DEPENDENCY-GRAPH MODELS OF EVIDENTIAL SUPPORT

John D. Lowrance

COINS Technical Report 82-26

September 1982

This work was supported in part by:

The National Science Foundation
under Grant Number
MCS79-18209,

The University of Massachusetts
under a University Fellowship,

and

The Artificial Intelligence Center,
SRI International.

DEPENDENCY-GRAPH MODELS

OF EVIDENTIAL SUPPORT

A Dissertation Presented

By

JOHN DOUGLAS LOWRANCE

Approved as to style and content by:

Edward M. Riseman, Chairperson of Committee

Michael A. Arbib, Member

Daniel P. Friedman, Member

Victor R. Lesser, Member

Edward M. Riseman, Department Head
Computer and Information Science

To Beryl, Rex, and Cynthia for their continual support

ACKNOWLEDGEMENTS

My sincere appreciation goes to my committee and Allen Hanson for their guidance throughout the course of this work. My thanks also goes to the members of the VISIONS project for their technical assistance and friendship, particularly Frank Glazer, John Prager, Lenny Wesley, and Bryant York. A deep debt of gratitude is owed to those at SRI, including Art Farley, Marty Fischler, Tom Garvey, and Nils Nilsson, for their patience, support, and assistance in this endeavor while I have been with the institute. Finally, I want to thank Susan Parker for an extremely smooth transition from my manuscript to this final document.

It is an old maxim of mine that when you have excluded the impossible, whatever remains, however improbably, must be the truth.

—Sir Arthur Conan Doyle

If a man will begin with certainties he will end with doubt, but if he will be content to begin with doubts he shall end in certainties.

— Francis Bacon

Not ignorance, but ignorance of ignorance, is the death of knowledge.

—Alfred North Whitehead

ABSTRACT

Dependency-Graph Models
of Evidential Support

September 1982

John Douglas Lowrance

A.B., Indiana University

M.S., University of Massachusetts

Ph.D., University of Massachusetts

Directed by: Professor Edward M. Riseman

Dependency-graph models of evidential support are formal systems capable of pooling and extending evidential information, while maintaining internal consistency. In this formalism, the likelihood of a proposition is represented as a subinterval of the unit interval. The lower bound represents the degree of "support" provided a proposition by a body of evidence, and the upper bound represents the extent to which it remains "plausible." The smaller this interval, the more precisely the probability of that proposition is known.

Evidential information, extracted from the environment by (indivisible) sources of knowledge, enters these models in the form of probability "mass" distributions, defined over sets of propositions common to both them and the model. These mass distributions are combined through Dempster's rule of combination [Dempster 1967]. The result is a new mass distribution representing their consensus. Next, this pooled information is extended from those propositions it directly bears upon, to those it indirectly bears upon, and converted to the interval repre-

sentation. Prior probabilities, frequently difficult or impossible to collect in artificial intelligence domains, but required by most other systems of inexact reasoning, are not needed. This form of evidential reasoning, based on [Shafer 1976], is more general than either a Boolean or Bayesian approach, providing for Boolean and Bayesian inferencing when the appropriate information is available.

Dependency graphs are formal representations of dependency relations. A dependency graph consists of a set of propositions (nodes), a covering assignment of confidences (node values), and a coordinated set of dependency relationships (connecting arcs) constraining the assignment of confidences. Confidences can be fully specified (a single value), partially specified (several values), or unspecified (all values). Similarly, a dependency graph can describe any degree of dependence/independence among its propositions. This freedom to express partial information makes dependency graphs suitable for modeling the degrees of belief one should accord a group of related propositions based on evidential information, a suitable host for Shafer's theory.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF DEFINITIONS

CHAPTER I

INTRODUCTION

The ability to extend a given body of information through inference is an important aspect of any intelligent system. This ability is based on an understanding of the dependencies within an environment of interest. Dependence implies predictability. If propositions are mutually dependent and the nature of that dependence is understood, then information concerning the truthfulness of some of these propositions can be translated into information about the truthfulness of some of the others.

An important type of inference involves the determination of the truthfulness of propositions. This type of inference is well understood when the truthfulness of the propositions can be expressed as certainties. However, intelligent systems often need to reason from uncertain and incomplete information. This reflects the evidential nature of their domains of application. Propositions may not be known to be true or false, but may only be attributed degrees of support based on bodies of evidence extracted from the environment of interest. Indeed, if the bodies of evidence are inconclusive, the exact degrees of support may not even be known, but only some bounding conditions on them. The propositional and predicate calculi are not sufficient in these evidential domains requiring inexact reasoning from incomplete information.

Initially, the need for inexact reasoning led to the abandonment of formally founded systems and a move towards informal systems based on ad hoc scoring functions. These models frequently consisted of a delicately "balanced" system of weights, arrived at through a "tuning" process of repeated adjustments. Typically, the complexity of these models thwarted all attempts to analytically understand or improve their performance. The levels of performance attained are a credit to the perseverance of their designers, though the reliability of these systems remains an article of faith.

More recently, systems have been developed based on formal logics of inexact reasoning in the hopes of improving their understandability. Though these systems are formally based, they still employ intuitively motivated techniques. This is necessary since the logics and the domains of application are not fully compatible. The introduction of these techniques has some undesirable effects. If inferencing were allowed to proceed in an unconstrained manner, inconsistent predictions might be made. That is, several incompatible predictions about the truthfulness of a single proposition might be simultaneously generated. To prevent this from occurring, constraints are placed on the inferencing, maintaining consistency at a cost of reduced flexibility.

Inference can be formalized in terms of a set of propositions, a covering assignment of confidences, and a set of dependency relationships that constrains the assignment of confidences. The assigned confidences reflect the perceived truthfulness of the propositions. The dependency relationships describe how the truthfulness of the propositions are interrelated. Inferential reasoning extends partial coverings towards full coverings that are consistent with the dependency relationships.

These dependency relationships collectively form a dependency relation. If they are mutually consistent, they form a simultaneously satisfiable set of constraints over the confidences of the propositions. However, if they are not properly coordinated, they simultaneously describe some dependency relationships in several incompatable ways, resulting in an unsatisfiable set of constraints. Predictions, based on such an inconsistent dependency relation, are likewise inconsistent. This is the major problem with most of the previously developed systems for inexact reasoning. They are not internally consistent.

This thesis provides an internally consistent system capable of reasoning from evidential information. The inferential constraints that compensate for the lack of consistency in other systems are not needed, thereby increasing inferential flexibility. This system is based on a formal logic of inexact reasoning better suited to evidential domains than those logics previously employed. The result is a system that is both intuitively and analytically more understandable.

Our approach was to develop a general theory of propositional dependence and then to specialize this to a theory of evidential dependence. This began with the formalization of the notion of a dependency relation and its graphical representation. Propositions are represented as nodes, dependency relationships as arcs, and assigned confidences as values of the nodes. Graphical representations have been used before and have proven well-suited to mechanized inferential reasoning. Confidence information is translated from proposition to proposition, predicting proposition confidences based on the confidences of neighboring propositions, with the graphical structure providing the appropriate indexing.

Unlike previously developed graphical representations of dependency information, <u>dependency graphs</u> have an associated set of consistency conditions that must be satisfied. These preserve the fundamental properties of dependence, guaranteeing that no dependency relationships are redundantly described in incompatible ways. This assures the integrity of the predictions based on these relationships.

A <u>dependency-graph model</u>, consisting of a dependency graph and an accompanying inference engine, makes predictions about the environment it models. The dependency graph describes how a set of propositions depend on one another and reflects the perceived dependencies in the environment. The inference engine makes predictions based on these perceived dependencies, taking incomplete information about the confidences of the propositions from some source of knowledge over the environment, and extending it through inferential reasoning. If the dependency graph and the initial confidence information accurately reflect the environment, so will the predictions. The internal consistency of a model is guaranteed if the dependency graph consistency conditions are satisfied. Therefore, inaccuracies must be attributed to external, not internal, inconsistencies. The theory of dependency-graph models makes no statement about how to achieve external consistency for any particular class of environmental situations.

A body of information might not provide the exact confidence of a proposition, but only provide some partial information concerning it; that is, only provide some set of confidences that the true confidence must be within. Similarly, dependence is not always complete; it too can be partial. Knowing the exact confidence of one of two partially dependent propositions may translate into a set of possible confidences for the other. In either case, dependency graphs represent such partial information by assigning to propositions sets of confidence values instead of single confidence values. This freedom to express partial information within dependency graphs is critical to their application as models of evidential support. Partial information is the rule, not the exception, in evidential domains. Without it, varying degrees of ignorance cannot be properly represented or reasoned about.

Though the theory of dependency-graph models makes no statement about how to achieve external consistency for any particular class of environmental situations, a mathematical theory of evidence by Glenn Shafer provides just the formal foundation necessary to construct dependency-graph models of evidential support. The freedom to express partial information within dependency-graph models makes them a suitable host for Shafer's theory. The adoption of Shafer's theory leads to the adoption of a subset of dependency-graph models as appropriate models of evidential support.

A dependency-graph model of evidential support takes a single body of evidential information (at a time) and extends it through inferential reasoning, translating that information from the propositions the evidence directly bears upon, to those propositions the evidence indirectly bears upon. When knowledge sources are errorful - the typical situation in artificial intelligence applications - it is imperative that the distinct bodies of evidential information they produce be combined to compensate for their individual failings. Dempster's rule of combination, an integral part of Shafer's theory, does exactly that: it pools distinct bodies of evidential information. With its addition, the results of multiple knowledge source applications can be combined and the appropriate logical conclusions drawn.

Figure 1 illustrates the basic architecture of a dependency-graph model of evidential support. The heart of the system is the dependency graph. It describes how the truthfulness of a number of propositions are interrelated relative to the environment of interest. It is based strictly upon information delimiting the possibilities in the environment, making it far easier to construct and verify than those models requiring probabilistic estimates. Each knowledge source, after having examined the environment, relates its findings to the model through a "mass distribution." These mass distributions each partition a unit of belief among a subset of the propositions in the dependency graph, mass being attributed to those propositions for which there is direct supporting evidence. To compensate for the individual errors of the knowledge sources, their findings are pooled, according to Dempster's rule of combination and the information in the dependency graph, resulting in a new mass distribution that represents the consensus of their individual opinions. Based upon this mass distribution and the dependency graph, the inference engine derives an "evidential interval" for

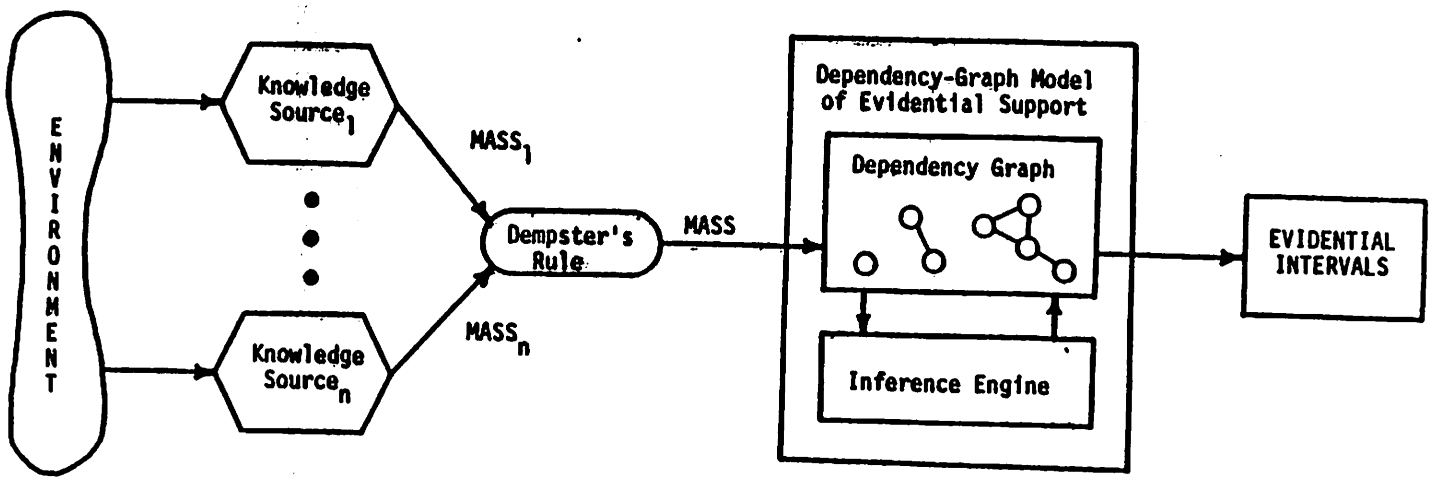


Figure 1. Basic system architecture.

each proposition in the graph. These intervals are subintervals of [0,1]. The lower bound of an evidential interval represents the total weight of supporting evidence for a proposition, the upper bound represents the extent to which that proposition remains plausible, and the width represents the degree to which that proposition's likelihood remains unknown. So long as the knowledge sources do not completely contradict one another, this procedure is guaranteed to produce a covering of evidential intervals, that is consistent with Shafer's theory and invariant with respect to the order of the inferential steps.

This thesis is both a description of a general representation of dependency information and its use as a basis for inferential reasoning, as well as a description of a specific representation of evidential support and its use as a basis for evidential reasoning (Figure 2).

The remainder of this report begins with a more detailed discussion of the previous approaches taken to the mechanization of inexact reasoning. This is followed by definitions of dependency relations and their representation as dependency graphs. A taxonomy for classifying dependency relations and dependency graphs according to the specificity and order of the dependency relationships contained within them allows there to be an incremental introduction. They vary from those restricted to total, binary relationships, to those that are unconstrained. This leads into a discussion of inferential reasoning over dependency graphs, including the introduction of a distinct type of inference engine for each class of dependency graph. Dependency-graph models are similarly classified and discussed. The next chapter reviews Shafer's

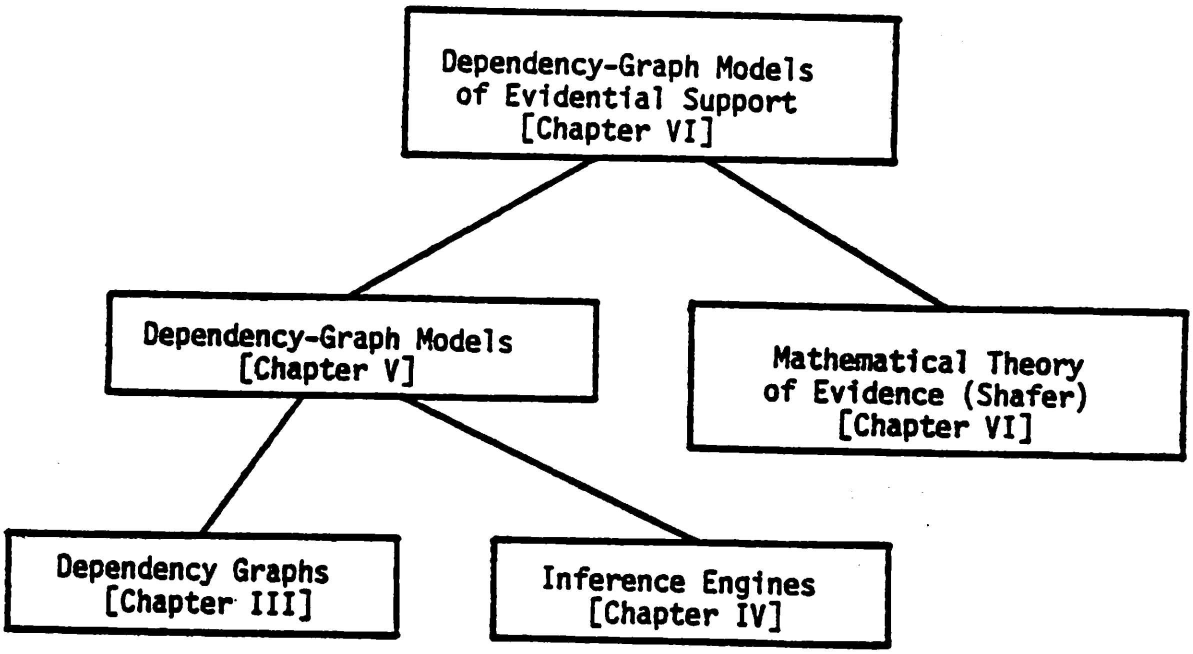


Figure 2. Thesis components.

mathematical theory of evidence, and describes how it can be used as a guide to the construction of dependency-graph models of evidential aup-port. This chapter concludes with an example application, highlighting the strengths of this form of inexact reasoning. Finally, some summary comments and suggestions for further investigation round out this discussion.

CHAPTER II

RELATED WORK

When artificially intelligent systems were applied to real-world situations, and not just those found in idealized theoretical worlds, Boolean logics were no longer sufficient. Truth and falsity needed to be replaced by degrees of belief; precision, by degrees of ignorance. Inferential reasoning could no longer be based on the propositional and predicate calculi.

Initially, ad hoc techniques replaced formal logics as the bases of these systems. With the abandonment of formality came an accompanying lack of understanding, making system performance difficult to predict or improve. Experimental testing and tuning replaced analysis and correction. But the simultaneous correction of both internal and external inconsistencies, performed through repeated adjustments, proved difficult. Even when this tuning process seemed to have produced a "balanced" system, some simple additions or modifications might completely disrupt its performance. Trial-and-error tuning proved to be no substitute for formal understanding.

Frustration led designers back to formality. Statistical decision theory seemed particularly attractive, especially in the form of Bayes' rule of conditioning. Indeed, several early programs, based on Bayes' rule, successfully modeled the medical diagnostic process [Gorry 1968, 1973; Warner 1964]. This success was largely due to the vast amounts of data available pertaining to their respective diagnostic domains, thus allowing the required probabilistic measures to be accurately estimated. Unfortunately, this kind of data is not generally available in many other domains of interest.

A Bayesian approach in domains where the appropriate data is not available necessitates the use of numerous approximations and assumptions. Subjective probabilities, provided by an "expert," replace statistically estimated probabilities. Inconsistencies in these estimates and oversimplifying assumptions can render this entire approach worthless, realizing none of the benefits that originally justified its use.

Historically, the inadequacies of Bayesian probability in the analysis of real-world problems has led to a variety of alternative formal approaches. These include the theory of "fuzzy sets" [Zadeh 1965; Goguen 1968], the theory of "choice" [Tversky 1972; Luce 1965], the logic of "surprise" [Shackle 1952, 1955], the theory of "confirmation" [Carnap 1950; Hempel 1945], the theory of "upper and lower probabilities" [Dempster 1967, 1968], and the theory of "evidence" [Shafer 1973, 1975, 1976]. Each of these provides an alternative basis for mechanized inferential reasoning, several having already served in this capacity.

MYCIN [Shortliffe 1974, 1976; Shortliffe and Buchanan 1975], one of the most successful and influential systems, is based on an alternative theory, the theory of confirmation. The approach was to develop a system that reflects the observations of philosophers who have dealt with the theory of confirmation, but not to be completely constrained by their results. Whenever this theory proved inadequate, intuitively motivated techniques were added. MYCIN's knowledge-base consists of a set of rules, each consisting of one or more stimulus proopsitions, a response proposition, and a "certainty factor" that quantifies the degree to which belief in the stimulus propositions confirms the response proposition. This knowledge-base is provided by an expert, with the certainty factors playing the same role as the estimated conditional probabilities in a Bayesian approach. A certainty factor combines two distinct measures from the theory of confirmation, measures of "Belief" and "Disbelief,"[1] into a single number. The claim is that this number more closely corresponds to the number an expert gives when asked to quantify the strength of a judgmental rule. The inference rule utilized was conceived purely on intuitive grounds, but satisfies some theoretically motivated criteria [Törnebohm 1966]. Though this system is not formally sound, it was carefully constructed with an eye towards formal concerns, making it analytically more understandable. The justification for their approach did not rest with "a claim of improving on Bayes' Theorem, but rather with the development of a mechanism whereby judgmental knowledge can be efficiently represented and utilized for the modeling of medical decision making, especially in contexts where (a) statistical data are lacking, (b) inverse probabilities are not known, and (c) conditional independence can be assumed in most cases."[2]

---

[1]These terms are used by Shortliffe in explanation of the theory of confirmation. They are not part of the standard terminology of the theory.

[2]Shortliffe 1976, p. 185.

MYCIN's success in diagnosing bacterial infections, despite a lack of statistical data, suggested that similar techniques might be advantageously employed in other domains with a shortage of statistical data. MYCIN's use of production rules to represent judgmental knowledge and its inclusion of formally based mechanisms for handling uncertainty were the dominant influences in the design of PROSPECTOR [Duda, Hart, Nilsson, and Sutherland 1977], a geological consultant system intended to help geologists in evaluating the mineral potential of exploration sites. However, PROSPECTOR is not based on the theory of confirmation, but on a subjective Bayesian technique that retains, insofar as possible, the well-understood methods of Bayesian probability theory, introducing only those modifications needed to compensate for the subjectivity of the probabilities. Subjective probabilities are interpreted as measuring degrees of belief rather than long-run relative frequencies of occurrence [Fine 1973].

PROSPECTOR reformulated the problem of rule-based inexact reasoning in terms of a directed graph. The utility of graphical representations for mechanized inferential reasoning had been previously demonstrated [Erman and Lesser 1975; Trigoboff 1976]. Judgmental rules typically are not independent, but can be linked together to form what they term an "inference network." A link explicitly occurs when the response proposition of one rule is a stimulus proposition of another. In this representation, propositions are represented as nodes, judgmental rules as directed arcs between nodes, probabilities associated with the nodes indicate degrees of belief in those propositions, and conditional probabilities associated with the arcs indicate the strength of those rules.[3] Inferencing is defined in terms of propagating probabilities from node to node, according to the information along the arcs.

PROSPECTOR used approximations to overcome many of the problems of dealing with subjective probabilities. Some of these approximations include: the use of a piecewise linear interpolation formula to correct for inconsistent probabilities (i.e., probabilities that do not conform with Bayes' rule); the assumption that evidence combines either independently or as a logical function (i.e., conjunction, disjunction, or negation); an interpolation formula to account for the combination of uncertain independent evidence; the use of simple formulas from fuzzy set theory to combine dependent evidence. These approximations led to a computationally simple method for updating probabilities that has proven very successful [Duda, Hart, Nilsson 1976].

PROSPECTOR's formulation offers several advantages over that of MYCIN. Since Bayesian probability theory is the most widely known, PROSPECTOR's formulation often creates fewer conceptual barriers. PROSPECTOR's graphical representation tends to ease the visualization of the knowledge-base, and thereby of the inference process. Unlike MYCIN, PROSPECTOR can utilize a rule regardless of the level of support that exists for its stimulus propositions. MYCIN can only use a rule

---

[3] The actual values associated with the arcs are quotients of conditional probabilities and are better suited to directed inferencing in their formulation.

once its stimulus propositions have attained a level of support above a preset, empirically selected threshold. Thus, PROSPECTOR makes more complete use of the available information than does MYCIN.

Although both MYCIN and PROSPECTOR represent giant steps towards a well-founded theory of mechanized inexact reasoning, they share some common problems. A number of these problems center around their inherent lack of internal consistency. This is not the Bayesian type of inconsistency previously alluded too, but a functional inconsistency. Chains of inferences across several rules, beginning and ending with the same proposition, form "loops." Inferences around these loops cannot be handled in a reasonable manner. If such a loop were permitted to occur, the probability on which all of the inferences are based, that is the probability of the initial proposition, would likely be refuted upon completion of the loop. Both MYCIN and PROSPECTOR are forced to eliminate the possibility of these loops at a cost of reduced flexibility.

A closer look at PROSPECTOR will better illustrate this problem. Imagine two propositions A and B. Bayes' rule describes how their marginal probabilities P(A) and P(B) are related to their conditional probabilities $P(A|B)$, $P(A|\bar{B})$, $P(B|A)$, and $P(B|\bar{A})$. If B is known to be true, then $P(A|B)$ is the probability of A; if B is false, then its $P(A|\bar{B})$; $P(B|A)$ is the probability of B given A; and $P(B|\bar{A})$ is the probability of B given $\bar{A}$. But what if the best estimate of the truthfulness of A is $P_t(A)$? What is the corresponding probability of B, $P_t(B)$?

PROSPECTOR uses the following equations, derivable from Bayes' rule, as a basis for linearly relating $P_t(A)$ and $P_t(B)$:

$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B}), \tag{2.1}$$

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}). \tag{2.2}$$

$P_t(A)$ and $P_t(B)$ are substituted for $P(A)$ and $P(B)$ in equation (2.1), producing the following mapping from $P_t(B)$ to $P_t(A)$:

$$P_t(A) := P(A|B)P_t(B) + P(A|\bar{B})[1 - P_t(B)]. \tag{2.3}$$

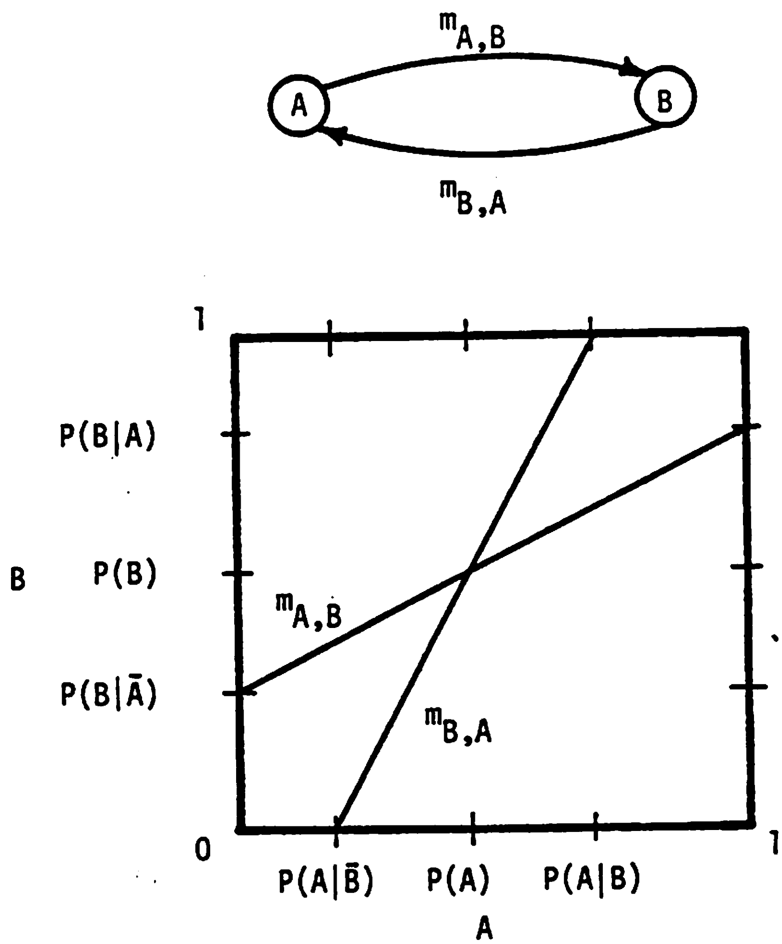
If the same substitutions are made in (2.2), an analogous mapping from $P_t(A)$ to $P_t(B)$ is derived:

$$P_t(B) := P(B|A)P_t(A) + P(B|\bar{A})[1 - P_t(A)]. \tag{2.4}$$

The problem becomes apparent if these are considered as the mappings relating propositions A and B by two distinct rules, one with A as the stimulus and B as the response, the other with B as the stimulus and A as the response.

These mappings are not mutual inverses. In general, their only point of agreement about the relationship between $P_t(A)$ and $P_t(B)$ is at the marginals (Figure 3). If unconstrained inferencing were allowed in spite of this symmetric inconsistency, feedback between A and B would eventually drive $P_t(A)$ and $P_t(B)$ to the marginals (Figure 4).

In actuality, PROSPECTOR uses more complicated mappings (ones that are piecewise linear, broken at the marginals) to correct for Bayesian inconsistencies within the subjectively specified marginal and condi-



$$P_t(B) := m_{A,B}[P_t(A)] = P(B|A)P_t(A) + P(B|\bar{A})[1 - P_t(A)]$$

$$P_t(A) := m_{B,A}[P_t(B)] = P(A|B)P_t(B) + P(A|\bar{B})[1 - P_t(B)]$$

Figure 3. Inconsistent Bayesian based mappings

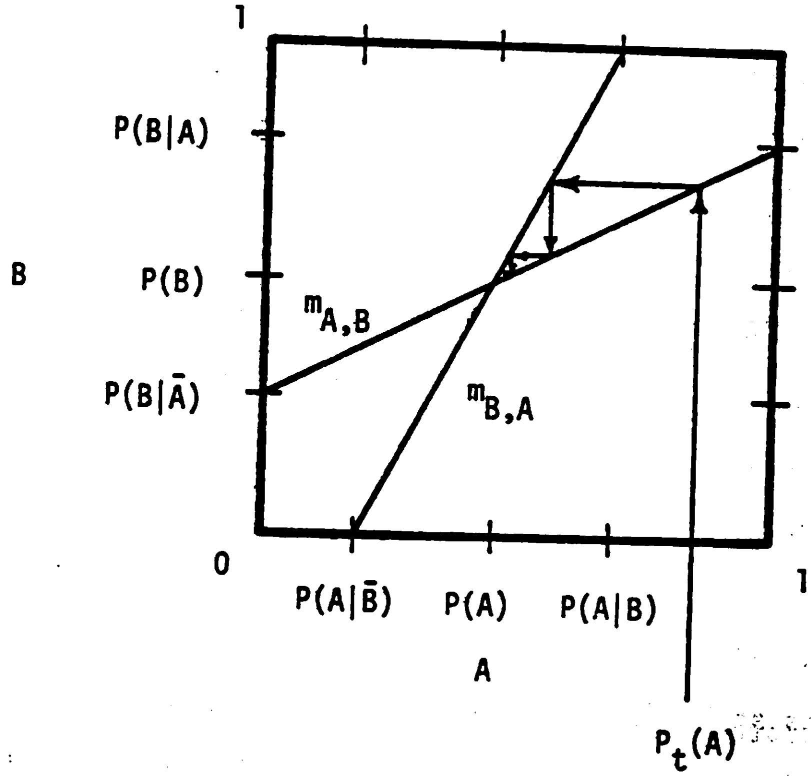


Figure 4. Feedback of Bayesian based mappings

tional probabilities. But, the characterization of the problem remains the same. The root cause is that Bayes' rule is a statement about the relationship between marginal and conditional probabilities, and cannot be parameterized to make predictions based on partial beliefs.

Both MYCIN and PROSPECTOR partially correct for these problems by severely constraining their inferencing process. Inferences are restricted to a single, predetermined direction along each arc within the graph. And still, care must be taken that no sequence of inferential steps forms a loop. The inconsistent feedback problem is eliminated, but at a high cost in terms of flexibility since no single set of directed inferential steps can be the most informative in all cases.

Recently, modifications have been proposed for both MYCIN and PROSPECTOR that would allow inferential reasoning in both directions along each arc [Friedman 1980; Konolige 1979]. Propositions would no longer be predetermined to always serve as the stimulus or response of a given rule, but could be dynamically selected to play either role. However, loops still must be prevented; rules can only be used in a single direction at a time. The inconsistent feedback problem remains.

A mathematical theory of evidence and an accompanying theory of probable reasoning by Glenn Shafer [Shafer 1973, 1975, 1976] provides an alternative foundation for the construction of mechanized systems of inexact reasoning. Shafer's work is an extension of Arthur Dempster's work on partial beliefs [Dempster 1967, 1968]. Shafer's theory departs from the more traditional Bayesian theory [Bayes 1763], avoiding several of its documented difficulties [Shafer 1976, Boole 1854, Fisher 1956]: of particular significance, its inability to represent ignorance and its insistence that new evidence be expressible as a certainty.

Shafer's theory is fundamentally different from those theories underlying MYCIN and PROSPECTOR. Unlike the theory of confirmation and the Bayesian theory, Shafer's theory does not rely on prior probabilities. It takes a conservative view. Inferences are made by eliminating the impossible, not by assuming the probable, much in the spirit of "constraint-satisfaction" and "relaxation" [Waltz 1972; Rosenfeld, Hummel, and Zucker 1976]. Sets of confidences are associated with propositions. It is presumed that the true confidence of each proposition is an element of its associated set. Inferencing consists of reducing these sets by eliminating those elements that are inconsistent with the sets assigned to related propositions. However, Shafer's theory must be viewed as a specialization of constraint-satisfaction and relaxation, since it also prescribes what confidence assignments are consistent with each propositional relationship in a way that guarantees inferential stability. The reasoning is from information about the possibility of co-occurrence, not the probability.

This is not to say that Shafer's theory bears no resemblance to the theory of confirmation or the Bayesian theory. In fact, Shafer's theory can be viewed as a direct generalization of the Bayesian theory [Shafer 1976]. This generalization is in the direction of the theory of confirmation. Shafer rejects the idea that the support afforded a proposition and its negation, based on a body of evidence, must sum to one.

He, and the theory of confirmation, maintains that an adequate summary of the impact of a body of evidence on a proposition must include the degrees to which it supports and refutes that proposition. Their sum is bounded by one, but is not constrained to equal one. This is the key to the representation of ignorance. When there is little evidence bearing on a proposition, frank agnosticism is expressed by according both that proposition and its negation very low degrees of support. Thus, unlike the Bayesian theory, Shafer's theory carefully avoids equating the lack of belief in a proposition with disbelief in that proposition.

We propose the adoption of Shafer's theory as a basis for mechanized inexact reasoning. Shafer argues for its suitability in evidential domains where information is uncertain and incomplete [Shafer 1976]. This is not to say that this theory is without its critics [Zadeh 1979]. However, its suitability is supported by the disappearance of the inconsistent feedback problem without the introduction of needless or inflexible constraints on the inferential process.

CHAPTER III

DEPENDENCY RELATIONS AND THEIR GRAPHICAL REPRESENTATIONS

Dependence is the foundation of inferential reasoning. This chapter is devoted to the properties and representation of propositional dependence.

Dependence/Independence

Two propositions are totally dependent on one another if the confidence (i.e., truthfulness) of one exactly determines the confidence of the other. If the confidence of one proposition leaves the confidence of another proposition completely in doubt, they are said to be totally independent. But dependence/independence need not be total. In this case, predictions are not always exact, and it may only be possible to predict a set of confidences within which the true confidence must lie. Thus, dependence/independence ranges from total dependence, to partial dependence/independence, to total independence.

Propositions that are mutually dependent, whether it be partially or totally, constitute a dependency relationship. These relationships need not be binary. A proposition's confidence may depend on the collective confidences of a group of propositions, rather than the confidence of a single proposition. The number of propositions involved in a dependency relationship is its order. A proposition's dependence on a group of propositions neither supports nor refutes that proposition's dependence on a subset of those propositions.

A coordinated collection of dependency relationships constitutes a dependency relation. It describes how the confidences of a set of propositions co-vary. Each n-ary dependency relationship has n associated predictive mappings. Each of these mappings predicts the confidence of one proposition based on the confidences of the other propositions in that relationship. These mappings can be arbitrarily complex. However, they cannot be arbitrarily assigned. They must be coordinated within a dependency relation; otherwise they might simultaneously describe a single dependency relationship in several incompatible ways. When such incompatibilities exist, the dependency relation is said to be inconsistent. A consistent set of dependency relationships forms a simultaneously satisfiable set of constraints over the proposition confidences. The solution space of these constraints defines the space of consistent coverings.

Dependency relations and their corresponding graphical representations, dependency graphs, are classified according to the order and specificity of the relationships contained within them: by order, whether or not all of the dependency relationships are binary, by specificity, whether or not all of the dependency relations are total. Subscripts denote their order ("2" for binary and "+" for higher orders) and superscripts denote their specificity ("1" for total and "+" for partial). Four classifications result: for dependency relations, D_2^1, D_2^+, D_+^1, and D_+^+; for dependency graphs, G_2^1, G_2^+, G_+^1, and G_+^+.

At times, these classes are interpreted as being exclusive. A dependency graph that represents only total dependency relationships with order no greater than two, necessarily falls in class G_2^1. However, these also can be interpreted hierarchically, based on their representational power. A G_2^1 dependency graph can be represented as a G_2^+ or G_+^1 dependency graph. In turn, G_2^+ and G_+^1 dependency graphs can be represented as G_+^+ dependency graphs. The reverse is not necessarily true. G_+^+ dependency graphs cannot, in general, be represented as G_2^+ or G_+^1 dependency graphs and they cannot, in general, be represented as G_2^1 dependency graphs. This hierarchy is summarized (Figure 5). The discussion that follows explains these classes in detail.

Dependency relations and dependency graphs are incrementally introduced in the remainder of this chapter, the simpler, more restrictive, providing the groundwork for the more general.

Binary Dependency Relations and Their Graphical Representations

D_2 dependency skeletons. Dependency skeletons are the frameworks of dependency relations. They describe where dependency relationships exist, without describing the nature of those dependencies. If all of the described dependency relationships are binary, the skeleton is called a D_2 dependency skeleton. Such a skeleton is defined as a binary relation over a set of propositions P. However, not all binary relations over P can be legitimate descriptions of dependence. A D_2 dependency skeleton is an equivalence relation. That is, it must be reflexive, symmetric, and transitive to reflect the nature of dependence: every proposition is dependent on itself; if proposition p_i is depen-

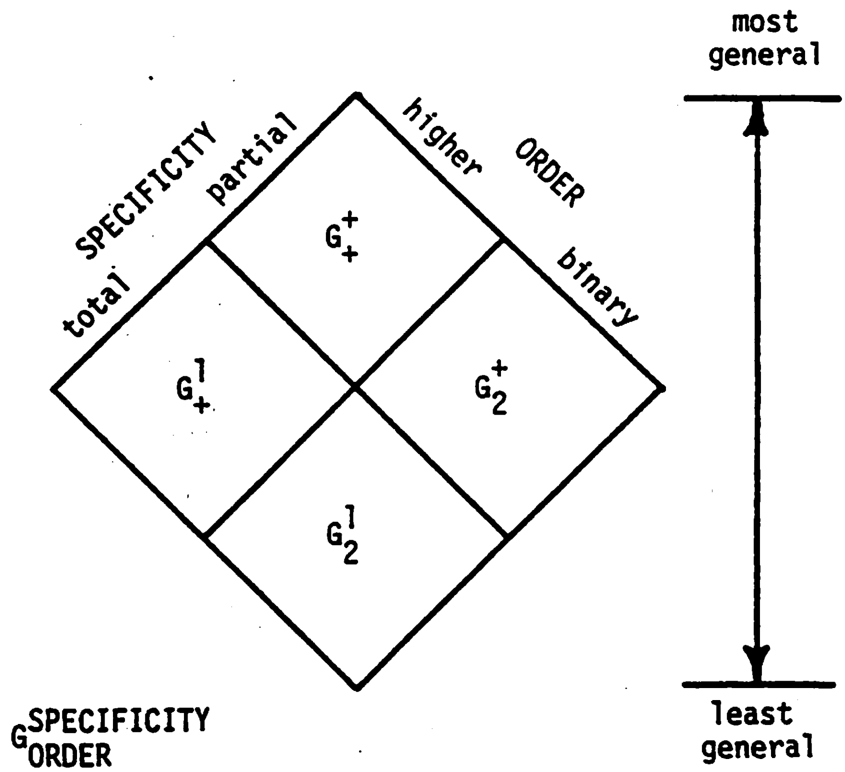


Figure 5. Summary of dependency-graph taxonomy

dent on p_j, then p_j is dependent on p_i; if p_i depends on p_j and p_j on p_k, p_i depends on p_k. This is captured by the following definition, where P is a set of propositions and D is a set of binary dependency relationships over P.

