 (*Qualitative influence edges*). $S^{\delta}(a, b, G) \equiv (a, b, \delta) \in Q$, where δ is one of $+$, $-$, 0, or ?.

By convention, S^0 links are left implicit in graphical displays of the network. They would also typically be left implicit—inferable via a closed-world assumption—in data structures representing qualitative networks. For example, a representation of the QPN of Fig. 1 would explicitly record $(dig, con, -)$ in Q but would leave $(dig, vf, 0)$ implicit in the absence of a signed link.

The *pred* function selects only the predecessors exerting nonzero influences on a variable.

Definition 3.2 (*Predecessors*).

$$pred_G(b) \equiv \{a \mid (a, b, \delta) \in Q, \text{ for some } \delta \in \{+, -, ?\}\},$$

$$pred_G^*(b) \equiv \bigcup_{c \in pred_G(b)} [\{c\} \cup pred_G^*(c)].$$

Note that for all $d \in D$, $pred_G(d) = \emptyset$. The subscript G is omitted when its value is clear from context.

3.2. Probabilistic semantics for qualitative influences

Consider two variables, a and b. Informally, when a and b denote boolean events, a qualitative influence is a statement of the form "a makes b more (or less) likely." This binary case is easy to capture in a probabilistic assertion. Let A and $\bar{A}$ denote the assertions $a = \textbf{true}$ and $a = \textbf{false}$, respectively, and similarly, B and $\bar{B}$.

Definition 3.3 (*Binary S^+*). We say "a positively influences b" (stochastically) and write $S^+(a, b, G)$, if and only if (iff) for all $x \in X(pred_G(b) - \{a\})$ such that x is consistent with both A and $\bar{A}$,[3]

$$\Pr(B \mid Ax) \geqslant \Pr(B \mid \bar{A}x) . \tag{1}$$

In Definition 3.3, the *context* x ranges over all consistent assignments to the variables other than a that influence b. (Henceforth, x in a formula will be understood as universally quantified over the values of predecessor variables.) Thus, S^+ is analogous to Forbus' qualitative proportionality, α_{Q^+} [13], which is an inequality on partial derivatives, also universally quantified over contexts. We need to include the *ceteris paribus* condition here and in the definitions below so that qualitative relations will be applicable in situations where x is partially or totally known. If we had stated the S^+ definition in marginal terms ("on average, a positively influences b"), it would not be valid to apply it in specific contexts.

Because its definition refers to a specific predecessor set, S^+ holds in a particular network; programs that alter the structure of the network may exhibit nonmonotonicity in S^+ relative to its first two arguments [17]. In the following I omit the third argument only when the intended network is unambiguous or inessential.

Conditions analogous to (1) and those following define negative and zero influences; I omit them for brevity. S^0, an assertion that (1) holds with equality, is the familiar concept of conditional independence of a and b given b's direct influences. We could rule out the independent case with strict versions of S^+ and S^-, but discussion is limited to non-strict influences for this paper.

$S^?$ always holds. It is included explicitly only so that we can represent S^0 implicitly in the lack of an influence assertion.

For dichotomous variables, it is not hard to show that Bayes' rule implies that (1) is equivalent to

$$\Pr(A \mid Bx) \geqslant \Pr(A \mid \bar{B}x) . \tag{2}$$

In the terminology of Bayesian belief revision, (1) is a condition on *posteriors*, while (2) is a condition on *likelihoods*. Notice that $S^+(a, b)$ is simply an assertion that the *likelihood ratio* is greater than or equal to unity.

Formalizing the intuitive idea that "higher values of a make higher values of b more likely" is not quite as straightforward when a and b take on more than two values. An obvious prerequisite for such statements is some interpretation of "higher." Therefore, we require that each random variable appearing in an S^+ and S^- assertion be associated with an order $\geqslant$ on its values. This relation has the usual interpretation for numeric variables such as "potassium concentration"; for variables like "automaticity," an ordering relation must be contrived.

[3] We can safely ignore cases where the conditional probabilities are undefined because these are impossible contexts.

The more troublesome part of defining positive influences is specifying what it means to "make higher values of b more likely." Intuitively, we want a statement that the probability distribution for b shifts toward higher values as a increases. To make such a statement, we need an ordering on cumulative probability distribution functions (CDFs) F_b over b that captures the notion of "higher."

However, probability distributions cannot be straightforwardly ordered according to the size of the random variable. Different rankings result from comparison of distributions by median, mean, or mean-log, for example. We require an ordering that is robust to changes of these measures because the random variables need be described by merely *ordinal* scales [27]. An assertion that calcium concentration positively influences automaticity should hold whether calcium is measured on an absolute or logarithmic scale, and regardless of how we measure automaticity.

An ordering criterion with the robustness we desire is *first-order stochastic dominance* (FSD) [65]. FSD holds for CDFs F_b and F_b' iff for any given value b_0 of b, the probability of obtaining b_0 or less is smaller for F_b than for F_b'. That is, F_b FSD F_b' iff

$$\forall b_0 \; F_b(b_0) \leq F_b'(b_0)\,. \tag{3}$$

A necessary and sufficient condition for (3) is that for all monotonically increasing (that is, order-preserving) functions ϕ,

$$\int \phi(b_0)\,\mathrm{d}F_b(b_0) \geq \int \phi(b_0)\,\mathrm{d}F_b'(b_0)\,. \tag{4}$$

That is, the mean of F_b is greater than the mean of F_b' for *any* monotonic transform of b. For further discussion and a proof that (3) is equivalent to (4), see [12].

We are now ready to define qualitative influences.

Definition 3.4 (S^+). Let $F_b(\cdot\,|a_i x)$ be the CDF for b given $a = a_i$ and context x. Then $S^+(a, b)$ iff

$$\forall a_1,a_2.\; a_1 \geq a_2 \;\Rightarrow\; F_b(\cdot\,|a_1 x)\ \mathrm{FSD}\ F_b(\cdot\,|a_2 x)\,. \tag{5}$$

Definition 3.4 is a generalization of Definition 3.3 under the convention that **true** > **false** for binary events.

Like (1), (5) is a condition on posteriors. A comparable definition of S^+ in terms of likelihood must imply FSD of the posteriors for *any prior distribution* F_b. That is, we allow that there may be a context x inducing any distribution on b. Milgrom [32] proves that the following condition is necessary and sufficient for (5) to hold for any $F_b(\cdot\,|x)$.

$$\forall a_1, a_2, b_1, b_2.$$

$$(a_1 \geq a_2) \wedge (b_1 \geq b_2) \Rightarrow \frac{f_a(a_1 \mid b_1 x)}{f_1(a_1 \mid b_2 x)} \geq \frac{f_a(a_2 \mid b_1 x)}{f_a(a_2 \mid b_2 x)}\,. \tag{6}$$

In (6), $f_a(\cdot \mid b_i x)$ is the probability density function for a given b_i and x.

This condition is known in statistics as the Monotone Likelihood Ratio Property (MLRP) [2]. The necessity of MLRP for (5) is established by the special case of dichotomous events. That (6) is a generalization of (2) is more clearly seen by rewriting the latter as

$$\frac{\Pr(A \mid Bx)}{\Pr(A \mid \bar{B}x)} \geq 1 \geq \frac{\Pr(\bar{A} \mid Bx)}{\Pr(\bar{A} \mid \bar{B}x)}\,. \tag{7}$$

For a demonstration of the sufficiency of MLRP, see [32].

It is convenient to adopt special notation for influences on the value node v. The value node is related to its predecessors by a *utility function* $u : X(pred(v)) \to \mathbb{R}$ [25].

Definition 3.5 (U^+). The variable a positively influences utility, $U^+(a)$, iff

$$\forall a_1, a_2.\ a_1 \geq a_2 \Rightarrow u(a_1, x) \geq u(a_2, x)\,. \tag{8}$$

The definition of $U^+(a)$ is a special case of $S^+(a, v)$ taking into account the deterministic relation (a degenerate probability distribution) between v and its predecessors in the network.

4. Indirect Relationships

Edges in a graph of influence links constrain the direct relationship between pairs of variables. Our next step is to design inference mechanisms to derive the indirect relationships that follow from patterns of local influences.

First, let us define the *canonical direction* between two variables to be the strongest qualitative influence derivable from those explicitly appearing in Q. The canonical direction can be easily computed from Q by preferring an explicit 0 to the other δ's (which are always consistent with 0 because the conditions are non-strict), preferring + or − to ?, and replacing the conjunction of + and − with 0.

Definition 4.1 (*dir*). Let $\Delta = \{\delta \mid (a, b, \delta) \in Q\}$. The canonical direction of influence of a on b, $dir(a, b, G)$, is given by

$$dir(a,b,G)=\begin{cases}\text{undefined}, & \text{if } b\in pred_G^*(a)\,,\\ 0\,, & \text{if } \Delta=\emptyset,\ 0\in\Delta,\ \text{or } \{+,-\}\subseteq\Delta\,,\\ +\,, & \text{if } \Delta=\{+\} \text{ or } \Delta=\{+,?\}\,,\\ -\,, & \text{if } \Delta=\{-\} \text{ or } \Delta=\{-,?\}\,,\\ ?\,, & \text{otherwise}\,.\end{cases}$$

If $dir(a, b) = ?$ then a and b are dependent in an unknown, varying, or context-dependent direction.

4.1. Probabilistic dependence in graph representations

Definition 4.2 (*dep*). The dependency graph, $dep(G)$, of $G = (V, Q)$ is

$$dep(G) = (V, E)\,,$$

where

$$(a, b) \in E \ \text{ iff } \ dir(a, b, G) \in \{+, -, ?\}\,.$$

The dependency graph simply encodes the pattern of nonzero influences without distinguishing the signs on the links. Pearl et al. [41] have characterized the expressiveness of these graphs with respect to the dependency structure of probability distributions. Some results of this work and terminology developed there will prove useful in analyzing the properties of QPNs.

In a directed acyclic graph representation, two variables are conditionally independent given any set of other variables that *d-separates* them in the graph.

Definition 4.3 (*d-separation*, Pearl [37]). Two variables a and b are *d-separated* by a set of variables S in a directed acyclic graph iff for every *undirected* path from a to b either:

(1) there is a node $s \in S$ on the path with at least one of the incident edges leading out of s, or
(2) there is a node t on the path with both incident edges leading in, and neither t nor any of its successors are in S.

The concept of d-separation is illustrated by the network of Fig. 2.

The following implication of Definition 4.3 is useful in justifying the inference rules for QPNs presented below.

Lemma 4.1. *If $b \not\in pred_G^*(a)$, then a and b are d-separated in $dep(G)$ by any S such that $pred_G(b) \subseteq S \subseteq \{s \mid b \not\in pred_G^*(s)\}$.*

For convenience, all proofs are relegated to Appendix A.

Taking $S = pred_G(b)$, this result is the basis for our closed-world assumption that $dir(a, b, G) = 0$ if there are no explicit influences in Q. If in addition there

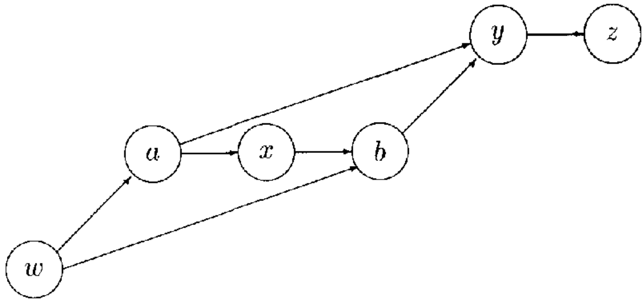

Fig. 2. Variables a and b are d-separated by $\{w, x\}$ but by no other subset of $\{w, x, y, z\}$.

are no directed paths from b to a, we adopt the default influence $S^0(a, b)$. In Pearl's terminology, this assumption is valid when $dep(G)$ is an *I-map*—a graph for which all d-separations are true conditional independencies.

4.2. Network transformations

We answer queries about relations among separated variables in a QPN by transforming the graph into one where the variables of interest are related directly. The method is based on Shachter's algorithm for evaluating numeric influence diagrams [47] by repeated reductions and arc reversals. Each manipulation preserves the probabilistic relationships—qualitative in our case—holding among variables in the possibly smaller set V. It is possible via sequences of these manipulations to answer queries about the relationships among any subset of variables in the network [48].

The two basic network transformation operators are reduction (*red*) and reversal (*rev*). The reduced network $red(b, G)$ is the qualitative probabilistic network obtained by splicing variable b out of G and adjusting qualitative influences as dictated by Theorem 4.3 below. The reversed network $rev(a, b, G)$ is obtained from G by replacing $(a, b, \delta) \in Q$ with the influence (b, a, δ) and updating other influences as specified in Theorem 4.4.

Let $G' = (V', Q')$ be the result of one of these transformation operations. For simplicitly we adopt the convention that Q' contains only the canonical directions. The relationship between Q and Q' is described in Section 4.3 below. Both the *red* and *rev* operations preserve essential properties of the networks:

- $dep(G')$ is acyclic.
- $dep(G')$ is an I-map.

4.3. Variable reductions

It can be demonstrated for the binary case that, in the absence of direct links from a to c,

$$S^+(a, b, G) \wedge S^+(b, c, G) \Rightarrow S^+(a, c, red(b, G)) .$$

The ability to perform inference across influence chains is an essential property of a qualitative algebra. From the digitalis model, for example, we would like to deduce that increasing the dose of digitalis decreases the heart rate but increases the likelihood of ventricular fibrillation. Indeed, most programs with models like this would make such an inference. Fortunately, the definition offered above for S^+ implies transitivity for multi-valued as well as binary variables.

Theorem 4.2.

$$S^{\delta_1}(a, b, G) \wedge S^{\delta_2}(b, c, G) \wedge S^0(a, c, G) \Rightarrow$$
$$S^{\delta_1 \otimes \delta_2}(a, c, red(b, G)) ,$$

where $\delta_i \in \{+, -, 0, ?\}$ *and* $\otimes$ *denotes sign multiplication, described in Fig.* 3.

Application of Theorem 4.2 requires that no direct influences exist between a and c. A more general specification of the result of variable reduction is the following:

Theorem 4.3.

$$S^{\delta_1}(a, b, G) \wedge S^{\delta_2}(b, c, G) \wedge S^{\delta_3}(a, c, G) \Rightarrow$$
$$S^{(\delta_1 \otimes \delta_2) \oplus \delta_3}(a, c, red(b, G)) , \tag{9}$$

where $\oplus$ *denotes sign addition, also described in Fig.* 3.

Theorem 4.2 is really a corollary of Theorem 4.3 with $\delta_3 = 0$, the identity element for $\oplus$.

This result tells us how to update the direction between pairs of predecessors and successors of a reduced variable. When the reduced variable has at most one successor, Theorem 4.3 covers all necessary changes to the network. When there are multiple successors, however, the removal may render them dependent, as the modified network does not contain the original d-separating set.

$\otimes$	+	−	0	?
+	+	−	0	?
−	−	+	0	?
0	0	0	0	0
?	?	?	0	?

$\oplus$	+	−	0	?
+	+	?	+	?
−	?	−	−	?
0	+	−	0	?
?	?	?	?	?

Fig. 3. The $\otimes$ operator for combining influence chains and the $\oplus$ operator for combining parallel influences. For example, $+ \otimes - = -$. The operations commute, associate, and distribute like ordinary multiplication and addition.

To reflect these dependencies, we must add new influences among these variables. One way to compute the new relations is to reverse all but one of the successor links according to the procedure of the next section, then perform reduction as described above.

We can solve some special problems using reduction transformations exclusively. If the variables can be ordered for reduction so that they have at most one successor when reduced, the updating rule of Theorem 4.3 is sufficient. For example, to find the qualitative influence of a on b given a set of variables W for any $a \in pred^*(b)$ and $pred^*(a) \subseteq W \subseteq pred^*(b)$ we need only splice out all other variables in the network. The restrictions on a and W ensure that the variables can be ordered for simple reduction. Because each application of reduction rule (9) reduces the number of influence edges (including zeros) in the network by one, the complexity of this procedure is $O(|V|^2)$.[4] This contrasts with the corresponding problem for numeric probabilistic networks, which is NP-hard [6]. Some sample reduction sequences are displayed in Fig. 4.

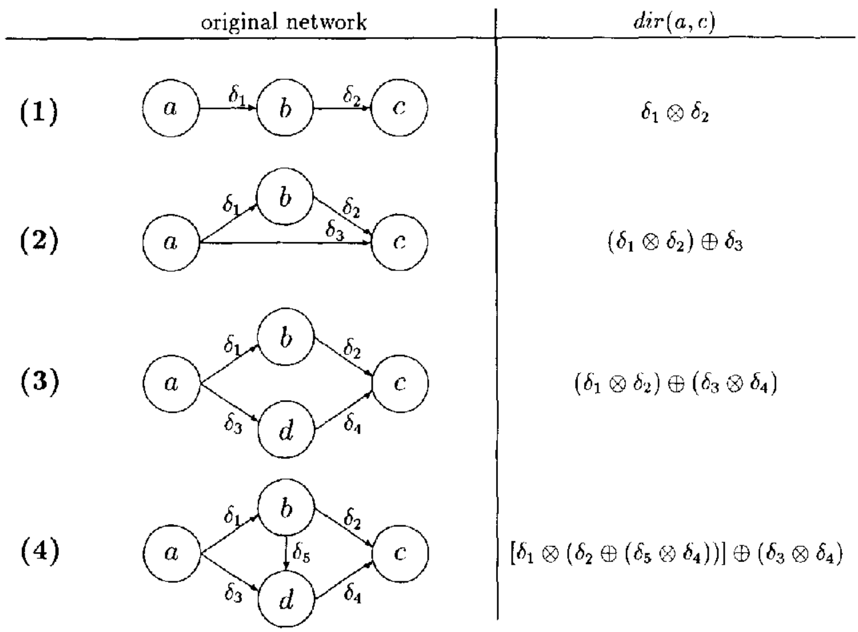

Fig. 4. Some sample reduction sequences. The right column contains the expression for $dir(a, c)$ in the network obtained by removing nodes between a and c. Fragments (1) and (2) correspond to the situations of Theorems 4.2 and 4.3, respectively. When $a \in pred^*(b)$, the relation with all intermediate variables reduced is simply the sign sum of all directed paths from a to b.

[4] Since it is possible to construct cases where this algorithm requires $\Omega(|V|^2)$ operations, its overall complexity is $\Theta(|V|^2)$.

4.4. Influence reversals

The procedure developed above is valid when a precedes b in the network and the variables of W are intermediate between the two. Often, however, the network is such that the qualitative influence of a on b cannot be determined by any sequence of single-successor reductions. In such cases we need to perform one or more reversals in the network before or after applying the methods of the previous section.

In reversing a qualitative influence link, we must preserve the essential properties mentioned in Section 4.2 above. To ensure acyclicity, we can reverse the influence from a to b only if there are no other directed paths between them. Reversal is also precluded if a is a decision variable. To guarantee that the dependency structure is valid after reversal (that is, $G' = rev(a, b, G)$ is an I-map), we generally have to insert additional links. As demonstrated by Shachter [48], it is sufficient that each of the two variables gain the other's predecessors:

$$pred_{G'}(a) = pred_G(a) \cup pred_G(b) \cup \{b\} - \{a\} ,$$

$$pred_{G'}(b) = pred_G(a) \cup pred_G(b) - \{a\} .$$

Definition 3.4 for $S^{\delta}(a, b)$ explicitly refers to the predecessors of b. Therefore, when the predecessor structure changes we need to recompute the influences that may be affected. The following result describes the influences holding after reversal.

Theorem 4.4. *Let $G' = rev(a, b, G)$. G' inherits all the qualitative influences of G except*:

(1) $dir(a, b, G')$ *is undefined.*

(2) $dir(b, a, G') = dir(a, b, G)$.

(3) $\forall w \in pred_{G'}(b)$,

$$dir(w, b, G') = [dir(w, a, G) \otimes dir(a, b, G)] \oplus dir(w, b, G) .$$

(4) $\forall w \in pred_{G'}(a) - \{b\}$,

$$dir(w, a, G') = \begin{cases} dir(w, a, G) , & \textit{if } dir(w, b, G) = 0 \\ ? & \textit{otherwise} \end{cases}$$

$$= dir(w, a, G) \oplus (dir(w, b, G) \otimes ?) .$$

This transformation is illustrated in Fig. 5.

Some information may be lost in the process of reversing influences. For example, let $G'' = rev(a, b, rev(a, b, G))$, the network obtained by reversing an influence then reversing it again. Application of Theorem 4.4 twice yields the result depicted in Fig. 6. Although the link from a to b is correct, the reversal process weakens the other links. More generally, the prospect of information

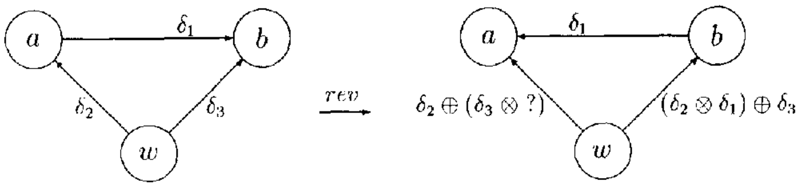

Fig. 5. Influence reversal.

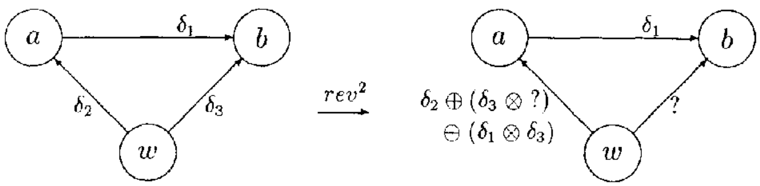

Fig. 6. Information lost in a double reversal of the influence from a to b.

lost suggests that the strategy for transforming a network may have significant impact on the strength of conclusions obtained. Analysis of this and related issues can be found in another paper focusing on inferential properties of QPNs [63].

5. Necessity Results

The preceding sections establish that the FSD condition for S^+ (Definition 3.4) is sufficient to support essential inferences such as the chaining of influences. In this section I present some simple desiderata for a qualitative influence definition that entail the *necessity* of FSD for these properties.

5.1. Posterior conditions

I start by specifying the form such definitions must take. To capture the intent of "higher values of a make higher values of b more likely" in a probabalistic semantics, it seems reasonable to restrict attention to conditions on the posterior distribution of b for increasing values of a. Therefore, I postulate that a definition of $S^+(a, b)$ must be of the form

$$\forall a_1, a_2 .\, a_1 \geqslant a_2 \Rightarrow F_b(\cdot \,|a_1 x)\, R\, F_b(\cdot \,|a_2 x)\,, \tag{10}$$

where R is some relation on CDFs. This condition is exactly (5) with FSD replaced by the more abstract relation.

There are two basic desiderata that severely restrict the possible relations R. First, S^+ must satisfy Theorem 4.2. Without the ability to chain inferences, the qualitative influence formalism has little computational value. Second, the condition must be a generalization of the original specification of S^+ for dichotomous variables (Definition 3.3). With only two possible values this appears to be a minimal monotonicity condition. These criteria lead to a sharp conclusion.

Theorem 5.1. *Let $S^+(a, b)$ be defined by* (10). *Given the following conditions:*
(1) *Theorem* 4.2,
(2) *for binary b, $a_1 \geq a_2$, and x,*

$$F_b(\cdot\,|a_1 x)\; R\; F_b(\cdot\,|a_2 x) \Leftrightarrow \Pr(B\,|\,a_1 x) \geq \Pr(B\,|\,a_2 x)\,,$$

the weakest R is FSD.

The force of this result is weakened somewhat by the a priori restriction of definitions to those having the form of (10). Many statistical concepts of directional relation (based on correlation or joint expectations, for example) do not fit (10) yet appear to be plausible candidates for a definition of qualitative influence. *Quadrant dependence* [30] holds between a and b when[5]

$$\forall a_1, a_2.\ a_1 \geq a_2 \Rightarrow F_b(\cdot\,|a \leq a_1)\ \mathrm{FSD}\ F_b(\cdot\,|a \leq a_2)\,. \tag{11}$$

Lehmann proves that quadrant dependence is necessary but not sufficient for *regression dependence*, which is his terminology for (5) without the quantification over contexts x. As quadrant dependence is weaker, yet still exhibits transitivity,[6] it seems to be an attractive alternative to regression dependence. To justify our choice of the latter, we must consider the decision-making implications of probabilistic models.

5.2. Decision-making with qualitative influences

The prime motivation for adopting a probabilistic semantics is so that the behavior of our programs can be justified by Bayesian decision theory [46]. A decision of d_1 over d_2 (that is, such a choice of assignments to decision variables) is valid with respect to a QPN if the network entails greater expected utility for the former. The most useful distinctions to make in designing a qualitative representation are those that will support inferences about properties of the valid decisions.

For example, if a positively influences utility (Definition 3.5) and there are no *indirect* paths from decision variable a to the value node, then a choice of a_1

[5] This is actually the condition Lehmann proposes as a strengthening of quadrant dependence. The basic quadrant dependence fixes a_1 at a's maximal value.

[6] For transitivity we need to quantify over contexts in (11). The proof parallels that for Theorem 4.2.

over a_2 is valid iff $a_1 \geqslant a_2$.[7] Decision-making power is enhanced if we can deduce new influences on utility from chains of influences in the network. Our definition of qualitative influence is necessary as well as sufficient for such inferences.

Theorem 5.2. *Suppose $U^{\delta_2}(b, G)$ and $U^0(a, G)$. A necessary and sufficient condition for $U^{\delta_1 \otimes \delta_2}(a, red(b, G))$ is $S^{\delta_1}(a, b, G)$ as in Definition* 3.4.

Figure 7 depicts this situation with $\delta_2 = +$.

Theorem 5.2 demonstrates that while conditions weaker than S^+, such as quadrant dependence, may be sufficient for propagating influences across chains, they are not adequate to justify *decisions* across chains. For choosing among alternatives, the relevant parameter is the utility function evaluated at a point; utilities conditioned on intervals of the decision variable (as in quadrant dependence) do not have the same decision-making import.

5.3. Simpson's paradox

Because qualitative influences are based on simple intuitive relationships, they may provide insight into qualitatively counterintuitive situations. One celebrated example is Simpson's paradox, in which a factor is shown to have positive impact on some result in all contexts (precisely the definition of S^+), yet its overall influence is negative.

In an instance presented by Blyth [3], patients given an experimental treatment have an increased chance of survival in each of two test cities. However, when the statistics from the cities are pooled, it turns out that patients with the treatment have a *decreased* survival probability. How is this possible? In this example, the population of $city_1$ have a significantly better prognosis and patients from $city_2$ are more likely to be treated. Thus, a treated patient is more likely to come from $city_2$ and is therefore less likely to survive.