DEFINITION 1. D_2 dependency skeletons.

> A D_2 dependency skeleton is an ordered pair (P,D) where P is a set of propositions and D is a set of two element subsets of P subject to the constraint that
>
> - if $d_i, d_j \in D$ and $\{p\} = d_i \cap d_j$, then $((d_i \cup d_j) - \{p\}) \in D$. □

D_2 dependency skeletons can be straightforwardly represented as undirected graphs with nodes representing porpositions and arcs between nodes representing dependency relationships between propositions. Not every dependency relationship need be represented by an arc in the graph. Since D_2 skeletons are transitive, all connected nodes represent pairwise dependent propositions. Arcs connecting nodes that already have a path between them are redundant and can be left out so long as the transitive nature of connectivity is recognized. Similarly, arcs connecting nodes to themselves, due to reflexivity, are excluded. Thus, several distinct graphs may represent the same D_2 dependency skeleton (Figure 6).

Dependence and independence are easily interpreted in terms of this graphical representation. Propositions are mutually dependent if their associatied nodes are connected (i.e., there is a path between them), otherwise they are independent. Each connected portion of the

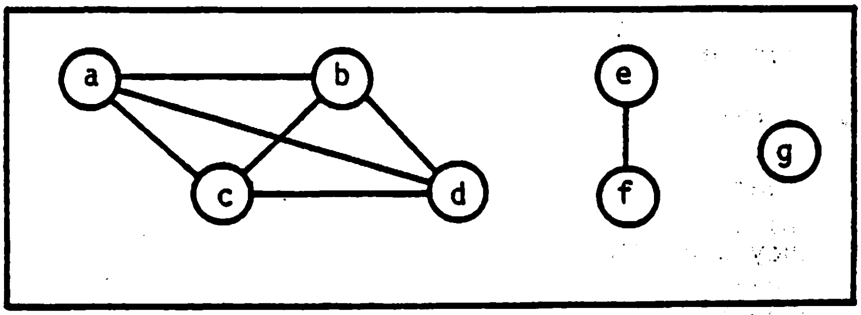
Figure 6. An example D_2 dependency skeleton and some alternative graphical representations.

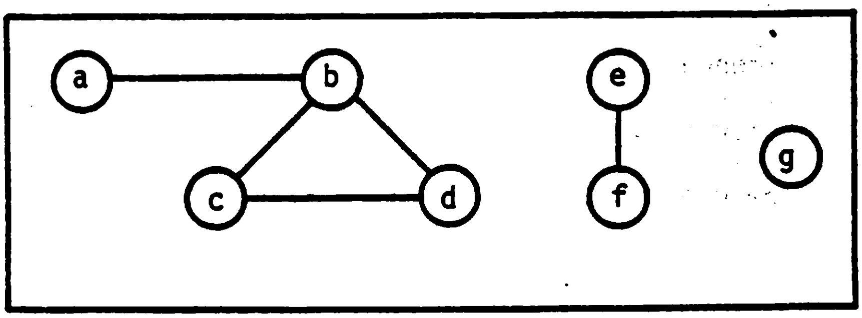
(P,D) = ({a,b,c,d,e,f,g},

{{a,b}, {a,c}, {a,d}, {b,c}, {b,d}, {c,d},{e,f}}).

Equivalence classes of (P,D): {a,b,c,d}, {e,f} & {g}.

Alternative graphs representing (P,D):





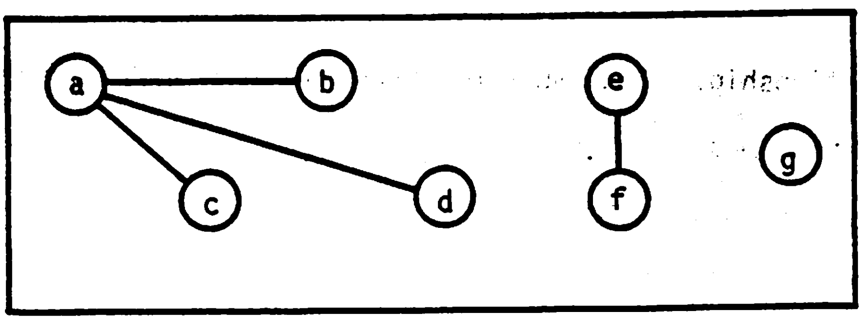


Figure 6

graph delineates an equivalence class of the dependency skeleton.

D_2 dependency relations. D_2 dependency skeletons and their graphs describe where binary dependencies exist, but they do not describe the nature of those dependency relationships. Without this information, inferences cannot be made. A D_2 dependency relation is a D_2 dependency skeleton with this added information. Formally, a D_2 dependency relation is an ordered triple (P,D,M). The first two elements constitute a D_2 dependency skeleton and the last element is a function that maps binary relationships from D into sets of confidence pairs selected from a range of confidences C. Each such pair represents a possible, simultaneous, confidence assignment to the propositions taking part in the selected relationship. In other words, each such set tabularly describes how the confidences in the propositions are related.

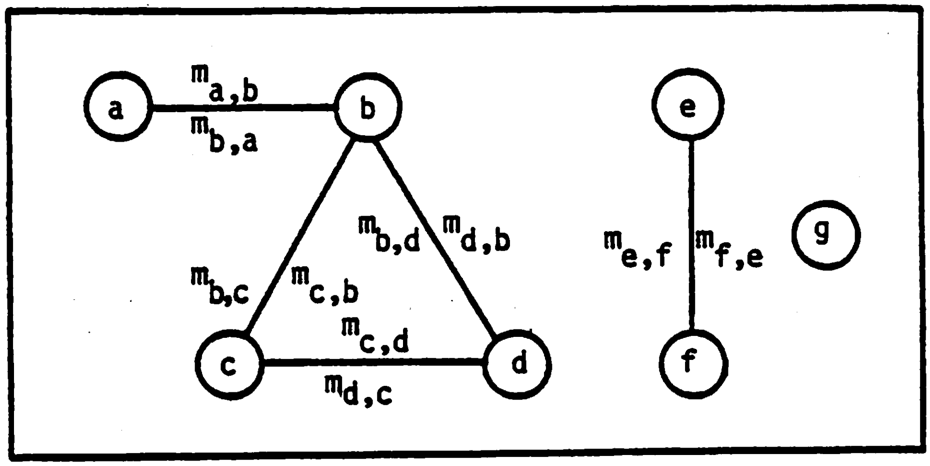
Graphically, we take a directed, fragmented view of M. First, M is divided into a set of functions, each of which is defined over exactly one of the binary relationships in D and is identical to M over that relationship. Then each of these functions are divided into two directed mappings: if the original function is defined over the propositions p_i and p_j, then one of the two mappings $m_{i,j}$ maps confidences associated with p_i to compatible confidences for p_j, and the other mapping $m_{j,i}$ maps confidences associated with p_j to compatible confidences for p_i. When these mappings are placed along the appropriate arcs of a graph representing the D_2 dependency skeleton of a D_2 dependency relation, a G_2 dependency graph for that relation is formed. Each arc has exactly two associated mappings relating the propositions it connects, one for each direction (Figure 7).

These added mappings describe the dependency relationships in a directed form well-suited to mechanized inferencing. One can imagine propagating confidences from node to node, along the arcs, according to these mappings, and thus inferring unknown confidences from known confidences.

Of course, M cannot be arbitrarily selected to form a D_2 dependency relation. If it were not well-formed, it might simultaneously describe a single dependency relationship in several incompatible ways, and this could lead to incompatible inferences. To ensure that D_2 dependency relations are consistent, their definition includes three consistency conditions: the first guarantees that the dependency relation is symmetrically consistent, the second guarantees that it is transitively consistent, and the third guarantees that no confidence assignment is unilaterally excluded.

DEFINITION 2. D_2 dependency relations.

> A D_2 dependency relation is an ordered triple (P,D,M), where (P,D) is a D_2 dependency skeleton and M is a function from ordered elements of D into ordered pairs from the confidence range C (i.e., for each $\{p_1,p_2\} \in D$, $M[(p_1,p_2)] \subseteq C^2$), subject to the constraints:
>
> 1. for every $\{p_1,p_2\} \in D$,
>
> $$(c_1,c_2) \in M[(p_1,p_2)] \leftrightarrow (c_2,c_1) \in M[(p_2,p_1)];$$
>
> 2. if $\{p_1,p_2\}$, $\{p_2,p_3\}$, $\{p_1,p_3\} \in D$,
>
> $$(c_1,c_2) \in M[(p_1,p_2)], \text{ and } (c_2,c_3) \in M[(p_2,p_3)],$$



$$(c_i, c_j) \in M[(p_i, p_j)] \leftrightarrow c_j \in m_{i,j}[c_i]$$

Figure 7. An example G_2 dependency graph.

then $(c_1,c_3) \in M[(p_1,p_3)]$;

3. for every $\{p_1,p_2\} \in D$ and every $c_1 \in C$, there exists $c_2 \in C$ such that

$$(c_1,c_2) \in M[(p_1,p_2)]. \quad \square$$

These consistency conditions for D_2 dependency relations can be related to their graphical representations. Symmetric consistency (1) guarantees that the two mappings placed along any single arc describe exactly the same correspondence, while transitive consistency (2) guarantees that all composite mappings implied by any sequence of arcs connecting the same two nodes are compatible. The final condition (3) simply states that every confidence represents a potential assignment for every node.

D_2^1 dependency relations. If a D_2 dependency relation has only total dependency relationships, then it is called a D_2^1 dependency relation. The superscript signifies that given any of these dependency relationships and a confidence assignment for one of its propositions, there is exactly one corresponding confidence value that can be assigned the other proposition. In other words, M is functional.

DEFINITION 3. D_2^1 dependency relations.

A D_2^1 dependency relation (P,D,M) is a D_2 dependency relation with the additional constraint that M is functional. In other words,

- for every $\{p_1,p_2\} \in D$, $c_1 \in C$, $|m_{1,2}[c_1]| = 1$, where $m_{1,2}[c_1] = \{c_2 | (c_1,c_2) \in M[(p_1,p_2)]\}$. $\square$

This means that the mappings placed along the arcs of a G_2^1 dependency graph are one-to-one; any two mappings along a single arc are mutual inverses; and the ordered combination of mappings around any circuit (a sequence of arcs along a path which begin and end at the same node) results in the identity mapping.

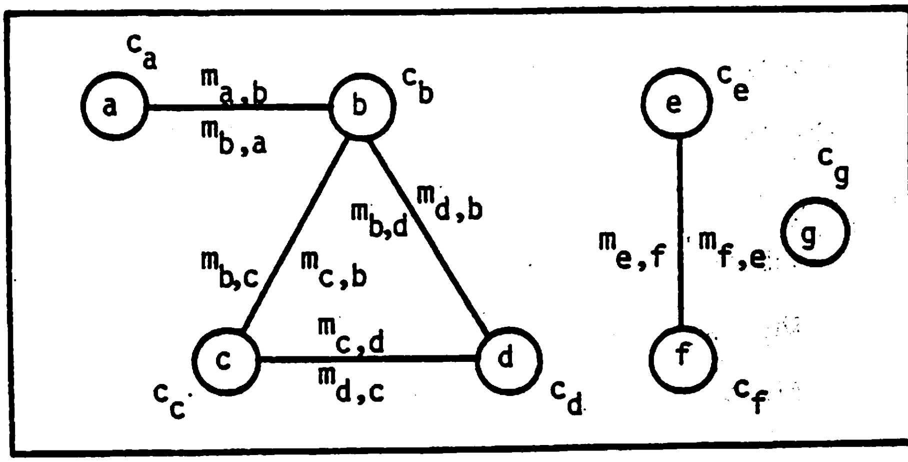
Clearly, a well-informed D_2^1 dependency relation describes a simultaneously satisfiable set of constraints over the confidences of a set of propositions. An assignment of confidences to propositions that satisfies these constraints is the object of the inferential process. Such an assignment is described by a C^1 covering.

DEFINITION 4. C^1 coverings for D_2 dependency relations.

> A C^1 covering for a D_2 dependency relation (P,D,M) is a function from P into a range of confidences C, $C^1:P \rightarrow C$, subject to the constraint that
>
> - for every $\{p_1,p_2\} \in D$,
> $(C^1[p_1], C^1[p_2]) \in M[(p_1,p_2)]$. □

Graphically, a C^1 covering sets the value of each node to its assigned confidence. The covering is consistent if all of the mappings along the arcs are satisfied. If any mapping is unsatisfied, the covering is inconsistent. In the case of a G_2^1 dependency graph, a mapping $m_{i,j}$ from the confidence c_i of a proposition p_i to the confidence c_j of a proposition p_j is satisfied if the image of c_i under $m_{i,j}$ is c_j (Figure 8).

Some example G_2^1 dependency graphs will help to illustrate. The



$$m_{i,j}[c_i] = c_j \leftrightarrow \{(c_i,c_j)\} = M[(p_i,p_j)],$$

where $c_i = C^1[p_i]$ and $c_j = C^1[p_j]$.

Figure 8. An example C^1 covering for a G_2^1 dependency graph.

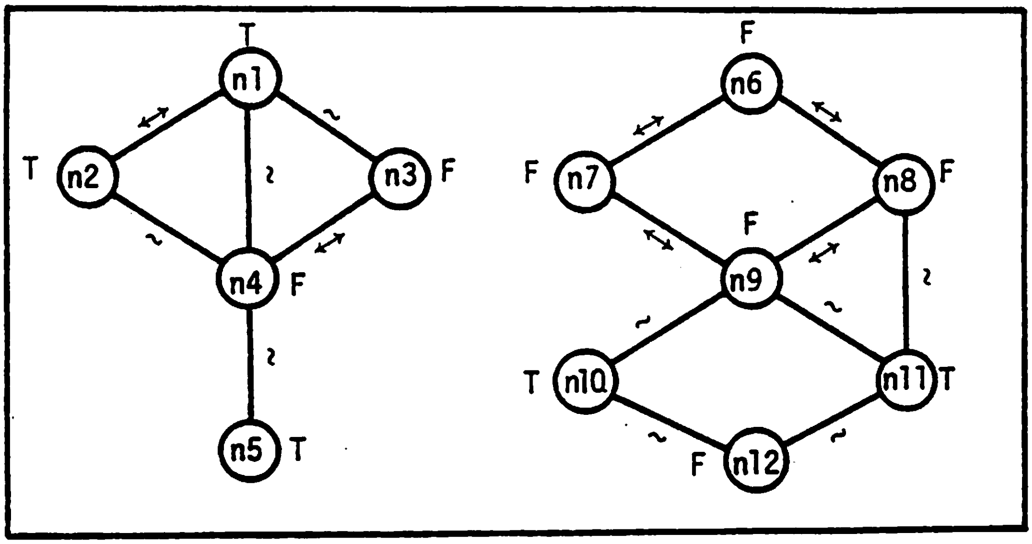
first example (Figure 9) mimics the propositional calculus restricted to the propositional connectives $\leftrightarrow$ and $\sim$. Confidences range over the Boolean values T and F; each dependency mapping is either the identity or complimentary mapping. Such a graph is consistent if and only if each arc has the same mapping for both directions and if every circuit contains an even number of complimentary mappings.

The second example (Figure 10) is similar to the first save that $C = [0,1]$. The mappings are still the identity and complimentary mappings but defined over $[0,1]$. The conditions for consistency remain the same.

In the third example (Figure 11) confidences range over the positive real numbers, $C = \mathbb{R}^+$. The mappings are lines through the origin with different positive slopes. A combination of these mappings is a linear mapping with a slope equal to the product of the slopes of its component mappings. The identity mapping is the linear mapping with slope equal to 1. This graph is transitively consistent since the products of the slopes along each circuit are equal to 1. It is symmetrically consistent since the product of the slopes of the mappings associated with each arc are equal to 1.

The fourth example (Figure 12) is like the third except that the confidences can be negative, as can the slopes of the mappings. Everything else remains the same.

<u>D_2^+ dependency relations</u>. If not all of the dependency relationships in a D_2 dependency relation are total, it is a D_2^+ dependency relation. In



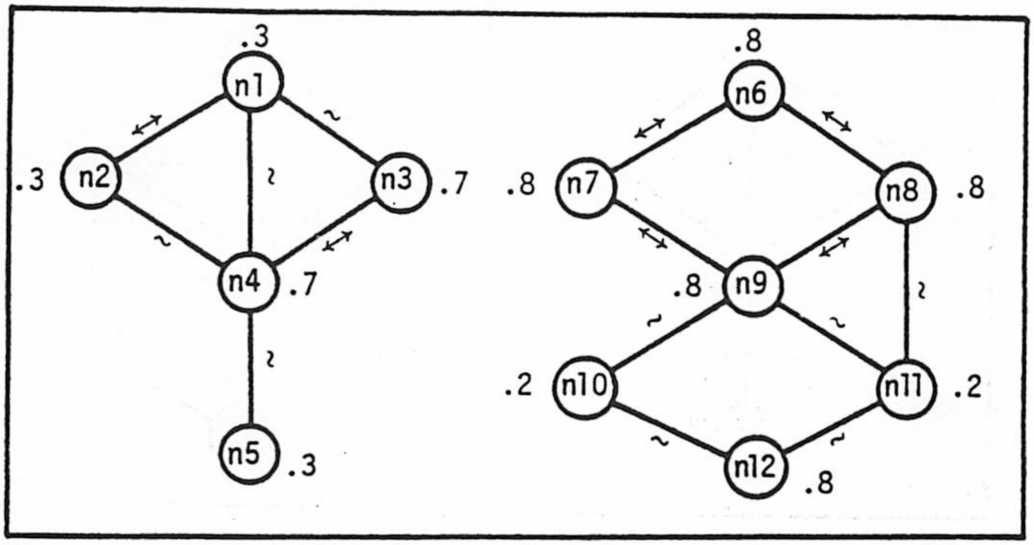
Note: Each arc is marked with a single mapping that is used in both directions.

C = {T,F}.

$\leftrightarrow$ = identity mapping: $\leftrightarrow[T] =_{def} T, \leftrightarrow[F] =_{def} F.$

$\sim$ = complementary mapping: $\sim[T] =_{def} F, \sim[F] =_{def} T.$

Figure 9. An example G_2^1 dependency graph: Boolean.



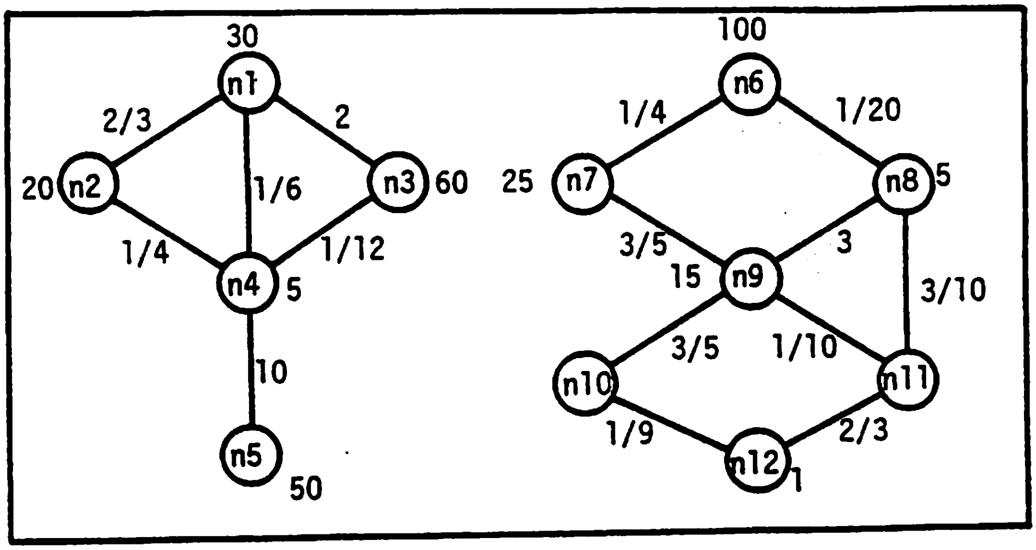
Note: Each arc is marked with a single mapping that is used in both directions.

C = [0,1]

↔ = identity mapping: ↔[c] $=_{def}$ c, for all c ∈ C.

~ = complementary mapping: ~[c] $=_{def}$ 1 - c, for all c ∈ C.

Figure 10. An example G_2^1 dependency graph: probabilistic.



Note: Each arc is marked with the slope of the mapping from the confidence of the lower numbered node to the higher numbered node.

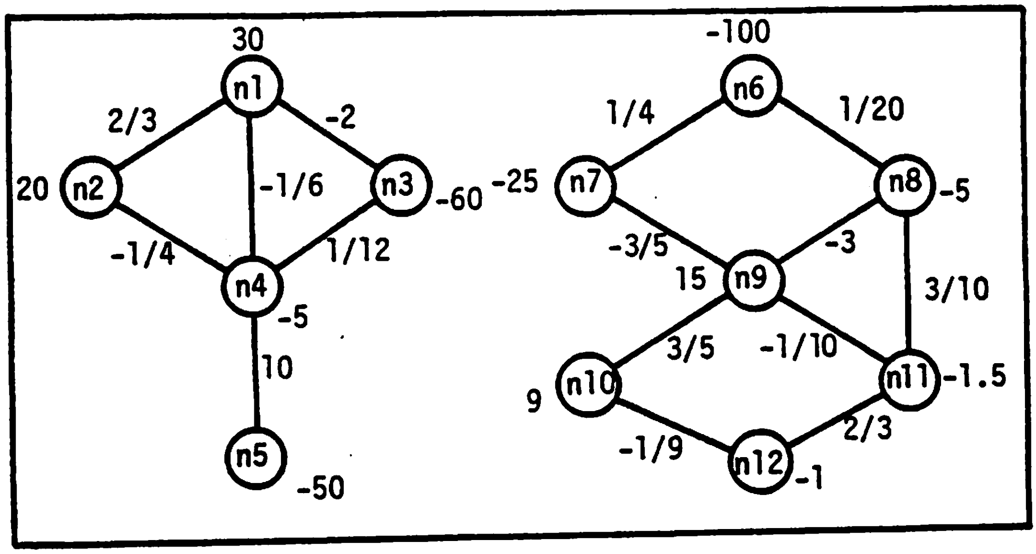
$C = [0,\infty)$.

For all $c_i, c_j \in C$: if $i < j$,

$$\text{then } m_{ni,nj}[c_i] =_{def} c_i \cdot \text{slope-on-arc},$$

$$m_{nj,ni}[c_j] =_{def} c_j \div \text{slope-on-arc}.$$

Figure 11. An example G_2^1 dependency graph: weighted.



Note: Each arc is marked with the slope of the mapping from the confidence of the lower numbered node to the higher numbered node.

$C = (-\infty,\infty)$.

For all $c_i, c_j \in C$: if $i < j$,

$$\text{then } m_{ni,nj}[c_i] =_{def} c_i \cdot \text{slope-on-arc},$$

$$m_{nj,ni}[c_j] =_{def} c_j \div \text{slope-on-arc}.$$

Figure 12. An example G_2^1 dependency graph: ± weighted.

this case the superscript signifies that given a single confidence value for one proposition in a dependency relationship, there are one or more compatible confidence values for the other proposition. This comes directly from the definition of partial dependence. When a dependency relationship is total, the confidence of one proposition uniquely determines the confidence of the other. However, when a relationship is partial, the confidence of one proposition may only determine a set of possible confidences for the other.

DEFINITION 5. D_2^+ dependency relations.

> A D_2^+ dependency relation is a D_2 dependency relation where M is not functional. □

The partial nature of D_2^+ dependency relations motivates the use of coverings that partially specify confidence assignments. C^+ coverings accomplish this by assigning each proposition a (nonempty) set of confidences, with the true confidence of each proposition being an element of its assigned set. The smaller the set assigned a proposition, the more precisely its confidence is known.

DEFINITION 6. C^+ coverings for D_2 dependency relations.

> A C^+ covering for a D_2 dependency relation (P,D,M) is a function from P into nonempty subsets of confidences with range C, $C^+: P \rightarrow (2^C - \{\phi\})$, subject to the constraint that
>
> - for every $\{p_1,p_2\} \in D$,
>   if $c_1 \in C^+[p_1]$ then there exists $c_2 \in C^+[p_2]$ such that $(c_1,c_2) \in M[(p_1,p_2)]$. □

Graphically, D_2^+ dependency relations and their accompanying C^+ coverings can be represented the same way as D_2^1 relations with C^1 coverings. The only difference is that the assigned confidences are sets and therefore the mappings are from sets to sets. A mapping $m_{i,j}$ from proposition p_i to proposition p_j is satisfied by the assignment of confidence sets c_i to p_i and c_j to p_j if c_j is contained within the image of c_i under $m_{i,j}$.

Connected propositions within a G_2^+ dependency graph are not necessarily dependent. If the mapping relating two connected propositions describes an all-to-all correspondence--every confidence of one proposition corresponds to every confidence of the other--then the propositions are independent; in other words, each in no way constrains the other. Such independent relationships may be described by a single arc in a G_2^+ graph or inferred from several arcs. All G_2^+ graphs can be converted to fully connected graphs through the addition of such independent arcs.

Like G_2^1 graphs, a G_2^+ graph need not include every relationship in the dependency relation it represents. Those relationships that can be inferred through combinations of the other relationships in the graph need not appear. Indeed, if less precision can be tolerated, even relationships that cannot be inferred from the others can be left out. This has the effect of expanding the space of consistent coverings. All those coverings that are consistent with D_2^+ are consistent with a less precise G_2^+ graph, but so are some coverings that are inconsistent with

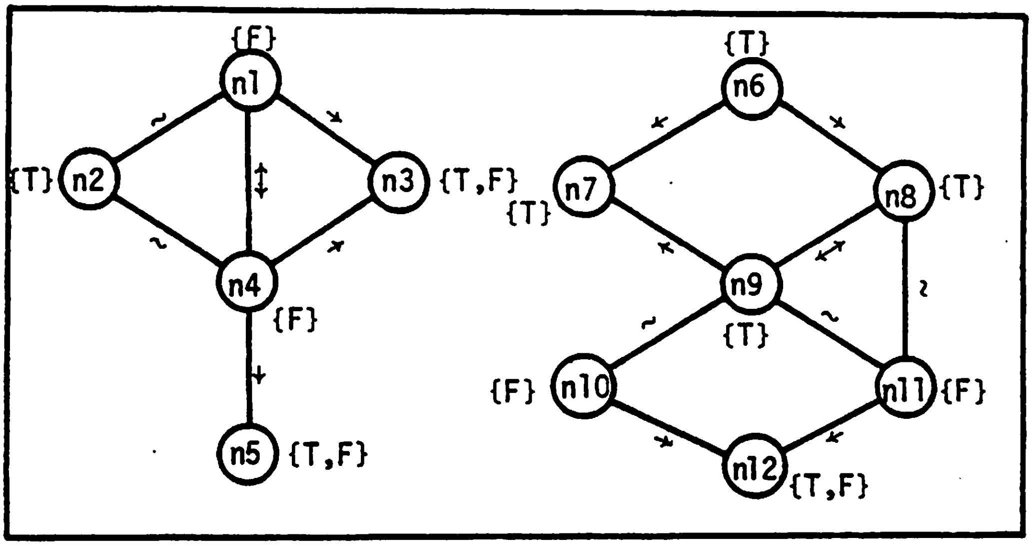
D_2^+. A less precise graph can also result from the weakening of some of the mappings. If a mapping in the graph is weaker than that indicated by M (i.e., that mapping's image is sometimes larger than that of M), additional coverings might be consistent with that graph. In the extreme, replacing a mapping with another that takes every subset of confidences to the set of all confidences, is equivalent to the removal of that relationship. Some example G_2^+ graphs follow.

The first example (Figure 13) mimics the propositional calculus restricted to the propositional connectives $\leftrightarrow$, ~, and $\rightarrow$. $\leftrightarrow$ and ~ correspond to the identity and complementary mappings respectively; they are total. The remaining connectives, $\rightarrow$ and $\leftarrow$, are partial. Confidences are restricted to the Boolean values T and F.

The second example (Figure 14) is similar to the first except that the Boolean valued confidences have been replaced with confidences that range between 0 and 1. $\leftrightarrow$ and ~ still correspond to the identity and complementary mappings. $\rightarrow$ corresponds to the mappings that take a confidence to all confidences greater-than-or-equal to the given confidence, and $\leftarrow$ to all the confidences less-than-or-equal to the given confidence.

Higher Order Dependency Relations and Their Graphical Representations

The dependency relations and graphs introduced thus far are sufficient for describing dependencies so long as they are binary. But when dependencies exist among larger groups of propositions, more general de-



$C = \{T,F\}.$

$\leftrightarrow$ = identity mapping: $\leftrightarrow[\{T\}] =_{def} \{T\},$

$\leftrightarrow[\{F\}] =_{def} \{F\},$

$\leftrightarrow[\{T,F\}] =_{def} \{T,F\}.$

$\sim$ = complementary mapping: $\sim[\{T\}] =_{def} \{F\},$

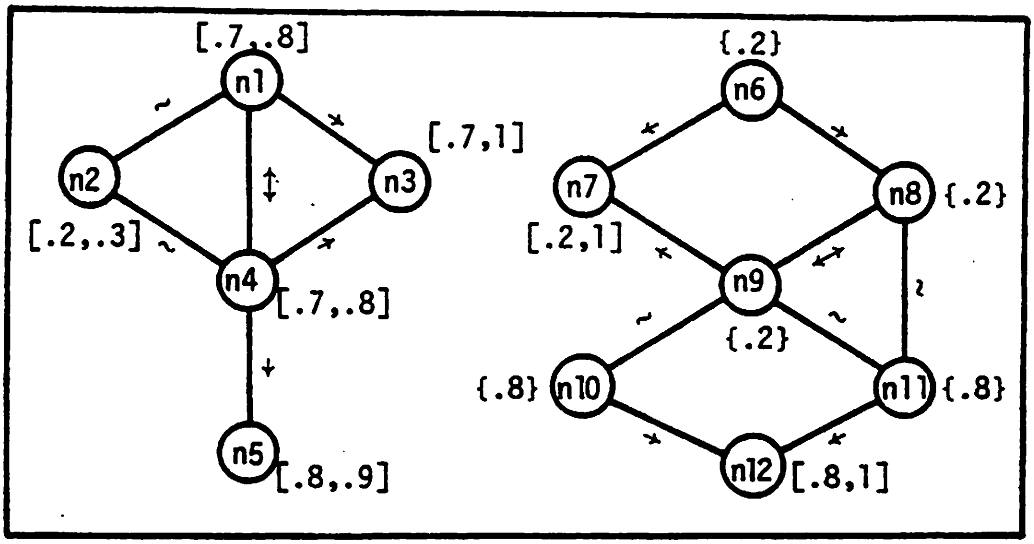
$\sim[\{F\}] =_{def} \{T\},$

$\sim[\{T,F\}] =_{def} \{T,F\}.$

$\rightarrow[\{T\}] =_{def} \{T\}, \rightarrow[\{F\}] =_{def} \{T,F\}, \rightarrow[\{T,F\}] =_{def} \{T,F\}.$

$\leftarrow[\{T\}] =_{def} \{T,F\}, \leftarrow[\{F\}] =_{def} \{F\}, \leftarrow[\{T,F\}] =_{def} \{T,F\}.$

Figure 13. An example G_2^+ dependency graph: Boolean.



$C = [0,1]$.

$\leftrightarrow$ = identity mapping: $\leftrightarrow[c_i] =_{def} c_i$.

$\sim$ = complementary mapping: $\sim[c_i] =_{def} \{c_j' | c_j' = 1 - c_i',\ c_i' \in c_i\}$.

$\rightarrow [c_i] =_{def} \{c_j' | c_i' \leq c_j' \leq 1,\ c_i' \in c_i\}$.

$\leftarrow [c_j] =_{def} \{c_i' | 0 \leq c_i' \leq c_j',\ c_j' \in c_j\}$.

Figure 14. An example G_2^+ dependency graph: probabilistic.

pendency relations and graphs are required. This section introduces dependency relations and graphs that are capable of describing dependency relationships of arbitrary order.

D_+ dependency skeletons. D_2 dependency relations are collections of binary dependency relationships. The natural extension is to drop the restriction that the relationships be binary, which leads to D_+ dependency relations. These are collections of dependency relationships of arbitrary order, limited only by the number of propositions over which the relations are defined. A D_+ dependency skeleton, the framework of a D_+ dependency relation, is a fairly obvious extension of a D_2 dependency skeleton, except for the added condition that it is closed under union. This additional constraint captures the notion that lower order dependency relationships can be arbitrarily combined to form higher order ones. The definition follows.

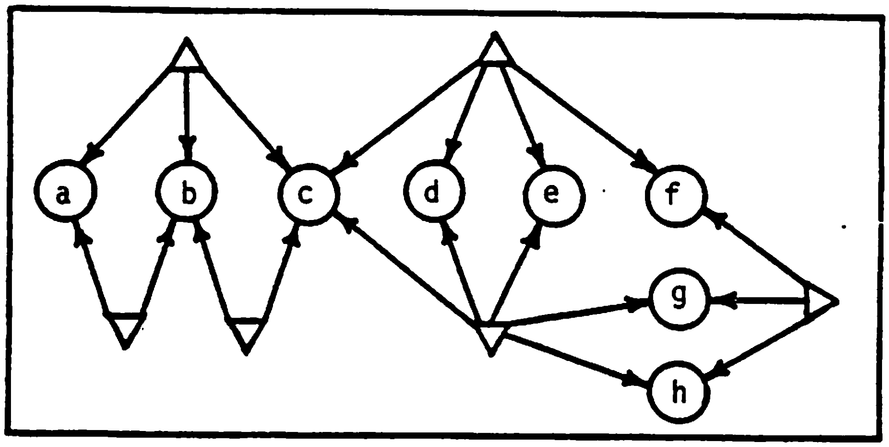
DEFINITION 7. D_+ dependency skeletons.

> A D_+ dependency skeleton is an ordered pair (P,D), where P is a set of propositions and D is a set of nonempty, nonunit subsets of P that is closed under union and is subject to the constraint that
>
> - if $d_i, d_j \in D$, $d_i \neq d_j$, and $p \in d_i \cap d_j$, then $((d_i \cup d_j) - \{p\}) \in D$. □

Graphically, a binary relationship from a D_2 dependency skeleton was represented as an undirected arc between two nodes. This was a straightforward representation since arcs are inherently binary. But

D_+ relationships cannot be similarly represented since they may be of higher orders. Instead, a D_+ dependency relationship is graphically represented by a relationship-node and a set of directed arcs. Each of these arcs points from the relationship-node to a proposition node. The set of proposition-nodes connected by (inpointing) arcs to a relationship-node represents the propositions taking part in that relationship. If n nodes are so connected, that relationship is of order n. Relationship-nodes are distinguishable from proposition-nodes by the directionality of their connecting arcs. Relationship-nodes always have outpointing arcs emitting from them. Arcs connected to proposition-nodes are always inpointing, though proposition-nodes may have no connecting arcs. In drawings of these graphs, relationship-nodes are further distinguished by their triangular shape versus the circular (or eliptical) shape of proposition-nodes. Just as before, not all of the relationships in a D_+ dependency skeleton need to be graphically represented: those that can be inferred, need not be included (Figures 15 and 16).

D_+ dependency relations. D_+ dependency relations are analogous to D_2 dependency relations only they are capable of describing n-ary dependency relationships. As previously described, a group of n propositions are dependent if the confidence of each proposition can be partially or totally predicted from the confidences of the other propositions in the group. Therefore, M is a function from ordered elements of D to sets of confidence vectors, the length of each vector being equal to the order of the selected relationship. Just as before, each



Elements of D directly represented in the graph:

{a,b}, {b,c}, {a,b,c}, {f,g,h}, {c,d,e,f}, {c,d,e,g,h}.

Some additional elements of D that can be inferred from the graph:

{a,c}, {b,d,e,f}, {a,b,f,g,h}, {a,b,c,d,e,f,g,h}.

Figure 15. An example G_+ dependency-graph skeleton.

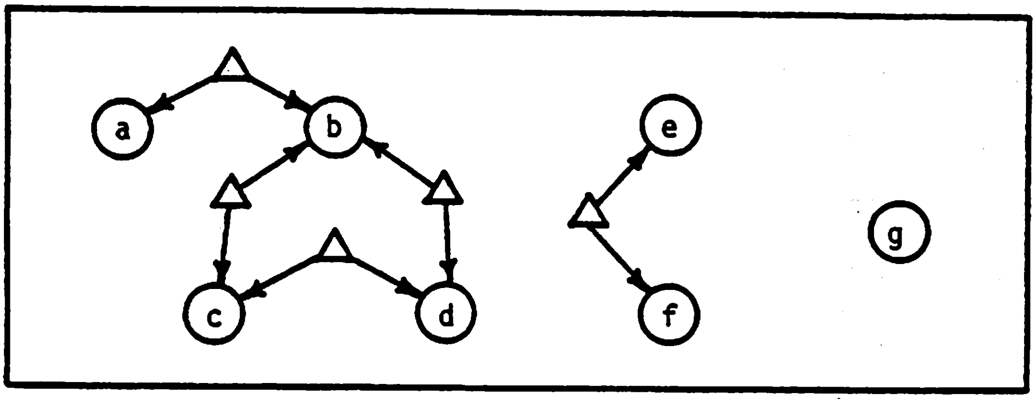


Figure 16. An example G_+ dependency-graph skeleton. This example is a direct translation of the G_2 dependency-graph skeleton in Figure 6.

vector represents a possible, simultaneous, confidence assignment to the propositions taking part in the relationship.

In the definition that follows, all but the second constraint are trivial extensions of those found in the D_2 definition. The second constraint is substantially changed since there are more complicated ways for these higher order dependency relationships to overlap. In general, two relationships sharing some common propositions need to be coordinated so that they do not describe the relationship among those propositions in incompatible ways. In the definition, this is guaranteed by ensuring that all lower order relationships are compatible with higher order relationships containing them. Thus, each higher order relationship plays the coordinating role for the lower order relationships it subsumes.

DEFINITION 8. D_+ dependency relations.

> A D_+ dependency relation is an ordered triple (P,D,M), where (P,D) is a D_+ dependency skeleton and M is a function, from ordered elements of D into equal length vectors of confidence values with range C (i.e., for each $\{p_1,\ldots,p_n\} \in D$, $M[(p_1,\ldots,p_n)] \subseteq C^n$), subject to the constraints:
>
> 1. for every $\{p_1,\ldots,p_i,\ldots,p_j,\ldots,p_n\} \in D$,
> $(c_1,\ldots,c_i,\ldots,c_j,\ldots,c_n) \in M[(p_1,\ldots,p_i,\ldots,p_j,\ldots,p_n)]$
> $\leftrightarrow (c_1,\ldots,c_j,\ldots,c_i,\ldots,c_n) \in M[(p_1,\ldots,p_j,\ldots,p_i,\ldots,p_n)]$;
>
> 2. if $d_m, d_n \in D$, $d_m \subseteq d_n$,
> $d_m = \{p_1,\ldots,p_m\}$,
> $d_n = \{p_1,\ldots,p_m,\ldots,p_n\}$,

then for each $(c_1,\ldots,c_m) \in C^m$,

$(c_1,\ldots,c_m) \in M[(p_1,\ldots,p_m)]$

$\leftrightarrow$ there exists $c_{m+1},\ldots,c_n \in C$ such that

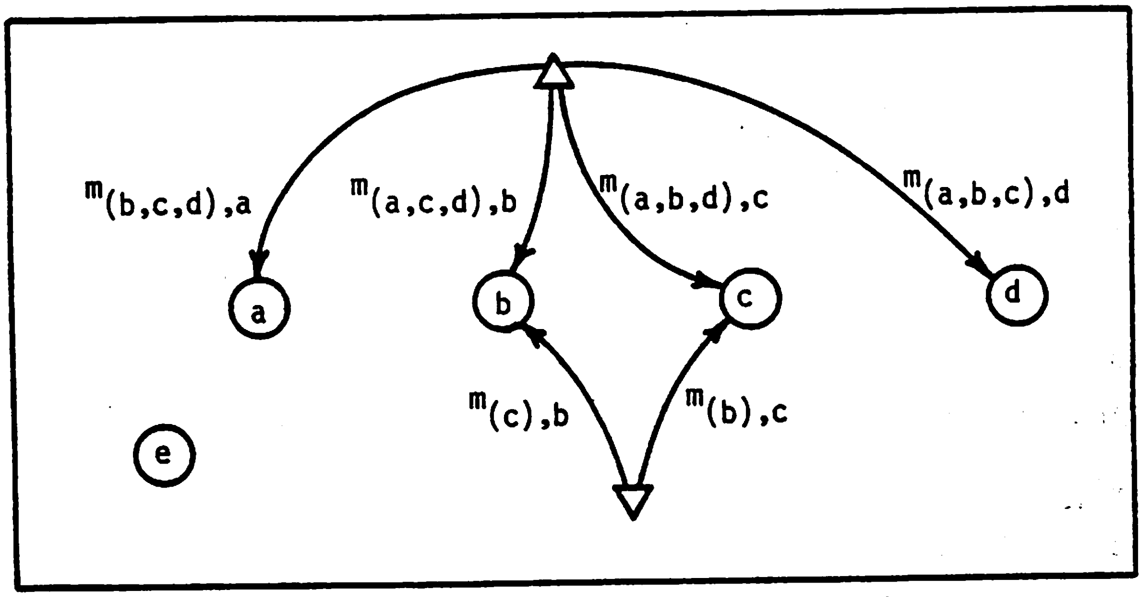
$(c_1,\ldots,c_m,\ldots,c_n) \in M[(p_1,\ldots,p_m,\ldots,p_n)]$

3. for every $\{p_1,\ldots,p_n\} \in D$ and every $c_1 \in C$, there exists $c_2,\ldots,c_n \in C$ such that

   $(c_1,\ldots,c_n) \in M[(p_1,\ldots,p_n)]$. □

Again, when a dependency relation is represented by a dependency graph, M is fragmented into local mappings and distributed across the arcs in the graph. But in the case of G_+ dependency graphs, each arc has one associated mapping that describes how the confidence of the proposition at the head of the arc varies with respect to the confidences of the other propositions connected to the relationship-node at the tail of the arc. For a relationship over propositions $p_1,\ldots,p_n$, the mapping $m_{(1,\ldots,n-1),n}$, placed along the arc pointing to p_n, predicts the confidence of p_n given the ordered confidences of $p_1,\ldots,p_{n-1}$. Thus, each n-ary dependency relationship is described by n mappings of degree n-1 (Figure 17).

D^1_+ dependency relations. Before moving on to the most general form of dependency relation, let us consider the extension of D^1_2 dependency relations to higher order dependencies. This gives us D^1_+ dependency relations that describe total dependency relationships over arbitrary numbers of propositions. M is functional and therefore, C^1 coverings apply.



$$(c_1,\ldots,c_n) \in M[(p_1,\ldots,p_n)] \leftrightarrow c_n \in m_{(1,\ldots,n-1),n}[c_1,\ldots,c_{n-1}].$$

Figure 17. An example G_+ dependency graph.

DEFINITION 9. D_+^1 dependency relations.

> A D_+^1 dependency relation (P,M,D) is a D_+ dependency relation with the additional constraint that M is functional. In other words,
>
> - for every $\{p_1,\ldots,p_n\} \in D$, $c_1,\ldots,c_n \in C$,
> $|m_{(1,\ldots,n-1),n}[c_1,\ldots,c_{n-1}]| \leq 1$,
> where $m_{(1,\ldots,n-1),n}[c_1,\ldots,c_{n-1}]$
> $= \{c_n | (c_1,\ldots,c_n) \in M[(p_1,\ldots,p_n)]\}$. □

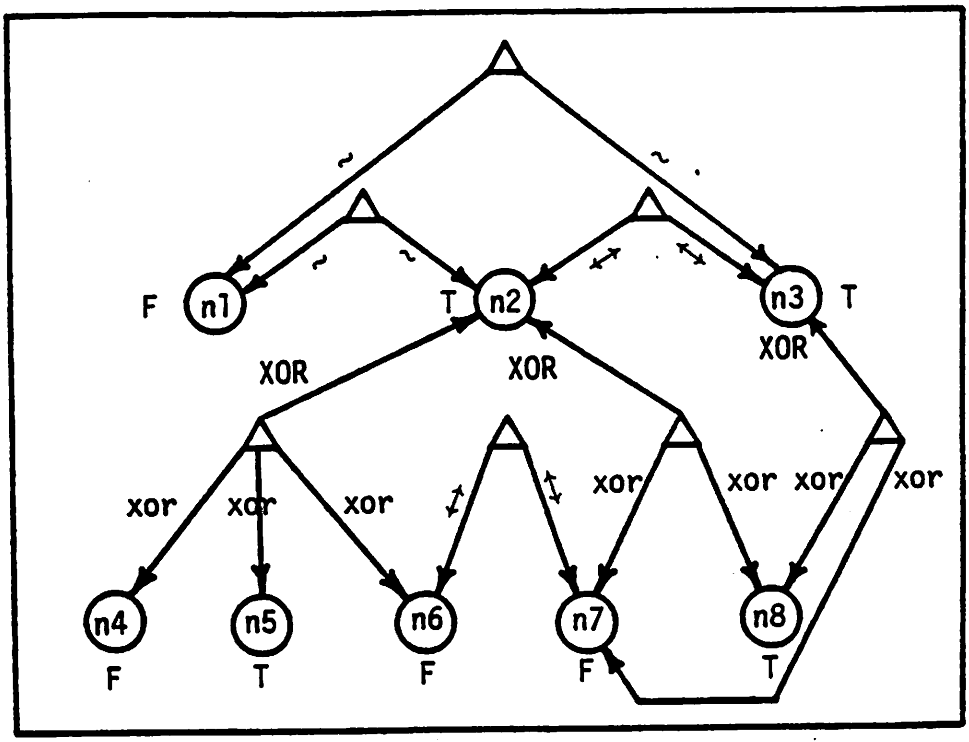
DEFINITION 10. C^1 coverings for D_+ dependency relations.

> A C^1 covering for a D_+ dependency relation (P,D,M) is a function from P into a range of confidences C, $C^1: P \rightarrow C$, subject to the constraint that
>
> - for every $\{p_1,\ldots,p_n\} \in D$,
> $(C^1[p_1],\ldots,C^1[p_n]) \in M[(p_1,\ldots,p_n)]$. □

Two example G_+^1 dependency graphs follow (Figures 18 and 19). They mimic the propositional calculus restricted to the propositional connectives $\leftrightarrow$ and $\sim$, with one additional n-ary connective, XOR. This additional connective is defined as follows:

$$XOR[p_0,\ldots,p_n] =_{def} (p_0 \leftrightarrow p_1 \vee \ldots \vee p_n) \wedge \sim \bigvee_{\substack{1 \leq i,j \leq n \\ i \neq j}} (p_i \wedge p_j).$$

Note that the definition differs from the standard definition of "exclusive or." It is motivated by those situations where a class, represented by p_0, is broken into n subclasses, represented by $p_1,\ldots,p_n$, based on an equivalence relation. The distinguished proposition p_0 is true if and only if exactly one of the indistinguished propositions is



C = {T,F}.

$\leftrightarrow$ = identity mapping: $\leftrightarrow[T] =_{def} T, \leftrightarrow[F] =_{def} F$.

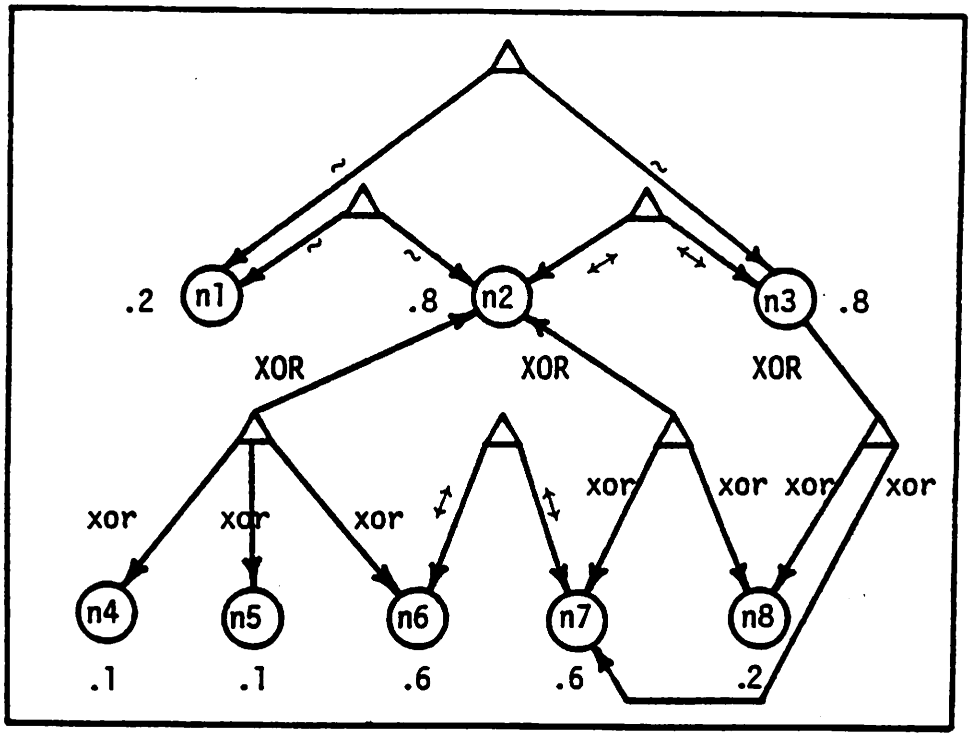
$\sim$ = complementary mapping: $\sim[T] =_{def} F$, $\sim[F] =_{def} T$.

$$XOR[c_1,\ldots,c_n] =_{def} \begin{cases} T, & \exists_{1 \le i \le n}\ c_i = T,\ \ c_j = F,\ 1 \le j \le n,\ i \ne j, \\ F, & c_i = F,\ 1 \le i \le n. \end{cases}$$

$$xor[c_0,c_1,\ldots,c_{i-1},c_{i+1},\ldots,c_n] =_{def} \begin{cases} T, & c_0 = T,\ c_j = F,\ 1 \le j \le n,\ i \ne j, \\ F, & c_0 = F,\ c_j = F,\ 1 \le j \le n,\ i \ne j. \end{cases}$$

where c_0 corresponds to the distinguished proposition.

Figure 18. An example G_{+}^{1} dependency graph: Boolean.



$C = [0,1]$

$\leftrightarrow$ = identity mapping: $\leftrightarrow[c] =_{def} c$.

$\sim$ = complementary mapping: $\sim[c] =_{def} 1 - c$.

$$XOR[c_1,\ldots,c_n] =_{def} \sum_{i=1}^{n} c_i.$$

$$xor[c_0,c_1,\ldots,c_{i-1},c_{i+1},\ldots,c_n] =_{def} c_0 - \sum_{\substack{j=1 \\ i\neq j}}^{n} c_j,$$

where c_0 corresponds to the distinguished proposition.

Figure 19. An example G_+^1 dependency graph: probabilistic.

true. No two of the indistinguished propositions are ever to be simultaneously true; if two of the indistinguished propositions were simultaneously true, the connective would be undefined. The confidences in the first example range over the Boolean values T and F, while in the second example they range over the real interval [0,1]. Throughout the examples, "XOR" is the mapping that predicts the confidence of the distinguished proposition from the confidences of the indistinguished propositions, and "xor" is the mapping that predicts the confidence of an indistinguished proposition from the confidences of all of the other propositions.

D_+^+ dependency relations. D_+^1 dependency relations are restricted to total dependency relationships. The relaxing of this final constraint, to include partial dependency relationships, results in D_+^+ dependency relations. These may contain dependency relationships over any number of propositions and with any degree of dependence/independence. When viewed hierarchically, all of the previously described relations are subsumed by D_+^+ dependency relations.

DEFINITION 11. D_+^+ dependency relations.

> A D_+^+ dependency relation is a D_+ dependency relation where M is not functional. □

When graphically represented, M is fragmented and distributed over the dependency graph in the same way as for G_+^1. Of course, the difference is that these fragments map sets to sets, just like those in G_2^+

graphs. Also like G_2^+, C^+ coverings are used, allowing confidence assignments to be partially specified. A covering is consistednt with a graph if all of its mappings are satisfied. In this case, a mapping $m_{(1,\ldots,n-1),n}$ from propositions $p_1,\ldots,p_{n-1}$ to proposition p_n is satisfied by the assignments of confidence sets c_i to p_i, $1 \le i \le n$, if c_n is contained within the image of $c_1,\ldots,c_{n-1}$ under $m_{(1,\ldots,n-1),n}$.

DEFINITION 12. C^+ coverings for D_+ dependency relations.

> A C^+ covering for a D_+ dependency relation (P,D,M) is a function from P into nonempty subsets of confidences with range C, $C^+: P \to (2^C - \{\phi\})$, subject to the constraint that
>
> - for every $\{p_1,\ldots,p_n\} \in D$,
>   if $c_1 \in C^+[p_1]$ then there exists $c_2 \in C^+[p_2],\ldots,c_n \in C^+[p_n]$, such that $(c_1,\ldots,c_n) \in M[(p_1,\ldots,p_n)]$.

Like all dependency graphs, if some of the mappings in G_+^+ are weaker than their counterparts in M, the graph may only approximate the underlying D_+^+ dependency relation, resulting in a larger space of consistent covering.

The propositional calculus examples of the previous sections are here extended into G_+^+ graphs, with the additional connectives $\wedge$ and $\vee$. $\wedge$ biconditionally relates the truthfulness of a distinguished proposition, to a conjunction of indistinguished propositions; $\vee$, to a disjunction:

$$\wedge[p_0,\ldots,p_n] =_{def} (p_0 \leftrightarrow p_1 \wedge\ldots\wedge p_n),$$

$$\vee[p_0,\ldots,p_n] =_{def} (p_0 \leftrightarrow p_1 \vee\ldots\vee p_n).$$

The mappings in the first example (Figure 20), defined over the Boolean values T and F, are straightforward. The mappings in the second example (Figure 21), defined over [0,1], require some explanation. But this will be deferred (pp. 107-123), for it is these mappings that are justified by Shafer's theory of inexact reasoning. The reader is encouraged to verify that these examples are legitimate G_+^+ dependency graphs with C^+ coverings, while awaiting further discussion for their motivation.

Both of these examples are (demonstratively) approximations of their underlying dependency relations. The V-relationship, found in each, could be strengthened to an XOR-relationship. This is known since the disjuncts n4 and n5 are already participating in an XOR-relationship and therefore are known to be exclusive. This imprecision, in the second example, is cause for imprecision in the confidence assignment of the proposition represented by n2. If the graph were precise, n2 would have an associated confidence interval of [.7,.9], as would n6. Of course, the intervals that appear are not incorrect; they do delimit the correct confidences; they are merely less precise than they might be.

This concludes our journey from D_2^1 dependency relations to D_+^+ dependency relations, a journey from a highly constrained, but simple,

Figure 20. An example G_+^+ dependency graph: Boolean.

$$C = \{T,F\}.$$

$$\leftrightarrow[c_i] =_{def} c_i.$$

$$\sim [c_i] =_{def} \begin{cases} \{T\}, & c_i = \{F\}, \\ \{F\}, & c_i = \{T\}, \\ \{T,F\}, & \text{else}. \end{cases}$$

$$\rightarrow [c_i] =_{def} \begin{cases} \{T\}, & c_i = \{T\}, \\ \{T,F\}, & \text{else}. \end{cases}$$

$$\leftarrow [c_j] =_{def} \begin{cases} \{F\}, & c_j = \{F\}, \\ \{T,F\}, & \text{else}. \end{cases}$$

$$XOR[c_1,\ldots,c_n] =_{def} \begin{cases} \{T\}, & \exists_{i=1}^{n} c_i = \{T\}, \\ \{F\}, & c_i = \{F\},\ 1 \le i \le n\ , \\ \{T,F\}, & \text{else}. \end{cases}$$

$$xor[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n] =_{def} \begin{cases} \{T\}, & c_0 = \{T\},\ c_j = \{F\},\ 1 \le j \le n,\ i \ne j \\ \{F\}, & c_0 = \{F\} \text{ or } \exists_{j=1,\ i \ne j}^{n} c_j = \{T\}, \\ \{T,F\}, & \text{else}. \end{cases}$$

$$\Lambda[c_1,\ldots,c_n] =_{def} \begin{cases} \{T\}, & c_i = \{T\},\ 1 \le i \le n, \\ \{F\}, & \exists_{i=1}^{n} c_i = \{F\}, \\ \{T,F\}, & \text{else}. \end{cases}$$

$$\wedge[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n] =_{def} \begin{cases} \{T\}, & c_0 = \{T\}, \\ \{F\}, & c_0 = \{F\},\ c_j = \{T\},\ 1 \le j \le n,\ i \ne j \\ \{T,F\}, & \text{else} \end{cases}$$

$$\bigvee[c_1,\ldots,c_n] =_{def} \begin{cases} \{T\}, & \exists_{i=1}^{n} c_i = \{T\}, \\ \{F\}, & c_i = \{F\},\ 1 \le i \le n\ , \\ \{T,F\}, & \text{else}. \end{cases}$$

$$\vee[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n] =_{def} \begin{cases} \{T\}, & c_0 = \{T\},\ c_j = \{F\},\ 1 \le j \le n,\ i \ne j \\ \{F\}, & c_0 = \{F\}, \\ \{T,F\}, & \text{else}. \end{cases}$$

where c_0 always corresponds to a distinguished proposition.

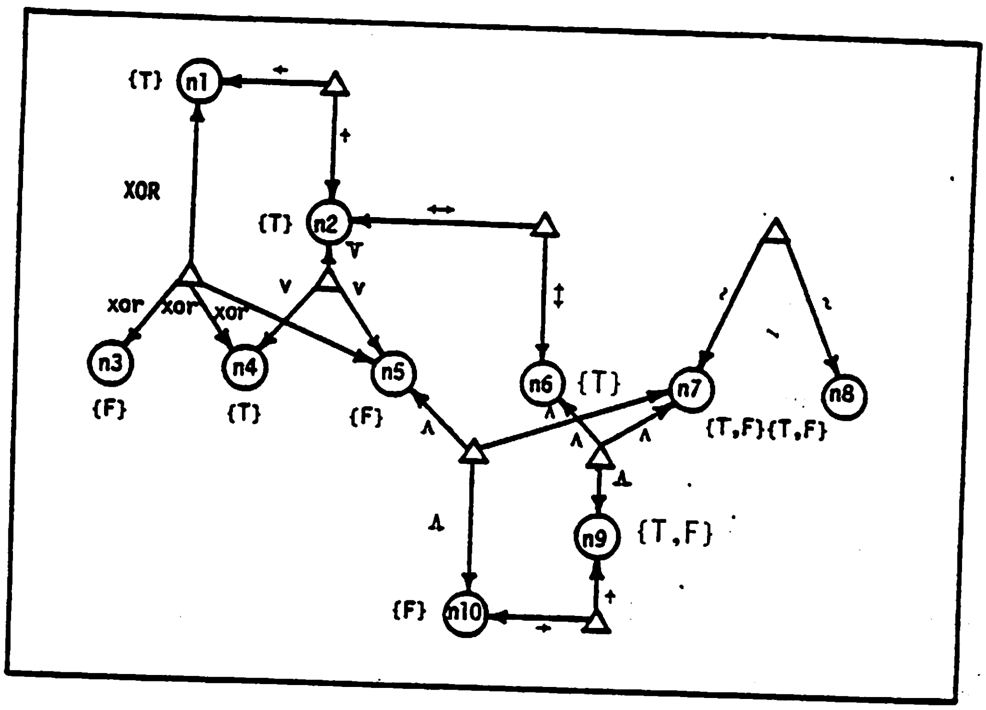


Figure 20

Figure 21. An example G_+^+ dependency graph: probabilistic.

$C = [0.1].$

$\leftrightarrow[c_i] =_{def} c_i.$

$\sim[c_i] =_{def} \{c_j' \mid c_j' = 1 - c_i',\ c_i' \in c_i\}.$

$\rightarrow[c_i] =_{def} \{c_j' \mid c_i' \le c_j' \le 1,\ c_i' \in c_i\}.$

$\leftarrow[c_j] =_{def} \{c_i' \mid 0 \le c_i' \le c_j',\ c_j' \in c_j\}.$

$$\mathrm{XOR}[c_1,\ldots,c_n] =_{def} \{c_0' \mid c_0' = \sum_{i=1}^{n} c_i',\ c_0' \in C,\ c_i' \in c_i,\ 1 \le i \le n\}.$$

$$\mathrm{xor}[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n]$$
$$=_{def} \{c_i' \mid c_i' = c_0' - \sum_{\substack{j=1\\ i\ne j}}^{n} c_j',\ c_i' \in C,\ c_j' \in c_j,\ 0 \le j \le n,\ i \ne j\}.$$

$$\bigwedge[c_1,\ldots,c_n] =_{def} \{c_0' \mid 1 + (\sum_{i=1}^{n} c_i' - 1) \le c_0' \le \operatorname*{MIN}_{i=1}^{n} c_i',\ c_0' \in C,$$
$$c_i' \in c_i,\ 1 \le i \le n\}.$$

$$\wedge[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n]$$
$$=_{def} \{c_i' \mid c_0' \le c_i' \le c_0' - (\sum_{\substack{j=1\\ i\ne j}}^{n} c_j' - 1),\ c_i' \in C,\ c_j' \in c_j,\ 0 \le j \le n,\ i \ne j\}.$$

$$\bigvee[c_1,\ldots,c_n] =_{def} \{c_0' \mid \operatorname*{MAX}_{i=1}^{n} c_i' \le c_0' \le \sum_{i=1}^{n} c_i',\ c_0' \in C,\ c_i' \in c_i,\ 1 \le i \le n\}.$$

$$\vee[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n]$$
$$=_{def} \{c_i' \mid (c_0' - \sum_{\substack{j=1\\ i\ne j}}^{n} c_j') \le c_i' \le c_0',\ c_i' \in C,\ c_j' \in c_j,\ 0 \le j \le n,\ i \ne j\}.$$

where c_0 always corresponds to a distinguished proposition.

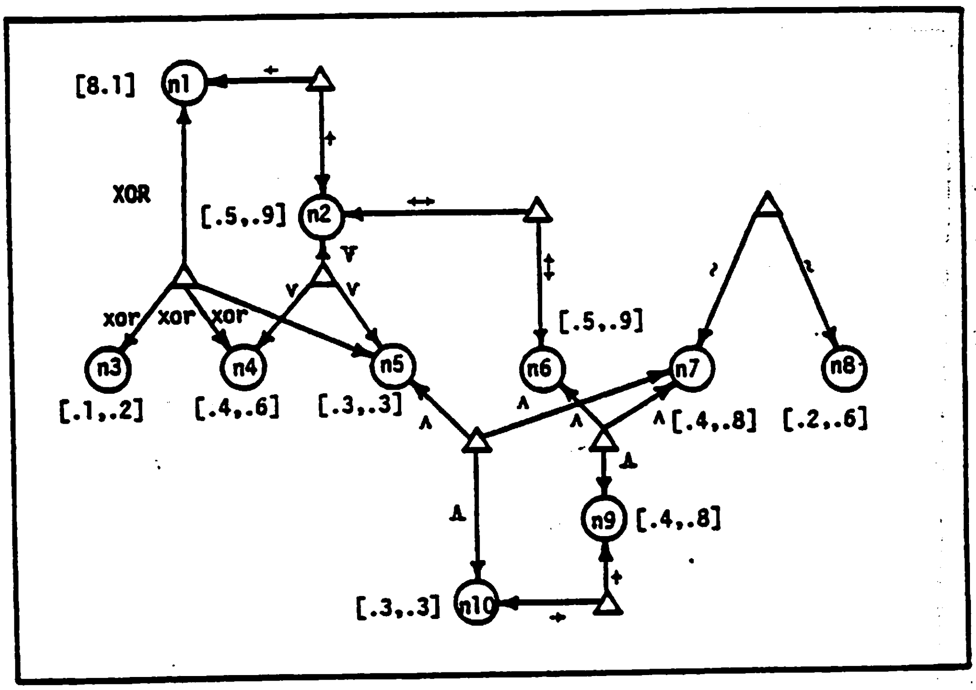

[8.1]
n1
XOR
[.5,.9]
n2
xor
xor
xor
n3
n4
n5
n6
n7
n8
[.1,.2]
[.4,.6]
[.3,.3]
[.5,.9]
[.4,.8]
[.2,.6]
n9
[.4,.8]
[.3,.3]
n10

Figure 21

form of propositional dependence, to a more complex, general form. At each step, consistency conditions, included in the definitions of dependency relations, ensure that multiple descriptions of relationships within dependency relations are consistent. The basis for all of the preceding definitions is that redundant descriptions need to be consistent. Dependency graphs, compact representations of dependency relations, permit some imprecision in representation so long as the consistency of the underlying dependency relation is preserved. As will be seen in the following chapter, well-formed dependency graphs provide a sound foundation for automated inferential reasoning.

CHAPTER IV

DEPENDENCY-GRAPH INFERENCE-ENGINES

When a proposition is dependent on a group of other propositions, information about its confidence can be inferred from information about the confidences of those propositions in that group. Within a dependency graph, an n-ary dependency relationship is described by n (coordinated) mappings, each capable of translating information about the confidence of n-1 of the propositions to information about the confidence of the remaining proposition. In this chapter, inferential reasoning is formally defined as the process by which this predictive capability is exploited to extend and refine incomplete information about proposition confidences. This definition takes the form of four inference rules, corresponding to the four classes of dependency relations, and the inference engines that apply them.

To this point, all of the coverings that have been defined preclude any inferential reasoning. They have all been complete in that a confidence was associated with every proposition, and totally refined in that all of the mappings were satisfied by the confidences. Based on the information embodied in a dependency relation, these coverings cannot be made more precise. They are the goals of the inferential process, and are what one would hope to discover; they are not the starting points, but the fixed points. Less informative coverings, analogous to C^1 and C^+ coverings, are $\hat{C}^1$ and $\hat{C}^+$ coverings. A <u>$\hat{C}^1$ covering</u> has a (possibly) reduced domain relative to a C^1 covering: a <u>$\hat{C}^+$</u>

<u>covering</u> has a (possibly) expanded image relative to a C^+ covering. A $\hat{C}^1$ covering is (possibly) incomplete and a $\hat{C}^+$ covering is (possibly) unrefined.

DEFINITION 13. $\hat{C}^1$ coverings.

A $\hat{C}^1$ covering for a dependency relation (P,D,M) is identical to a C^1 covering for that same dependency relation, but (possibly) with a reduced domain.

$\hat{C}^1 : \hat{P} \to C$,

where $\hat{P} \subseteq P$ and

there exists $C^1 : P \to C$, such that for all $\hat{p} \in \hat{P}$,

$C^1[\hat{p}] = \hat{C}^1[\hat{p}]$. □

DEFINITION 14. $\hat{C}^+$ coverings.

A $\hat{C}^+$ covering for a dependency relation (P,M,D) is identical to a C^+ covering for that same dependency relation, but (possibly) with an expanded image.

$\hat{C}^+ : P \to (2^C - \{\phi\})$,

where there exists $C^+ : P \to (2^C - \{\phi\})$, such that for all $p \in P$,

$C^+[p] \subseteq \hat{C}^+[p]$. □

A wide range of informedness is expressible through these coverings. At one extreme, the vacuous $\hat{C}^1$ and $\hat{C}^+$ coverings provide no information. The <u>vacuous $\hat{C}^1$ covering</u> associates no confidence with any proposition. The <u>vacuous $\hat{C}^+$ covering</u> associates the entire range of confidences with each proposition thus, merely reaffirming the minimal limiting conditions. At the other extreme, they may associate a single confidence value with each proposition, exactly determining the confidence of every proposition. This range of description is much the same as that found in dependency-graph representations of dependency relations, where dependency relationships can be left out of the representation or replaced by weaker mappings.

Given a dependency relation and a $\hat{C}^1$ or $\hat{C}^+$ covering (of the appropriate type), there is a potential for extending or refining that covering. This is formally defined in terms of four <u>inference rules</u>, one for each class of dependency relation. Each rule is a function of three arguments, a dependency relation, a covering for that dependency relation, and a directed, dependency relationship selected from that dependency relation. The result is a (possibly more informative) covering. If the inferential step provides new information, the resultant covering is the original covering augmented by this new information. Otherwise, there is either insufficient information to make an inferential step or the inferential step simply reaffirms a portion of the original covering; in either of these cases, the resultant covering is the original covering. The R_2^1 and R_+^1 inference rules (potentially) augment $\hat{C}^1$ coverings, with confidence assignments for previously unassigned propositions, based on the assignments of neighboring propositions. The R_2^+ and R_+^+ inference rules (potentially) reduce the size of the assigned confidence sets in $\hat{C}^+$ coverings, discarding those confidences that are inconsistent with the confidence sets assigned neighboring propositions. Throughout the following definitions, references are made to $\hat{C}$ coverings as both functions and sets. These two views are interchangeable:

$$\hat{C}^1 : P \to C \quad \leftrightarrow \{(p,c) \mid p \in P,\ c = \hat{C}^1[p]\};$$

$$\hat{C}^+ : P \to (2^C - \{\phi\}) \leftrightarrow \{(p,c) \mid p \in P,\ c \in \hat{C}^+[p]\}.$$

DEFINITION 15. R_2^1 inference rule.

R_2^1 is a function of three arguments, a D_2^1 dependency relation (P,D,M), a $\hat{C}^1$ covering for that relation, and a directed element from D, (p_1,p_2). R_2^1 returns a (possibly more complete) $\hat{C}^1$ covering.

$$R_2^1[(P,D,M),\hat{C}^1,(p_1,p_2)] = \begin{cases} \{(p_2,c_2)\} \cup \hat{C}^1, & (\hat{C}^1[p_1],c_2) \in M[(p_1,p_2)]; \\ \hat{C}^1, & \text{else.} \end{cases} \square$$

DEFINITION 16. R_+^1 inference rule.

R_+^1 is a function of three arguments, a D_+^1 dependency relation (P,D,M), a $\hat{C}^1$ covering for that relation, and a directed element from D, $(p_1,\ldots,p_n)$. R_+^1 returns a (possibly more complete) $\hat{C}^1$ covering.

$$R_+^1[(P,D,M,)\ \hat{C}^1,(p_1,\ldots,p_n)] = \begin{cases} \{(p_n,c_n)\} \cup \hat{C}^1, & (\hat{C}^1[p_1],\ldots,\hat{C}^1[p_{n-1}],c_n) \in M[(p_1,\ldots,p_n)]; \\ \hat{C}^1, & \text{else.} \end{cases} \square$$

DEFINITION 17. R_2^+ inference rule.

R_2^+ is a function of three arguments, a D_2^+ dependency relation (P,D,M), a $\hat{C}^+$ covering for that relation, and a directed element from D, (p_1, p_2). R_2^+ returns a (possibly more refined) $\hat{C}^+$ covering.

$$R_2^+[(P,D,M),\hat{C}^+,(p_1,p_2)] = \hat{C}^+ - \{(p_2,\not{c}_2) \mid \not{c}_2 \in \hat{C}^+[p_2] - C_2\},$$

$$\text{where } C_2 = \{c_2 \mid (c_1,c_2) \in M[(p_1,p_2)],\ c_1 \in \hat{C}^+[p_1],\ c_2 \in \hat{C}^+[p_2]\}. \ \square$$

DEFINITION 18. R_+^+ inference rule.

R_+^+ is a function of three arguments, a D_+^+ dependency relation (P,D,M), a $\hat{C}^+$ covering for that relation, and a directed element from D, $(p_1,\ldots,p_n)$. R_+^+ returns a (possibly more refined) $\hat{C}^+$ covering.

$$R_+^+[(P,D,M),\hat{C}^+,(p_1,\ldots,p_n)] = \hat{C}^+ - \{(p_n,\not{c}_n) \mid \not{c}_n \in \hat{C}^+[p_n] - C_n\},$$

$$\text{where } C_n = \{c_n \mid (c_1,\ldots,c_n) \in M[(p_1,\ldots,p_n)],$$

$$c_i \in \hat{C}^+[p_i],\ 1 \le i \le n\}. \ \square$$

Given a dependency relation D and a $\hat{C}$ covering (i.e., a $\hat{C}^1$ or $\hat{C}^+$ covering) for D, there are multiple inferences that might be drawn. Every element of D is a candidate for an inferential step and several of these might provide additional information beyond that already available in $\hat{C}$. To reap the collective benefits of multiple inferential steps, an <u>inference engine</u> serially applies its inference rule to selected elements of D, using the result of each application as an argument to the next.[4] Ordered elements of D are selected according to a

[4]Here, inference engines with parallel search strategies are excluded to simplify the discussion. They require that a merging operation be defined over multiple coverings.

search strategy. When the search strategy returns the null set, the inference engine halts. An inference rule and search strategy fully determine an inference engine. The class of the inference rule determines the class of the engine (e.g., an inference engine that applies R_+^+ is an E_+^+ inference engine) and thereby the class of the dependency relations and coverings it is defined over (e.g., an E_+^+ inference engine is defined over D_+^+ dependency relations and $\hat{C}^+$ coverings).

DEFINITION 19. Inference engines.

> An inference engine E is a procedure of two arguments, a dependency relation (P,D,M) and a covering for that relation $\hat{C}$. E selects an ordered element from D in accordance with a search strategy S; appplies its inference rule R to the selected element; and loops with the resultant covering. This continues until S returns the null set, at which point, the final covering is returned. E is not (necessarily) a function since its search strategy might depend on some additional (heuristic) information H.

$$E[(P,D,M),\hat{C}] = \begin{cases} \hat{C}, & S[(P,D,M),\hat{C},H] = \phi; \\ E[(P,D,M),\ R[(P,D,M),\hat{C},\vec{d}]], & \vec{d} \in S[(P,D,M),\hat{C},H]; \end{cases}$$

where $S[(P,D,M),\hat{C},H] \subseteq \{(p_1,\ldots,p_n) \mid \{p_1,\ldots,p_n\} \in D\}$. □

As with any proof procedure, questions of soundness and completeness come to mind. Will these inference engines discover only that which follows and all of that which follows from a dependency relation and a $\hat{C}$ covering for it? Instead of addressing these questions separately for each class of inference engine, we will only consider them for the most general class E_+^+ and be satisfied with some limited afterthoughts about the others. Relative to E_+^+ these questions can be rephrased as one: does an E_+^+ inference engine, given a D_+^+ dependency relation and a $\hat{C}^+$ covering for D_+^+, always return the least precise C^+ covering for D_+^+ over which $\hat{C}^+$ can be defined? The answer is yes provided the search strategy does not prevent informative inferences from eventually being drawn. Such a search strategy is said to be complete.

DEFINITION 20. Complete search strategies.

> A search strategy S, of an inference engine E, is complete if it returns the null set only when the given $\hat{C}$ covering is everywhere invariant to further application of E's inference rule R.
>
> $$E[(P,D,M),\hat{C}] = \hat{C} \leftrightarrow \begin{cases} \text{for all } \{p_1,\ldots,p_n\} \in D, \\ R[(P,D,M),\hat{C},(p_1,\ldots,p_n)] = \hat{C}. \;\Box \end{cases}$$

THEOREM 1. Soundness and completeness of E_+^+ inference engines.

> E_+^+ inference engines are sound and, with complete search strategies, they are also complete; i.e., an E_+^+ inference engine with a complete search strategy is guaranteed to minimally reduce any $\hat{C}^+$ covering to a C^+ covering.
>
> $$E_+^+[D_+^+, \hat{C}^+] = C^+ \subseteq \hat{C}^+$$
>
> and there does not exist another C^+ covering $\mathcal{C}^+$ for D_+^+ such that $C^+ \subset \mathcal{C}^+ \subseteq \hat{C}^+$. $\Box$

PROOF 1.

I. A complete E_+^+ inference engine, restricted to a single dependency relationship, discovers a minimally reduced $\hat{C}^+$ covering that satisfies that relationship.

Let $D_+^+ = (P,D,M,);$

$d^+ = \{p_1,\ldots,p_n\} \in D;$

$\vec{d}_i = (p_1,\ldots,p_{i-1},p_{i+1},\ldots,p_n,p_i),\ 1 \le i \le n;$

$\hat{C}^+_{(i)} = R_+^+[D_+^+,\hat{C}^+,\vec{d}_i],\ 1 \le i \le n;$

$\hat{C}^+_{(i,j)} = R_+^+[D_+^+,\hat{C}^+_{(i)},\vec{d}_j],\ 1 \le i,j \le n.$

To prove that $\hat{C}^+_{(i,j)} = \hat{C}^+_{(j,i)},\ 1 \le i,j \le n,$

first let n=2 and prove that $\hat{C}^+_{(1,2)} = \hat{C}^+_{(2,1)}.$

For each $c_1 \in \hat{C}^+_{(1)}[p_1],$

there exists $c_2 \in \hat{C}^+[p_2]$ such that

$(c_1,c_2) \in M[(p_1,p_2)],$

which implies that $c_2 \in \hat{C}^+_{(2)}[p_2].$

Therefore, $\hat{C}^+_{(2,1)}[p_1] = \hat{C}^+_{(1)}[p_1].$

Clearly, $\hat{C}^+_{(1,2)}[p_1] = \hat{C}^+_{(1)}[p_1]$ giving

$\hat{C}^+_{(1,2)}[p_1] = \hat{C}^+_{(2,1)}[p_1]$ and symmetrically

$\hat{C}^+_{(1,2)}[p_2] = \hat{C}^+_{(2,1)}[p_2].$

Therefore, $\hat{C}^+_{(1,2)} = \hat{C}^+_{(2,1)}$, when n=2.