A QPN modeling this example would have qualitative influences $S^+(treat, survive)$, $S^+(city, survive)$ (adopting the convention $city_1 > city_2$), and $S^-(treat,$

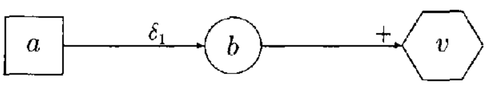

Fig. 7. Chaining utility influences. The influence $\delta_1 = +$ in G is necessary and sufficient for $U^+(a, red(b, G))$.

[7] The existence of other paths from a to utility would leave open the possibility that the net influence of a is negative. For example, we could summarize the therapeutic effect of digitalis through conduction and heart rate as a direct positive influence. But this might be outweighed by the indirect negative influence of digitalis via automaticity.

city). Reducing *city* according to (9) leads to the ambiguous conclusion $S^?(treat, survive)$, indicating that Simpson's paradox is a possibility in this circumstance.

If there had been no interaction between the likelihood of treatment and the patient's residence, or if the interaction had been in the other direction, Simpson's paradox could not arise. The following is a direct consequence of the QPN update rules, Theorems 4.3 and 4.4.

Corollary 5.3. *Suppose* $V = \{a, b, c\}$ *and* $S^+(a, c, G)$, *and let* $\delta_{a,b}$ *be* $dir(b, a, G)$ *or* $dir(a, b, G)$, *whichever is defined. A necessary condition for Simpson's paradox to apply for a and c with respect to b is*:

$$\delta_{a,b} \otimes dir(b, c, G) \in \{-, ?\} .$$

Although the phenomena surrounding Simpson's paradox are well understood, QPNs provide a convenient language for expressing these enabling conditions.[8] With a, b, and c standing for *treat*, *city*, and *survive*, respectively, the result applies directly to the example above. (Incidentally, if the residence of the patient is known at the time of a treatment decision, the model correctly mandates that the treatment should be administered.)

6. Back to the Digitalis Model

To summarize the discussion of qualitative influences thus far, let us return to the digitalis example presented in Section 2.2. We are interested in computing the effect of the decision variable, *dig*, on utility. The network of Fig. 1 reduces to the one depicted in Fig. 8(a), which further reduces to that of Fig. 8(b).

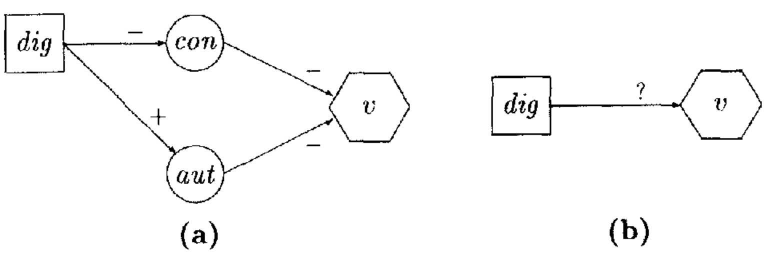

Fig. 8. Reduction of the digitalis model. (a) Digitalis is therapeutic in its effect on conduction but toxic via the influence on automaticity. (b) The overall effect of digitalis cannot be resolved with qualitative influences.

[8] Whether it obtains in a given case depends on the numeric probabilities, that is, we cannot express *sufficient* conditions for Simpson's paradox in terms of qualitative influences. See Neufeld and Horton [35] for a discussion of Simpon's paradox in the context of another formalism based on probabilistic inequalities.

The result, not surprisingly, is ambiguous. Purely qualitative influences are too weak to determine optimal decisions in the presence of true tradeoffs. Nevertheless, the QPN is sufficient to determine some influences (for example, calcium on ventricular fibrillation), and uncovers the source of indeterminacy in others.

In the next section, a second type of qualitative relationship is introduced: qualitative synergy. Synergies complement influences by constraining the interactions among probabilistic influences. Although synergies cannot resolve the tradeoff of Fig. 8(b), they can provide useful facts about the relation of the optimal digitalis dosage to other variables in the model.

7. Qualitative Synergy

Swartout's XPLAIN knowledge base includes the "domain principle" that if a state variable acts synergistically with the drug to induce toxicity, then smaller doses should be given for higher observed values of the variable [55]. This fact could be derived by a *domain-independent* inference procedure given a suitable definition for qualitative synergy. Two variables synergistically influence a third if their joint influence is greater (in the sense of FSD) than separate, statistically independent influences. In the digitalis example, we need to assert that digitalis acts at least independently with Ca and K deviations in increasing automaticity. For the desired result, we also need the fact that heart rate and ventricular fibrillation are synergistic in their influence on utility. (This synergy is due to our indifference to heart rate—indeed it is undefined—for patients in fibrillation. The relation of this indifference to synergy is clarified in Section 7.7 below.)

Figure 9 illustrates the QPN for digitalis enhanced by synergy assertions. Potassium (K) is omitted for simplicity; its implications are analogous (with sign reversal) to those for calcium. Qualitative synergies are indicated by a boxed sign with multiple inputs and a single output. The input variables are synergistic in the designated direction in their influence on the output variable.

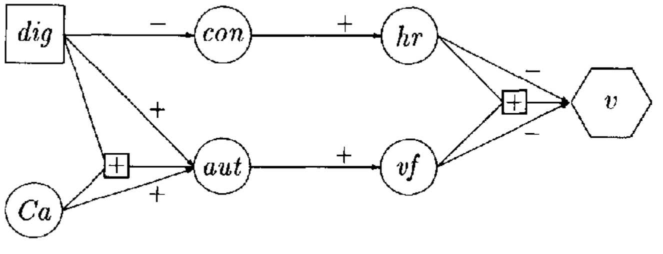

Fig. 9. The digitalis model with synergy. A boxed sign indicates that the inputs are qualitatively synergistic in their influence on the output.

7.1. Synergy notation

Qualitative synergies are the second type of qualitative relationship represented in Q for a QPN $G = (V, Q)$. As qualitative influences are directed edges augmented by sign, qualitative synergies are directed hyper-edges with a sign label. A qualitative synergy assertion that the variables in $T \subset V$ are synergistic in direction δ on variable w is written $Y^{\delta}(T, w, G)$.

Definition 7.1 (*Qualitative synergy hyper-edges*).

$$Y^{\delta}(T, w, G) \equiv (T, w, \delta) \in Q .$$

7.2. Qualitative synergy defined

A formal definition of qualitative synergy must capture the informal intuition expressed above that the "joint influence is greater than separate statistically independent influences." This will be the case when the effect of varying one variable is enhanced by simultaneous variation of the other.

The joint influence of two variables a and b on a third, c, is expressed by the conditional cumulative distribution for c, $F_c(\cdot\,|ab)$. To compare magnitude of "influence," we need some reference points. One way to measure a difference in influence is to take the difference of two conditional CDFs. Two variables are synergistic if the difference associated with raising one is greater (in the sense of FSD) for higher values of the second.

Definition 7.2 (*Qualitative synergy*, Y^{δ}). Variables a and b are synergistic on c in network G, written $Y^{+}(\{a, b\}, c, G)$ iff

$$\begin{aligned} &\forall a_1, a_2, b_1, b_2, c_0 \,. \\ &\quad (a_1 \geq a_2) \wedge (b_1 \geq b_2) \Rightarrow \\ &\quad F_c(c_0 \,|\, a_1 b_1 x) - F_c(c_0 \,|\, a_2 b_1 x) \leq F_c(c_0 \,|\, a_1 b_2 x) - F_c(c_0 \,|\, a_2 b_2 x) \,. \end{aligned} \tag{12}$$

Replacing $\leq$ in condition (12) by $\geq$ or $=$ defines subsynergy or zero synergy (Y^{-} and Y^{0}), respectively. If the variable set T in $Y^{\delta}(T, w, G)$ contains more than two elements, the condition above holds for all *pairs* of variables in T.

As usual, x ranges over assignments to the other predecessors of c.

The inequality (12) quantified over c_0 can be viewed as stochastic dominance of the respective distributions of CDF differences. The condition means that raising a from a_2 to a_1 has a greater effect for higher values of b. Note that the inequality is symmetric in a and b.

If $S^{0}(a, c)$, then $Y^{0}(\{a, w\}, c)$ follows immediately for any variable $w \in$ *pred*$^*(c)$ because of conditional independence. With conditional independence,

$F_c(\cdot\,|a_1wx) = F_c(\cdot\,|a_2wx)$ for all w and x, therefore both sides of equation (12) are zero.

Lacking an explicit synergy assertion for two or more variables that are predecessors of another, the prudent closed-world assumption is $Y^?$: no constraint on their interaction.[9] Although it is reasonable to assume S^0 in the absence of knowledge to the contrary, in this case, the variables are tied by a common immediate successor. They are *not* d-separated by this successor, and interactions in situations with this pattern are quite common.

Fortunately, there are several prototypical patterns of systematic interaction that might alleviate the burden of specifying qualitative synergies. One that has attracted some interest in the literature on numeric probabilistic networks is the "noisy OR gate" model proposed by Pearl [39, Chapter 4].

In the noisy OR model, the binary-valued predecessors of a binary "effect" variable are considered separate possible causes of the effect. Each "cause" variable is associated with a parameter p_i representing the probability of the effect given that this variable is true and all other predecessors are false. We can compute the rest of the conditional probabilities for y under the assumption that the "inhibiting events" that prevent Y given each Z_i are independent. For effect variable y with predecessors $z_1, \ldots, z_n$, the conditional probabilities are:

$$\Pr(Y\,|\,z_1 \ldots z_n) = 1 - \prod_{\{i\,|\,Z_i\}} (1 - p_i)\,. \tag{13}$$

Regardless of the magnitudes of the parameters p_i, the noisy OR model entails subsynergy, Y^-. To see this, consider the Y^- condition ((12) with the inequality reversed) for the special case of binary variables. $Y^-(\{z_j, z_k\}, y)$ iff:

$$\begin{aligned}&\forall x \in X(\{z_i\,|\,(i \neq j) \wedge (i \neq k)\})\\ &\quad \Pr(Y\,|\,Z_jZ_kx) - \Pr(Y\,|\,\bar{Z}_jZ_kx) \leq \Pr(Y\,|\,Z_j\bar{Z}_kx) - \Pr(Y\,|\,\bar{Z}_j\bar{Z}_kx)\,.\end{aligned} \tag{14}$$

Let

$$p_x = \prod_{Z(x)} (1 - p_i)\,,$$

where

$$Z(x) = \{i\,|\,Z_i \text{ in assignment } x\}\,.$$

Then from the noisy OR model (13),

$$\Pr(Y\,|\,Z_jZ_kx) - \Pr(Y\,|\,\bar{Z}_jZ_kx) = p_x\,p_j(1 - p_k) \tag{15}$$

and

$$\Pr(Y\,|\,Z_j\bar{Z}_kx) - \Pr(Y\,|\,\bar{Z}_j\bar{Z}_kx) = p_x\,p_j\,. \tag{16}$$

[9] In the examples of this section, all synergies are specified explicitly.

Because $0 \leq p_k \leq 1$, the expression in (15) is not greater than that of (16), satisfying the binary Y^- condition (14). Figure 10 illustrates the relation between a numeric probabilistic network using the noise OR model and its corresponding QPN.

It is also easy to verify that Henrion's generalizations of the noisy OR model [21] entail Y^-. Intuitively, a noisy OR is subsynergistic because, as with deterministic OR gates, raising an input has less effect when other inputs are already raised. In contrast, a model based on a probabilistic generalization of "gating conditions" (see Rieger and Grinberg [44]) would be synergistic because an increase in one variable enables the effect of the other. More generally, we should expect non-? synergy results from canonical models because any representation that specifies an n-way influence in terms of $O(n)$ parameters must employ some systematic assumption about interactions.[10]

7.3. Supermodularity, Y^δ, and monotone decisions

The Y^δ definition relates closely to the concept of *supermodular* functions [45, 56].

Definition 7.3 (*Supermodularity*, Ross [45]). A function g such that, for all $a_1 \geq a_2$ and $b_1 \geq b_2$:

$$g(a_1, b_1) + g(a_2, b_2) \geq g(a_1, b_2) + g(a_2, b_1) \tag{17}$$

is called *supermodular*. If (17) holds with equality, then g is *modular*, and if the inequality is reversed, g is *submodular*.

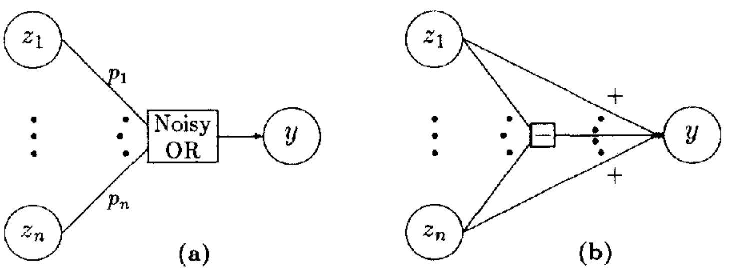

Fig. 10. (a) The "noisy OR" model, and (b) its corresponding qualitative abstraction.

[10] Dempster's rule of combination is also subsynergistic under an analogous definition of synergy in terms of belief functions [49]. A demonstration of this requires further assumptions regarding how to interpret conditioning as evidence combination.