With n>2, it's clear that

$$\hat{C}^{+}_{(1,2)}[p_k] = \hat{C}^{+}_{(2,1)}[p_k] = \hat{C}^{+}[p_k],\ 2<k<n.$$

Therefore, $\hat{C}^{+}_{(i,j)} = \hat{C}^{+}_{(j,i)}$, $1\le i,j\le n$.

From this it follows that $E^{+}_{+}[(P,\{d\},M),\ \hat{C}^{+}]$ is independent of the search strategy, so long as it is complete. Clearly, $E^{+}_{+}[(P,\{d\},M),\hat{C}^{+}]$ minimally reduces $\hat{C}^{+}$ relative to d since the only elements deleted from $\hat{C}^{+}$ are those that are inconsistent with M as it applies to d.

II. A $\hat{C}^{+}$ covering, minimally reduced to be consistent with the highest order dependency relationship in a D^{+}_{+} dependency relation, is the minimally reduced C^{+} covering for that dependency relation.

Let $D^{+}_{+} = (P,D,M)$ and $P = \{p_1,\ldots,p_n\}$.

Since $E^{+}_{+}[(P,\{P\},M),\hat{C}^{+}]$ has been proven to be the minimal reduction of $\hat{C}^{+}$ consistent with $M[(p_1,\ldots,p_n)]$ (part I of this proof), and since

$$(c_1,\ldots,c_i,\ldots,c_n) \in M[(p_1,\ldots,p_i,\ldots,p_n)]$$
$$\rightarrow (c_1,\ldots,c_i) \in M[(p_1,\ldots,p_i)]$$

where $P \supset \{p_1,\ldots,p_i\} \in D$,

it follows that $E^{+}_{+}[(P,\{P\},M),\hat{C}^{+}]$ is the minimal reduction of $\hat{C}^{+}$ satisfying all of M.

Therefore, $E^{+}_{+}[(P,\{P\},M),\hat{C}^{+}]$ is a C^{+} covering for D^{+}_{+} that is the minimal reduction of $\hat{C}^{+}$.

III. A complete E^{+}_{+} inference engine, operating over an entire D^{+}_{+} dependency relation, minimally reduces any $\hat{C}^{+}$ covering to a C^{+} covering for D^{+}_{+}. An inductive proof follows.

<u>Basis</u>. Given a $\hat{C}^{+}$ covering and a single relationship from a D^{+}_{+} dependency relation, a complete E^{+}_{+} inference engine will discover a minimally reduced $\hat{C}^{+}$ covering that satisfies that relationship (part I of this proof).

Inductive step. Assume that a complete E_+^+ inference engine, given a $\hat{C}^+$ covering and all but one of the relationships from D_+^+, is guaranteed to return a minimally reduced $\hat{C}^+$ covering that satisfies all of those relationships. Prove that with the addition of this excluded relationship, E_+^+ returns a minimally reduced C^+ covering for D_+^+.

Let $D_+^+ = (P,D,M)$ and $d = \{p_1,\ldots,p_n\} \in D$.

If d is the excluded relationship and $d \neq P$, then $E_+^+[(P,D-\{d\},M),\hat{C}^+]$ is the minimal reduction consistent with the highest order relationship and this has been proven to be a C^+ covering for all of D_+^+ (part II of this proof). Therefore, the addition of d has no effect.

If d is the excluded relationship and $d = P$, then $E_+^+[(p,D-\{d\},M),\hat{C}^+]$ is the minimal reduction of $\hat{C}^+$ satisfying the relationships in D-{d}. If this is a $\hat{C}^+$ covering for all of D_+^+, then an E_+^+ inference engine can further reduce it to a C^+ covering through application of R_+^+ relative to {P} (parts I and II of this proof). This would be the minimally reduced C^+ coverings for all of D_+^+.

It remains to be proven that

$$E_+^+[(P,\{P\},M),\hat{C}^+] \subseteq E_+^+[(P,D-\{P\},M),\hat{C}^+].$$

If this were not true, then there would have to be some $d_i \in D$,

$$d_i = \{p_1,\ldots,p_i\} \subset P = \{p_1,\ldots,p_i,\ldots,p_n\}, \text{ and}$$
$$c_j \in \hat{C}^+[p_j],\ 1 \le j \le n,$$

such that

$$(c_1,\ldots,c_i,\ldots,c_n) \in M[(p_1,\ldots,p_i,\ldots,p_n)] \text{ and}$$
$$(c_1,\ldots,c_i) \notin M[(p_1,\ldots,p_i)].$$

But this contradicts the second constraint on D_+ depencency relations. Therefore,

$$E_+^+[(P,\{P\},M),\hat{C}^+] \subseteq E_+^+[(P,D-\{P\},M),\hat{C}^+] \text{ and}$$

$$E_+^+[(P,\{P\},M),\ E_+^+[(P,D-\{P\},M),\hat{C}^+]] = C^+.$$

Therefore, $E_+^+[D_+^+,\hat{C}^+] = C^+$, by induction. □

In terms of a G_+^+ representation of a D_+^+ dependency relation, an E_+^+ inference engine makes inferences along arcs selected by its search strategy. If the search strategy is complete, E_+^+ will draw all of the inferences possible given the information in G_+^+. But if G_+^+ is only an approximation of D_+^+, it may be that some of what follows from D_+^+ will not be discovered. Completeness may be sacrificed, though soundness is guaranteed provided G_+^+ _is_ an approximation of D_+^+. For G_+^+ to be an approximation of D_+^+, each of its mappings must contain the corresponding portion of M in D_+^+; thus less precise predictions might be made relative to G_+^+, but they will be consistent with those that would be made relative to D_+^+. Figures 22 and 23 show sequences of inferences that a complete E_+^+ inference engine might make given some $\hat{C}^+$ coverings for the example G_+^+ graphs of Chapter III.

In general, all of the other classes of inference engines can be viewed as restricted versions of E_+^+. Like E_+^+, the other inference engines E_2^1, E_2^+, and E_+^1 are sound and, given complete search strategies, also complete. However, in the case of E_2^1 and E_+^1, those that extend $\hat{C}^1$ coverings, they do not always produce a C^1 covering. This is not to say that they are incomplete, but only that completeness does not guarantee the production of a C^1 covering. For example, in a G_2^1 dependency graph, connected propositions are totally dependent on one another; unconnected propositions are totally independent. This means that an E_2^1 inference engine can predict the confidence of a proposition only if the confidence of some connected proposition is known. Connected portions of a G_2^1 graph correspond to equivalence classes; information

Figure 22. An example of $E^{+}_{\ddagger}$ inference activity. At each step, the choice of a highlighted arc by $E\ddagger$ for application of $R\ddagger$ will reduce the $\hat{C}^{+}$ covering; * marks the selected arch.

$$C = \{T,F\}.$$

$$\leftrightarrow[c_i] =_{def} c_i.$$

$$\sim[c_i] =_{def} \begin{cases} \{T\}, & c_i = \{F\}, \\ \{F\}, & c_i = \{T\}, \\ \{T,F\}, & \text{else.} \end{cases}$$

$$\rightarrow[c_i] =_{def} \begin{cases} \{T\}, & c_i = \{T\}, \\ \{T,F\}, & \text{else.} \end{cases}$$

$$\leftarrow[c_j] =_{def} \begin{cases} \{F\}, & c_j = \{F\}, \\ \{T,F\}, & \text{else.} \end{cases}$$

$$XOR[c_1,\ldots,c_n] =_{def} \begin{cases} \{T\}, & \exists_{i=1}^{n} c_i = \{T\}, \\ \{F\}, & c_i = \{F\},\ 1 \le i \le n, \\ \{T,F\}, & \text{else.} \end{cases}$$

$$xor[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n] =_{def} \begin{cases} \{T\}, & c_0 = \{T\},\ c_j = \{F\},\ 1 \le j \le n,\ i \ne j \\ \{F\}, & c_0 = \{F\} \text{ or } \exists_{j=1,\, i \ne j}^{n} c_j = \{T\}, \\ \{T,F\}, & \text{else.} \end{cases}$$

$$\bigwedge[c_1,\ldots,c_n] =_{def} \begin{cases} \{T\}, & c_i = \{T\},\ 1 \le i \le n, \\ \{F\}, & \exists_{i=1}^{n} c_i = \{F\}, \\ \{T,F\}, & \text{else.} \end{cases}$$

$$\wedge[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n] =_{def} \begin{cases} \{T\}, & c_0 = \{T\}, \\ \{F\}, & c_0 = \{F\},\ c_j = \{T\},\ 1 \le j \le n,\ i \ne j \\ \{T,F\}, & \text{else} \end{cases}$$

$$\bigvee[c_1,\ldots,c_n] =_{def} \begin{cases} \{T\}, & \exists_{i=1}^{n} c_i = \{T\}, \\ \{F\}, & c_i = \{F\},\ 1 \le i \le n, \\ \{T,F\}, & \text{else.} \end{cases}$$

$$\vee[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n] =_{def} \begin{cases} \{T\}, & c_0 = \{T\},\ c_j = \{F\},\ 1 \le j \le n,\ i \ne j \\ \{F\}, & c_0 = \{F\}, \\ \{T,F\}, & \text{else.} \end{cases}$$

where c_0 always corresponds to a distinguished proposition.

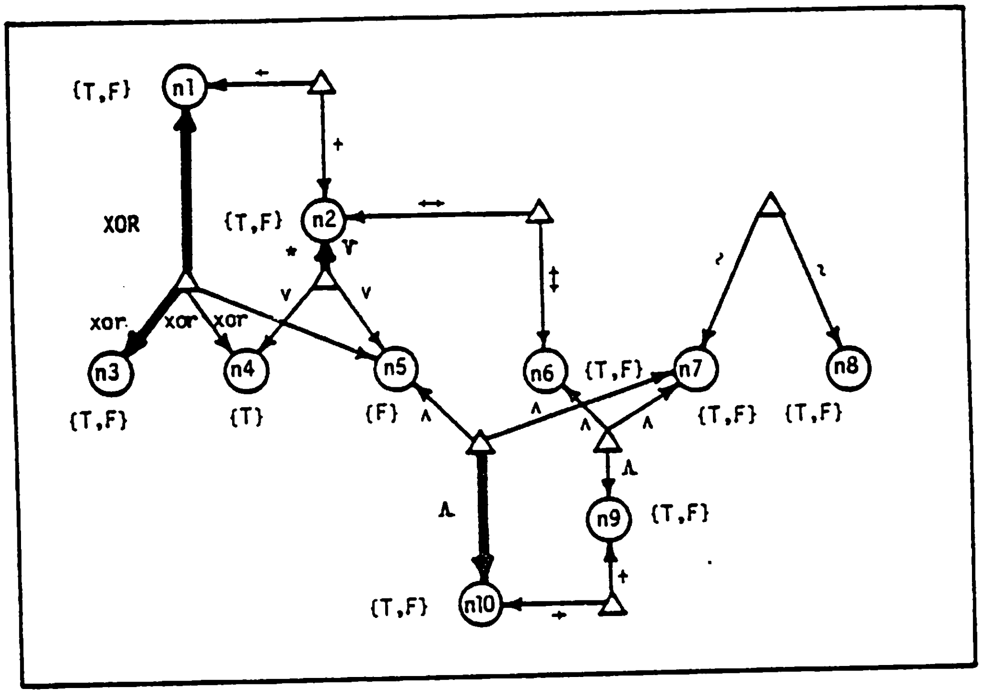


Figure 22 (a). Initial state$_0$

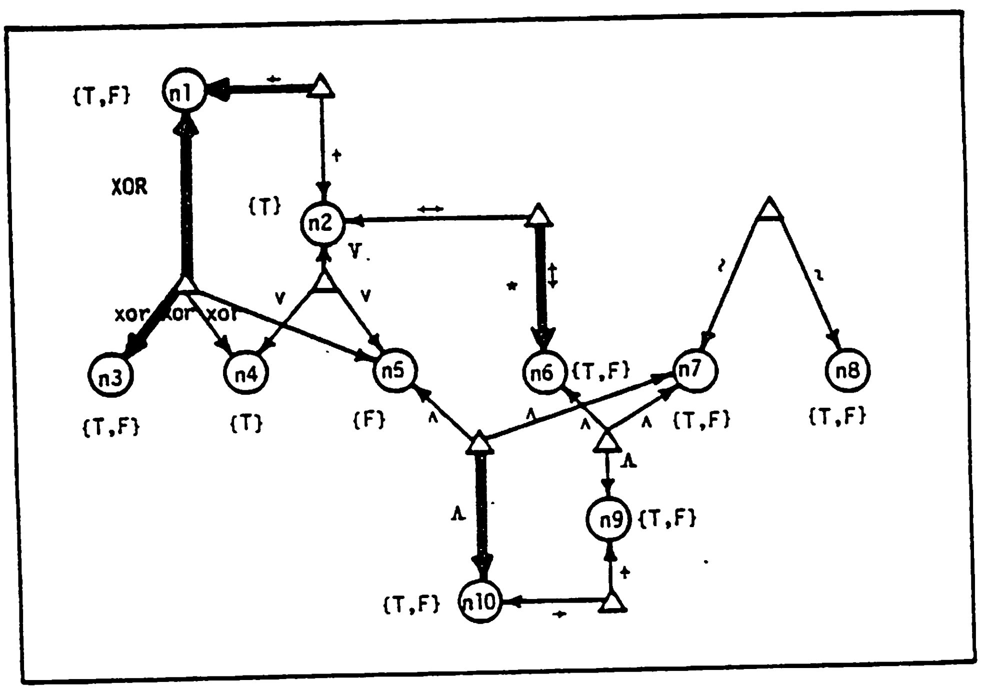


Figure 22 (b). State $_1$

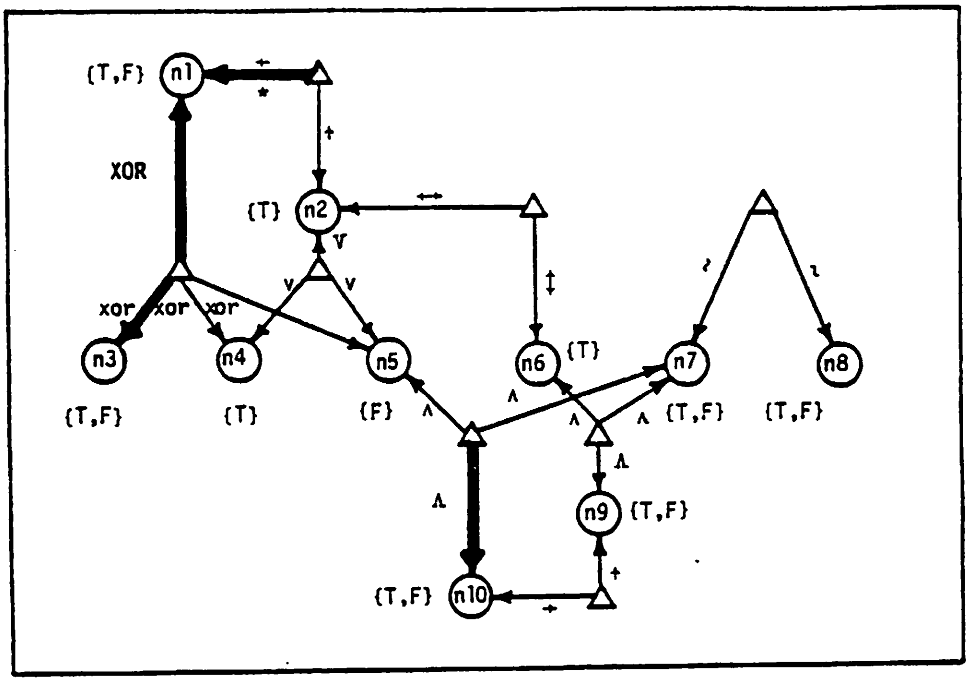


Figure 22 (c). $State_2$

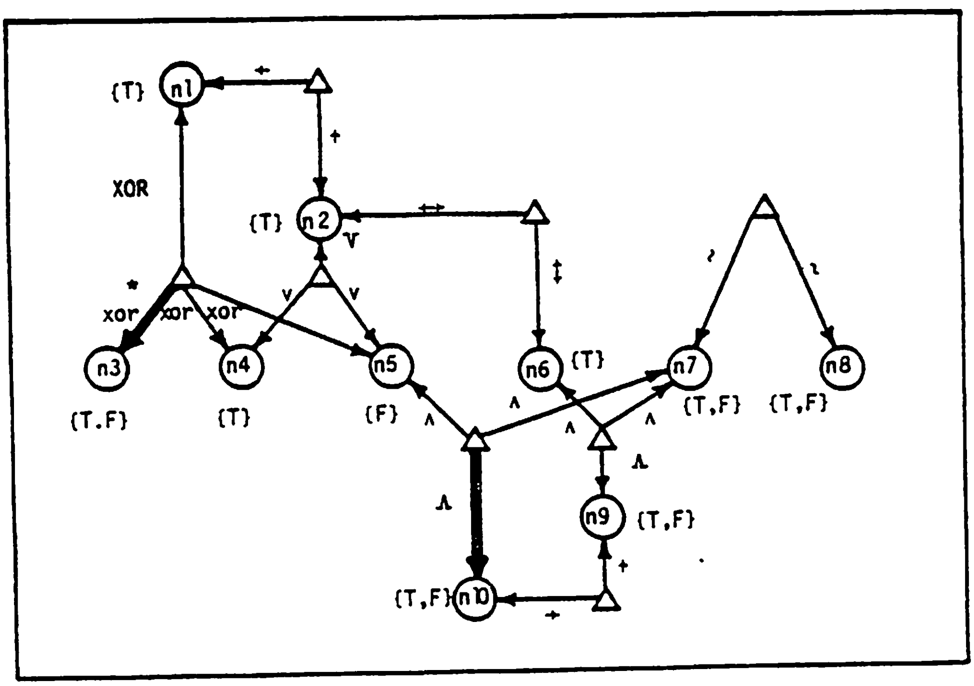


Figure 22 (d). $State_3$

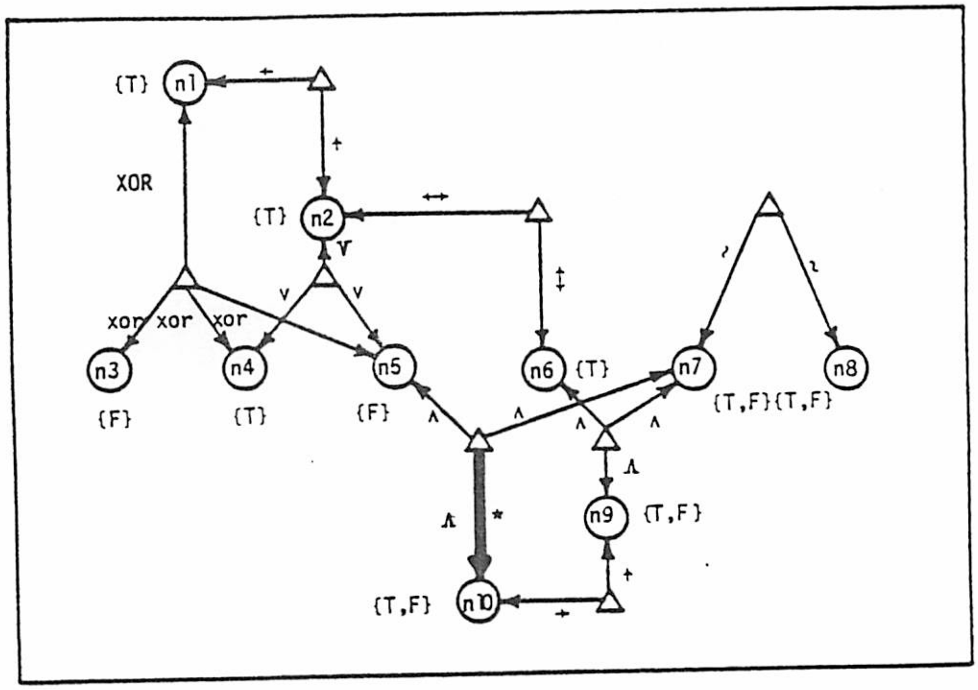


Figure 22 (e). State$_4$

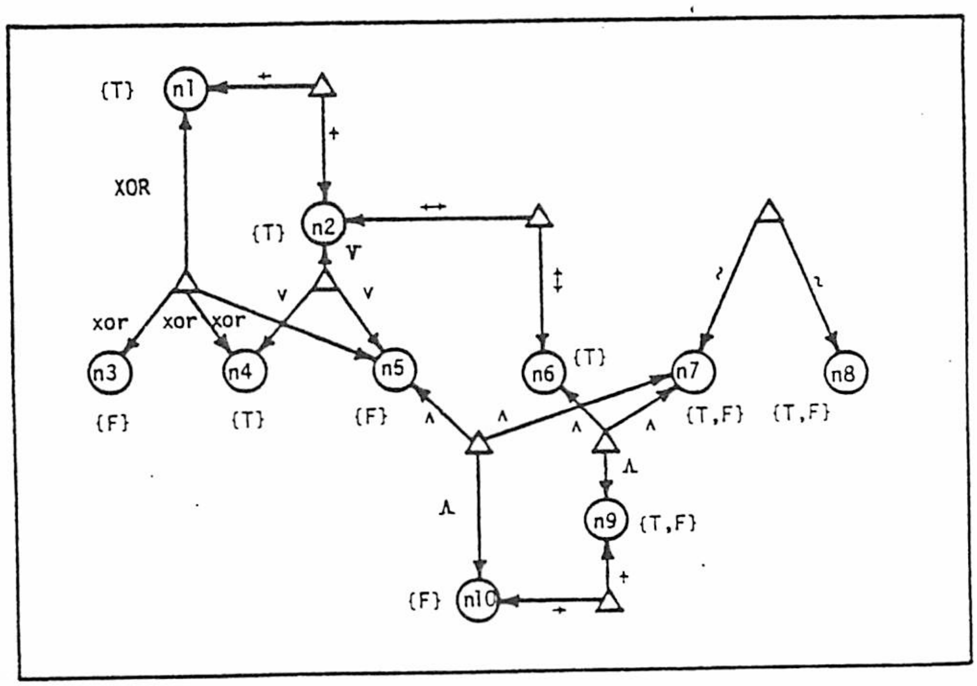


Figure 22 (f). Final State$_5$

Figure 23. An example of E_+^+ inference activity. At each step, the choice of a highlighted arc by E_+^+ for application of R_+^+ will reduce the C^+ covering; * marks the selected arc.

$$C = [0.1].$$

$$\leftrightarrow[c_i] =_{def} c_i.$$

$$\sim[c_i] =_{def} \{c_j' \mid c_j' = 1 - c_i',\ c_i' \in c_i\}.$$

$$\rightarrow[c_i] =_{def} \{c_j' \mid c_i' \le c_j' \le 1,\ c_i' \in c_i\}.$$

$$\leftarrow[c_j] =_{def} \{c_i' \mid 0 \le c_i' \le c_j',\ c_j' \in c_j\}.$$

$$\mathrm{XOR}[c_1,\ldots,c_n] =_{def} \{c_0' \mid c_0' = \sum_{i=1}^{n} c_i',\ c_0' \in C,\ c_i' \in c_i,\ 1 \le i \le n\}.$$

$$\mathrm{xor}[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n]$$
$$=_{def} \{c_i' \mid c_i' = c_0' - \sum_{\substack{j=1 \\ i \ne j}}^{n} c_j',\ c_i' \in C,\ c_j' \in c_j,\ 0 \le j \le n,\ i \ne \underline{j}\}.$$

$$\bigwedge[c_1,\ldots,c_n] =_{def} \{c_0' \mid 1 + (\sum_{i=1}^{n} c_i' - 1) \le c_0' \le \mathrm{MIN}_{i=1}^{n}\, c_i',\ c_0' \in C,$$
$$c_i' \in c_i,\ 1 \le i \le n\}.$$

$$\wedge[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n]$$
$$=_{def} \{c_i' \mid c_0' \le c_i' \le c_0' - (\sum_{\substack{j=1 \\ i \ne j}}^{n} c_j' - 1),\ c_i' \in C,\ c_j' \in c_j,\ 0 \le j \le n,\ i \ne j\}.$$

$$\bigvee[c_1,\ldots,c_n] =_{def} \{c_0' \mid \mathrm{MAX}_{i=1}^{n}\, c_i' \le c_0' \le \sum_{i=1}^{n} c_i',\ c_0' \in C,\ c_i' \in c_i,\ 1 \le i \le n\}.$$

$$\vee[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n]$$
$$=_{def} \{c_i' \mid (c_0' - \sum_{\substack{j=1 \\ i \ne j}}^{n} c_j') \le c_i' \le c_0',\ c_i' \in C,\ c_j' \in c_j,\ 0 \le j \le n,\ i \ne j\}.$$

where c_0 always corresponds to a distinguished proposition.

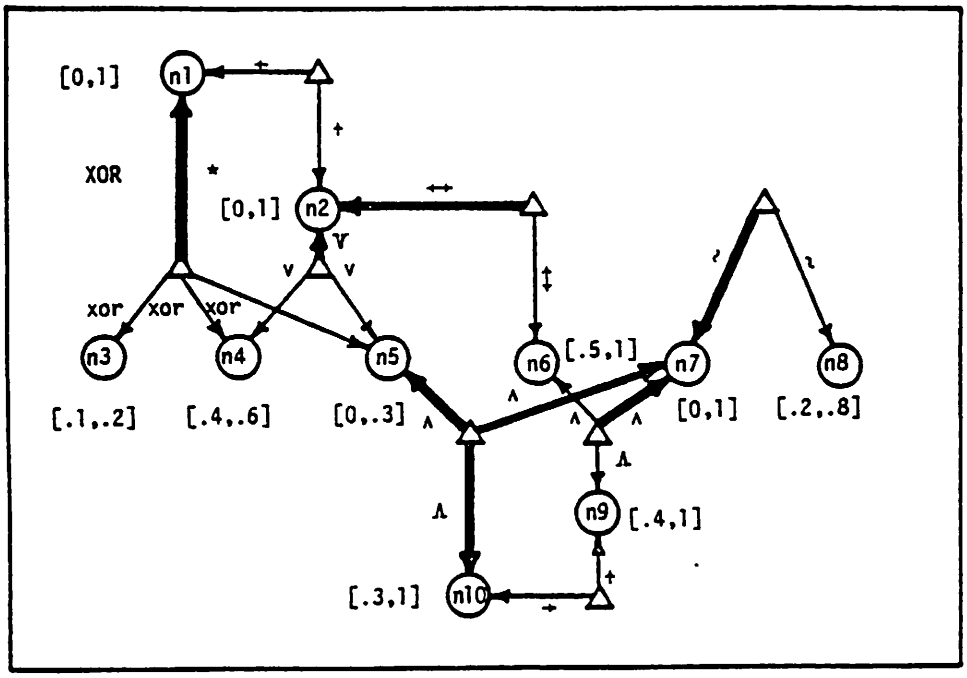


Figure 23 (a). Initial $state_0$

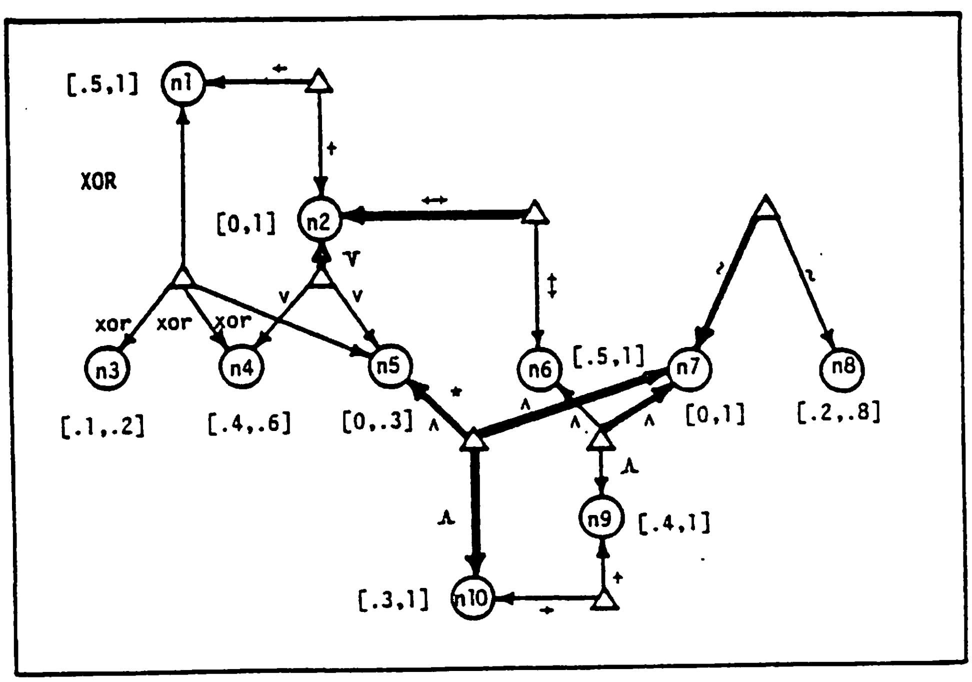


Figure 23 (b). $State_1$

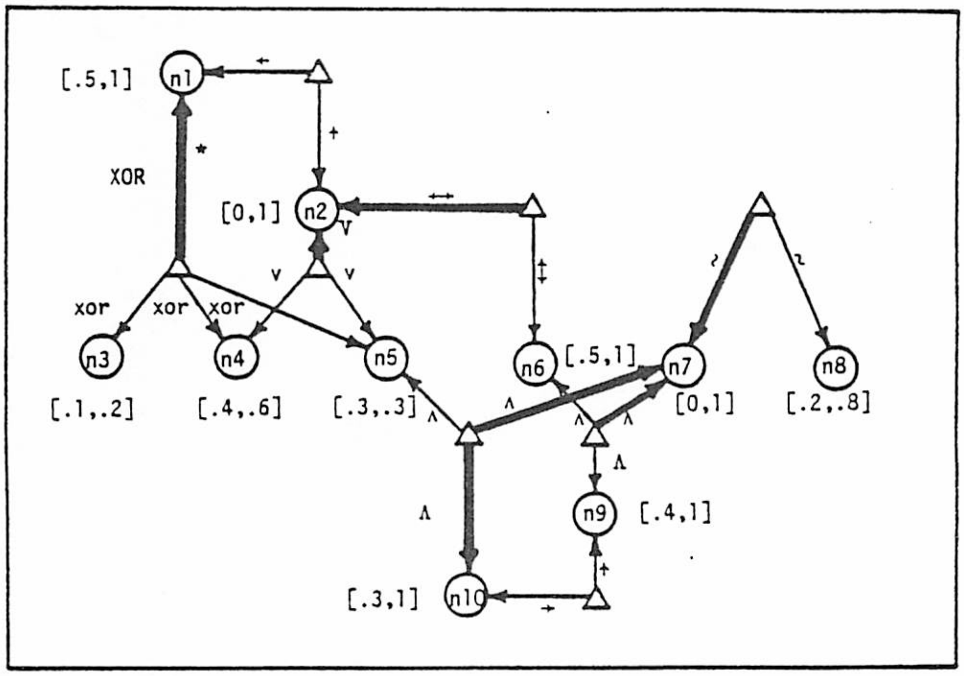


Figure 23 (c). $State_2$

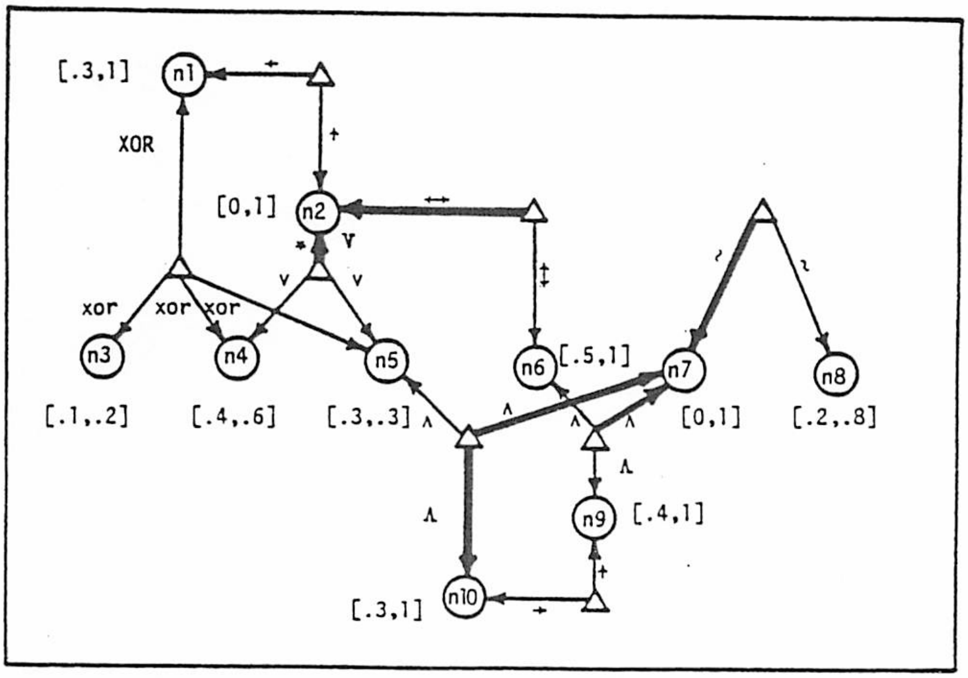


Figure 23 (d). $State_3$

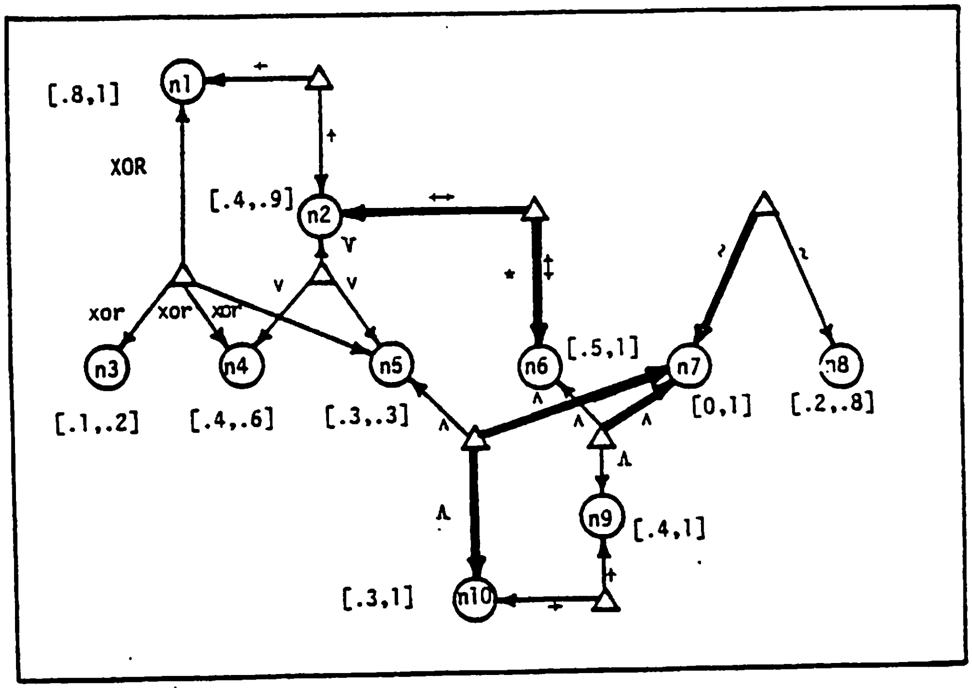


Figure 23 (e). $State_4$

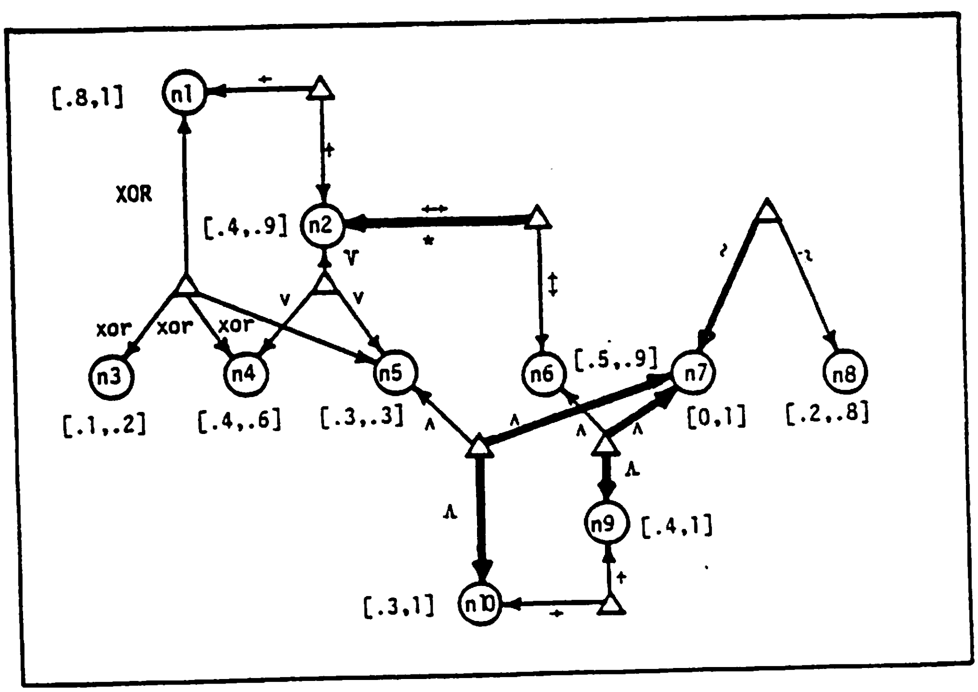


Figure 23 (f). $State_5$

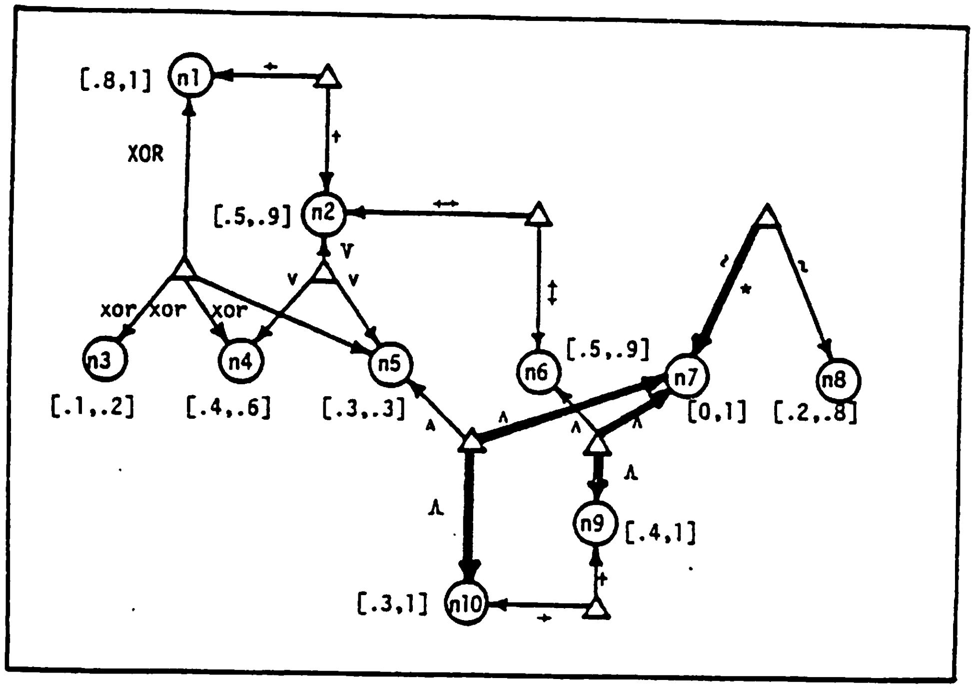


Figure 23 (g). State_6

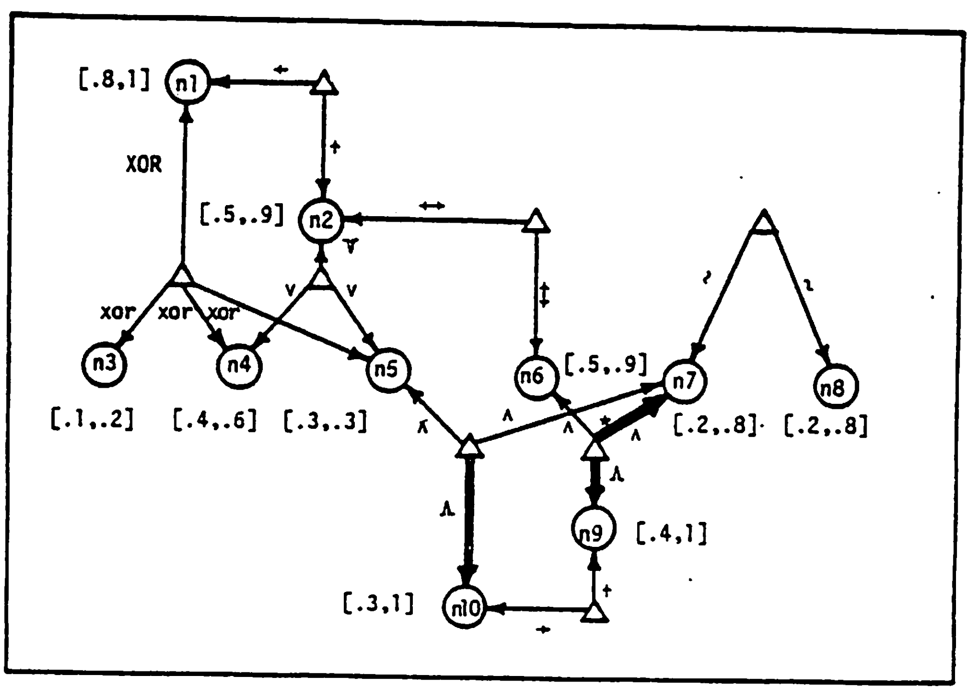


Figure 23 (h). State_7

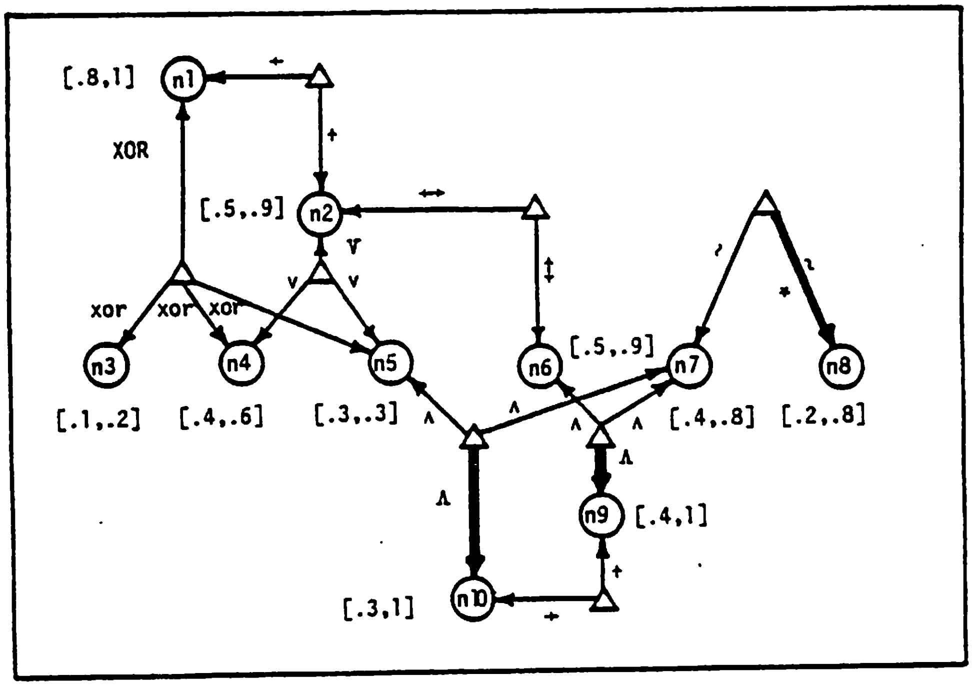


Figure 23 (i). $State_8$

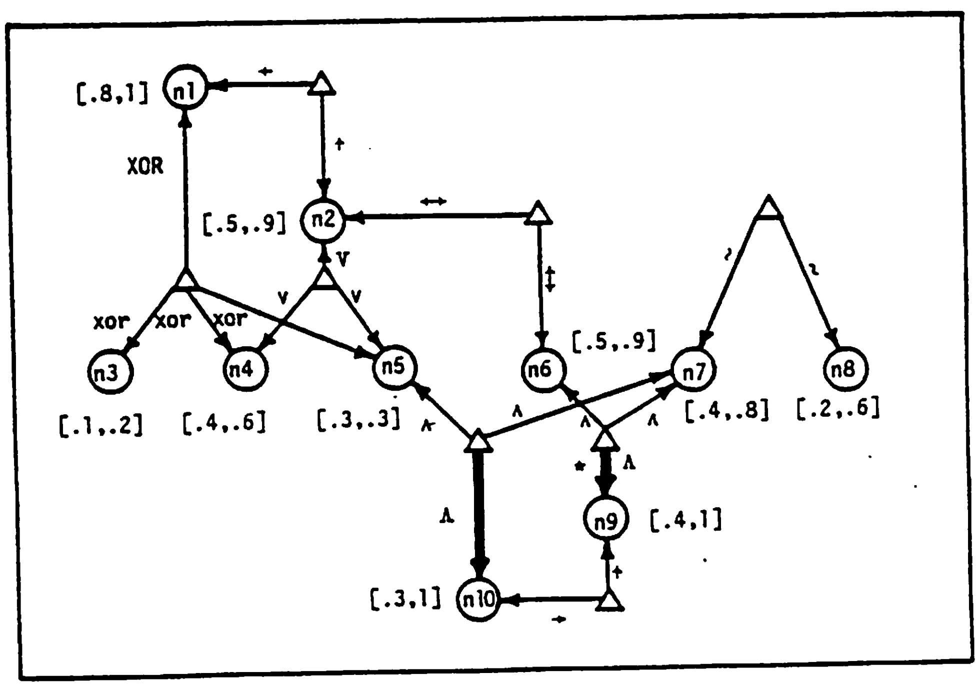


Figure 23 (j). $State_9$

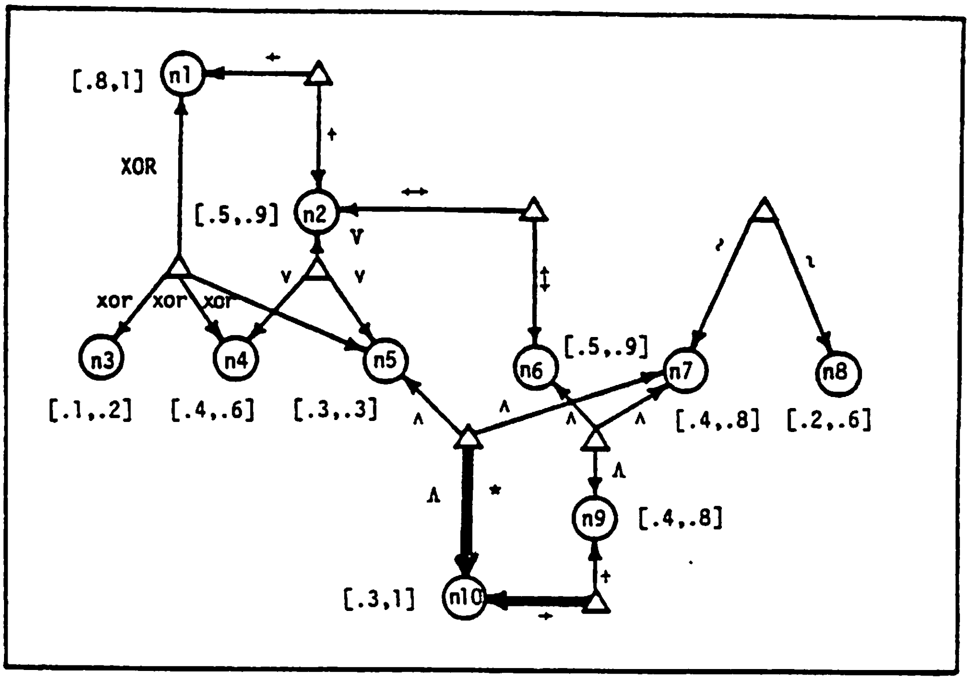


Figure 23 (k). $State_{10}$

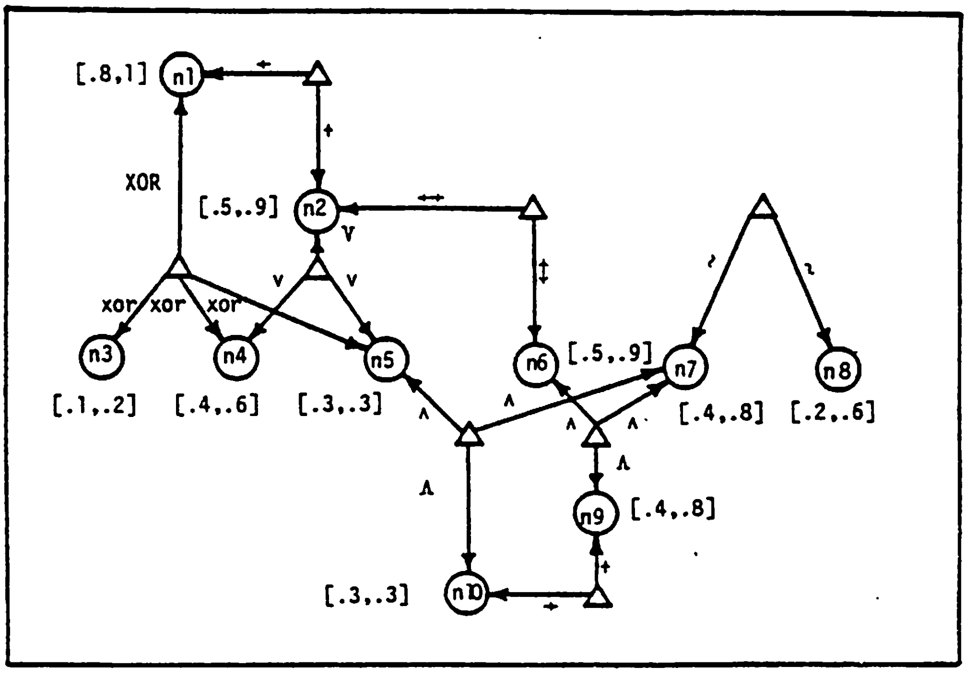


Figure 23 (l). Final $State_{11}$

about a proposition in an equivalence class is equivalent to information about any other proposition in that equivalence class. If a $\hat{C}^1$ covering assigns a confidence to at least one proposition in each equivalence class of a G_2^1 dependency graph, then an E_2^1 inference engine can extend that incomplete covering to a complete C^1 covering; otherwise, the best E_2^1 can do is to extend it to a more complete $\hat{C}^1$ covering. Similar comments pertain to E_+^1 inference engines.

Clearly, the effectiveness of any inference engine is intimately tied to its search strategy. And the best search strategy for a given situation depends on a number of factors. For example, if some kind of parallel hardware is to be used, the best search strategy might be to continually apply the inference rule everywhere until quiescence. If serial hardware is to be employed, search strategies that take advantage of the inherent indexing structure of graphs (i.e., strategies based on graph walking algorithms) are probably better suited. But here the best choice varies depending on the connectivity of the graph, where the initial information is located within the graph, and what information is desired. If there is a single path from one proposition to another and confidence information is provided about one and desired about the other, the optimal inference ordering is obvious. But if several paths exist between these propositions, each with a different degree of specificity, the best ordering is not as obvious. As with all search situations, there is a trade off between doing extensive analysis, to select optimal paths, and making blind selections. A detailed analysis of inference-engine search-strategies is beyond the scope of this thesis, but the standard body of search techniques is applicable [Nilsson 1980].

Theorem 1 guarantees that a dependency-graph inference-engine, given a well-formed dependency-graph and $\hat{C}$ covering, is a sound inference procedure. This is true even when the ground information is partial. Models, based on dependency graphs, are thereby guaranteed to be internally consistent. Dependency-graph models are the topic of the next chapter.

CHAPTER V

DEPENDENCY-GRAPH MODELS

Dependency graphs and their accompanying inference engines might be of theoretical interest to mathematicians, but not until they are interpreted as models of environmental situations do they fall within the realm of artificial intelligence. In this capacity, information concerning the confidences of some environmental propositions can be transformed into predictions about other environmental propositions. An interpretation is established by identifying the nodes of a dependency graph with propositions relative to some environment. Given accurate confidence information, an accurate dependency-graph model will make accurate predictions through the application of its inference engine.

Confidence information, from some source of knowledge over the environment, enters a dependency-graph model in the form of an (incomplete) $\hat{C}^1$ covering or an (unrefined) $\hat{C}^+$ covering. The model's inference engine extends or refines this covering, translating the information from proposition to proposition within the dependency graph, yielding more complete information about the confidences of the propositions represented in the graph (Figure 24). The accuracy of these predictions depends on both the accuracy of the initial confidence information and how accurately the dependency graph reflects the dependency relationships in the environment. If either of these is in error, the integrity of the predictions is impugned.

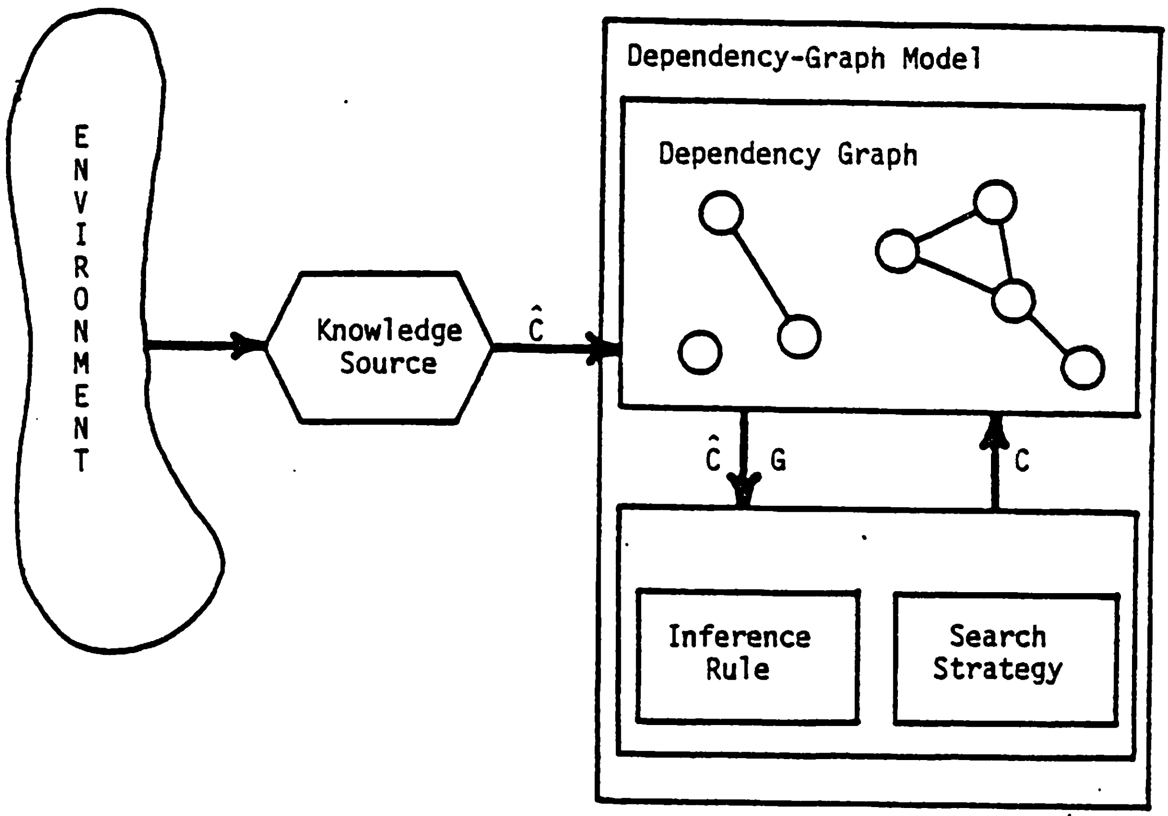


Figure 24. Dependency-graph model.

A model is accurate so long as it does not contradict the reality of the environment it propounds to model. A model that fails to contradict its environment is said to be <u>externally consistent</u> i.e., consistent with the external environment it models. <u>Internal consistency</u>, the absence of internal contradiction within a model, is a prerequisite for external consistency. If a model contains contradictory descriptions, it necessarily contradicts its environment! Inferencing based on an internally inconsistent model leads to logical contradictions, and thereby model failure. Of course, internal consistency does not guarantee external consistency; it is a necessary, but not a sufficient condition. Since a well-formed dependency-graph model is guaranteed to be internally consistent, inaccuracies must be attributed to external, and not internal, inconsistencies.

A dependency-graph model can be externally inconsistent due to either its dependency graph or its covering. Either might contradict the environment, leaving the model open to failure. The most graceful failures are those that simply lead to incorrect predictions about the environment. More catastrophic failures can lead to complete <u>model breakdown</u>. Model breakdown can occur if the initial confidence information is not a $\hat{C}$ covering for the dependency graph i.e., there is no underlying C covering over which $\hat{C}$ could be defined. Applications of an inference engine to such a noncovering $\hat{\mathcal{C}}$ can lead to incompatible predictions, necessarily so if the inference-engine's search-strategy is complete. In the case of $\hat{\mathcal{C}}^1$ noncoverings, a complete E^1 inference engine eventually attempts to reset the confidence of some proposition already assigned; in the case of $\hat{\mathcal{E}}^+$ noncoverings, a complete E^+ inference engine eventually assigns some proposition the null set of confidences, indicating that no confidence assignment for that proposition is consistent with all of its neighbors' assignments. This situation is analogous to a theorem prover provided an inconsistent set of axioms; the null clause follows. In either case, the search strategy can be set up to halt when breakdown occurs, but the results are largely meaningless. A noncovering and dependency graph are incompatible relative to any environment. One or both must be externally inconsistent.

One important property of dependency-graph models is that a covering which only provides information about a single proposition is guaranteed to be a $\hat{C}$ covering. Dependency graphs were carefully defined to prevent unilateral exclusions of confidence values from propositions. The significance of this will become apparent when dependency-graph models of evidential support are explored in the next chapter.

An externally consistent model still may be largely uninformative. External consistency only guarantees that a model does not contradict its environment. Clearly a model, that never makes any predictions other than reaffirming that each proposition is to be assigned some confidence out of the possible range of confidences, is always correct, but uninformative. To some degree all models of real-world situations are merely approximations of their environments; infinite precision cannot be expected. Imprecision is an inherent nemesis of the modeling process and, paradoxically, also its salvation.

Approximations can be found at several levels in dependency-graph models. To begin with, dependency graphs are defined as approximations of underlying dependency relations. The assumed dependency relations, in turn, approximate the environments they model. Finally, the coverings allow confidence assignments to be specified with varying degrees of precision. Although these approximations lead to less precise predictions, they also allow available partial information to be gainfully employed. A truthful set of partial statements leads to truthful partial conclusions; a false set of precise statements leads to useless contradiction. Since partial knowledge is the rule when modeling any real-world situation, it is imperative that any inference mechanism, on which such models are to be based, be capable of performing inexact reasoning from incomplete information. External consistency can only be achieved within a modeling environment tolerant of imprecision, one that does not demand overstatement of what is known. Dependency-graph models provide such an environment without a loss of logical consistency. Thus, model failures are less likely to occur and are never attributable to internal inconsistencies.

In general, achieving external consistency remains an open problem. We have not described how externally consistent dependency-graph models can be constructed, nor have we described how, when given such a model, we could reason from confidence information that contradicts that model. Although the theory of dependency-graph models makes no statements about how to achieve external consistency for any particular class of environmental situations, the freedom to express partial information within dependency-graph models makes them a suitable host for a mathematical theory of evidential reasoning that does describe how external consistency is achieved within this limited context. The next chapter explores this application in detail.

CHAPTER VI

DEPENDENCY-GRAPH MODELS OF EVIDENTIAL SUPPORT

Frequently, the situations of interest in artificial intelligence applications are evidential. Knowledge sources operating over an environment extract bodies of evidence that attribute degrees of support, not truth or falsity, to selected propositions. In general, these knowledge sources cannot provide the exact degree of support that should be accorded each proposition, but rather provide partial information prone to occasional errors. To counter the fallibility of these knowledge sources, bodies of evidence from several knowledge sources with distinct perspectives on the environment often need to be pooled; the consensus of several independent opinions are generally more reliable than any single one. This scenario is particularly evident in those applications involving perceptual reasoning and situation assessment.

A mathematical theory of evidence by Glenn Shafer [Shafer 1976] provides the appropriate foundation for the construction of dependency-graph models of evidential support. These are a subset of the M_+^+ dependency-graph models (i.e., those models consisting of G_+^+ dependency graphs, $\hat{C}^+$ coverings, and E_+^+ inference engines). Shafer's theory dictates the appropriate D_+^+ dependency relations and C^+ coverings for representing the impact of a body of evidence on a set of propositions. Dempster's rule of combination [Dempster 1976, 1968], an integral part of Shafer's theory, is a rule for pooling distinct bodies of evidence. The information needed for its application is embodied within the dependency-graph models constructed in accordance with Shafer's theory. Thereby, dependency-graph models of evidential support provide an integrated framework for the combination and extrapolation of evidential information in a way that is both sound and amenable to mechanization.

The next few sections of this chapter introduce Shafer's theory of evidence and describe how dependency-graph models are constructed in accordance with it. This is followed by an introduction to Dempster's rule of combination and its application with respect to dependency-graph models of evidential support. This chapter closes with some examples.

A Mathematical Theory of Evidence

Descriptions of evidential impact. Shafer's theory of evidence begins with the familiar idea of using a number between zero and one to indicate the degree of support a body of evidence provides for a proposition.

$$0 \leq SPT[p] \leq 1. \qquad (6.1)$$

However, Shafer maintains that an adequate summary of the impact of a body of evidence on a proposition also must include how well the negation of that proposition is supported, i.e., the dubiety of the proposition.

$$SPT[\bar{p}] = DBT[p] \; . \qquad (6.2)$$

Unlike the Bayesian theory, the support for a proposition and the support for its negation are not required to sum to one; their sum is merely bounded by one.

$$SPT[p] + SPT[\bar{p}] \leq 1, \quad \text{i.e., } SPT[p] + DBT[p] \leq 1. \tag{6.3}$$

This is the key to the representation of ignorance. When there is little evidence bearing on a proposition, frank agnosticism is expressed by according both that proposition and its negation very low degrees of support, zero if the evidence has no bearing whatsoever on the proposition.

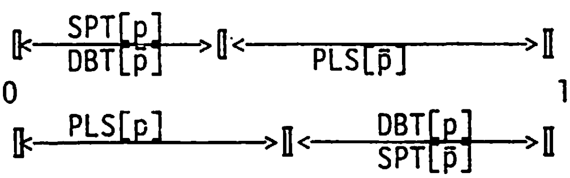
$$\text{State of Ignorance: } SPT[p] = DBT[p] = 0. \tag{6.4}$$

Consider the difficulty of expressing ignorance within the Bayesian framework. Suppose that for a given situation we have two propositions p_1 and p_2, and we know that exactly one of these is true, but we have no idea which one. The usual technique is to set the probability of p_1, $\bar{p}_1$, p_2, and $\bar{p}_2$ all to .5; thus all of the propositions are equally likely. However, what if p_1 can be more finely described in terms of propositions $p_{1,1}$ and $p_{1,2}$, exactly one of which is true if and only if p_1 is true? The Bayesian theory does not allow us to assign .5 to all of these propositions. If p_1 and p_2 are to remain at .5, then the sum of $p_{1,1}$ and $p_{1,2}$ must be .5. So we might assign them each .25. But now p_2 (at .5) is twice as likely as $p_{1,1}$ (at .25) and 2/3 as likely as $\bar{p}_{1,1}$ (at .75). An alternative approach might be to assign 1/3 to each of the propositions $p_{1,1}$, $p_{1,2}$, and p_2, but then none of these propositions are as likely as their negations (at 2/3) or p_1 (at 2/3). How can one conclude from no information that some things are more likely than others? The problem is that the Bayesian theory does not distinguish between disbelief and the absence of belief. Any belief not attributed to a proposition is automatically attributed to its negation. Within Shafer's theory the support and dubiety of all of these propositions can be zero, favoring no one proposition over any other.

One final measure for the expression of evidential impact remains, plausibility. A proposition is plausible to the extent that one fails to doubt it, which is the amount dubiety differs from one.

$$\begin{aligned} PLS[p] &= 1 - DBT[p], \\ &= 1 - SPT[\bar{p}]. \end{aligned} \qquad (6.5)$$

Clearly support, dubiety, and plausibility all convey the same information.



Thus, a summary of the impact of a body of evidence on a proposition can be expressed in several equivalent ways. But a complete summary must include at least two values, one from each of the following two sets:

$$\{SPT[p], DBT[\bar{p}], PLS[\bar{p}]\},$$

$$\{SPT[\bar{p}], DBT[p], PLS[p]\}.$$

Frames of discernment. The remainder of Shafer's theory is most easily introduced utilizing the familiar formalism whereby propositions are represented as subsets of a given set, here referred to as the frame of discernment or θ. When a proposition corresponds to a subset of a frame of discernment, that frame is said to discern that proposition. If the task is to determine the true value of some variable v, then θ is the set of all the possible values for that variable and the propositions of interest are precisely those of the form , "The true value of v is in p," where p is a subset of θ. Thus, the propositions of interest are in correspondence with the subsets of θ. For example, if the task were character recognition, the frame of discernment would be the set of possible characters and the propositions of interest would correspond to subsets of those characters. The proposition, "The character has two parallel vertical line-segments," might correspond to the set {H,M,N}. The primary advantage of this formalism is that it translates the logical notions of conjunction, disjunction, implication, and negation into the more graphic set-theoretic notions of intersection, union, inclusion, and complementation.

The interdependence of evidential support. The interdependence of evidential support relative to a frame of discernment is based on two assumptions:

1. The chosen frame of discernment contains the true value of the variable of interest;
2. Any support committed to a proposition is thereby committed to any other proposition it implies.

The first assumption dictates that the proposition corresponding to the frame of discernment always receives full support, and its negation, no support.

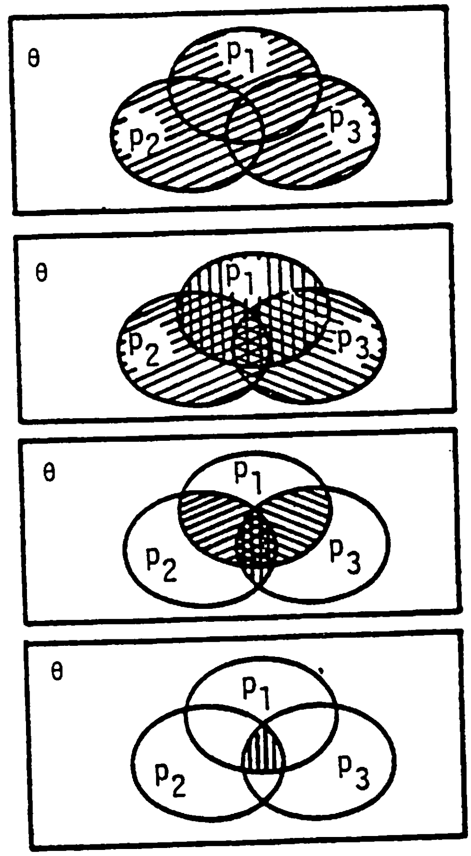
$$\begin{aligned} &SPT[\theta] = 1, \\ DBT[\theta] = &SPT[\bar{\theta}] = SPT[\phi] = 0. \end{aligned} \tag{6.6}$$

The second assumption dictates that any support committed to one subset of the frame of discernment is thereby committed to any subset containing it. One proposition implies another if it is a subset of that proposition in the frame of discernment. Of the total support commited to a given proposition p relative to a frame of discernment, some may be commited to one or more proper subsets of p, while the rest is committed exactly to p--to p and to no smaller subset.

$$SPT[p_1 \cup \ldots \cup p_n] \geq \sum_{\substack{J \subseteq \{1,\ldots,n\} \\ J \neq \phi}} (-1)^{|J|+1} \, SPT[\bigcap_{j \in J} p_j]. \tag{6.7}$$

Let us explain this inequality with a descriptive example. Expanding expression (6.7) for the case where n = 3, we see that it endeavors to count the support given each portion of the frame of discernment in $(p_1 \cup p_2 \cup p_3)$ exactly once: first summing across the subsets p_1, p_2, p_3, including their overlapping portions more than once; then excluding these overlapping portions $(p_1 \cap p_2)$, $(p_1 \cap p_3)$, and $(p_2 \cap p_3)$; finally adding back the overlapping portion excluded once too often in the last step $(p_1 \cap p_2 \cap p_3)$.

$$SPT[p_1 \cup p_2 \cup p_3] \geq SPT[p_1]+SPT[p_2]+SPT[p_3] - SPT[p_1 \cap p_2]-SPT[p_1 \cap p_3] -SPT[p_2 \cap p_3] + SPT[p_1 \cap p_2 \cap p_3]. \quad (6.8)$$



Expression (6.8) is an inequality since there may be some support for $(p_1 \cup p_2 \cup p_3)$ that is not attributable to any smaller subset.

An alternative view of evidential support focuses on the measure of direct support a body of evidence provides each proposition, excluding any support that evidence might indirectly provide. This measure, called a proposition's basic probability <u>mass</u>, is understood to be the support committed exactly to a proposition, not the total support committed to it. A mass function is defined by a support function.

$$MASS[p] = \sum_{p_i \subseteq p} (-1)^{|p-p_i|} SPT[p_i]. \quad (6.9)$$

A mass function provides exactly the same information as a support function: the support for a proposition being equal to the sum of the mass attributed it or any of its subsets.

$$SPT[p] = \sum_{p_i \subseteq p} MASS[p_i] \tag{6.10}$$

The total mass distributed over the frame of discernment is always equal to one. This is obvious since all propositions are subsets of the frame of discernment and the frame of discernment always receives full (unit) support. From this it is clear that the mass assigned the null set is identically zero. Thus, a mass function partitions a unit of support among the subsets of θ, assigning to each subset p that portion committed to p and to nothing smaller.

$$MASS[\phi] = 0; \tag{6.11}$$

$$\sum_{p \subseteq \theta} MASS[p] = 1.$$

Intuitively, mass is attributed to the most precise propositions a body of evidence supports. If a portion of mass is attributed to a proposition, it represents a minimal commitment to that proposition, and all of the propositions it implies. Additional mass suspended above that proposition, at propositions that imply neither it nor its negation, represents a potential commitment. This mass neither supports nor denies that proposition at present, but might later contribute either way based on additional information. The amount of mass so suspended above a proposition accounts for the difference between its support and plausibility, the latitude remaining in that proposition's probability. Thus, mass associated with the disjunction of p_1 and p_2 represents potential commitments to p_1 and to p_2 that are not yet realized, and an immediate commitment to the proposition $(p_1 \vee p_2)$ and all that it implies. Mass directly attributed to θ is noncommittal with respect to all of the propositions that θ discerns. It provides an equal potential for each and represents the degree to which the evidence fails to determine anything beyond the initial assumption that θ holds.

The impact of a body of evidence on a set of propositions discerned by a frame of discernment can be described by a support function or a mass function, each being a notational variant of the other. A few examples follow.

If nothing is known about the propositions discerned by a frame of discernment, complete ignorance is expressed by the <u>vacuous support</u> function. It supports no propositions other than the frame of discernment. The entire unit of mass is assigned to the frame of discernment.

$$SPT[p_i] = \begin{cases} 0, & p_i \neq \theta \\ 1, & p_i = \theta. \end{cases} \qquad MASS[p_i] = \begin{cases} 0, & p_i \neq \theta \\ 1, & p_i = \theta. \end{cases}$$

A <u>simple support</u> function distributes its mass between the frame of discernment and one other proposition (e.g., p_1). This represents the situation where a body of evidence points precisely and unambiguously to a single proposition. The mass attributed to this proposition (e.g., m_1) reflects the strength of the evidence. All propositions implied by this one receive an equal measure of support; all other propositions receive no support.

$$SPT[p_i] = \begin{cases} 0, & p_1 \not\subseteq p_i \\ m_1, & p_1 \subseteq p_i \neq \theta \\ 1, & p_i = \theta. \end{cases} \qquad MASS[p_i] = \begin{cases} 0, & p_i \neq p_1,\ p_i \neq \theta \\ m_1, & p_i = p_1 \\ 1 - m_1, & p_i = \theta. \end{cases}$$

In general, the basic support mass can be distributed over several propositions (e.g., m_1 and m_2 over p_1 and p_2). This situation occurs when the evidence is <u>ambiguous</u>, pointing not to a single proposition, but to several.

$$SPT[p_i] = \begin{cases} 0, & p_1 \not\subseteq p_i,\ p_2 \not\subseteq p_i \\ m_1, & p_1 \subseteq p_i,\ p_2 \not\subseteq p_i \\ m_2, & p_1 \not\subseteq p_i,\ p_2 \subseteq p_i \\ m_1 + m_2, & p_1 \subseteq p_i,\ p_2 \subseteq p_i,\ p_i \neq \theta \\ 1, & p_i = \theta. \end{cases}$$

$$MASS[p_i] = \begin{cases} 0, & p_i \neq p_1,\ p_i \neq p_2,\ p_i \neq \theta \\ m_1, & p_i = p_1 \\ m_2, & p_i = p_2 \\ 1 - m_1 - m_2, & p_i = \theta. \end{cases}$$

However, if this represents an evidential situation, Shafer maintains that the following condition must hold:

$$\mathrm{MASS}[\bigcup_{\substack{\mathrm{MASS}[p]>0 \\ \& p\subseteq\theta}} p] > 0. \qquad (6.12)$$

This guarantees that all partially supported propositions have unequal support and plausibility assignments, reflecting the indeterminant nature of evidential information. If this condition does not hold, the function does not represent an evidential situation since it determines the confidences of some propositions with seemingly infinite precision. In that case it is called a pseudo-support function.

All Bayesian functions are pseudo-support functions (except in the trivial case where every proposition or its negation receives unit support). The Bayesian theory dictates that any support not attributed to a proposition is necessarily attributed to its negation. Therefore, the support and plausibility of every proposition are equal.

$$\mathrm{SPT}[p] + \mathrm{SPT}[\bar{p}] = 1; \qquad (6.13)$$

$$\mathrm{SPT}[p] = 1 - \mathrm{SPT}[\bar{p}];$$

$$\mathrm{SPT}[p] = \mathrm{PLS}[p].$$

This means that the entire unit of mass is distributed exclusively over the unit subsets of the frame of discernment.

$$\sum_{q\in\theta} \mathrm{MASS}[\{q\}] = 1. \qquad (6.14)$$

Every proposition's probability is precisely determined with seemingly infinite precision which is the kind of information that could only be provided by an onmipotent source. Otherwise there would have to be some admitted possibility of error; that is, there would have to be some factors associated with these estimates by which support and plausibility would differ.

Although Shafer rejects pseudo-support functions as being representative of evidential situations, his general theory of belief takes these into account. When perfect information is available (e.g., a Bayesian chance function), it can be quite gainfully employed. Dependency-graph models of evidential support (the models proposed in the next few sections of this chapter) preserve this ability of Shafer's general theory.

M_+^+ Models of Evidential Support

The freedom to express partial information within C^+ coverings and G_+^+ dependency graphs makes them a suitable host for Shafer's theory. This section describes the transition from the adoption of Shafer's theory to the adoption of a subset of M_+^+ dependency-graph models as models of evidential support. Throughout this section and the next the reader might wish to refer to the examples in the final section of this chapter (pp. 140-188).

Descriptions of evidential impact as C^+ coverings. According to Shafer's theory, the impact of a body of evidence on a proposition is summarized by its support and plausibility values. These can be interpreted as the lower and upper bounds delimiting the possible probabilities of a proposition[5]: the lower bound corresponding to Shafer's degree of support and the upper bound, Shafer's degree of plausibility. Thus, the impact of a body of evidence on a set of propositions can be represented as a C^+ covering that assigns each proposition a subinterval of the unit interval [0,1].

For $c_i = C^+[p_i]$,

$$MIN[c_i] = SPT[p_i],$$

$$MAX[c_i] = PLS[p_i],$$

$$c_i = \{c_i' \mid SPT[p_i] \le c_i' \le PLS[p_i],\ c_i' \in C\},$$

$$= [SPT[p_i],\ PLS[p_i]].$$

$$\underset{0}{[}\text{———}[SPT[p_i],\ PLS[p_i]]\text{———}\underset{1}{]}$$

The wider these subintervals are, the smaller the impact of the evidence, and the less known. In the extreme case of complete ignorance, where there is no support for any propositions other than the frame of discernment, the vacuous C^+ covering applies. When a proposition's probability is known with precision (i.e., its support and plausibility values are equal), it is properly assigned a degenerate subinterval collapsed about that probability

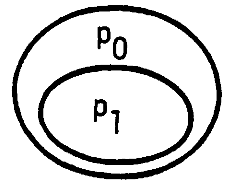
---

[5]This is the interpretation introduced by Dempster [Dempster 1976], later adapted by Shafer.