The most important property of supermodular functions, from our perspective, is that they imply *monotone decisions*. Let the function $a_g(b)$ choose the value of a that maximizes g for the given b.

$$a_g(b) = arg \max_a g(a, b) .$$

It can be shown that $a_g(\cdot)$ increases monotonically in b if g is supermodular (see Ross [45, p. 6]).

The following result clarifies the connection between Y^δ and supermodularity.

Lemma 7.1. *$Y^+(\{a, b\}, c)$ (respectively Y^- and Y^0) holds iff the function*

$$e_\phi(a, b \mid x) = \int \phi(c_0) f_c(c_0 \mid abx)\, dc_0$$

is supermodular (submodular, modular) in a and b for all increasing functions ϕ and contexts x.

The function e_ϕ is the expectation of c under the monotonic transform ϕ. The equivalence between *sub*modularity of F_0 for all c (Definition 7.2, the Y^+ condition) and *super*modularity of expectations for all ϕ is reminiscent of the correspondence between the FSD condition (3) and increasing expectations for all ϕ (4).

Once again, it is useful to define special notation for synergistic influences on the value node.

Definition 7.4 (Y_U^δ). Variables a and b are synergistic on utility, $Y_U^\delta(\{a, b\}, G)$, for $\delta = +, -, 0$, iff u is supermodular, submodular, or modular, respectively, in a and b.

Note that $Y_U^\delta(T, G)$ is weaker than $Y^\delta(T, v, G)$, as the condition on u need not hold for all monotonic transformations.

In the terminology of utility theory, δ-modularity expresses *multi-attribute risk aversion*, *proneness*, or *neutrality* as δ is $-$, $+$, or 0, respectively [9, 43]. Multi-attribute risk neutrality is equivalent to additive separability for u [11], as suggested by the form of the modularity condition (17).

The correspondence between Y_U^δ and supermodularity is useful because of the monotone decision property of supermodular functions. Consider the situation of Fig. 11. There we have $Y_U^+(\{a, b\})$ even though $dir(a, v) = dir(b, v) = ?$. Qualitative influences alone tell us nothing about which value we should choose for the decision variable a. Positive synergy, on the other hand, implies that if b is observable, our policy should be to choose higher values of a

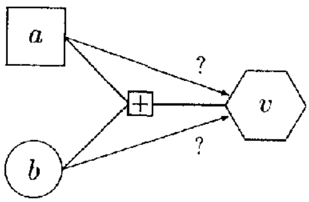

Fig. 11. Synergistic influence on utility. Even though $U^?(a)$ and $U^?(b)$ we can deduce that the optimal choice of a is increasing in b.

for greater values of the observed b. While this still does not reveal the exact value of the optimal a, it dictates the *form* that our strategy should take.

7.4. Propagation of synergies in networks

The mechanisms for deducing indirect synergies that hold in a QPN are analogous to the network transformation techniques for qualitative influences developed in Section 4. In particular, we can extend qualitative synergies through qualitative influences by variable reduction.

Theorem 7.2. *Synergies can be extended along qualitative influences by reduction according to the following*:

$$Y^{\delta_1}(\{a, b\}, c, G) \wedge S^{\delta_2}(c, d, G) \wedge S^0(a, d, G) \wedge S^0(b, d, G) \Rightarrow$$
$$Y^{\delta_1 \otimes \delta_2}(\{a, b\}, d, red(c, G)) .$$

This reduction is depicted in Fig. 12.

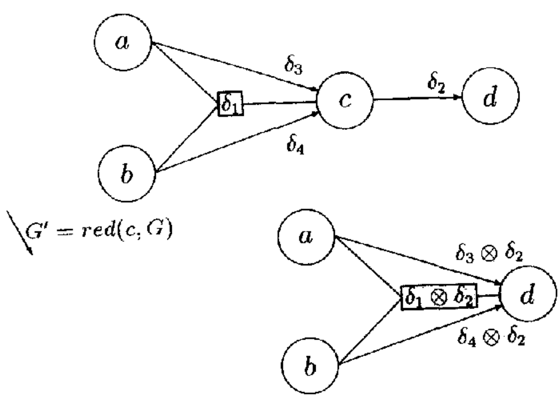

Fig. 12. Propagation of synergy through qualitative influences. Values for $dir(a, d, G')$ and $dir(b, d, G')$ follow from Theorem 4.2. The new synergy $Y^{\delta_1 \otimes \delta_2}$ is the result of Theorem 7.2.

Like Theorem 4.2, Theorem 7.2 requires that there be no direct influences among the variables newly linked in the reduced QPN. The next result provides the reduction rule for the more general case.

Theorem 7.3.

$$\begin{aligned}&Y^{\delta_1}(\{a,b\},c,G)\wedge S^{\delta_2}(c,d,G)\wedge Y^{\delta_3}(\{a,c\},d,G)\wedge\\&Y^{\delta_4}(\{b,c\},d,G)\wedge S^{\delta_5}(a,c,G)\wedge S^{\delta_6}(b,c,G)\wedge Y^{\delta_7}(\{a,b\},d,G)\Rightarrow\\&Y^{(\delta_1\otimes\delta_2)\oplus(\delta_3\otimes\delta_6)\oplus(\delta_4\otimes\delta_5)\oplus\delta_7}(\{a,b\},d,\mathit{red}(c,G))\,.\end{aligned}\tag{18}$$

Theorem 7.3 generalizes Theorem 7.2 because

$$S^0(a,d,G)\wedge S^0(b,d,G)\Rightarrow\delta_3=\delta_4=\delta_7=0$$

by conditional independence.

Note that the signs of direct influences from a and b to d do not affect the synergy propagation, though the signs of influences on c do. This more complicated situation is illustrated in Fig. 13.

A special case of the foregoing results demonstrates how to propagate synergies *backwards* through qualitative influences. Upon reduction, a vari-

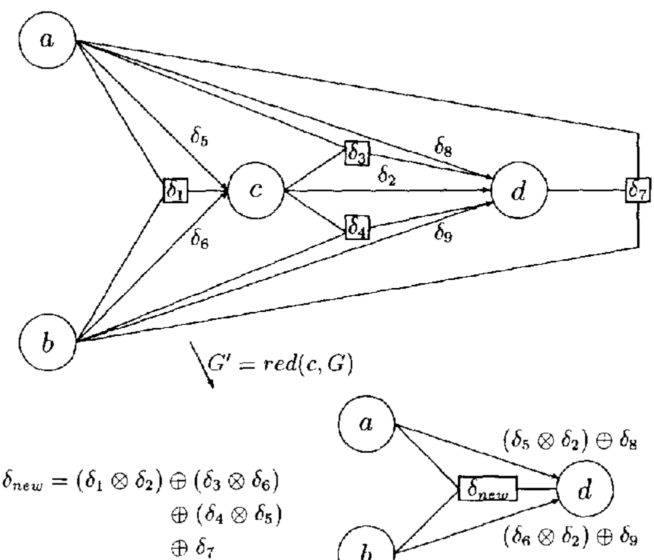

Fig. 13. Variable reduction with parallel synergies.

able's predecessors assume its role in all synergies, with modified signs reflecting the direction of the predecessor's influence.

Corollary 7.4.

$$Y^{\delta_3}(\{a, c\}, d, G) \wedge S^{\delta_6}(b, c, G) \wedge S^0(a, c, G) \Rightarrow$$
$$Y^{\delta_3 \otimes \delta_6}(\{a, b\}, d, red(c, G)) .$$

The results follows from the assignment $\delta_1 = \delta_5 = \delta_7 = 0$ in Theorem 7.3. (The zero synergy of a and b on c, δ_1, follows from the zero influence of a on c, δ_5.) Application of Corollary 7.4 is illustrated in Fig. 14.

For an example of the use of backwards propagation, consider a synergy relation from the digitalis model. In the more detailed model of Fig. 15, the effects of variables *dig* (digitalis dosage) and *Ca* (measured serum calcium) would be mediated by *dig'* and *Ca'*, the actual concentrations of digitalis and calcium in the bloodstream. Even though the synergy assertion is in terms of the physiological parameters, we can deduce synergy on the practically relevant proxy variables by reduction according to Corollary 7.4.

Though the definition for Y_U^δ differs from Y^δ, the synergy update rule (18) also holds when d is the value node and Y_U is substituted for Y as appropriate. In fact, for backwards propagation the Y_U^δ condition is *exactly* preserved.

Theorem 7.5. *Given $S^{\delta_6}(b, c, G)$ and $S^0(a, c, G)$, $Y_U^{\delta_3}(\{a, c\}, G)$ is both necessary and sufficient for $Y_U^{\delta_3 \otimes \delta_6}(\{a, b\}, red(c, G))$.*

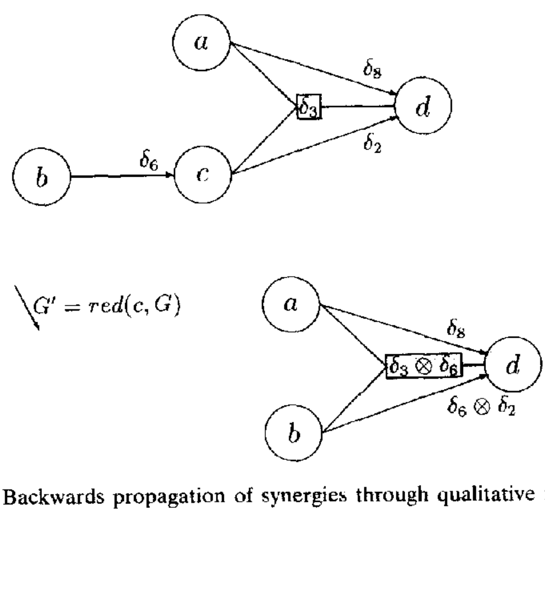

Fig. 14. Backwards propagation of synergies through qualitative influences.

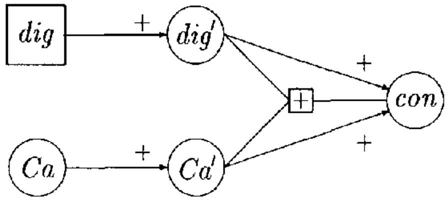

Fig. 15. An elaboration of a digitalis model fragment. Variables *dig* and *Ca* represent dosage and measurement, respectively, while the primed versions are actual concentrations. The unprimed variables are synergistic by reduction of *dig'* and *Ca'*.

A canonical decision situation with the above form is the estimation problem from statistics. The problem is to choose an estimate a of the true "state of nature" θ given only an observation z that is statistically related to θ. Karlin and Rubin [24] demonstrate that if

(1) the optimal estimate is increasing in θ (the monotone decision property of Section 7.3),
(2) utility decreases away from the optimum, and
(3) z is related to θ by the MLRP (the likelihood condition for S^+ (6), Section 3.2),

then a and z also satisfy the monotone decision property.

By representing the estimation problem as the QPN of Fig. 16, we see that the sufficiency part of Theorem 7.5 is a similar result, with the monotone decision property replaced by the stronger condition of qualitative synergy. Synergy seems justified for the estimation problem because the relative value of a higher estimate increases with the state of nature.

The applicability of the setup in Fig. 16 goes well beyond estimation. Suppose the state of nature θ represents an unobservable disease severity and the decision variable a the aggressiveness of therapy. Choosing a therapy level is similar to estimating the severity of disease, as more serious conditions call for stricter treatments. It is essential that a program be capable of inferring the qualitative implications for therapy of any symptom z related to disease severity in a known direction.

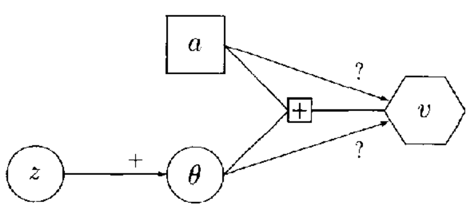

Fig. 16. A qualitative probabilistic network for the estimation problem.

7.5. Synergy reversal

Synergies must also be updated upon reversal of a link. Consider a reversal of the influence from c to d in the network of Fig. 13 (top half). Synergies on d are revised (or newly created) according to the following rule.

Theorem 7.6.

$$Y^{\delta_1}(\{a, b\}, c, G) \wedge S^{\delta_2}(c, d, G) \wedge Y^{\delta_3}(\{a, c\}, d, G) \wedge$$

$$Y^{\delta_4}(\{b, c\}, d, G) \wedge S^{\delta_5}(a, c, G) \wedge S^{\delta_6}(b, c, G) \wedge Y^{\delta_7}(\{a, b\}, d, G) \Rightarrow$$

$$Y^{(\delta_1 \otimes \delta_2) \oplus (\delta_3 \otimes \delta_6) \oplus (\delta_4 \otimes \delta_5) \ominus \delta_7}(\{a, b\}, d, rev(c, d, G)) .$$

After reversal, the possibility of interactions with d render all synergies on c ambiguous. Synergies on variables other than c or d are unaffected by the operation.

7.6. Landmark values

The monotone decision property can be used to develop a concept of *landmark values* for QPNs analogous to the landmark value concept in qualitative simulation [28]. A landmark value is any distinguished point in the domain of a variable. Their usefulness to qualitative reasoning accrues when landmark values of several variables correspond in a meaningful way or the point has some other qualitative significance for the application.

In QPNs, the interesting landmarks are optimal values of decision variables and the corresponding values of observable non-decision variables. Suppose that in the disease-severity interpretation of Fig. 16, the variable z represents an observable symptom with a specially designated "normal" value of z^*. There is a corresponding landmark value of the decision variable, a^*, representing the optimal level of therapy given $z = z^*$. The value of a^* may be known to the program, especially if there is documented experience with z-normal patients, everything else being equal. Even if its exact value is not known, or if it depends on other variables, the a^* concept has meaning as a landmark value in terms of its optimality property.