The interdependence of evidential support as G_+^+ dependency graphs. When propositions are interpreted as subsets of a frame of discernment, support relationships (by expression 6.7) are in direct (one-to-one) correspondence with set-theoretic relationships. In turn, set-theoretic relationships (e.g., inclusion, intersection, union, and complementation) have direct logical correlates (e.g., implication, conjunction, disjunction, and negation). Thus, any set-theoretic or logical relationship is directly translatable into a support relationship. In particular, any complete set of primitive set-theoretic or logical relationships translates into a complete set of primitive support relationships, from which G_+^+ dependency graphs, representing the interdependence of evidential support, can be constructed. Given a consistent set-theoretic or logical description of the relationships among a set of propositions, a consistent support relation follows.

Such a set of support relationships are here defined. Each of these definitions begin with two statements of the relationship being defined: the first is a logical statement and the second is a set-theoretic statement corresponding to the accompanying Venn diagram. This is followed by a drawing of the relationship's graphical representation (superset nodes appearing above subset nodes) and the definitions of its mappings. Each mapping is defined by two expressions: the first with a domain of continuous subintervals of [0,1], the lower and upper bounds corresponding to Shafer's measures of support and plausibility; the second with a domain of arbitrary subsets of confidence values from [0,1]. These latter definitions appeared in some of the previous examples (Figures 23 and 25) and are direct generalizations of the interval based mappings. Each definition is immediately followed by proofs that the intervals its mappings predict are true bounds on the intervals justified by Shafer's theory. These proofs rely on set theory and the previously introduced expressions of Shafer's theory (referenced by parenthetical expression numbers).

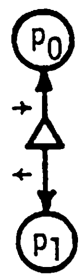
DEFINITION 21. IF support relationships.



IMPLICATION: $p_1 \rightarrow p_0$.

INCLUSION: $p_1 \subseteq p_0$.

IF:



$$\rightarrow [c_1] =_{def} [SPT[p_1],1],$$

$$=_{def} \{c_0' | c_1' \leq c_0',\ c_0' \in C,\ c_1' \in c_1\}.$$

$$\leftarrow [c_0] =_{def} [0,PLS[p_0]],$$

$$=_{def} \{c_1' | c_1' \leq c_0',\ c_1' \in C,\ c_0' \in c_0\}. \ \square$$

$$SPT[p_0] = SPT[p_1 \cup (p_0 - p_1)] \ :p_1 \subseteq p_0$$

$$\geq SPT[p_1] + SPT[p_0 - p_1] - SPT[p_1 \cap (p_0 - p_1)] \ :(6.7)$$

$$\geq SPT[p_1] + SPT[p_0 - p_1] - SPT[\phi]$$

$$\geq SPT[p_1] + SPT[p_0 - p_1] \ :(6.6)$$

$$\geq SPT[p_1] \ :(6.1)$$

$$PLS[p_0] \leq 1 \ :(6.1)\&(6.5)$$

Therefore, $\rightarrow[c_1] \supseteq [SPT[p_0],PLS[p_0]]$, where $c_1 = [SPT[p_1],PLS[p_1]]$. $\square$

$SPT[p_1] \geq 0$:(6.1)

$PLS[p_1] = 1 - SPT[\bar{p}_1]$:(6.5)

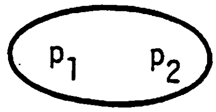
$\leq 1 - SPT[\bar{p}_0]$:$(\bar{p}_0 \subseteq \bar{p}_1)\&(SPT[\bar{p}_0] \leq SPT[\bar{p}_1])$

$\leq PLS[p_0]$:(6.5)

Therefore, $\leftarrow[c_0] \supseteq [SPT[p_1],PLS[p_1]]$, where $c_0 = [SPT[p_0],PLS[p_0]]$. □

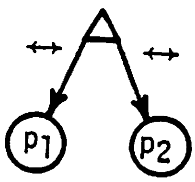
DEFINITION 22. IFF support relationships.

BIIMPLICATION: $p_1 \leftrightarrow p_2$.



EQUALITY: $p_1 = p_2$.

IFF:



$\leftrightarrow [c_1] =_{def} [SPT[p_i],PLS[p_i]],$

$=_{def} c_i$. □

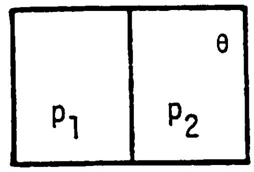
$SPT[p_i] = SPT[p_j] \quad :p_i = p_j$

$PLS[p_i] = PLS[p_j] \quad :p_i = p_j$

Therefore, $\leftrightarrow [c_i] = [SPT[p_j],PLS[p_j]]$, where $c_i = [SPT[p_i],PLS[p_i]]$,

$i,j \in \{1,2\}, i \neq j$. □

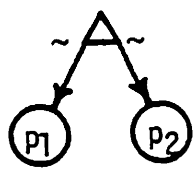
DEFINITION 23. NOT support relationships.



NEGATION: $\sim p_1 \leftrightarrow p_2$.

COMPLEMENTATION: $\bar{p}_1 = p_2 = \theta - p_1$.

NOT:



$\sim[c_i] =_{def} [1\text{-PLS}[p_i], 1\text{-SPT}[p_i]],$

$=_{def} \{c_j' \mid c_j' = 1 - c_i',\ c_i' \in c_i\}$. □

$\text{SPT}[p_i] = \text{DBT}[\bar{p}_i]$:(6.2)

$= \text{DBT}[p_j]$ $:\bar{p}_i = p_j$

$= 1 - \text{PLS}[p_j]$:(6.5)

$\text{PLS}[p_i] = 1 - \text{DBT}[p_i]$:(6.5)

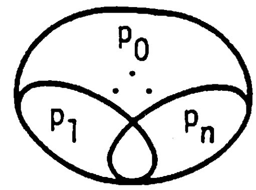
$= 1 - \text{SPT}[\bar{p}_i]$:(6.2)

$= 1 - \text{SPT}[p_j]$ $:\bar{p}_i = p_j$

Therefore, $\sim[c_i] = [\text{SPT}[p_j], \text{PLS}[p_j]]$, where $c_i = [\text{SPT}[p_i], \text{PLS}[p_i]]$,

$i, j \in \{1,2\},\ i \neq j$. □

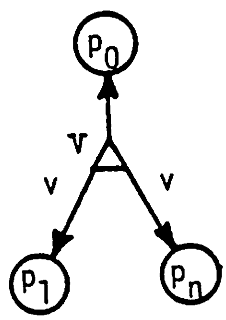
DEFINITION 24. OR support relationships.



DISJUNCTION; $p_0 \leftrightarrow \bigvee_{1\le i\le n} p_i$.

UNION: $p_0 = \bigcup_{1\le i\le n} p_i$.

OR:



$$V[c_1,\ldots,c_n]$$

$$=_{def} [\max_{1\le i\le n} [SPT[p_i]], MIN[1, \sum_{1\le i\le n} PLS[p_i]]],$$

$$=_{def} \{c_0' \mid \max_{1\le i\le n} [c_i'] \le c_0' \le \sum_{1\le i\le n} c_i',\ c_0' \in C,$$

$$c_i' \in c_i,\ 1\le i\le n\}.$$

$$v[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n]$$

$$=_{def} [MAX[0,SPT[p_0] - \sum_{\substack{1\le j\le n\\ i\ne j}} PLS[p_j]],PLS[p_0]],$$

$$=_{def} \{c_i' \mid (c_0' - \sum_{\substack{1\le j\le n\\ i\ne j}} c_j') \le c_i' \le c_0',\ c_i' \in C,$$

$$c_j' \in c_j,\ 0\le j\le n,\ i \ne j\}.\ \square$$

$SPT[p_0] \ge SPT[p_i],\ 1\le i\le n \quad :(p_i \subseteq p_0)\&(IF)$

$\ge \max_{1\le i\le n} [SPT[p_i]]$

$PLS[p_0] = PLS[\bigcup_{1\le i\le n} p_i] \quad :p_0 = \bigcup_{1\le i\le n} p_i$

$\le \sum_{1\le i\le n} PLS[p_i] \quad :1 \ge SPT[\bar{a} \cup \bar{b}] \quad :(6.1)$

$:1-SPT[\bar{a}] \geq SPT[\bar{a} \cup \bar{b}] - SPT[\bar{a}]$

$:PLS[a] \geq SPT[\bar{a} \cup \bar{b}] - SPT[\bar{a}]$:(6.5)

$:SPT[\bar{a} \cup \bar{b}] \geq SPT[\bar{a}] + SPT[\bar{b}] - SPT[\bar{a} \cap \bar{b}]$
:(6.7)

$:-SPT[\bar{a} \cap \bar{b}] \leq SPT[\bar{a} \cup \bar{b}] - SPT[\bar{a}] - SPT[\bar{b}]$

$:-SPT[\bar{a} \cap \bar{b}] \leq PLS[a] - SPT[\bar{b}]$:(above)

$:1-SPT[\bar{a} \cap \bar{b}] \leq PLS[a] + 1-SPT[\bar{b}]$

$:PLS[a \cup b] \leq PLS[a] + PLS[b]$

$PLS[p_0] \leq 1$:(6.1)&(6.5)

$PLS[p_0] \leq MIN[1, \sum_{1 \leq i \leq n} PLS[p_i]]$

Therefore, $V[c_1,\ldots,c_n] \geq [SPT[p_0],PLS[p_0]]$, where $c_j = [SPT[p_j],$
$PLS[p_j]]$, $1 \leq j \leq n$. □

$SPT[p_i] \geq SPT[\bigcup_{1 \leq j \leq n} p_j] - PLS[\bigcup_{\substack{1 \leq j \leq n \\ i \neq j}} p_j]$

$:SPT[a] \geq SPT[a \cap \bar{b}]$:$(a \supseteq a \cap \bar{b})$&(IF)

$: \geq SPT[(a \cup b) \cap \bar{b}]$

$: \geq SPT[a \cup b] + SPT[\bar{b}] - SPT[(a \cup b) \cup \bar{b}]$:(6.7)

$: \geq SPT[a \cup b] + SPT[\bar{b}] - SPT[\theta]$

$: \geq SPT[a \cup b] + SPT[\bar{b}] - 1$:(6.6)

$: \geq SPT[a \cup b] - PLS[b]$:(6.5)

$$\geq SPT[p_0] - \sum_{\substack{1 \leq j \leq n \\ i \neq j}} PLS[p_j] \quad :(\text{above})$$

$$SPT[p_i] \geq 0 \quad :(6.1)$$

$$SPT[p_i] \geq MAX[0, SPT[p_0] - \sum_{\substack{1 \leq j \leq n \\ i \neq j}} PLS[p_j]]$$

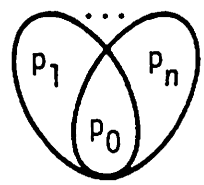
$$PLS[p_i] \leq PLS[p_0] \quad :(p_i \sqsubseteq p_0)\&(IF)$$

Therefore, $v[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n] \sqsupseteq [SPT[p_i],PLS[p_i]]$, $1 \leq i \leq n$,

where $c_j = [SPT[p_j],PLS[p_j]]$, $0 \leq j \leq n$. □

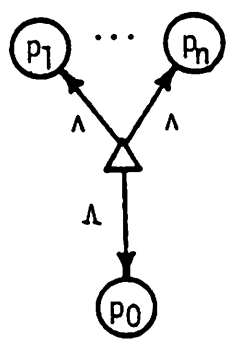
DEFINITION 25. AND support relationships.

CONJUNCTION: $p_0 \leftrightarrow \bigwedge_{1\le i\le n} p_i$.



INTERSECTION: $p_0 = \bigcap_{1\le i\le n} p_i$.

AND:



$$\wedge[c_1,\ldots,c_n]$$

$$=_{def}[\text{MAX}[0,1+\sum_{1\le i\le n}(\text{SPT}[p_i]-1)],$$

$$\min_{1\le i\le n}[\text{PLS}[p_i]]],$$

$$=_{def}\{c'_0 \mid 1+\sum_{1\le i\le n}(c'_i-1)\le c'_0\le \min_{1\le i\le n}[c'_i],\ c'_0\in C,$$

$$c'_i\in c_i,\ 1\le i\le n\}.$$

$$\wedge[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n]$$

$$=_{def}[\text{SPT}[p_0],$$

$$\text{MIN}[1,\text{PLS}[p_0]-\sum_{\substack{1\le j\le n\\ i\ne j}}(\text{SPT}[p_j]-1)]],$$

$$=_{def}\{c'_i \mid c'_0\le c'_i\le c'_0-\sum_{\substack{1\le j\le n\\ i\ne j}}(c'_j-1),$$

$$c'_i\in C,\ c'_j\in c_j,\ 0\le j\le n,\ i\ne j\}.\ \square$$

$$\text{SPT}[p_0] = \text{SPT}[\bigcap_{1\le i\le n} p_i] \quad :p_0=\bigcap_{1\le i\le n} p_i$$

$$\ge 1+\sum_{1\le i\le n}(\text{SPT}[p_i]-1)$$

$$:\text{PLS}[\bar{a}\cup\bar{b}]\le \text{PLS}[\bar{a}]+\text{PLS}[\bar{b}] \quad :(\text{OR})$$

$$:1-SPT[a \cap b] \leq (1-SPT[a]) + (1-SPT[b]) \quad :(6.5)$$

$$:SPT[a \cap b] \geq 1+(SPT[a]-1) + (SPT[b]-1)$$

$SPT[p_0] \geq 0 \quad :(6.1)$

$$SPT[p_0] \geq MAX[0,1 + \sum_{1 \leq i \leq n} (SPT[p_i] - 1)]$$

$PLS[p_0] \leq PLS[p_i],\ 1 \leq i \leq n \quad :(p_0 \subseteq p_i)\&(IF)$

$$PLS[p_0] \leq \underset{1 \leq i \leq n}{MIN} [PLS[p_i]]$$

Therefore, $\Lambda[c_1,\ldots,c_n] \supseteq [SPT[p_0],PLS[p_0]]$, where $c_j = [SPT[p_j],$ $PLS[p_j]],\ 1 \leq j \leq n.$ □

$SPT[p_i] \geq SPT[p_0] \quad :(p_0 \subseteq p_i)\&(IF)$

$$PLS[p_i] \leq PLS[\bigcap_{1 \leq j \leq n} p_j] - (SPT[\bigcap_{\substack{1 \leq j \leq n \\ i \neq j}} p_j] - 1)$$

$$:SPT[\bar{a}] \geq SPT[\bar{a} \cup \bar{b}]-PLS[\bar{b}] \quad :(OR)$$

$$:1-PLS[a] \geq 1-PLS[a \cap b]-1+SPT[b] \quad :(6.5)$$

$$:PLS[a] \leq PLS[a \cap b]-(SPT[b]-1)$$

$$\leq PLS[p_0] - (SPT[\bigcap_{\substack{1 \leq j \leq n \\ i \neq j}} p_j]-1) \quad :p_0 = \bigcap_{\substack{1 \leq j \leq n \\ i \neq j}} p_j$$

$$\leq PLS[p_0] - (1 + \sum_{\substack{1 \leq j \leq n \\ i \neq j}} (SPT[p_j]-1)-1) \quad :(above)$$

$$\leq PLS[p_0] - \sum_{\substack{1 \leq j \leq n \\ i \neq j}} (SPT[p_j]-1)$$

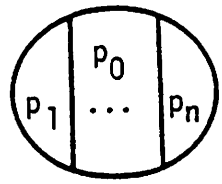
$PLS[p_i] \leq 1$:(6.1)&(6.5)

$$PLS[p_i] \leq MIN[1, PLS[p_0] - \sum_{\substack{1 \leq j \leq n \\ i \neq j}} (SPT[p_j]-1)]$$

Therefore, $\wedge[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n] \supseteq [SPT[p_i],PLS[p_i]]$, $1 \leq i \leq n$,

where $c_j = [SPT[p_j],PLS[p_j]]$, $0 \leq j \leq n$. □

DEFINITION 26. XOR support relationships.



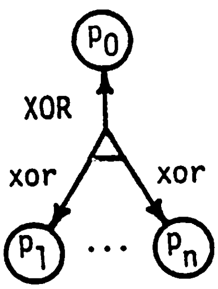
EXCLUSIVE DISJUNCTION:

$$(p_0 \leftrightarrow \bigvee_{1\le i\le n} p_i) \wedge \sim(\bigvee_{\substack{1\le i,j\le n\\ i\ne j}} (p_i \wedge p_j)).$$

DISJOINT UNION:

$$p_0 = \bigcup_{1\le i\le n} p_i,\ \phi = \bigcup_{\substack{1\le i,j\le n\\ i\ne j}} (p_i \cap p_j).$$

XOR:



$$\mathrm{XOR}[c_1,\ldots,c_n]$$

$$=_{def}[\sum_{1\le i\le n} \mathrm{SPT}[p_i],\ \mathrm{MIN}[1, \sum_{1\le i\le n} \mathrm{PLS}[p_i]]],$$

$$=_{def}\{c_0' \mid c_0' = \sum_{1\le i\le n} c_i',\ c_0' \in C,\ c_i' \in c_i,\ 1\le i\le n\}.$$

$$\mathrm{xor}[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n]$$

$$=_{def}[\mathrm{MAX}[0,\ \mathrm{SPT}[p_0] - \sum_{\substack{1\le j\le n\\ i\ne j}} \mathrm{PLS}[p_j]],$$

$$\mathrm{PLS}[p_0] - \sum_{\substack{1\le j\le n\\ i\ne j}} \mathrm{SPT}[p_j]],$$

$$=_{def}\{c_i' \mid c_i' = c_0' - \sum_{\substack{1\le j\le n\\ i\ne j}} c_j',\ c_i' \in C,\ c_j' \in c_j,$$

$$0\le j\le n,\ i \ne j\}.\ \square$$

$$SPT[p_0] = SPT[\bigcup_{1\le i\le n} p_i] \quad :p_0 = \bigcup_{1\le i\le n} p_i$$

$$\ge \sum_{1\le i\le n} SPT[p_i] - \sum_{\substack{1\le i,j\le n \\ i\ne j}} SPT[p_i \cap p_j] + \cdots \quad :(6.7)$$

$$\ge \sum_{1\le i\le n} SPT[p_i] - \Sigma\, SPT[\phi] + \cdots \quad :p_i \cap p_j = \phi$$

$$\ge \sum_{1\le i\le n} SPT[p_i] \quad :(6.6)$$

$$PLS[p_0] \le MIN[1, \sum_{1\le i\le n} PLS[p_i]] \quad :(OR)$$

Therefore, $XOR[c_1,\ldots,c_n] \supseteq [SPT[p_0],PLS[p_0]]$, where $c_j = [SPT[p_j], PLS[p_j]]$, $1\le j\le n$. □

$$SPT[p_i] \ge MAX[0,SPT[p_0] - \sum_{\substack{1\le j\le n \\ i\ne j}} PLS[p_j]] \quad :(OR)$$

$$PLS[p_i] = 1 - SPT[\bar{p}_i] \quad :(6.5)$$

$$= 1 - SPT[\bar{p}_0 \cup \bigcup_{\substack{1\le j\le n \\ i\ne j}} p_j] \quad :\bar{p}_i = \bar{p}_0 \cup \bigcup_{\substack{1\le j\le n \\ i\ne j}} p_j$$

$$\le 1 - SPT[\bar{p}_0] - SPT[\bigcup_{\substack{1\le j\le n \\ i\ne j}} p_j] + SPT[\bar{p}_0 \cap \bigcup_{\substack{1\le j\le n \\ i\ne j}} p_j] \quad :(6.7)$$

$$\leq PLS[p_0] - SPT[\bigcup_{\substack{1\leq j\leq n\\ i\neq j}} p_j] + SPT[\phi]$$

$$:(\phi = \bar{p}_0 \cap \bigcup_{\substack{1\leq j\leq n\\ i\neq j}} p_j)\&(6.5)$$

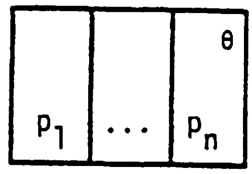
$$\leq PLS[p_0] - SPT[\bigcup_{\substack{1\leq j\leq n\\ i\neq j}} p_j] \quad :(6.6)$$

$$\leq PLS[p_0] - \sum_{\substack{1\leq j\leq n\\ i\neq j}} SPT[p_j] \quad :(\text{above})$$

Therefore, $xor[c_0,\ldots,c_{i-1},c_{i+1},\ldots,c_n] \supseteq [SPT[p_i],PLS[p_i]]$, $1\leq i\leq n$,

where $c_j = [SPT[p_j],PLS[p_j]]$, $0\leq j\leq n$. □

DEFINITION 27. X support relationships.



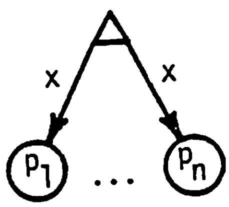
EXCLUSION: $\left(\bigvee_{1\le i\le n} p_i\right) \wedge \sim\left(\bigvee_{\substack{1\le i,j\le n\\ i\ne j}} (p_i \wedge p_j)\right)$.

PARTITION: $\theta = \bigcup_{1\le i\le n} p_i,\ \phi = \bigcup_{\substack{1\le i,j\le n\\ i\ne j}} (p_i \cap p_j)$.

X: $x[c_1,\ldots,c_{i-1},c_{i+1},\ldots,c_n]$,

$$=_{def}[\mathrm{MAX}[0,\ 1- \sum_{\substack{1\le j\le n\\ i\ne j}} \mathrm{PLS}[p_j]],$$

$$1 - \sum_{\substack{1\le j\le n\\ i\ne j}} \mathrm{SPT}[p_j]],$$



$$=_{def}\{c_i' \mid c_i' = 1 - \sum_{\substack{1\le j\le n\\ i\ne j}} c_j',\ c_i' \in C,$$

$$c_j' \in c_j,\ 1\le j\le n,\ i \ne j\}. \square$$

$x[c_1,\ldots,c_{i-1},c_{i+1},\ldots,c_n]$

$= \mathrm{xor}[[1,1],c_1,\ldots,c_{i-1},c_{i+1},\ldots,c_n]$

$:(\theta = \bigcup_{1\le i\le n} p_i),\ (\mathrm{SPT}[\theta] = \mathrm{PLS}[\theta] = 1):(6.6),\ \&(\mathrm{XOR})$

Therefore, $x[c_1,\ldots,c_{i-1},c_{i+1},\ldots,c_n] \supseteq [\mathrm{SPT}[p_i],\mathrm{PLS}[p_i]],\ 1\le i\le n,$

where $c_j = [\mathrm{SPT}[p_j],\mathrm{PLS}[p_j]],\ 1\le j\le n.\ \square$

The correspondence among set-theory, propositional logic, and Shafer's theory eases the transition from an evidential domain to its representation as a G_+^+ dependency graph. A dependency graph can be constructed using familiar set-theoretic or logical relationships as the base components. The appropriate support mappings then can be overlaid, resulting in a G_+^+ dependency graph representing evidential support. If a set-theoretic description of the interrelationships among a set of propositions is internally consistent (i.e., the propositions can be identified, relative to some frame of discernment, with the stated set-theoretic relationships), then the corresponding evidential dependency-graph is guaranteed to be internally consistent. This follows since all of the support relationships are specializations of the same dependency-relation schemata, which is defined by expressions 6.6 and 6.7, and proven by the previous derivations.[5] The consistency conditions of dependency graphs guarantee that redundant descriptions are compatible, a condition that is obviously satisfied when there is a single relational statement from which they all follow. So long as the set-theoretic information does not interject inconsistencies into these specializations, they are consistent.

Constructing M_+^+ models of evidential support. The construction of an M_+^+ model of evidential support might begin with an assessment of the

---

[5]The other expressions used in those derivations either follow directly from 6.6 and 6.7 (e.g., 6.1), or simply rename support (e.g., 6.2 and 6.5).

environment to determine the propositions of interest. These should include both those propositions whose truthfulness is of ultimate interest, and those that the available knowledge sources can make statements about. If each knowledge source responds to a different feature of the environment, each might require a distinct set of propositions relative to the feature space it explores. Additionally, some intermediate propositions might be included to better, or more easily, describe the interdependencies that exist among these propositions of primary interest.

The next step is to determine and represent the interrelationships among these propositions. This might be accomplished by the following steps: relating all of the propositions to a frame of discernment; discovering their set-theoretic relationships relative to this frame; describing these relationships in terms of set-theoretic primitives; translating these into their corresponding support primitives; and finally, representing all as a G_+^+ dependency graph.

As usual, the more fully the interrelationships of the propositions are described in a G_+^+ dependency graph, the better the M_+^+ model. For example, if a set of propositions are related by an OR support relationship in a model, where an XOR would be appropriate, the model may not be as informative as it might. Either way the predictions are consistent; it is their precision, not their accuracy, that is in question.

<u>Inferential reasoning from M_+^+ models of evidential support</u>. As previously described, knowledge sources communicate with M_+^+ models via $\hat{C}^+$

coverings. These are coverings that assign each proposition a set of confidence values bounding a set assigned by a C^+ covering. In the case of M_+^+ models of evidential support, $\hat{C}^+$ coverings assign subintervals of [0,1], the lower bound serving as a lower limit on the support accorded a proposition (thereby a lower limit on its plausibility), the upper bound serving as an upper limit on its plausibility (thereby an upper limit on its support).

$$\begin{aligned} \hat{SPT}[p] \le SPT[p] \le PLS[p] \le \hat{PLS}[p], \\ [\hat{SPT}[p], \hat{PLS}[p]] = \hat{C}^+[p], \\ [SPT[p], PLS[p]] = C^+[p]. \end{aligned} \quad (6.15)$$

In other words, a $\hat{C}^+$ covering for an evidential model is a <u>conservative</u> estimate of the impact of a body of evidence. Support may be understated, but not overstated. And from this it follows that plausibility may be overstated, but not understated.

$$\begin{aligned} \hat{SPT}[\bar{p}] \le SPT[\bar{p}]; \\ 1 - \hat{PLS}[p] \le 1 - PLS[p]; \\ PLS[p] \le \hat{PLS}[p]. \end{aligned} \quad (6.16)$$

Given a $\hat{C}^+$ covering which represents the estimated impact of a body of evidence discovered by a knowledge source, an E_+^+ inference engine can refine this inofrmation to a C^+ covering (Figure 25). The accuracy of these predictions depends on the accuracy of the knowledge source and the accuracy of the dependency graph. In particular, the description of the environment given by the dependency graph and the understanding of the environment by the knowledge source need to be consistent. As previously discussed, the absence of a consistent view of the environment can lead to model breakdown. This condition occurs when a knowledge source injects the model with a noncovering.

A partial solution to this problem is for a knowledge source to be more conservative in its estimates, reducing the possibility of contradiction. But the problem remains.

Clearly, for the best results, a knowledge source needs to be consistent with its dependency graph, and both must be consistent with the environment they model. However, if there are only minor inconsistencies, complete model breakdown can be avoided and relatively satisfactory predictions made, if the knowledge source expresses its information in terms of mass (instead of support and plausibility). That is, the model can be guaranteed to produce well-formed C^+ coverings through a slightly different inference technique, approximating those coverings that would follow if the knowledge source and model were mutually consistent. As will be seen in the next section, mass is also the preferred means of expression when combining information from several knowledge sources.

Given a mass function, there are three possibilities: vacuous support, simple support, or ambiguous support (pp. 102-103). Vacuous support and simple support lead directly to $\hat{C}^+$ coverings, guaranteeing that the standard inference technique can be applied without fear of model breakdown. In the case of vacuous support, the full unit of mass is attributed to the frame of discernment, it being the only proposition with non-zero support. This situation is straightforwardly represented by the vacuous C^+ covering, a fixed point for any E_+^+ inference engine.

For simple support, the unit of mass is distributed over one proposition p and (possibly) the frame of discernment. This is represented by a $\hat{C}^+$ covering that associates [MASS[p],1] with p and the full unit interval with all other propositions. This follows since the support for p is equal to the mass attributed it, and it is completely plausible. Application of an E_+^+ inference engine (potentially) increases the support of those propositions implied by p to MASS[p], and decreases the plausibility of those propositions whose negations are implied by p to 1 - MASS[p]. This follows from expression 6.7 and the assumption that the dependency graph is set-theoretically consistent i.e., all propositions and their negations are nonintersecting.

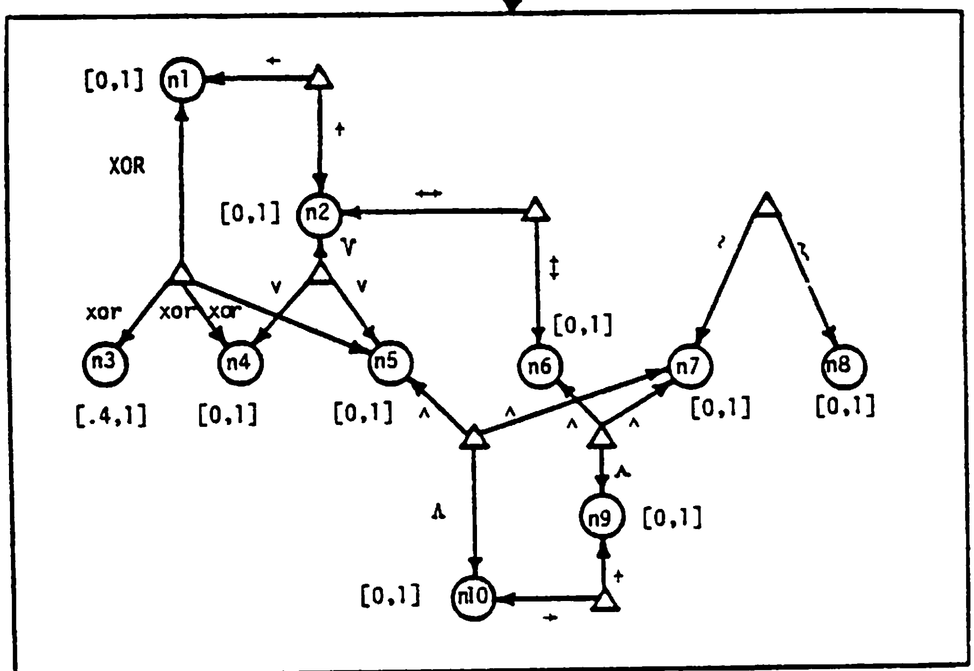
When support is ambiguous because mass is distributed across several (focal) propositions, the support attributed each proposition is equal to the sum of the mass attributed it and any propositions that imply it (expression 6.10). Consider a mass function with focal propositions $p_1,\ldots,p_n$ and (possibly) the frame of discernment. For each proposition p_i, a $\hat{C}^+$ covering can be initialized and refined to a C^+ covering, just as if p_i were the focus of a simple support function attributing MASS[p_i] to p_i and 1 - MASS[p_i] to the frame of discernment. Doing so for each p_i produces n C^+ coverings, each describing the direct and indirect support provided by a portion of the mass. The support committed to any proposition, based on all of the mass, is the total support attributed it in these n coverings. Remembering that the support for the negation of a proposition is the amount its plausibility differs from one, these dubiety measures can be similarly summed to determine the total support for the negation of each proposition, and then translated back into plausibility measures. Since no single portion of mass can provide simultaneous support for a proposition and its negation (assuming that the dependency graph is set-theoretically consistent and the previously described technique for simple support has been employed), the total support for a proposition and its negation is bounded by one, assuring that the support for any proposition is less than or equal to its plausibility. Thus, model breakdown cannot occur! And from expression 6.10 it follows that the result is always a $\hat{C}^+$ covering, and depending on the G_+^+ graph, it may be a C^+ covering. Either way, one final application of a complete E_+^+ inference engine ensures a C^+ covering (Figure 25).

Although this technique has the effect of smoothing over inconsistencies, it does not eliminate them. It merely relocates their ill effects in the interface between the model and the knowledge source. Since the inconsistencies are manifest externally, model breakdown is precluded. When a knowledge source provides information, it does so in the language of the model. If the support/plausibility dialect is utilized, partial information about the knowledge source's understanding of the interrelationships among the propositions is inadvertently communicated. This information is implicit. For example if the support provided one proposition p_i is greater than the plausibility provided p_j, then p_i must not imply p_j. If that is incompatible with the

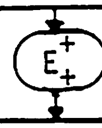
Figure 25. An example application from a mass function.

$$
\text{MASS}[n] = \begin{cases} .4, & n = n3 \\ .2, & n = n5 \\ .3, & n = n10 \\ .1, & n = \theta \\ 0.00, & \text{else.} \end{cases}
$$

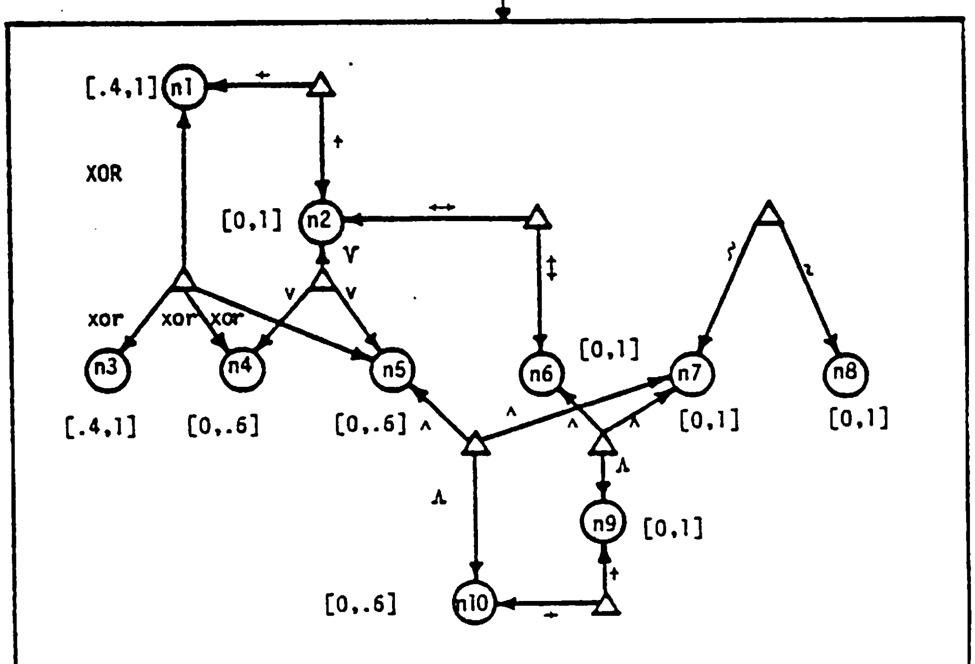
MASS[n3] = .4,



[0,1] n1
XOR
[0,1] n2
xor
xor
xor
n3
n4
n5
n6
n7
n8
[.4,1]
[0,1]
[0,1]
[0,1]
[0,1]
[0,1]
[0,1]
n9 [0,1]
[0,1] n10

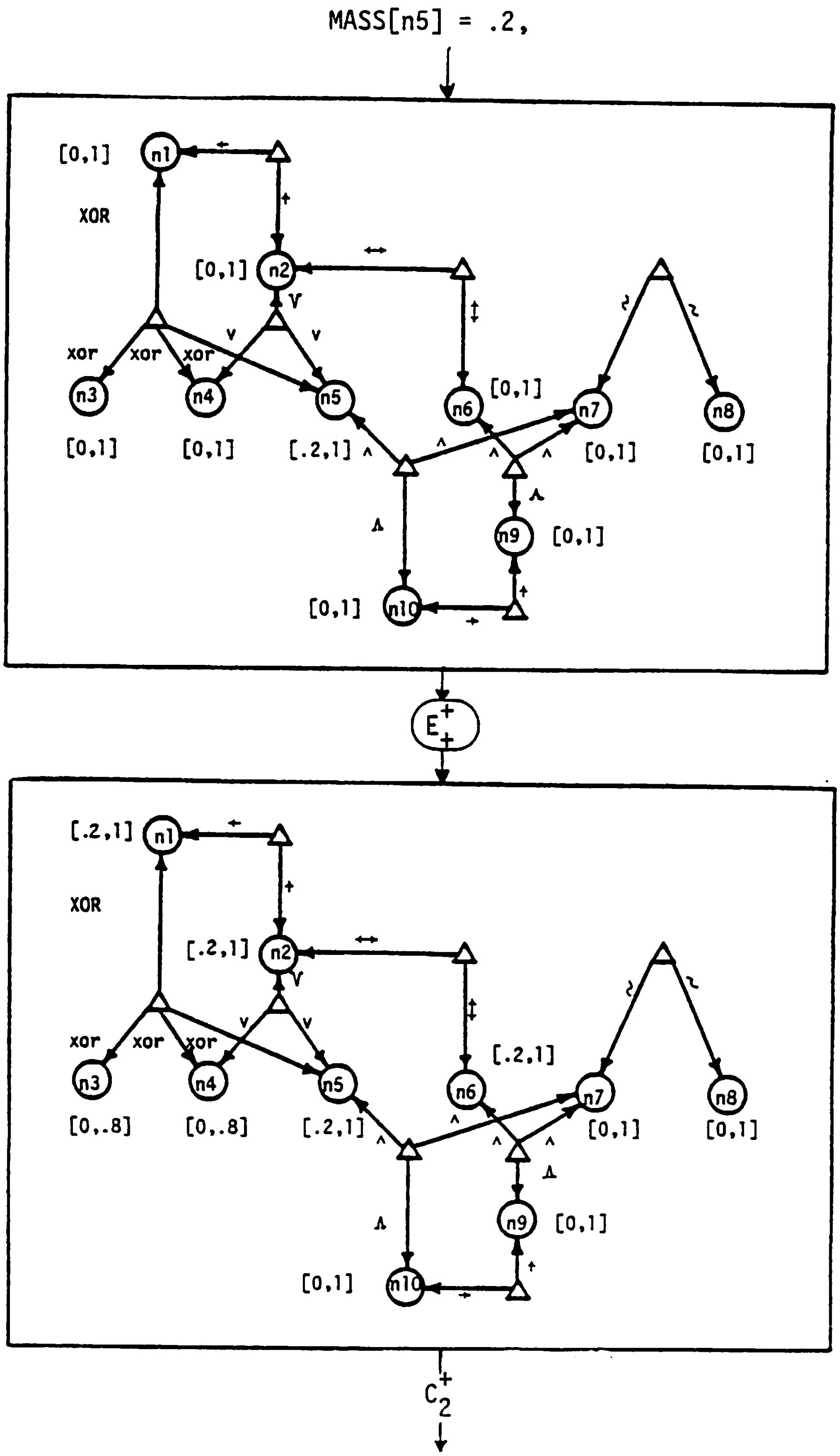


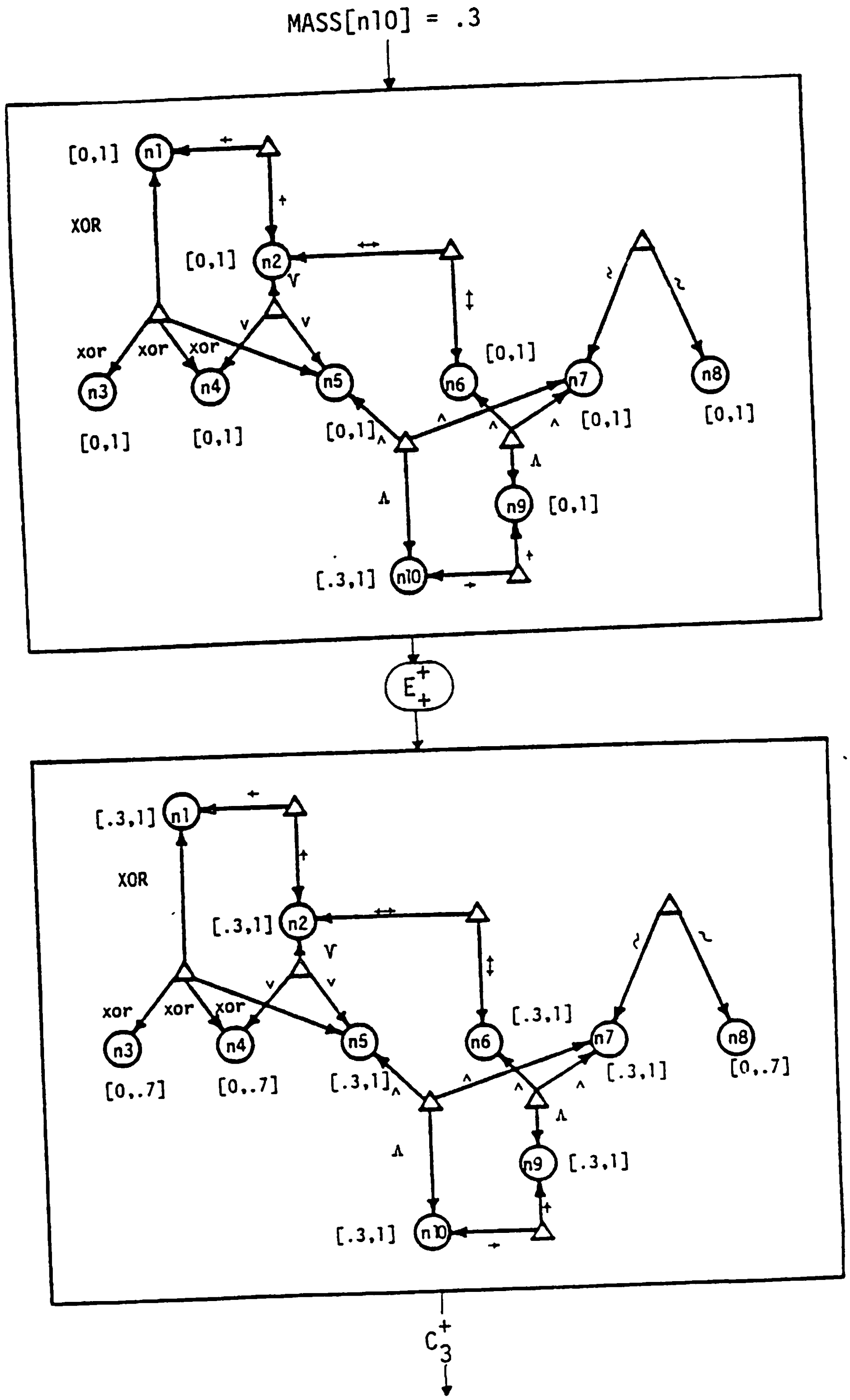
E

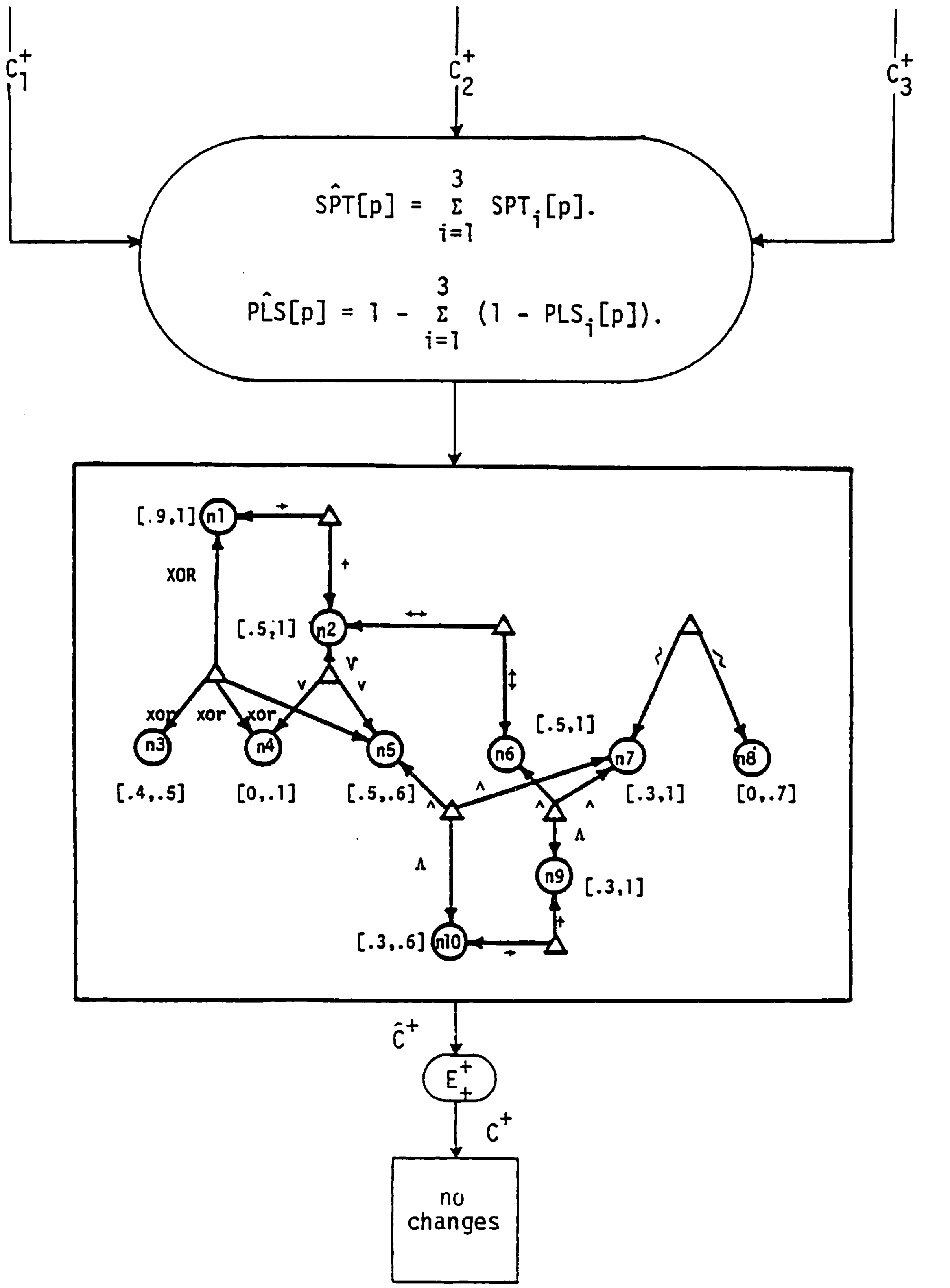


[.4,1] n1
XOR
[0,1] n2
xor
xor
xor
n3
n4
n5
n6
[0,1]
n7
n8
[.4,1]
[0,.6]
[0,.6]
[0,1]
[0,1]
n9 [0,1]
[0,.6] n10

C_1^+


MASS[n5] = .2,
[0,1]
n1
XOR
[0,1]
n2
xor
xor
xor
n3
n4
n5
n6
[0,1]
n7
n8
[0,1]
[0,1]
[.2,1]
[0,1]
[0,1]
n9
[0,1]
[0,1]
n10
E
[.2,1]
n1
XOR
[.2,1]
n2
xor
xor
xor
n3
n4
n5
n6
[.2,1]
n7
n8
[0,..8]
[0,..8]
[.2,1]
[0,1]
[0,1]
n9
[0,1]
[0,1]
n10
C


MASS[n10] = .3
[0,1] n1
XOR
[0,1] n2
xor xor xor
n3 n4 n5 n6 n7 n8
[0,1] [0,1] [0,1] [0,1] [0,1] [0,1] [0,1]
n9 [0,1]
[.3,1] n10
E
[.3,1] n1
XOR
[.3,1] n2
[0,.7] [0,.7] [.3,1] [.3,1] [.3,1] [0,.7]
n9 [.3,1]
[.3,1] n10
C
3



C_1^+
C_2^+
C_3^+
$\hat{SPT}[p] = \sum_{i=1}^{3} SPT_i[p].$
$\hat{PLS}[p] = 1 - \sum_{i=1}^{3} (1 - PLS_i[p]).$
[.9,1] n1
XOR
[.5,1] n2
xor
xor
xor
v
v
v
n3
n4
n5
n6
n7
n8
[.4,.5]
[0,.1]
[.5,.6]
[.5,1]
[.3,1]
[0,.7]
n9 [.3,1]
[.3,.6] n10
$\hat{C}^+$
E_+^+
C^+
no changes

model (i.e., p_i does imply p_j), it is a noncovering and may lead to model breakdown. When a knowledge source provides information in the mass dialect, no relational information is communicated. Since a mass function makes no relational statements, it cannot refute those in the model. The model interprets the information according to its understanding of the world, irrespective of the knowledge source's understanding. If the two are incompatible, the information is garbled during transmission. A C^+ covering is produced, but it just doesn't correspond to what the knowledge source was attempting to convey. If the differences are small, this presents no major problem. If the differences are large, the results are largely meaningless.

Before moving on to an example M_+^+ model of evidential support and its application, the next section describes Dempster's rule of combination, a rule for pooling distinct bodies of evidence.

Pooling Distinct Bodies of Evidence

A dependency-graph model of evidential support takes a single body of evidential information, expressed in terms of support/plausibility or mass, and extends it through inferential reasoning. This translates the information from those propositions the evidence directly bears upon, to those it indirectly bears upon. If the source of this evidential information is unreliable, so are the resulting predictions. This is the typical situation in artificial intelligence applications. The knowledge sources, being synthetic entities themselves, are unreliable i.e., they are prone to occasional errors. However, individual inaccuracies can be overcome when multiple independent opinions are available; one opinion is independent of another if their error likelihoods are unrelated. Pooling such a set of opinions will generally result in more accurate predictions. Dempster's rule of combination provides a formal foundation for this process in the context of dependency-graph models of evidential support. The next two sections describe Dempster's rule and its application within this context.

Dempster's rule of combination. Dempster's rule of combination [Dempster 1967, 1968; Shafer 1976] is a rule for pooling distinct bodies of evidential information. It is most easily described in terms of mass. Given two mass functions, representing two independent bodies of evidence, Dempster's rule produces a third mass function, representing the consensus of those disparate opinions. It is both commutative and associative, which leads to the clearly desirable property that any number of opinions can be combined, in whatever order is most convenient, and the result is guaranteed to be the same.

Mathematically, Dempster's rule is the orthogonal sum. Given two mass functions $MASS_1$ and $MASS_2$, their orthogonal sum, denoted $MASS_1 \oplus MASS_2$, is defined as follows:

$$MASS_1 \oplus MASS_2[p] = N \cdot \sum_{\substack{i,j \\ a_i \cap b_j = p}} MASS_1[a_i] \cdot MASS_2[b_j] \qquad (6.17)$$

where $N = (1 - k)^{-1}$,

$$k = \sum_{\substack{i,j \\ a_i \cap b_j = \phi}} MASS_1[a_i] \cdot MASS_2[b_j],$$

$$k < 1.$$

A partitioned unit square depicts this computation (Figure 26). The horizontal strips correspond to the mass $MASS_1$ attributes each of its focal propositions; the vertical strips similarly correspond to $MASS_2$. For example, a horizontal strip of measure $MASS_1[a_i]$ is committed to proposition a_i by $MASS_1$, and a vertical strip of measure $MASS_2[b_j]$ is committed to proposition b_j by $MASS_2$. The area of the rectangle at the intersection of these two strips, $MASS_1[a_i] \cdot MASS_2[b_j]$, is committed to a proposition p by $MASS_1 \oplus MASS_2$, where p is equivalent to $a_i \cap b_j$. Several of these rectangles may be committed to that same proposition, increasing the total area of the unit square committed to p. This accounts for the summation in expression 6.17, leaving only the normalization factor N to be interpreted.

It may be that propositions a_i and b_j are nonintersecting in the frame of discernment. If this is the case, there is no proposition p, equivalent to $a_i \cap b_j$, with which $MASS_1 \oplus MASS_2$ can associate the measure of mass $MASS_1[a_i] \cdot MASS_2[b_j]$. The sum of all such mass, denoted by k in expression 6.17, is a measure of the conflict in the combination. If this accounts for the entire unit square, the combination does not exist. Otherwise, k is proportionally redistributed over those propositions in the unit square that do exist; hence the normalization factor N.

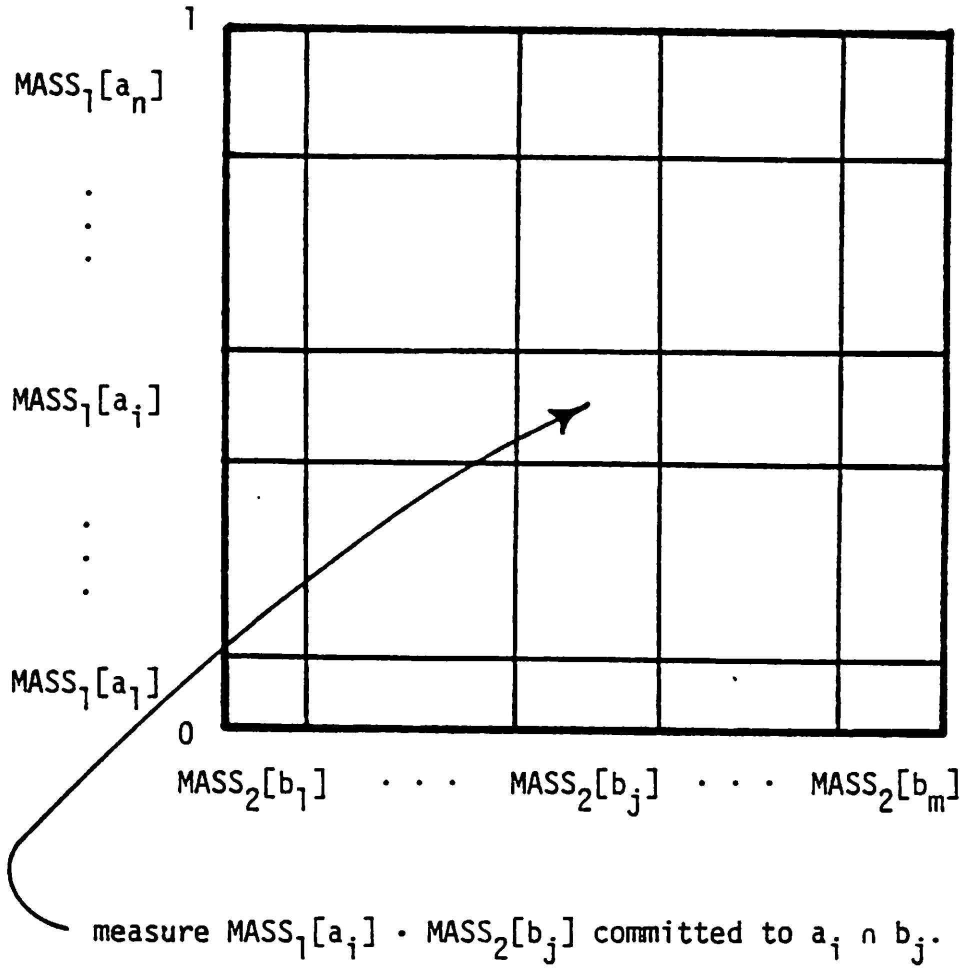


Figure 26. Depiction of Dempster's rule as applied to two mass functions $MASS_1$ and $MASS_2$.

Intuitively, Dempster's rule moves mass towards more specific propositions commensurate with both bodies of evidence, and away from propositions incompatible with either body of evidence. If both of the mass functions are Bayesian functions (p. 104), one of which attributes its full unit of mass to a single proposition a and the other distributes its mass over the propositions $b_1, \ldots, b_n$, then Dempster's rule has the same effect as Bayes' rule of conditional probabilities $P(b_i|a)$ [Shafer 1976]. However, when evidence cannot be expressed as a certainty, then Dempster's rule, unlike Bayes' rule, still applies. The only time Dempster's rule is not applicable is when the given bodies of evidence are completely contradictory.

Multiple bodies of evidence and M_+^+ models. Dempster's rule depends on the initial mass functions representing the bodies of evidence to be combined and the pairwise intersections of their focal propositions relative to the frame of discernment. These intersections can be directly determined from the information in an M_+^+ model of evidential support. As was previously described, a G_+^+ dependency graph representing evidential support can be interpreted in terms of set-theoretic relationships. Set intersection algorithms, over such graphical representations of set-theoretic relationships, have been previously developed [McSkimin 1976].

The only problem is that the dependency graph may not contain enough information to uniquely and completely determine some of these intersections. For example, there might not be a proposition in the graph that corresponds to a given intersection. When this occurs, Dempster's rule does not have the appropriate proposition with which to associate its mass product. This could be avoided if all pairwise intersections were explicitly included in the graph, along with enough dependency information to conclusively designate them as exactly the intersections they represent, but this seems far too stringent a requirement. Instead, the effect of Dempster's rule can be approximated by evenly distributing each mass product over those propositions in the graph that collectively best approximate the appropriate intersection. The approximation used includes those propositions that are mutual subsets of both propositions, but are not themselves subsets of any of the other propositions in the approximation. When no mutual subsets can be found, the propositions are assumed to be disjoint and their mass product attributed to k. The effect of these approximations is for Dempster's rule to overzealously jump towards some conclusions, when the graph lacks the appropriate information. If this causes a problem, the graph can be expanded by its designer, permitting those intersections most crucial to the model's accuracy, to be more precisely determined by the model.

Given several distinct bodies of evidence, each expressed as a mass function over an M_+^+ model of evidential support, Dempster's rule can be repeatedly applied based on that model, until a single mass function, representing the combination of all of that evidence, is produced. At that point this mass function can be fed to the model, the inference engine applied, and a C^+ covering, representing a support and plausibility estimate for each proposition, returned. This procedure is guaranteed to produce such a covering, so long as the original bodies of evidence share some common beliefs (i.e., at no time during the application of Dempster's rule is k equal to one). When a group of knowledge sources can be assumed to be operating independently of one another, their results can be pooled and interpreted in this way (Figure 27).

An Example Dependency-Graph Model of Evidential Support and Its Application

This section presents an example dependency-graph model of evidential support, and demonstrates how it can be applied to make predictions based on multiple sources of evidential information. The model domain is a simple one. Although a more complex example might have better demonstrated the sophistication of these techniques, it probably would not have been as instructive. The chosen example concerns the identification of convex, regular polygons based on evidential information about their characteristic features. In particular, there are eight nonoverlapping possible identifications: oblique kite, isoceles kite, oblique trapezoid, isoceles trapezoid, parallelogram, rhombus, rectangle, and square; and there are six feature spaces: the number and relative position of equal angles, the number and relative position of equal sides, the number of parallel sides, the equality of the diagonals, whether or not the diagonals intersect at right angles, and the number of bisecting diagonals. This information is summarized in Table

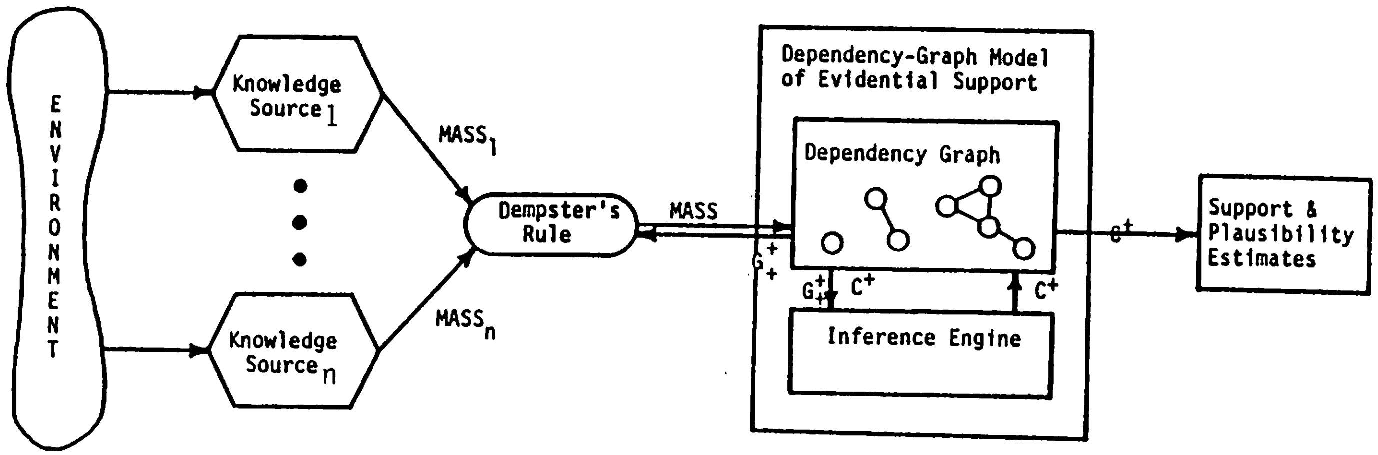


Figure 27. Dependency-graph model of evidential support: system architecture.

1. Six independent knowledge sources are assumed, one operating over each of these feature spaces. The hypothetical task is the identification of a presented polygon, based on the evidential information provided by these knowledge sources after having examined it.

Modelling begins with the determination of a frame of discernment. The key requirement is that all of the propositions of interest be in correspondence with its subsets. This includes those propositions of ultimate interest (i.e., those that identify a polygon as one of the eight generic types) and those included in the vocabularies of the knowledge sources. For the problem at hand, a frame of discernment is sufficient whose elements each consist of a polygonal-type paired with a feature vector representing one possible combination of features such a polygon could exhibit. Thus θ is a subset of the cross product of polygonal-types (PT), equal-angles (EA), equal-sides (ES), parallel-sides (PS), equal-diagonals (ED), right-diagonals (RD), and bisecting-diagonals (BD), where each of these feature sets consists of the entries found in the similarly named column of Table 1.

$$\theta \subset PT \times EA \times ES \times PS \times ED \times RD \times BD.$$

Every element of θ corresponds to a row in the table. There is exactly one element for each row, except the oblique kite and oblique trapezoid rows. These each have four associated elements, one for each possible combination of the listed choices for equal-angles and equal-sides. Within the context of an oblique kite or oblique trapezoid, the equal-angles feature is independent of the equal-sides feature (i.e., knowing

POLYGONS \ FEATURES	EQUAL ANGLES	EQUAL SIDES	PARALLEL SIDES	EQUAL DIAGONALS	RIGHT DIAGONALS	BISECTING DIAGONALS
OBLIQUE KITE	0 or 1 adj. pair	0 or 1 adj. pair	0	no	no	1
ISOCELES KITE	2 opp. pairs	2 adj. pairs	0	no	yes	1
OBLIQUE TRAPEZOID	0 or 1 adj. pair	0 or 1 adj. pair	1 pair	no	no	0
ISOCELES TRAPEZOID	2 adj. pairs	1 opp. pair	1 pair	yes	no	0
PARALLELOGRAM	2 opp. pairs	2 opp. pairs	2 pairs	no	no	2
RHOMBUS	2 opp. pairs	4	2 pairs	no	yes	2
RECTANGLE	4	2 opp. pairs	2 pairs	yes	no	2
SQUARE	4	4	2 pairs	yes	yes	2

Table 1. Regular convex polygons and some of their characteristic features.

one doesn't allow the other to be predicted with any greater precision), therefore all of the combinations are included.

The next step is to construct a dependency graph interrelating the propositions of interest with respect to this frame of discernment. Beginning with the response propositions, there are the eight identifying polygonal-types. The most important relationship among these is that they are all nonoverlapping in θ. An X relationship over a set of nodes, one for each element of PT, describes this. Six of these nodes can be identified as primitive elements of θ. The other two, OKT and OTR (i.e., oblique kite and oblique trapezoid), can be identified with four primitive elements each. Since the number of primitive elements is so small, they can all be represented: the six indirectly by their associated PT nodes, the remaining eight directly. Two XOR relationships, relating OKT and OTR to their identifying primitive elements, fill out this representation (Figure 28). Other nodes and relationships might be included. For example, some nodes representing disjunctions of these polygonal-types might be added along with the appropriate X, XOR, and OR relationships. In general, so long as any additional nodes and relationships are consistent with this frame of discernment, they can be included. However, no other propositions are of direct interest to this hypothetical task, and thus no others are included.

The next step is to represent the propositions in the vocabularies of the knowledge sources. It is assumed that these include the entries in the columns of Table 1 along with some other disjunctive proposi-

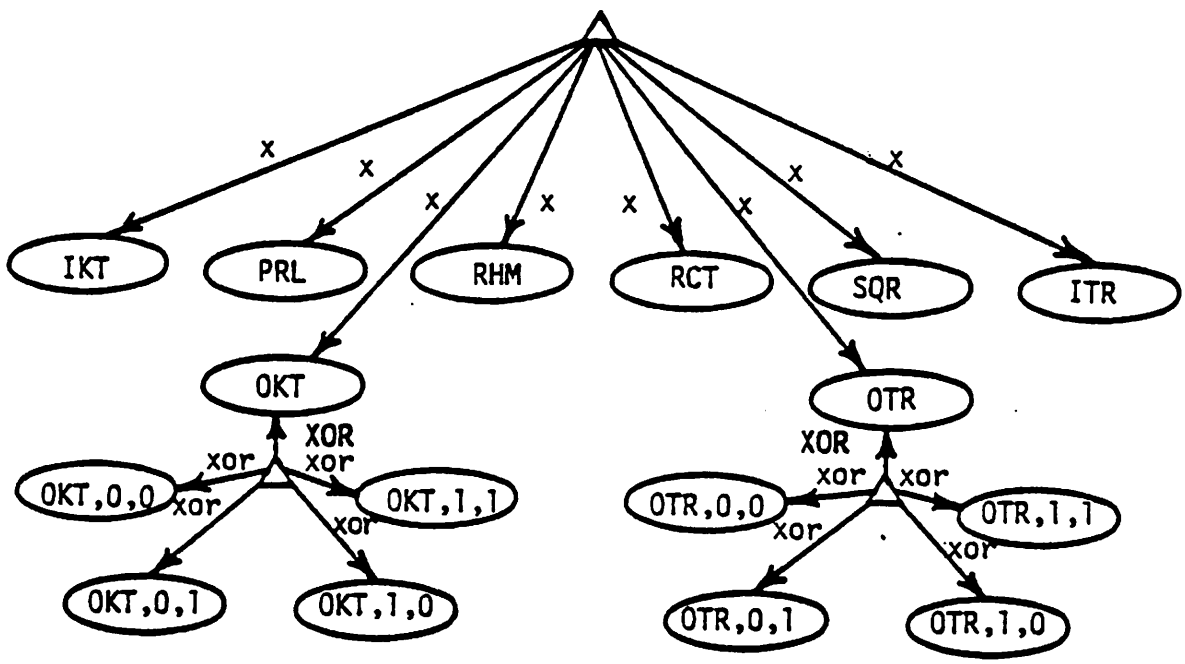


Figure 28. Polygonal-types (PT) sub-graph.

tions. Simple hierarchies adequately represent the interrelationships among the propositions within each knowledge source's vocabulary (Figures 29-34). This is not to say that more complex, tangled hierarchies could not be used. For example, the three subgraphs describing various aspects of diagonals could be interwoven as in Figure 35. Again, so long as the integrity of the underlying frame of discernment is not impugned, any set of propositions and relationships can be included.

The only remaining task is to interrelate these subgraphs. For this example it is sufficient to independently relate each of the feature subgraphs to the polygonal-type subgraph, as illustrated by Figures 36-42. This is fairly straight forward, relating the leaves of each feature hierarchy to the nodes in the polygonal-type subgraph, except for the inclusion of an additional X relationship in each of the equal-angles and equal-sides interfaces. These relationships tie off OKT and OTR directly to the feature subgraphs. Without these, they would still be indirectly related to the feature subgraphs, through the primitive elements of the frame of discernment they encompass. But tighter connections are desirable since these are the propositions of primary interest to this system. Considering all of these subgraphs collectively, a dependency graph representing evidential support within this domain is defined.

This dependency graph can be used as a basis for both the combination and extrapolation of multiple bodies of evidence relative to this domain. Assuming the appropriate mechanism for the application of Dempster's rule and an E_+^+ inference engine with a complete search

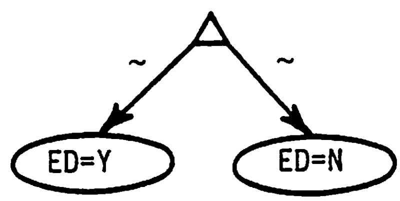


Figure 29. Equal-diagonals (ED) sub-graph.

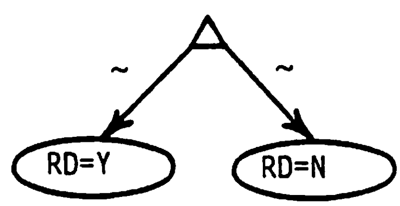


Figure 30. Right-diagonals (RD) sub-graph.

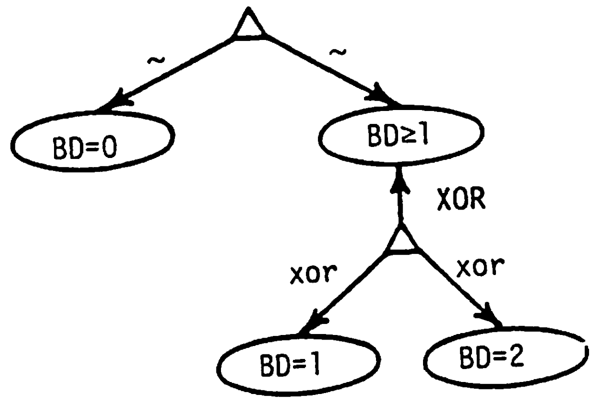


Figure 31. Bisecting-diagonals (BD) sub-graph.

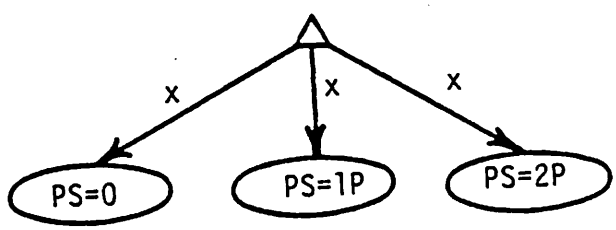


Figure 32. Parallel-sides (PS) sub-graph.

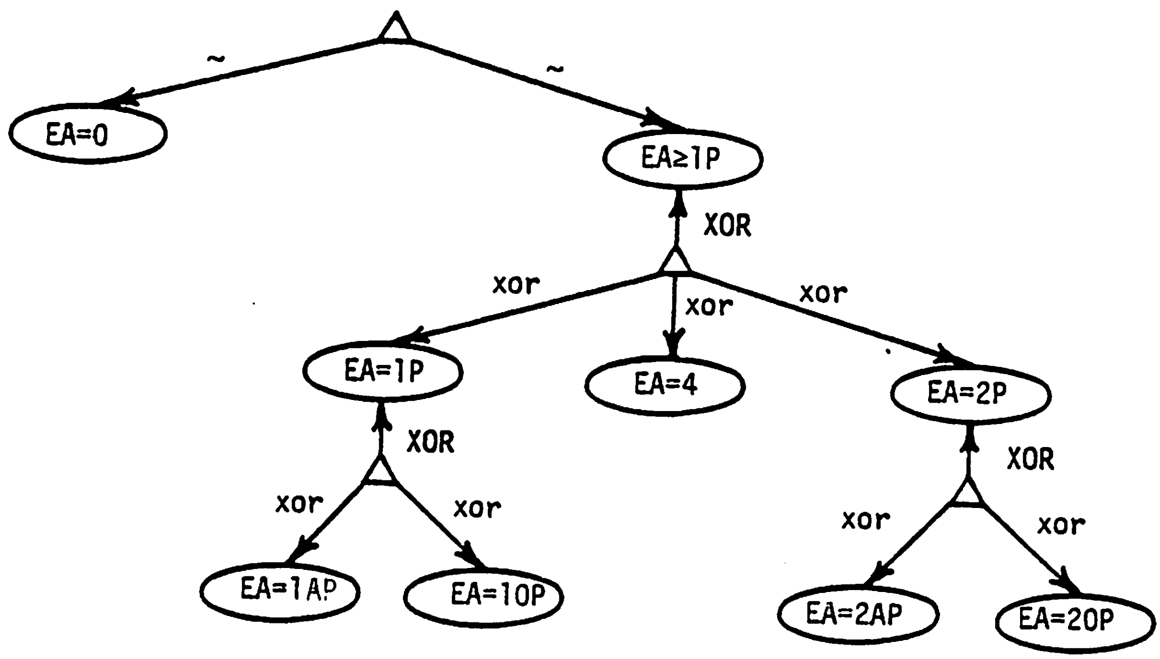


Figure 33. Equal-angles (EA) sub-graph.

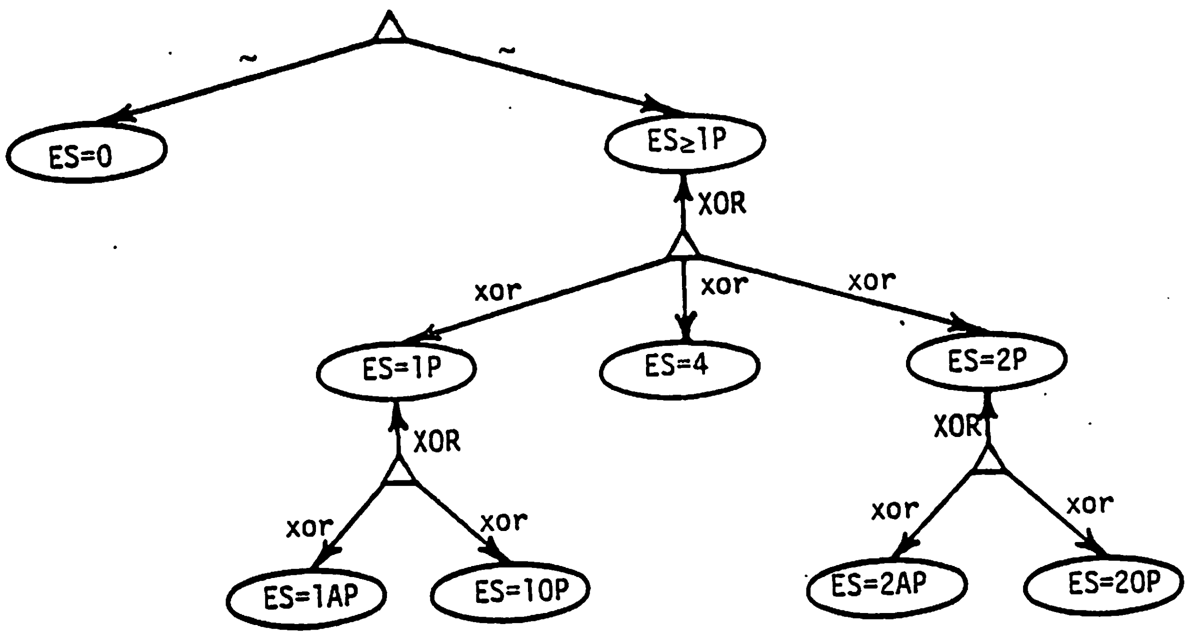


Figure 34. Equal-sides (ES) sub-graph.

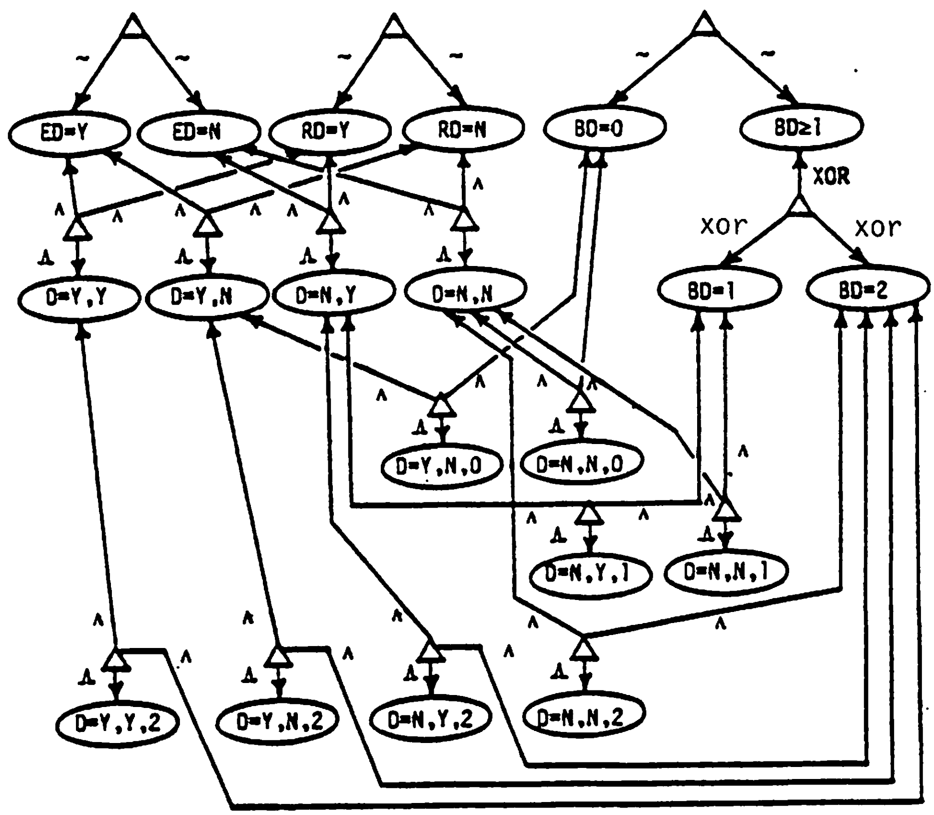


Figure 35. Diagonals (D) sub-graph.