Suppose further that a patient presents with an elevated z-value of $z' > z^*$. The qualitative implication drawn from our model is that the corresponding optimal therapy a' is increased, $a' \geqslant a^*$, all else being equal. As correspondences in the quantity space [13] are known in finer detail, the program can determine optimal strategies with increasing precision.

7.7. Synergy in the digitalis example

To complete our discussion of qualitative synergy, let us return to the digitalis model of Fig. 9. As promised, I start by justifying the synergy relation between *hr* and *vf*.

Consider two heart rates, $hr_1 \geq hr_2$, and the two values of the binary variable *vf*. The synergy condition, $Y_U^+(\{hr, vf\})$, is an instantiation of Definition 7.4:

$$u(hr_1, VF) - u(hr_2, VF) \geq u(hr_1, \overline{VF}) - u(hr_2, \overline{VF}) . \tag{19}$$

Given *VF*, the heart rate is irrelevant (and ill-defined because ventricular fibrillation is a state where the heart is not contracting regularly). Therefore, the left-hand side of (19) is zero. For patients *not* in fibrillation, lower heart rates are preferable, by $U^-(hr)$, at least within the range considered here. This implies that the right-hand side of (19) is negative, satisfying the inequality.

By applying the results of Section 7.4, we can successively reduce any variables positioned between the ones of interest. Figure 17 shows the result of removing all but *dig*, *Ca*, and *v*. The final step, transformation from the fragment of Fig. 17(a) to that of Fig. 17(b), requires parallel combination of synergies using Theorem 7.3.

The final result of the exercise is that while the value of administering digitalis is ambiguous, by $U^?(dig)$, we can deduce that the optimal dosage is a decreasing function of calcium, by $Y_U^-(\{dig, Ca\})$. The more detailed model of Fig. 15 showed us that this result holds whether we are talking about the actual substance concentrations in the bloodstream or about the amounts administered and measured by imperfect means.

Inferences of this sort play a central role in therapy planning and in development of consultation systems via automatic programming [33, 55]. For planning, this type of result is a constraint on the class of admissible plans, significantly pruning the search space [59]. This is an especially useful kind of constraint for the automatic generation of a consultation system because the

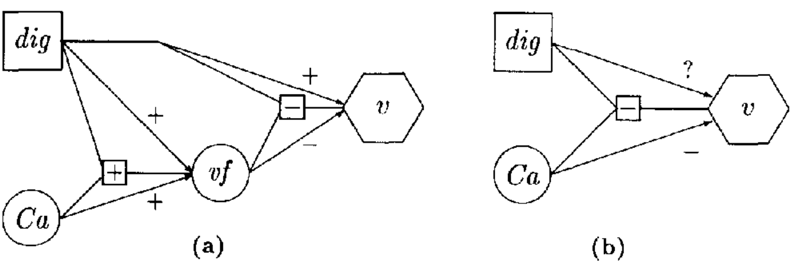

Fig. 17. Transformation of the digitalis model with synergy: (a) collapsing the therapeutic pathway and consolidating the toxic, (b) final situation after reduction of *vf*.

qualitative form of the solution corresponds to the structure of part of the target code.

The digitalis dosage d^* for patients with normal calcium—a distinguished point in the quantity space for Ca—is a landmark value as described in Section 7.6. Subsynergy implies that the dosage for a patient with calcium above normal should be lower than d^*. This is essentially the strategy of the digitalis program produced by Swartout's XPLAIN system [55], where a domain principle mandates that dosage should be adjusted according to "drug sensitivities." QPNs provide a more general and principled language for encoding domain knowledge, from which policies such as this can be derived.

8. Related Work

The QPN representation and reasoning techniques presented here borrow many concepts from other work in AI and decision theory. The most obvious debt is to research in numeric probabilistic networks, especially that of Pearl [39] and Shachter [47]. This work also relates to other efforts by similarity of purpose. In the following sections I compare it with research in qualitative probability, ordering relations on random variables, and nonmonotonic and qualitative reasoning.

8.1. Qualitative probability

The central task in designing a qualitative probability representation—indeed in the design of a qualitative representation for anything—is choosing the important qualitative distinctions to make. For example, a straightforward mapping of techniques from qualitative physics might suggest that we carve up the [0, 1] probability scale into a quantity space by choosing a small set of designated reference points. For example, the set of points {0.01, 0.05, 0.5, 0.95, 0.99} might be chosen as especially significant.

Such a scheme is a "non-starter" because it is only by coincidence that the important qualitative thresholds for any problem will align themselves with the fixed boundaries in the probabilistic quantity space. Furthermore, it is not clear that the types of manipulations typically performed on probabilities will respect these boundaries in a systematic fashion. For example, Bacchus' inheritance reasoner [1] cannot chain inferences about typicality. Attempts to construct qualitative notions of *absolute* probability (see, for example, the work of Halpern and colleagues [18, 19]) are likely to encounter similar problems.[11] Unlike the scales of physical parameters, the probability interval does not appear to have values (except the endpoints) that are universally interesting or even of special significance within a domain. And the *qualification problem* [50] is inevitably important here because one can almost always think of conditions

[11] For a more fundamental argument about the limitations of this approach, see the recent work of Xiang et al. [66].

that would bring the probability of any nonanalytic event outside any given nonuniversal range.

This suggests that it might be more appropriate to base qualitative probability concepts on *relative* likelihoods. A relative likelihood logic permits statements that one formula is more likely than another [10, 15]. Absolute probability is subsumed by a scheme of this type given a set of special formulas corresponding to canonical chance situations (such as experiments with an idealized coin) of all probabilities.

The qualitative relationships presented here can be viewed as a special case of relative likelihood where only assertions about the comparative probability of particular conditional events are permitted. Both S^δ and Y^δ are limited to comparisons of the likelihood of a given event under different conditions. For the binary case, S^δ induces a quantity space on the *likelihood ratio* (7) with a distinguished value of one.

There are three primary advantages to restricting the formalism to these special likelihood comparisons. First, information in the constraints is substantially—though not completely—preserved by the transformation operations presented in Sections 4 and 7, a necessary prerequisite for tractable inference. (See Blyth [4] for examples of difficulties with some other seemingly reasonable qualitative likelihood comparisons.) Second, the ability to deduce decision properties suggests that these comparisons are making some of the significant qualitative distinctions. And third, the *ceteris paribus* condition in the definitions reduces the impact of the qualification problem, as does the embedding of the formalism in closed-world networks.

The enterprise of qualitative probability is not necessarily hostile to quantitative probability. In Savage's axiomatization of Bayesian decision theory [46, Chapter 3], the qualitative likelihood ordering logically precedes development of quantitative probability measures.[12] The existence of a numeric representation for likelihood is only a convenient fact that simplifies much of the theory and supports some direct applications. The emphasis to date on numerical probability representations in applied decision theory and AI is due in part to technological history; there is no fundamental requirement of probability or utility theory that we focus exclusively on the precise extreme of the representation spectrum.

8.2. Relations on random variables

Philosophers have long attempted to develop mathematical definitions of causality, occasionally producing probabilistic interpretations. Motivated by a

[12] The same is true of an earlier treatment by Koopman [26]. Strictly speaking, the qualitative theory is more general than the quantitative one, which typically requires some sort of additivity axiom. This is not, however, a motiviation for the present work (indeed, the proofs assume additivity), which stresses advantages for knowledge representation and computation.

more limited set of concerns, I have ignored in this treatment temporal properties, mechanisms, and other issues salient to causality. These matters aside, Suppes [54] proposes a probabilistic condition for binary events that is equivalent to S^+ (1) without the context quantification. For multi-valued variables, Suppes suggests quadrant dependence (11). A cause is considered spurious if the probabilistic relation can be explained by a prior common cause. The concept of spuriousness can be partially captured in QPNs by distinguishing qualitative influences inferred via arc reversals (spurious) from those derivable solely from reductions along influence chains (genuine). This is similar in spirit to the approach of Simon [52], and is equivalent to the distinction emphasized by Pearl [38] between causal and evidential support.

As suggested previously, ordering of random variables has also attracted considerable interest in statistics [2, 30, 45] and decision theory [65]. Milgrom [32] demonstrates the application of MLRP to theoretical problems in informational economics.

The key difference between the S^+ definition proposed here and previous work is that we obtain transitivity by requiring the condition to hold in all contexts. Humphreys [23] proves a special case of Theorem 4.2 to the effect that binary qualitative influences along Markov chains (graphs where each node has a single predecessor, thereby eliminating context) can be combined by sign multiplication. In contrast, Suppes demonstrates that the causal algebra induced by his condition—defined only at the margin—does not possess the transitive property. A causal algebra either lacking sound reduction rules like those of Section 4.3 or restricted to simple Markov chains would have little value for knowledge representation.

Considerably less attention has been devoted to relations of probabilistic synergy. The supermodularity concept of Section 7.3 has not, to my knowledge, previously been interpreted in a probabilistic context. However, a constraint similar in spirit to sub-synergy was exploited by NESTOR [5, p. 102], a diagnostic program based on probabilistic inequalities. (NESTOR used qualitative influences to bound probability intervals as well.) And we saw in Section 7.2 that several canonical probabilistic models proposed for AI programs are special cases of Y^δ.

8.3. Nonmonotonic reasoning

There has been considerable interest of late in probabilistic accounts of nonmonotonic reasoning [40]. Recently, Neufeld [34] proposed a probabilistic semantics for defaults based on a relation equivalent to strict binary S^+ without the context quantification. His reasoner derives consequences of an *inference graph* of defaults and logical relations by applying properties of the probabilistic relation and conditional independencies implicit in the graph's structure.

Although the use of qualitative probabilistic relations for nonmonotonic inference is interesting, I am skeptical about the ultimate potential of any purely probabilistic approach. Likelihood is only one of many criteria for believing [8]; a satisfactory semantics for defaults must encompass the full range of factors determining whether adopting a particular state of belief is cognitively and computationally rational [7, 51].

8.4. Qualitative reasoning

It might appear at first glance that the very imprecision sanctioned by qualitative mechanisms obviates the need to consider explicitly uncertainty underlying the models. This position, however, confounds the weakness of inferences and input specifications with other kinds of variability in the model. The distinction is crucial because the latter might undermine the soundness of conclusions drawn from qualitative knowledge bases.

The interpretation of a set of qualitative physical relationships as "qualitative differential equations" (see Kuipers [28], for example) treats each relationship as a constraint on some "true" functional relationship that holds over time. To assert that $b = M^+(a)$ (in Kuipers' notation) is to claim that there exists an increasing function f such that $b_t = f(a_t)$ for all t. This is incompatible with a probabilistic interpretation, even though f is only loosely constrained. A qualitative influence assertion of $S^+(a, b)$, on the other hand, leaves open the possibility that the relationship is non-deterministic (f might vary over time) and does not prohibit an increase in a from coinciding with a decrease in b.

Application of qualitative-physics inference mechanisms in a probabilistic environment is dangerous because they tend to take as impossible what is merely unlikely. For example, Forbus' measurement interpretation algorithm for qualitative process theory [14] prunes away the qualitative behaviors that are inconsistent with observations of the system. If the dynamics of the system are really probabilistic (I do not claim that this is the case for Forbus' application), then this step is not valid because no behaviors are truly inconsistent. In such a situation, measurements serve to change the likelihoods of various behaviors but never to rule them out. This difference is vital in a critical application because some highly unlikely behaviors may nevertheless be important enough to demand attention from the reasoner.

Though we cannot prune measurement interpretations, we might be able to perform some pruning on the plan space using the techniques presented above. A particular measurement does not in general reveal any facts about the other model variables with certainty, yet it may allow us to deductively conclude that some decision variables (perhaps dials in the control room) should be adjusted in particular directions.

9. Conclusion

9.1. Summary

A QPN model represents qualitative constraints on the probabilistic relationships among a set of variables. In this paper I have defined and analyzed two basic constraint types: qualitative influences that express direct relationships between variables, and qualitative synergies that express interactions among influences. The probabilistic definitions justify sound graph-based inference procedures that answer queries about the qualitative relationship of any subset of variables in the model. Qualitative relationships involving the special value variable v dictate structural properties of the optimal assignment to decision variables.

Despite the ubiquity of constructs similar to qualitative influences in knowledge representation mechanisms, there has been little study of the semantics of these statements. Previous work either denies the probabilistic nature of the relationships among variables in the model or takes for granted the ability to draw inferences by chaining influences in the network. I have defined a positive qualitative influence of a on b as an assertion that, in all contexts, the posterior probability distribution for b given a is stochastically increasing (in the sense of FSD) in a. A series of results provides theoretical support for this S^{δ} definition:

- S^{δ} justifies reduction of variables by influence chaining. Reduction of any subset of variables can be performed in $O(|V|^2)$ time.
- S^{δ} permits some nontrivial conclusions upon influence reversal.
- S^{δ} is the weakest posterior condition that justifies chaining of influences.
- S^{δ} is necessary and sufficient for chaining *decisions* across influences.

Two variables a and b are positively synergistic on c if the posterior distribution for c is increased more (in the sense of FSD) upon a positive change in a for higher values of b. This Y^{δ} definition has several computationally and decision-theoretically useful properties:

- Canonical models such as the "noisy OR" often entail Y^{δ}.
- Y^{δ} is equivalent to supermodularity on expectation with respect to all monotonic transformations.
- Y^{δ}_{U} implies the monotone decision property.
- Synergies may be propagated forwards or backwards along qualitative influences. They also may be nontrivially updated upon influence reversal.
- Any nonredundant sequence of reductions and reversals is computable in polynomial time.

Together, the two qualitative relationships provide a simple yet powerful modeling language. A planner is often able to derive important facts about the qualitative structure of optimal strategies from only weak premises on the qualitative relationships in the domain.

9.2. Discussion

Though powerful in some respects, the qualitative relationships are also quite limited. First, they can only express monotone associations. Second, as we saw in Section 6, QPNs are unable to resolve true tradeoffs because parallel influences of different sign are indeterminate in combination ($+ \oplus - = ?$). Indeed, "unresolvable in a QPN" might be the best available formal definition of a tradeoff situation.

Thus, a QPN decision model can support planning "up to tradeoffs." Indeed, SUDO-PLANNER uses dynamically generated QPNs to establish constraints on the admissible plan space for a medical therapy domain [61]. Admissibility is defined with respect to the qualitative relations in the domain. To proceed beyond that point, we would need more precise knowledge of these relations. I see no insurmountable barriers to the development of hybrid representations that augment QPNs with stronger constraints, up to and including constraint to exact numeric values. As mentioned above, features of such a hybrid scheme were explored by Cooper in the NESTOR project [5]. While NESTOR's basic representation was probability intervals, it applied constraints similar to qualitative influences and synergies to bound the result of certain combination operations.

Another possibility for tradeoff resolution is in the "order of magnitude" techniques [42] developed in qualitative physics. In the case when one parallel influence can be declared *negligible* with respect to another—for example, the mildly unpleasant taste of an orally-administered drug relative to its curative powers—indeterminacy can be avoided by simply ignoring the former when in conflict with the latter.

Finally, evaluation of QPNs as a knowledge representation must also take into account the feasibility of constructing knowledge bases of reasonable complexity. For reasons of modularity and precision, QPNs should be substantially easier to generate than their numeric counterparts. Preliminary experience from the development of SUDO-PLANNER has confirmed the feasibility of automatic model generation for small networks (on the order of twenty nodes) [61]. Further research is necessary to develop techniques for constructing qualitative probabilistic networks of significantly greater scale.

Appendix A. Proofs

Lemma 4.1. *If $b \not\in pred_G^*(a)$ then a and b are d-separated in dep(G) by any S such that $pred_G(b) \subseteq S \subseteq \{s \mid b \not\in pred_G^*(s)\}$.*

Proof. Two variables are d-separated iff every undirected path between them is blocked according to one of the conditions of Definition 4.3. Every path between a and b must pass through one of b's predecessors or one of its successors. Because $pred_G(b) \subseteq S$, the paths through the predecessors are blocked by the first condition. Consider a path through a successor of b. Let t be the first variable on the path, starting from b, that has both incident edges leading in. Such a variable must exist because $b \not\in pred_G^*(a)$. Because it is the first, there is a directed path to it from b. But b has no directed paths to elements of S. Therefore, neither t nor any of its successors are in S and t blocks the path via the second condition of Definition 4.3. □

Theorem 4.2.

$$S^{\delta_1}(a, b, G) \wedge S^{\delta_2}(b, c, G) \wedge S^{0}(a, c, G) \Rightarrow$$

$$S^{\delta_1 \otimes \delta_2}(a, c, red(b, G)),$$

where $\delta_i \in \{+, -, 0, ?\}$ and $\otimes$ denotes sign multiplication, described in Fig. 3.

Proof. I will prove the case $\delta_1 = \delta_2 = +$; the others are analogous. Choose a_1 and a_2 such that $a_1 \geq a_2$, and an x_0 in $X(pred(b) \cup pred(c) - \{a, b\})$ that is consistent with a_1 and a_2.[13] Let F_c denote the conditional CDF for c and $\underline{c}$ the minimal value of the variable. By the definition of cumulative probability we have

$$F_c(c_0 \mid a_i x_0) = \int_{\underline{c}}^{c_0} \int f_{bc}(b_0 c_1 \mid a_i x_0)\, \mathrm{d}b_0\, \mathrm{d}c_1 .$$

Changing the order of integration and decomposing the joint probability yields[14]

$$F_c(c_0 \mid a_i x_0) = \int \int_{\underline{c}}^{c_0} f_c(c_1 \mid a_i b_0 x_0) f_b(b_0 \mid a_i x_0)\, \mathrm{d}c_1\, \mathrm{d}b_0 . \tag{A.1}$$

Because a and c are conditionally independent given b and x, by the S^0 premise and Lemma 4.1, we can remove a_i from the f_c expression. Rewriting the density function as the derivative of a cumulative, we get

$$F_c(c_0 \mid a_i x_0) = \int \int_{\underline{c}}^{c_0} f_c(c_1 \mid b_0 x_0)\, \mathrm{d}c_1\, \mathrm{d}F_b(b_0 \mid a_i x_0) . \tag{A.2}$$

The inner integral is simply the CDF for c given b_0.

[13] In all subsequent proofs, x is understood to range over assignments to relevant predecessor variables in a similar manner.

[14] If some values of b_0 are inconsistent with x_0, then distributions of c conditioned on b_0 and x_0 (and therefore the right-hand sides of equations (A.1)–(A.3)) are not well-defined. This has no consequence, however, because the value of $f_b(b_0 \mid a_i x_0)$ in such cases will always be zero.

$$F_c(c_0 \mid a_i x_0) = \int F_c(c_0 \mid b_0 x_0)\, \mathrm{d}F_b(b_0 \mid a_i x_0)\,. \tag{A.3}$$

Because b positively influences c, the pointwise FSD condition (3) implies that for any c_0, $F_c(c_0 \mid b_0 x_0)$ is a *decreasing* function of b_0. And $S^+(a, b)$ entails FSD of $F_b(b_0 \mid a_1 x_0)$ over $F_b(b_0 \mid a_2 x_0)$. Therefore, (4) applies with the inequality reversed (negating $F_c(c_0 \mid b x_0)$ yields an increasing function), leading to the conclusion

$$\forall c_0\ F_c(c_0 \mid a_1 x_0) \leqslant F_c(c_0 \mid a_2 x_0)\,,$$

implying FSD. Because a_1, a_2, and x_0 were chosen arbitrarily, we have finally $S^+(a, c)$. $\square$

Theorem 4.3

$$S^{\delta_1}(a, b, G) \wedge S^{\delta_2}(b, c, G) \wedge S^{\delta_3}(a, c, G) \Rightarrow$$
$$S^{(\delta_1 \otimes \delta_2) \oplus \delta_3}(a, c, red(b, G))\,,$$

where $\oplus$ denotes sign addition, also described in Fig. 3.

Proof. Proceed as for the proof of Theorem 4.2 to equation (A.1). Because δ_3 is not generally zero, we cannot remove a_i in the next two steps.

$$F_c(c_0 \mid a_i x_0) = \int F_c(c_0 \mid a_i b_0 x_0)\, \mathrm{d}F_b(b_0 \mid a_i x_0)\,.$$

Define $\hat{F}_c$ as a variant where a_i is fixed to a_1 in the first term

$$\hat{F}_c(c_0 \mid a_i x_0) = \int F_c(c_0 \mid a_1 b_0 x_0)\, \mathrm{d}F_b(b_0 \mid a_i x_0)\,.$$

Note that $\hat{F}_c(c_0 \mid a_1 x_0) = F_c(c_0 \mid a_1 x_0)$ and that

$$\delta_3 = +\,(-) \Rightarrow \forall c_0\ F_c(c_0 \mid a_1 b_0 x_0) \leqslant (\geqslant)\ F_c(c_0 \mid a_2 b_0 x_0)\,,$$

therefore

$$\forall c_0\ \hat{F}_c(c_0 \mid a_2 x_0) \leqslant (\geqslant)\ F_c(c_0 \mid a_2 x_0)\,. \tag{A.4}$$

When $\delta_3 = ?$ it is possible that the relation varies with c_0. Regardless of δ_3, $F_c(c_0 \mid a_1 b_0 x_0)$ is a decreasing/increasing/nonmonotonic function of b_0 as δ_2 is $+/-/?$. For concreteness, suppose $\delta_1 = \delta_2 = +$ (again, the other cases are analogous). Following the reasoning in the proof of Theorem 4.2 above, we get

$$\hat{F}_c(c_0 \mid a_1 x_0) = F_c(c_0 \mid a_1 x_0)\ \mathrm{FSD}\ \hat{F}_c(c_0 \mid a_2 x_0)\,.$$

If $\delta_3 = +$ (more generally if δ_3 agrees with the polarity of the FSD relation), this result combines with (A.4) to imply FSD of the corresponding unhatted functions F_c, thereby establishing the result. Without such agreement FSD may be violated, permitting us to conclude only $S^?(a, c, red(b, G))$. □

Theorem 4.4. *Let $G' = rev(a, b, G)$. G' inherits all the qualitative influences of G except*:

(1) *$dir(a, b, G')$ is undefined.*

(2) *$dir(b, a, G') = dir(a, b, G)$.*

(3) *$\forall w \in pred_{G'}(b)$,*

$$dir(w, b, G') = [dir(w, a, G) \otimes dir(a, b, G)] \oplus dir(w, b, G).$$

(4) *$\forall w \in pred_{G'}(a) - \{b\}$,*

$$\begin{aligned} dir(w, a, G') &= \begin{cases} dir(w, a, G), & \text{if } dir(w, b, G) = 0 \\ ?, & \text{otherwise} \end{cases} \\ &= dir(w, a, G) \oplus (dir(w, b, G) \otimes ?). \end{aligned}$$

Proof. First, note that all variables outside $pred_G(a) \cup pred_G(b)$ retain the same set of d-separations. Second, let us verify each relation above:

(1) There is no longer an influence from a to b.

(2) To show that the influence on the reserved link remains unchanged it is convenient to work with the likelihood form of S^δ, equation (6). Applying Bayes' formula:

$$f_b(b_j \mid a_i x) = \frac{f_a(a_i \mid b_j x) f_b(b_j \mid x)}{f_a(a_i \mid x)}.$$

Choose four values $a_1 \geq a_2$ and $b_1 \geq b_2$.

$$\frac{f_b(b_1 \mid a_i x)}{f_b(b_2 \mid a_i x)} = \frac{f_a(a_i \mid b_1 x) f_b(b_1 \mid x)}{f_a(a_i \mid b_2 x) f_b(b_2 \mid x)} = g(b_1, b_2, x) \frac{f_a(a_i \mid b_1 x)}{f_a(a_i \mid b_2 x)}.$$

Using the monotone likelihood property, $dir(a, b, G) = +\ (-)$ implies

$$\frac{f_b(b_1 \mid a_1 x)}{f_b(b_2 \mid a_1 x)} \geq (\leq) \frac{f_b(b_1 \mid a_2 x)}{f_b(b_2 \mid a_2 x)}.$$

Rearranging we get

$$\frac{f_b(b_1 \mid a_1 x)}{f_b(b_1 \mid a_2 x)} \geq (\leq) \frac{f_b(b_2 \mid a_1 x)}{f_b(b_2 \mid a_2 x)},$$

the MLRP for b given a. As noted above (and proven by Milgrom [32]), this is necessary and sufficient for our posterior FSD condition to hold for any prior $F_a(a_0 \mid x)$.

(3) In G, the influence of w on b is relative to a predecessor set that includes a. In G' the influence is not so conditioned and is therefore equivalent to the influence on b obtained by splicing a out of the network. Applying Theorem 4.3 with the original influences yields the expression above.

(4) Here the reversal transforms an unconditional relation to a conditional one. If $dir(w, b, G) = 0$, w and b are d-separated by $pred(b)$ in $dep(G)$ (by Lemma 4.1), therefore $f_w(w_0 \mid abx) = f_w(w_0 \mid ax)$ by conditional independence. In that case the MLRP obviously holds for the conditional density iff it holds for the marginal one. If w has nonzero influence on b in G, this independence does not hold. Because a and w may interact significantly in their influence on b we cannot say anything about their relation given b. For example, let the three variables be binary with a and w marginally independent (that is, $dir(w, a, G) = 0$),

$$\Pr(A) = \Pr(W) = 0.5 ,$$

$$\Pr(B \mid \bar{A}\bar{W}) = 0.1 ,$$

$$\Pr(B \mid \bar{A}W) = 0.2 ,$$

$$\Pr(B \mid AW) = 0.9 .$$

Then $dir(w, a, G')$ can be $+$ or $-$ depending on whether $\Pr(B \mid A\bar{W})$ is less than or greater than 0.45. Either possibility is consistent with an initial G with $dir(a, b) = dir(w, b) = +$. $\square$

Theorem 5.1. *Let* $S^+(a, b)$ *be defined by* (10). *Given the following conditions*:

(1) *Theorem* 4.2,

(2) *for binary* b, $a_1 \geq a_2$, *and* x,

$$F_b(\cdot \mid a_1 x) \; R \; F_b(\cdot \mid a_2 x) \Leftrightarrow \Pr(B \mid a_1 x) \geq \Pr(B \mid a_2 x) , \tag{A.5}$$

the weakest R *is FSD.*

Proof. First, note that FSD satisfies these conditions. Next, assume that R satisfies them but R does not entail FSD. We will start with an instantiation of Theorem 4.2 and derive a contradiction. Let a, b, and c be the only variables (so we can safely ignore x) with $S^-(a, b)$, $S^+(b, c)$, and no other direct links. For concreteness, let b range over the unit interval $[0, 1]$ and c be binary with $\Pr(C \mid ab) = \phi(b)$, for some $\phi : [0, 1] \to [0, 1]$ monotonic. The monotonicity of ϕ guarantees $S^+(b, c)$ and its independence from a validates $S^0(a, c)$ in the original network. By assumption, Theorem 4.2 applies, yielding the conclusion $S^+(a, c)$ and thereore $F_c(c_0|a_1)\ R\ F_c(c_0|a_2)$. Because c is binary, (A.5) must hold. Using

$$\Pr(C\,|\,a_i) = \int \Pr(C\,|\,a_i b_0)\,\mathrm{d}F_b(b_0\,|\,a_i)\,,$$

the right-hand side of (A.5) becomes

$$\int_0^1 \phi(b_0)\,\mathrm{d}F_b(b_0\,|\,a_1) \geq \int_0^1 \phi(b_0)\,\mathrm{d}F_b(b_0\,|\,a_2)\,. \tag{A.6}$$

Because ϕ may be any monotonic function, FSD is necessary for (A.6) and is therefore entailed by R. □

Theorem 5.2. *Suppose $U^{\delta_2}(b, G)$ and $U^0(a, G)$. A necessary and sufficient condition for $U^{\delta_1 \otimes \delta_2}(a, red(b, G))$ is $S^{\delta_1}(a, b, G)$ as in Definition* 3.4.

Proof. The expected utility of a_i with any x is given by

$$u(a_i, x) = \int u(b_0, x)\,\mathrm{d}F_b(b_0\,|\,a_i x)\,. \tag{A.7}$$

Let us prove the case $\delta_1 = \delta_2 = +$. $U^+(a)$ is satisfied in the reduced network iff $u(a_i, x)$ is increasing in a_i. From (8) we know that $u(b_0, x)$ is monotonically increasing in b_0. In fact, it can be *any* monotonic function. Therefore, (A.7) is increasing in a_i under the same conditions as (4), which is exactly the S^+ condition (5) of Definition 3.4. □

Lemma 7.1. *$Y^+(\{a, b\}, c)$ (respectively Y^- and Y^0) holds iff the function*

$$e_\phi(a, b\,|\,x) = \int \phi(c_0) f_c(c_0\,|\,abx)\,\mathrm{d}c_0 \tag{A.8}$$

is supermodular (submodular, modular) in a and b for all increasing functions ϕ and contexts x.

Proof. Choose arbitrary $a_1 \geq a_2$, $b_1 \geq b_2$, and x. By Definition 7.3, e_ϕ is supermodular iff

$$e_\phi(a_1, b_1\,|\,x) + e_\phi(a_2, b_2\,|\,x) \geq e_\phi(a_1, b_2\,|\,x) + e_\phi(a_2, b_1\,|\,x)\,.$$

Rearranging,

$$e_\phi(a_1, b_1\,|\,x) - e_\phi(a_2, b_1\,|\,x) \geq e_\phi(a_1, b_2\,|\,x) - e_\phi(a_2, b_2\,|\,x)\,.$$

Substituting the definition of e_ϕ (A.8) and combining the integrals,

$$\int \phi(c_0)[f_c(c_0 \mid a_1 b_1 x) - f_c(c_0 \mid a_2 b_1 x)]\, \mathrm{d}c_0$$

$$\geqslant \int \phi(c_0)[f_c(c_0 \mid a_1 b_2 x) - f_c(c_0 \mid a_2 b_2 x)]\, \mathrm{d}c_0 \,. \tag{A.9}$$

A necessary and sufficient condition for (A.9) to hold for any increasing function ϕ is that the bracketed distribution differences be related by FSD. (Recall the equivalence between (3) and (4) in Section 3.2.) That is,

$$\forall c_0 \; F_c(c_0 \mid a_1 b_1 x) - F_c(c_0 \mid a_2 b_1 x) \leqslant F_c(c_0 \mid a_1 b_2 x) - F_c(c_0 \mid a_2 b_2 x) \,.$$

This is exactly the Y^+ condition of Definition 7.2. □

Theorem 7.2. *Synergies can be extended along qualitative influences by reduction according to the following*:

$$Y^{\delta_1}(\{a, b\}, c, G) \wedge S^{\delta_2}(c, d, G) \wedge S^0(a, d, G) \wedge S^0(b, d, G) \Rightarrow$$
$$Y^{\delta_1 \otimes \delta_2}(\{a, b\}, d, red(c, G)) \,.$$

Proof. Let us assume that $\delta_1 = \delta_2 = +$; the other cases are analogous. We can describe the cumulative for d conditional on a and b by integrating over its counterpart for c.

$$F_d(d_0 \mid abx) = \int_{\underline{d}}^{d_0} \int f_d(d_1 \mid abc_0 x) f_c(c_0 \mid abx)\, \mathrm{d}c_0\, \mathrm{d}d_1 \tag{A.10}$$

$$= \int \int_{\underline{d}}^{d_0} f_d(d_1 \mid c_0 x)\, \mathrm{d}d_1 \; f_c(c_0 \mid abx)\, \mathrm{d}c_0 \tag{A.11}$$

$$= \int F_d(d_0 \mid c_0 x) f_c(c_0 \mid abx)\, \mathrm{d}c_0 \,. \tag{A.12}$$

In going from (A.10) to (A.11) I took advantage of the conditional independence between d and each of a and b given c implied by the S^0 conditions and Lemma 4.1. Because $S^+(c, d)$, $F_d(d_0 \mid c_0)$ is a decreasing function of c_0 for any d_0. Therefore, equation (A.12) and Lemma 7.1 imply that $F_d(d_0 \mid abx)$ is a *sub*modular function of a and b for all d_0 (a function g is submodular iff $-g$ is supermodular). By the definition of submodularity,

$$\forall d_0 \; F_d(d_0 \mid a_1 b_1 x) - F_d(d_0 \mid a_2 b_1 x) \leqslant F_d(d_0 \mid a_1 b_2 x) - F_d(d_0 \mid a_2 b_2 x)\,, \qquad \text{(A.13)}$$

which is the condition for $Y^+(\{a, b\}, d)$ of Definition 7.2. □

Theorem 7.3.

$$Y^{\delta_1}(\{a, b\}, c, G) \wedge S^{\delta_2}(c, d, G) \wedge Y^{\delta_3}(\{a, c\}, d, G) \wedge$$
$$Y^{\delta_4}(\{b, c\}, d, G) \wedge S^{\delta_5}(a, c, G) \wedge S^{\delta_6}(b, c, G) \wedge Y^{\delta_7}(\{a, b\}, d, G) \;\Rightarrow$$
$$Y^{(\delta_1 \otimes \delta_2) \oplus (\delta_3 \otimes \delta_6) \oplus (\delta_4 \otimes \delta_5) \oplus \delta_7}(\{a, b\}, d, red(c, G))\,.$$

Proof. Start as in the proof of Theorem 7.2, but do not use conditional independence.

$$F_d(d_0 \mid abx) = \int_{\underline{d}}^{d_0} \int f_d(d_1 \mid abc_0 x) f_c(c_0 \mid abx)\, \mathrm{d}c_0\, \mathrm{d}d_1$$
$$= \int F_d(d_0 \mid abc_0 x)\, \mathrm{d}F_c(c_0 \mid abx)\,.$$

As in the proof of Theorem 4.3, define $\hat{F}_d$ to be the CDF with the conditioning variables fixed in the first term, to a_1 and b_1 in this case.

$$\hat{F}_d(d_0 \mid abx) = \int F_d(d_0 \mid a_1 b_1 c_0 x)\, \mathrm{d}F_c(c_0 \mid abx)\,.$$

Regardless of a_1 and b_1, $F_d(d_0 \mid a_1 b_1 c_0 x)$ has monotonocity properties determined solely by δ_2. Following the reasoning of the proof of Theorem 7.2, we have the following fact about $\hat{F}_d$ (a hatted version of (A.13)):

$$\forall d_0 \; \hat{F}_d(d_0 \mid a_1 b_1 x) - \hat{F}_d(d_0 \mid a_2 b_1 x)$$
$$R\, \hat{F}_d(d_0 \mid a_1 b_2 x) - \hat{F}_d(d_0 \mid a_2 b_2 x)\,, \qquad \text{(A.14)}$$

with R the relation $\leqslant$, $\geqslant$, $=$, or ? as $\delta_1 \otimes \delta_2$ is $+$, $-$, 0, or ?. Henceforth I will refer to functions satisfying conditions of the form (A.14) as *R-modular*. Let $\hat{F}'_d$ be intermediate between F_d and $\hat{F}_d$ where only b is fixed

$$\hat{F}'_d(d_0 \mid abx) = \int F_d(d_0 \mid ab_1 c_0 x)\, \mathrm{d}F_c(c_0 \mid abx)\,.$$

Note that $\hat{F}'_d(d_0 \mid a_1 b_i x) = \hat{F}_d(d_0 \mid a_1 b_i x)$ for either b_i. Therefore $\hat{F}'_d$ is R-modular iff

$$\forall d_0\ \hat{F}_d(d_0 \mid a_1 b_1 x) - \hat{F}'_d(d_0 \mid a_2 b_1 x) \\ R\ \hat{F}_d(d_0 \mid a_1 b_2 x) - \hat{F}'_d(d_0 \mid a_2 b_2 x)\,. \tag{A.15}$$

Using (A.14) and a little rearrangement, a sufficient condition for (A.15) is

$$\forall d_0\ \Delta_1(d_0)\ R\ \Delta_2(d_0)\,, \tag{A.16}$$

where

$$\Delta_i(d_0) \equiv \hat{F}_d(d_0 \mid a_2 b_i x) - \hat{F}'_d(d_0 \mid a_2 b_i x)\,. \tag{A.17}$$

Expanding the definitions for $\hat{F}_d$ and $\hat{F}'_d$,

$$\Delta_i(d_0) = \int [F_d(d_0 \mid a_1 b_1 c_0 x) - F_d(d_0 \mid a_2 b_1 c_0 x)]\, \mathrm{d}F_c(c_0 \mid a_2 b_i x)\,. \tag{A.18}$$

The difference inside the integral of equation (A.18) is an increasing, decreasing, or constant function of c_0 as δ_3, the synergy of a and c, is $-$, $+$, or 0. The influence of b on c, δ_6, determines an FSD relation among the $F_c(c_0 \mid ab_i)$. Therefore, condition (A.16) holds if $\delta_3 \otimes \delta_6$ agrees with R, which was determined by $\delta_1 \otimes \delta_2$.

Another application of this line of reasoning with the roles of a and b reversed leads to the conclusion that $\hat{F}''_d$, where

$$\hat{F}''_d(d_0 \mid abx) = \int F_d(d_0 \mid a_1 b c_0 x)\, \mathrm{d}F_c(c_0 \mid abx)\,,$$

is R-modular if $\delta_4 \otimes \delta_5$ agrees with R. Thus, agreement among these pairwise products yields R-modularity of $\hat{F}_d$, $\hat{F}'_d$, and $\hat{F}''_d$.

Suppose that δ_7 also agrees with R. Then, from the Y^δ definition we have

$$\forall d_0\ F_d(d_0 \mid a_1 b_1 c_0 x) - F_d(d_0 \mid a_2 b_1 c_0 x) \\ R\ F_d(d_0 \mid a_1 b_2 c_0 x) - F_d(d_0 \mid a_2 b_2 c_0 x)\,,$$

which entails the following inequality when integrating over a positive function:

$$\forall d_0 \int [F_d(d_0 \mid a_1 b_1 c_0 x) - F_d(d_0 \mid a_2 b_1 c_0 x)]\, \mathrm{d}F_c(c_0 \mid a_2 b_2 x) \\ R \int [F_d(d_0 \mid a_1 b_2 c_0 x) - F_d(d_0 \mid a_2 b_2 c_0 x)]\, \mathrm{d}F_c(c_0 \mid a_2 b_2 x)\,.$$

Equivalently,

$$\forall d_0\ \hat{F}_d(d_0 \mid a_2 b_2 x) - \hat{F}'_d(d_0 \mid a_2 b_2 x) \\ R\ \hat{F}''_d(d_0 \mid a_2 b_2 x) - F_d(d_0 \mid a_2 b_2 x)\,. \tag{A.19}$$

We can transform (A.19) to a relation on F_d alone by applying some R-modularity conditions already known and taking advantage of the equivalences among the hatted and primed Fs for particular values of a and b. Combining (A.19) with R-modularity of $\hat{F}''_d$,

$$\begin{aligned}\forall d_0\ & F_d(d_0 \mid a_1b_1x) - \hat{F}''_d(d_0 \mid a_2b_1x) + \hat{F}_d(d_0 \mid a_2b_2x) - \hat{F}'_d(d_0 \mid a_1b_2x)\\ & R\ F_d(d_0 \mid a_1b_2x) - F_d(d_0 \mid a_2b_2x)\,.\end{aligned}$$

Applying R-modularity of $\hat{F}'_d$ yields

$$\begin{aligned}\forall d_0\ & 2F_d(d_0 \mid a_1b_1x) - F_d(d_0 \mid a_2b_1x) - \hat{F}''_d(d_0 \mid a_2b_1x) + \hat{F}_d(d_0 \mid a_2b_2x)\\ & R\ F_d(d_0 \mid a_1b_2x) - F_d(d_0 \mid a_2b_2x) + \hat{F}'_d(d_0 \mid a_1b_2x)\,,\end{aligned}$$

and finally, R-modularity of $\hat{F}$ leads to the result

$$\begin{aligned}\forall d_0\ & F_d(d_0 \mid a_1b_1x) - F_d(d_0 \mid a_2b_1x)\\ & R\ F_d(d_0 \mid a_1b_2x) - F_d(d_0 \mid a_2b_2x)\,.\end{aligned}$$

Therefore, unanimity among the terms in the new synergy expression given by the theorem statement implies R-modularity of F_d, the condition of interest. Dissent by any term results in a synergy of $Y^?$, vacuously true. □

Theorem 7.5. *Given $S^{\delta_6}(b, c, G)$ and $S^0(a, c, G)$, $Y_U^{\delta_3}(\{a, c\}, G)$ is both necessary and sufficient for $Y_U^{\delta_3 \otimes \delta_6}(\{a, b\}, red(c, G))$.*

Proof. By the expected utility property and the conditional independence of a and c we have

$$u(a, b, x) = \int u(a, c_0, x)\, \mathrm{d}F_c(c_0 \mid bx).$$

Let Δ_i represent the utility difference upon varying a between a_1 and a_2 when $b = b_i$, that is, $\Delta_i = u(a_1, b_i, x) - u(a_2, b_i, x)$. Expanding,

$$\Delta_i = \int [u(a_1, c_0, x) - u(a_2, c_0, x)]\, \mathrm{d}F_c(c_0 \mid b_ix)\,.$$

Note that δ_6 determines an FSD condition on $F_c(c_0 \mid b_ix)$. A corresponding relation on the Δ_i is entailed iff the term in brackets is monotone in the same direction with respect to c_0. This is exactly the condition for $Y_U^{\delta_3}(\{a, c\})$. □

Theorem 7.6.

$$Y^{\delta_1}(\{a,b\},c,G) \wedge S^{\delta_2}(c,d,G) \wedge Y^{\delta_3}(\{a,c\},d,G) \wedge$$
$$Y^{\delta_4}(\{b,c\},d,G) \wedge S^{\delta_5}(a,c,G) \wedge S^{\delta_6}(b,c,G) \wedge Y^{\delta_7}(\{a,b\},d,G) \Rightarrow$$
$$Y^{(\delta_1 \otimes \delta_2) \oplus (\delta_3 \otimes \delta_6) \oplus (\delta_4 \otimes \delta_5) \oplus \delta_7}(\{a,b\},d,rev(c,d,G))\,.$$

Proof. The post-reversal distribution for d is conditioned on all d's pre-reversal predecessors except c and therefore is the same as that obtained by reducing c from the network. The result is identical to the expression from Theorem 7.3. □

ACKNOWLEDGEMENT

This work was supported by National Institutes of Health Grant No. R01 LM04493 from the National Library of Medicine while the author was at the MIT Laboratory for Computer Science. Marty Chavez, Tony Cox, Jon Doyle, Yeona Jang, Elisha Sacks, Peter Szolovits, Kate Unrath, Alex Yeh, and the anonymous referees positively (though subsynergistically) influenced this paper.

REFERENCES

1. F. Bacchus, A modest, but semantically well founded, inheritance reasoner, in: *Proceedings IJCAI-89*, Detroit, MI (1989) 1104–1109.
2. J.O. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer, Berlin, 2nd ed., 1985).
3. C.R. Blyth. On Simpson's paradox and the sure-thing principle, *J. Am. Stat. Assoc.* **67** (1972) 364–366.
4. C.R. Blyth, Some probability paradoxes in choice from among random alternatives, *J. Am. Stat. Assoc.* **67** (1972) 366–381.
5. G.F. Cooper, NESTOR: A computer-based medical diagnostic aid that integrates causal and probabilistic knowledge, Ph.D. Thesis, Stanford University, Stanford, CA (1984).
6. G.F. Cooper, The computational complexity of probabilistic inference using Bayesian belief networks, *Artificial Intelligence* **42** (1990) 393–405.
7. J. Doyle, Constructive belief and rational representation, *Comput. Intell.* **5** (1989) 1–11.
8. J. Doyle and M.P. Wellman, Impediments to universal preference-based default theories, in: *First International Conference on Principles of Knowledge Representation and Reasoning*, (1989) 94–102.
9. R. Engelbrecht, A note on multivariate risk and separable utility functions, *Manage. Sci.* **23** (1977) 1143–1144. Note on [43].
10. R. Fagin, J.Y. Halpern and N. Megiddo, A logic for reasoning about probabilities, Research Rept. RJ 6190, IBM (1988).
11. P.C. Fishburn, Von Neumann-Morgenstern utility functions on two attributes, *Oper. Res.* **22** (1974) 35–45.
12. P.C. Fishburn and R.G. Vickson, Theoretical foundations of stochastic dominance, in: G.A. Whitmore and M.C. Findlay, eds., *Stochastic Dominance: An Approach to Decision Making under Risk* (Heath, Lexington, MA, 1978).
13. K.D. Forbus, Qualitative process theory, *Artificial Intelligence* **24** (1984) 85–168.
14. K.D. Forbus, Interpreting measurements of physical systems, in: *Proceedings AAAI-86*, Philadelphia, PA (1986) 113–117.
15. P. Gärdenfors, Qualitative probability as an intensional logic, *J. Philos. Logic* **4** (1975) 171–185.

16. G.A. Gorry, H. Silverman and S.G. Pauker, Capturing clinical expertise: A computer program that considers clinical responses to digitalis, *Am. J. Med.* **64** (1978) 452–460.
17. B.N. Grosof, Non-monotonicity in probabilistic reasoning, in: J.F. Lemmer and L.N. Kanal, eds., *Uncertainty in Artificial Intelligence* **2** (Elsevier Science Publishers (North-Holland), Amsterdam, 1988) 237–249.
18. J.Y. Halpern and D.A. McAllester, Likelihood, probability, and knowledge, in: *Proceedings AAAI-84*, Austin, TX (1984) 137–141.
19. J.Y. Halpern and M.O. Rabin, A logic to reason about likelihood, *Artificial Intelligence* **32** (1987) 379–405.
20. D.E. Heckerman and E.J. Horvitz, The myth of modularity in rule-based systems for reasoning with uncertainty, in: J.F. Lemmer and L.N. Kanal, eds., *Uncertainty in Artificial Intelligence* **2** (Elsevier Science Publishers (North-Holland), Amsterdam, 1988) 23–34.
21. M. Henrion, Practical issues in constructing a Bayes' belief network, in: *Proceedings Workshop on Uncertainty in Artificial Intelligence* (1987) 132–139.
22. R.A. Howard and J.E. Matheson, Influence diagrams, in: R.A. Howard and J.E. Matheson, eds., *The Principles and Applications of Decision Analysis* **2** (Strategic Decisions Group, Menlo Park, CA, 1984) 719–762.
23. P. Humphreys, Cutting the causal chain, *Pacific Philos. Q.* **61** (1980) 305–314.
24. S. Karlin and H. Rubin, The theory of decision procedures for distributions with monotone likelihood ratio, *Ann. Math. Stat.* **27** (1956) 272–299.
25. R.L. Keeney and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Tradeoffs* (Wiley, New York, 1976).
26. B.O. Koopman, The axioms and algebra of intuitive probability, *Ann. Math.* **42** (1940) 269–292.
27. D.H. Krantz, R.D. Luce, P. Suppes et al., *Foundations of Measurement* (Academic Press, New York, 1971).
28. B. Kuipers, Qualitative simulation, *Artificial Intelligence* **29** (1986) 289–338.
29. S.L. Lauritzen and D.J. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems, *J. Roy. Stat. Soc.* **B50** (1988) 157–224.
30. E.L. Lehmann, Some concepts of dependence, *Ann. Math. Stat.* **37** (1966) 1137–1153.
31. J.F. Lemmer and L.N. Kanal, eds., *Uncertainty in Artificial Intelligence* **2** (Elsevier Science Publishers (North-Holland), Amsterdam, 1988).
32. P.R. Milgrom, Good news and bad news: Representation theorems and applications, *Bell J. Econ.* **12** (1981) 380–391.
33. J. Mostow and K. Voigt, Explicit integration of goals in heuristic algorithm design, in: *Proceedings IJCAI-87*, Milan, Italy (1987) 1090–1096.
34. E. Neufeld, Defaults and probabilities; extensions and coherence, in: *First International Conference on Principles of Knowledge Representation and Reasoning* (1989) 312–323.
35. E. Neufeld and J.D. Horton, Conditioning on disjunctive knowledge: Defaults and probabilities, in: *Proceedings Workshop on Uncertainty in Artificial Intelligence* (1989) 272–278.
36. R.S. Patil, P. Szolovits and W.B. Schwartz, Causal understanding of patient illness in medical diagnosis, in: *Proceedings IJCAI-81*, Vancouver, BC (1981) 893–899.
37. J. Pearl, Fusion, propagation, and structuring in belief networks, *Artificial Intelligence* **29** (1986) 241–288.
38. J. Pearl, Embracing causality in default reasoning, *Artificial Intelligence* **35** (1988) 259–271.
39. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, San Mateo, CA, 1988).
40. J. Pearl, Probabilistic semantics for nonmonotonic reasoning: A survey, in: *First International Conference on Principles of Knowledge Representation and Reasoning* (1989) 505–516.
41. J. Pearl, D. Geiger and T. Verma, Conditional independence and its representations, *Kybernetika* **25** (1989) 33–44.

42. O. Raiman, Order of magnitude reasoning, in: *Proceedings AAAI-86*, Philadelphia, PA (1986) 100–104.
43. S.F. Richard, Multivariate risk aversion, utility independence and separable utility functions, *Manage. Sci.* **22** (1975) 12–21.
44. C. Rieger and M. Grinberg, The declarative representation and procedural simulation of causality in physical mechanisms, in: *Proceedings IJCAI-77*, Cambridge, MA (1977) 250–256.
45. S.M. Ross, *Introduction to Stochastic Dynamic Programming* (Academic Press, New York, 1983).
46. L.J. Savage, *The Foundations of Statistics* (Dover, New York, 2nd ed., 1972).
47. R.D. Shachter, Evaluating influence diagrams, *Oper. Res.* **34** (1986) 871–882.
48. R.D. Shachter, Probabilistic inference and influence diagrams, *Oper. Res.* **36** (1988) 589–604.
49. G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, NJ, 1976).
50. Y. Shoham, What is the frame problem? in: M.P. Georgeff and A.L. Lansky, eds., *Reasoning about Actions and Plans: Proceedings of the 1986 Workshop* (Morgan Kaufmann, San Mateo, CA, 1986) 83–98.
51. Y. Shoham, Nonmonotonic logics: Meaning and utility, in: *Proceedings IJCAI-87*, Milan, Italy (1987) 388–393.
52. H.A. Simon, Spurious correlation: A causal interpretation, *J Am. Stat. Assoc.* **49** (1954) 467–479.
53. D.J. Spiegelhalter, Probabilistic reasoning in predictive expert systems, in: L.N. Kanal and J.F. Lemmer, eds., *Uncertainty in Artificial Intelligence* (Elsevier Science Publishers (North-Holland), Amsterdam, 1986) 47–67.
54. P. Suppes, *A Probabilistic Theory of Causality* (North-Holland, Amsterdam, 1970).
55. W.R. Swartout, XPLAIN: A system for creating and explaining expert consulting programs, *Artificial Intelligence* **21** (1983) 285–325.
56. D.M. Topkis, Minimizing a submodular function on a lattice, *Oper. Res.* **26** (1978) 305–321.
57. S.M. Weiss, C.A. Kulikowski, S. Amarel et al., A model-based method for computer-aided medical decision making, *Artificial Intelligence* **11** (1978) 145–172.
58. D.S. Weld and J. de Kleer, eds., *Readings in Qualitative Reasoning about Physical Systems* (Morgan Kaufmann, San Mateo, CA, 1989).
59. M.P. Wellman, Dominance and subsumption in constraint-posting planning, in: *Proceedings IJCAI-87*, Milan, Italy (1987) 884–890.
60. M.P. Wellman, Probabilistic semantics for qualitative influences, in: *Proceedings AAAI-87*, Seattle, WA (1987) 660–664.
61. M.P. Wellman, *Formulation of Tradeoffs in Planning under Uncertainty* (Pitman and Morgan Kaufmann, San Mateo, CA, 1990).
62. M.P. Wellman, Qualitative probabilistic networks for planning under uncertainty (revised), in: G. Shafer and J. Pearl, eds., *Readings in Uncertain Reasoning* (Morgan Kaufmann, San Mateo, CA, 1990).
63. M.P. Wellman, Graphical inference in qualitative probabilistic networks, *Networks* (to appear).
64. M.P. Wellman and D.E. Heckerman, The role of calculi in uncertain reasoning, in: *Proceedings Workshop on Uncertainty in Artificial Intelligence* (1987) 321–331.
65. G.A. Whitmore and M.C. Findlay, eds., *Stochastic Dominance: An Approach to Decision Making Under Risk* (Heath, Lexington, MA, 1978).
66. Y. Xiang, M.P. Beddoes and D. Poole, Can uncertainty management be realized in a finite totally ordered probability algebra? in: *Proceedings Workshop on Uncertainty in Artificial Intelligence* (1989) 385–393.

Received January 1988; revised version received January 1990