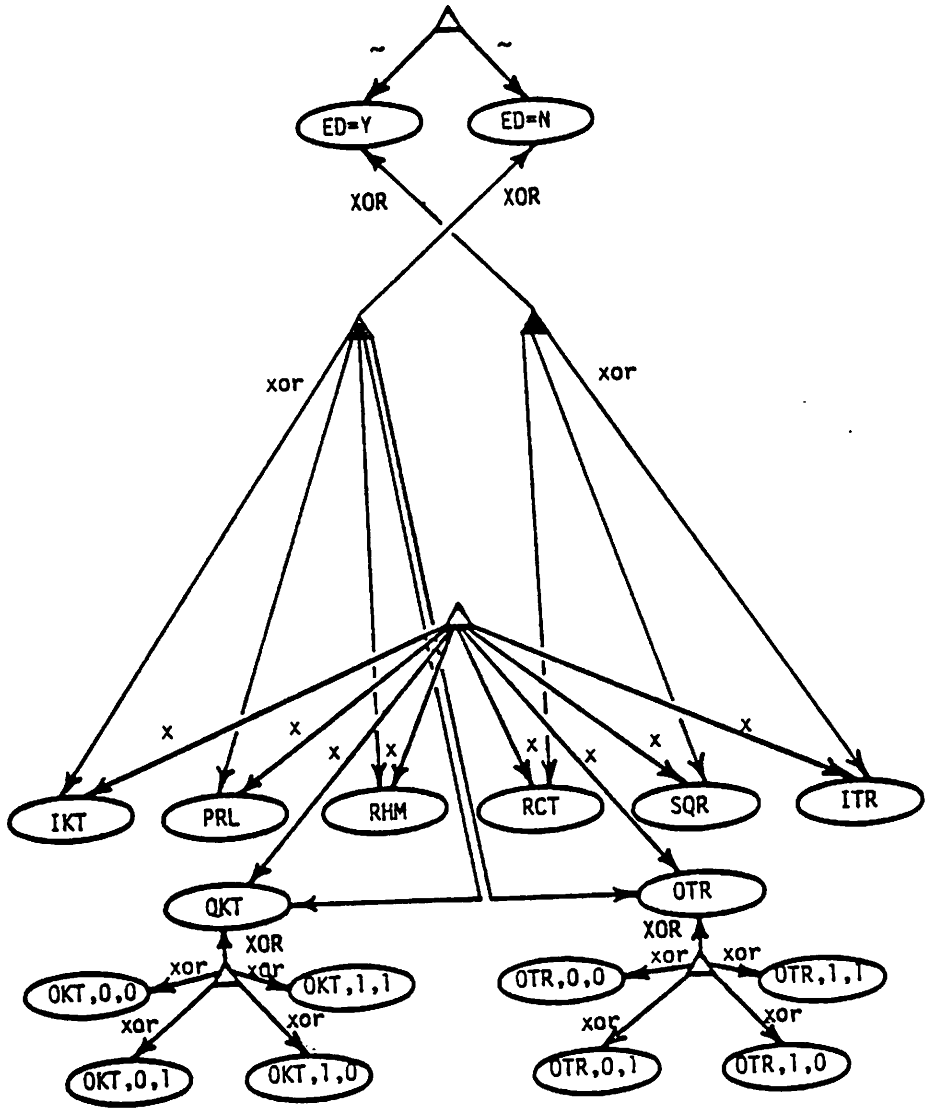


Figure 36. ED-PT interface.

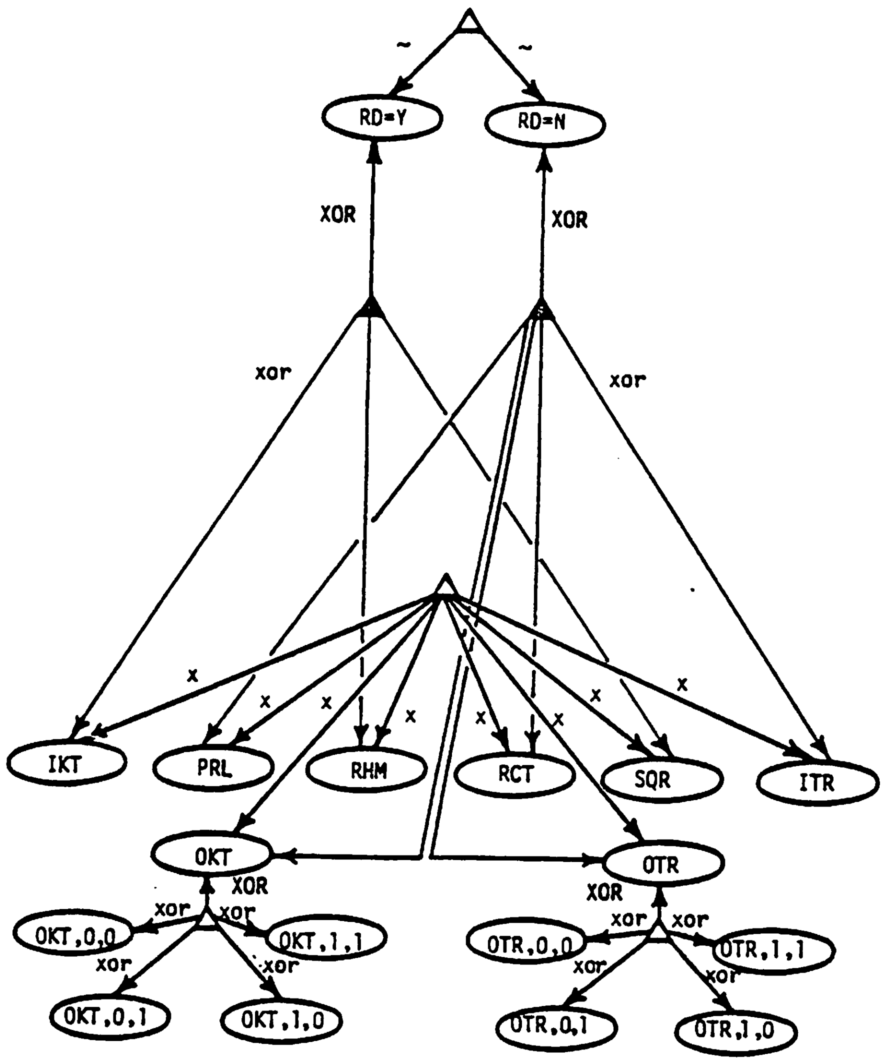


Figure 37. RD-PT interface.

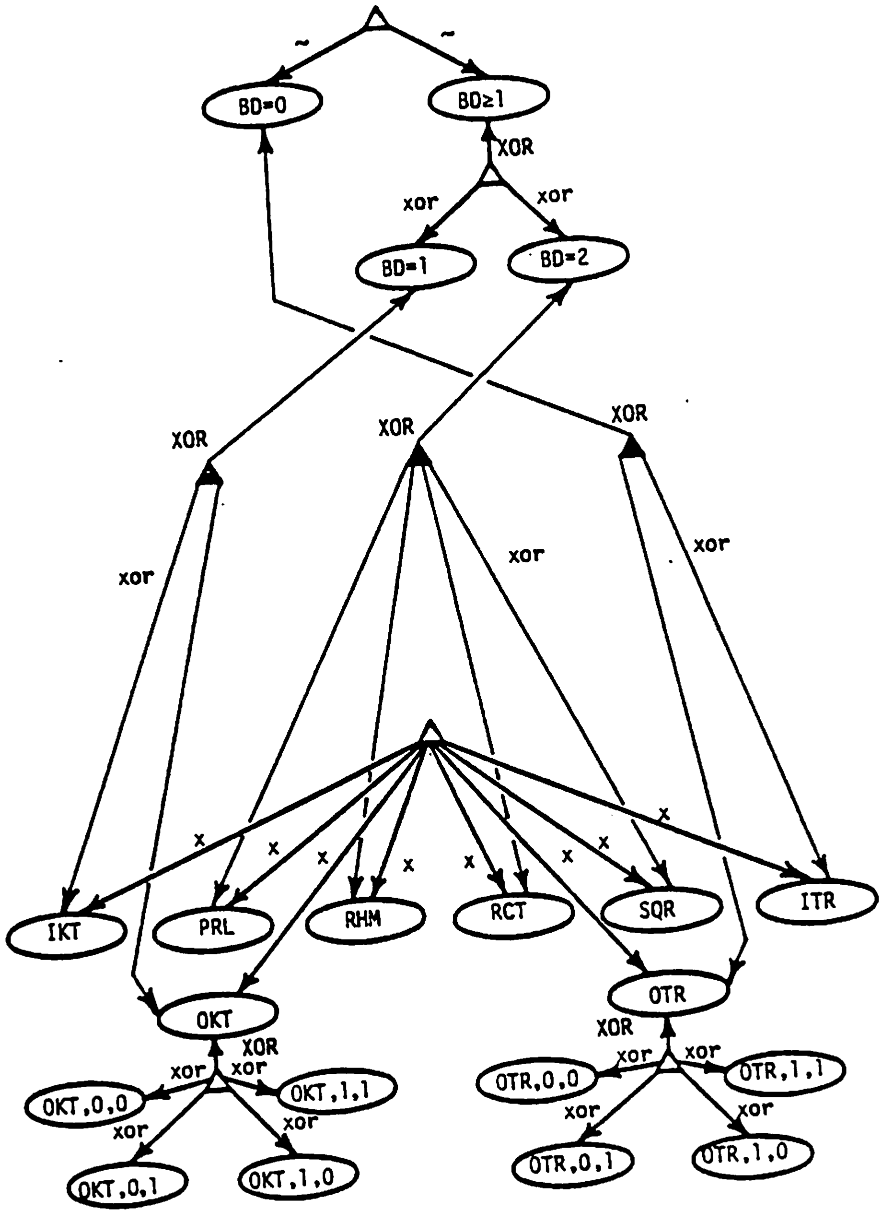


Figure 38. BD-PT interface.

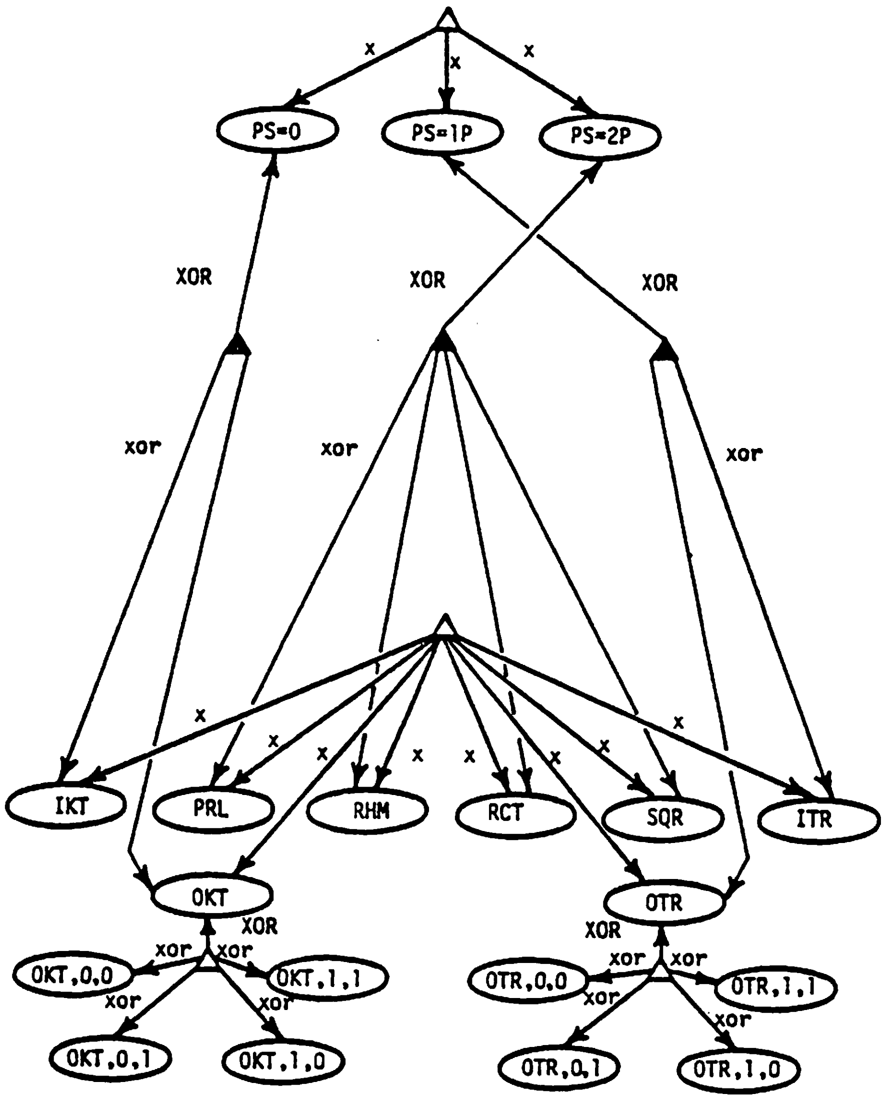


Figure 39. PS-PT interface.

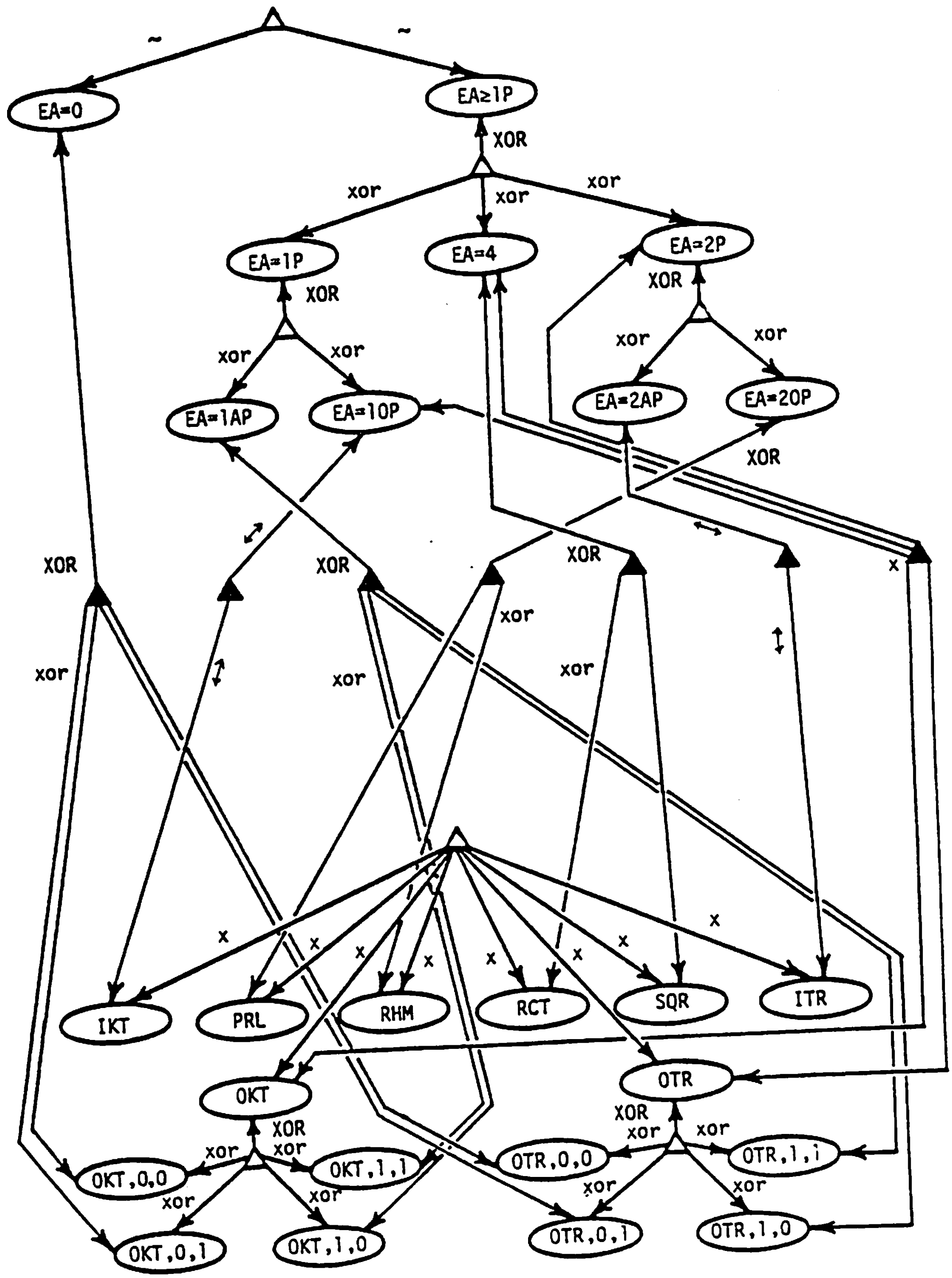


Figure 40. EA-PT interface.

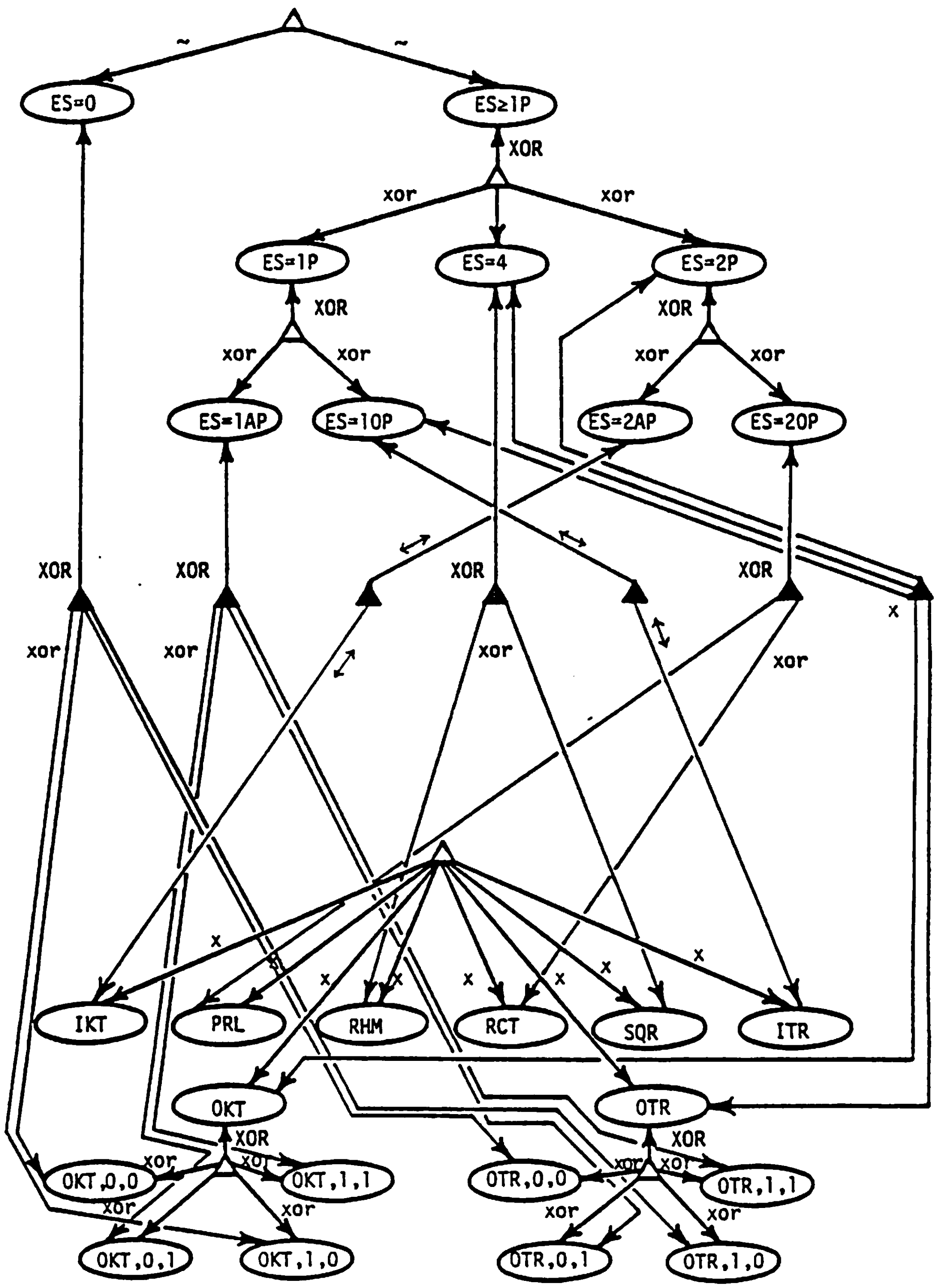


Figure 41. ES-PT interface.

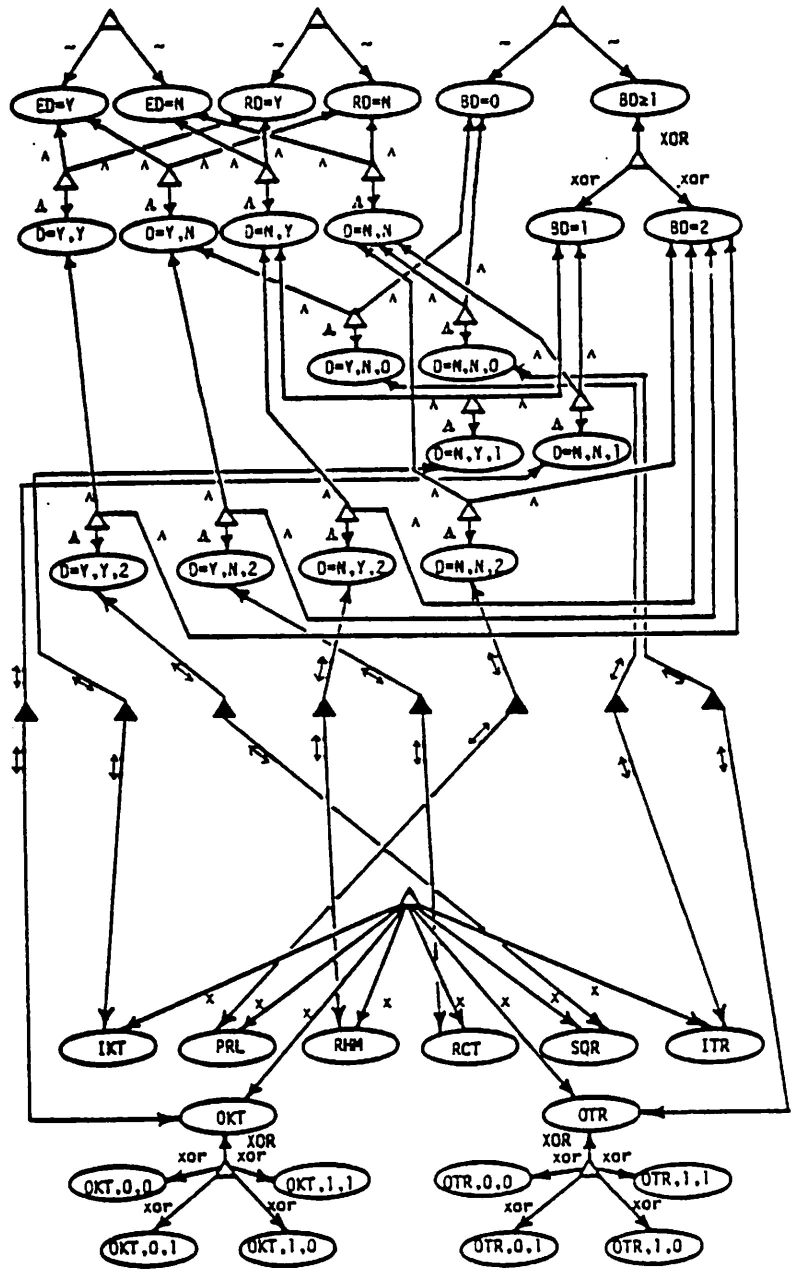


Figure 42. D-PT interface.

strategy, this dependency-graph model of evidential support is capable of the following examples of evidential reasoning.

Beginning with the equal-sides knowledge source, let us assume that it examines a given polygon and (unambiguously) determines that the polygon has two pairs of equal sides, with some doubt remaining. This might be represented by a mass function that attributes .80 to the proposition ES=2P and .20 to θ. This is an example of simple support (p. 102).

$$MASS_{ES}[p] = \begin{cases} .80, & p = ES{=}2P \\ .20, & p = \theta \\ 0.00, & \text{else} \end{cases}$$

Table 2 summarizes the model's conclusions given this information. Some propositions are supported, some are denied, and some remain unchanged, but no propositions are both supported and denied. Notice that those polygonal-types that can have two pairs of equal sides remain completely plausible, while the other polygonal-types have their plausibility reduced. Yet none of these are supported since any one of them could be false and still be consistent with this evidence.

Let us assume that the equal-angles knowledge source is even more vague, unable to make up its mind whether the presented polygon has four equal angles or just two (opposite) pairs of equal angles. It is not certain that either is true, but it leans towards all four being equal. This is represented by an ambiguous mass function (p. 103), possibly one that attributes .60 to EA=4, .30 to EA=2OP, and .10 to θ.

ES=0	[0.00, .20]	[]****----------------[]
ES=1AP	[0.00, .20]	[]****----------------[]
ES=1OP	[0.00, .20]	[]****----------------[]
ES=1P	[0.00, .20]	[]****----------------[]
ES=2AP	[0.00, 1.00]	[]********************[]
ES=2OP	[0.00, 1.00]	[]********************[]
ES=2P	[.80, 1.00]	[]----------------****[]
ES=4	[0.00, .20]	[]****----------------[]
ES≥1P	[.80, 1:00]	[]----------------****[]
IKT	[0.00, 1.00]	[]********************[]
ITR	[0.00, .20]	[]****----------------[]
OKT	[0.00, .20]	[]****----------------[]
OKT,0,0	[0.00, .20]	[]****----------------[]
OKT,0,1	[0.00, .20]	[]****----------------[]
OKT,1,0	[0.00, .20]	[]****----------------[]
OKT,1,1	[0.00, .20]	[]****----------------[]
OTR	[0.00, .20]	[]****----------------[]
OTR,0,0	[0.00, .20]	[]****----------------[]
OTR,0,1	[0.00, .20]	[]****----------------[]
OTR,1,0	[0.00, .20]	[]****----------------[]
OTR,1,1	[0.00 .20]	[]****----------------[]
PRL	[0.00, 1.00]	[]********************[]
RCT	[0.00, 1.00]	[]********************[]
RHM	[0.00, .20]	[]****----------------[]
SQR	[0.00, .20]	[]****----------------[]

Table 2. Inference results: $MASS_{ES}$.

$$MASS_{EA}[p] = \begin{cases} .60, & p = EA{=}4 \\ .30, & p = EA{=}2OP \\ .10, & p = \theta \\ 0.00, & \text{else} \end{cases}$$

This leads to some propositions being simultaneously supported by one of the focal propositions and denied by another. But still none of the polygonal-types receive any support, just differing degrees of plausibility (Table 3).

Assuming that these knowledge sources operate independently, their information can be combined by Dempster's rule. Figure 43 illustrates this combination. A rectangle has two pairs of equal sides and four equal angles; a parallelogram has two pairs of equal sides and two opposite pairs of equal angles; no other polygonal-types have these pairs of traits. Therefore the combined impact of this information is to support these two polygonal-types with some residual support going to the previous focal elements (Table 4). Since the equal-angles knowledge source is more confident of EA=4 than EA=2OP, rectangle receives more support than parallelogram. None of the other polygonal-types are supported, but their plausibilities vary. Thus, based on the combined evidence from these two knowledge sources, the probability of the presented polygon being a rectangle is at least .48 and possibly as high as .70, which is more likely than it being a parallelogram whose probability is bounded by .24 and .40, and both of these are far more likely than any other polygonal-types. This certainly is the result that one would expect.

EA=0	[0.00, .10]	[**------------------]
EA=1AP	[0.00, .10]	[**------------------]
EA=1OP	[0.00, .10]	[**------------------]
EA=1P	[0.00, .10]	[**------------------]
EA=2AP	[0.00, .10]	[**------------------]
EA=2OP	[.30, .40]	[------**------------]
EA=2P	[.30, .40]	[------**------------]
EA=4	[.60, .70]	[------------**------]
EA≥1P	[.90, 1.00]	[------------------**]
IKT	[0.00, .10]	[**------------------]
ITR	[0.00, .10]	[**------------------]
OKT	[0.00, .10]	[**------------------]
OKT,0,0	[0.00, .10]	[**------------------]
OKT,0,1	[0.00, .10]	[**------------------]
OKT,1,0	[0.00, .10]	[**------------------]
OKT,1,1	[0.00, .10]	[**------------------]
OTR	[0.00, .10]	[**------------------]
OTR,0,0	[0.00, .10]	[**------------------]
OTR,0,1	[0.00, .10]	[**------------------]
OTR,1,0	[0.00, .10]	[**------------------]
OTR,1,1	[0.00, .10]	[**------------------]
PRL	[0.00, .40]	[********------------]
RCT	[0.00, .70]	[**************------]
RHM	[0.00, .40]	[********------------]
SQR	[0.00, .70]	[**************------]

Table 3. Inference results: $MASS_{EA}$.

$MASS_{EA}$		ES=2P (.80)	θ (.20)
θ	.10	θ ∩ (ES=2P) = (ES=2P); (.10)(.80) = .08	θ ∩ θ = θ; (.1)(.2)=.02
EA=20P	.30	(EA=20P) ∩ (ES=2P) = PRL; (.30)(.80) = .24	(EA=20P) ∩ θ = (EA=20P); (.3)(.2)=.06
EA=4	.60	(EA=4) ∩ (ES=2P) = RCT; (.60)(.80) = .48	(EA=4) ∩ θ = (EA=4); (.6)(.2)=.12

$MASS_{ES}$

$$MASS_{ES} \oplus MASS_{EA}[p] = \begin{cases} .48, & p = RCT \\ .24, & p = PRL \\ .12, & p = EA{=}4 \\ .08, & p = ES{=}2P \\ .06, & p = EA{=}20P \\ .02, & p = \theta \\ 0.00, & \text{else} \end{cases}$$

Figure 43. Combination of $MASS_{ES}$ and $MASS_{EA}$.

EA=0	[0.00, .10]	[**------------------]
EA=1AP	[0.00, .10]	[**------------------]
EA=1OP	[0.00, .10]	[**------------------]
EA=1P	[0.00, .10]	[**------------------]
EA=2AP	[0.00, .02]	[*-------------------]
EA=2OP	[.30, .40]	[------**------------]
EA=2P	[.30, .40]	[------**------------]
EA=4	[.60, .70]	[------------**------]
EA≥1P	[.90, 1.00]	[------------------**]
ES=0	[0.00, .20]	[****----------------]
ES=1AP	[0.00, .20]	[****----------------]
ES=1OP	[0.00, .02]	[*-------------------]
ES=1P	[0.00, .20]	[****----------------]
ES=2AP	[0.00, .10]	[**------------------]
ES=2OP	[.72, 1.00]	[--------------******]
ES=2P	[.80, 1.00]	[----------------****]
ES=4	[0.00, .20]	[****----------------]
ES≥1P	[.80, 1.00]	[----------------****]
IKT	[0.00, .10]	[**------------------]
ITR	[0.00, .02]	[*-------------------]
OKT	[0.00, .10]	[*-------------------]
OKT,0,0	[0.00, .10]	[*-------------------]
OKT,0,1	[0.00, .10]	[*-------------------]
OKT,1,0	[0.00, .10]	[*-------------------]
OKT,1,1	[0.00, .10]	[*-------------------]
OTR	[0.00, .10]	[*-------------------]
OTR,0,0	[0.00, .10]	[*-------------------]
OTR,0,1	[0.00, .10]	[*-------------------]
OTR,1,0	[0.00, .10]	[*-------------------]
OTR,1,1	[0.00, .10]	[*-------------------]
PRL	[.24, .40]	[----****------------]
RCT	[.48, .70]	[---------*****------]
RHM	[0.00, .08]	[**------------------]
SQR	[0.00, .14]	[***-----------------]

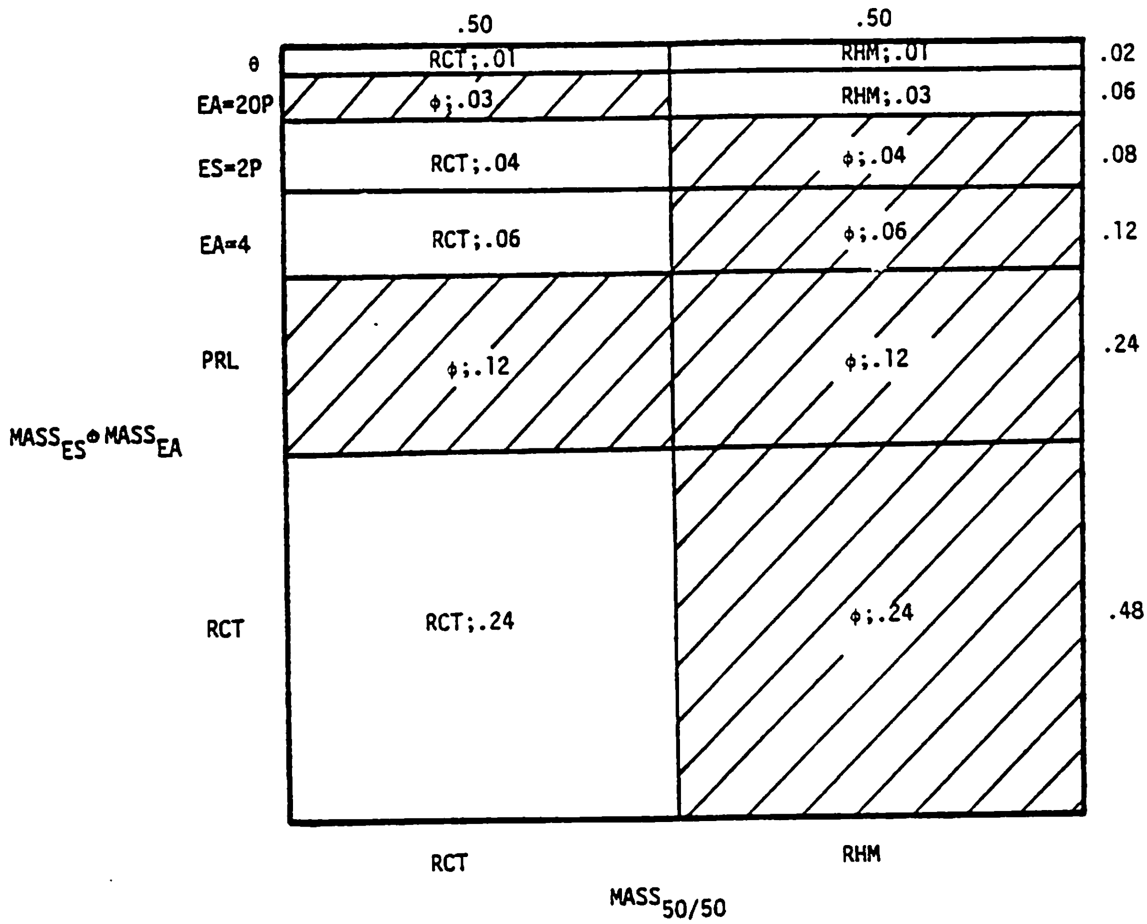
Table 4. Inference results: $MASS_{ES} \oplus MASS_{EA}$.

But suppose we have some valuable prior information concerning this presentation. Assume we know, without question, that it is either a rectangle or a rhombus and that there is an equal chance of either one i.e., a 50/50 Bayesian chance. Then this information can be gainfully employed by simply representing it as a mass function and combining it with the other (evidential) information.

$$MASS_{50/50}[p] = \begin{cases} .50, & p = RCT \\ .50, & p = RHM \\ 0.00, & else \end{cases}$$

The order of this combination is immaterial since the result is invariant with respect to the order. Figure 44 illustrates this combination. Notice that some of the intersections do not exist, causing Dempster's rule to renormalize over those that do. The impact of this information (Table 5) is to make rectangle even more favorable, and to eliminate all of the other polygonal-types, except rhombus, since they fall outside of the range of possibility. It also has the effect of collapsing all of the intervals to points i.e., the result is Bayesian. This is always the case when a Bayesian function is used in combination.

If a more elaborate Bayesian distribution were known to govern this situation, it could be used instead. For example, the following Bayesian mass function might represent the governing distribution.



$$MASS_{ES} \oplus MASS_{EA} \oplus MASS_{50/50}[p] = \begin{cases} N(.35) = .897, & p = RCT \\ N(.04) = .103, & p = RHM \\ 0.00, & \text{else} \end{cases}$$

where $N = (1 - k)^{-1} = (1 - .61)^{-1} = 2.564.$

Figure 44. Combination of $MASS_{ES}$, $MASS_{EA}$, and $MASS_{50/50}$.

EA=0	[0.00, 0.00]	[]*-----------------[]
EA=1AP	[0.00, 0.00]	[]*-----------------[]
EA=1OP	[0.00, 0.00]	[]*-----------------[]
EA=1P	[0.00, 0.00]	[]*-----------------[]
EA=2AP	[0.00, 0.00]	[]*-----------------[]
EA=2OP	[.10, .10]	[]--*---------------[]
EA=2P	[.10, .10]	[]--*---------------[]
EA=4	[.90, .90]	[]---------------*--[]
EA≥1P	[1.00, 1.00]	[]-----------------*[]
ES=0	[0.00, 0.00]	[]*-----------------[]
ES=1AP	[0.00, 0.00]	[]*-----------------[]
ES=1OP	[0.00, 0.00]	[]*-----------------[]
ES=1P	[0.00, 0.00]	[]*-----------------[]
ES=2AP	[0.00, 0.00]	[]*-----------------[]
ES=2OP	[.90, .90]	[]---------------*--[]
ES=2P	[.90, .90]	[]---------------*--[]
ES=4	[.10, .10]	[]--*---------------[]
ES≥1P	[1.00, 1.00]	[]-----------------*[]
IKT	[0.00, 0.00]	[]*-----------------[]
ITR	[0.00, 0.00]	[]*-----------------[]
OKT	[0.00, 0.00]	[]*-----------------[]
OKT,0,0	[0.00, 0.00]	[]*-----------------[]
OKT,0,1	[0.00, 0.00]	[]*-----------------[]
OKT,1,0	[0.00, 0.00]	[]*-----------------[]
OKT,1,1	[0.00, 0.00]	[]*-----------------[]
OTR	[0.00, 0.00]	[]*-----------------[]
OTR,0,0	[0.00, 0.00]	[]*-----------------[]
OTR,0,1	[0.00, 0.00]	[]*-----------------[]
OTR,1,0	[0.00, 0.00]	[]*-----------------[]
OTR,1,1	[0.00, 0.00]	[]*-----------------[]
PRL	[0.00, 0.00]	[]*-----------------[]
RCT	[.90, .90]	[]---------------*--[]
RHM	[.10, .10]	[]--*---------------[]
SQR	[0.00, 0.00]	[]*-----------------[]

Table 5. Inference results: $MASS_{ES} \oplus MASS_{EA} \oplus MASS_{50/50}$.[6]

[6] All entries have been rounded to two decimal places.

$$
MASS_B[p] = \begin{cases}
.05, & p = IKT \\
.10, & p = ITR \\
.0075, & p = OKT,0,0 \\
.0075, & p = OKT,0,1 \\
.0075, & p = OKT,1,0 \\
.0075, & p = OKT,1,1 \\
.03, & p = OTR,0,0 \\
.03, & p = OTR,0,1 \\
.03, & p = OTR,1,0 \\
.03, & p = OTR,1,1 \\
.20, & p = PRL \\
.15, & p = RHM \\
.10, & p = RCT \\
.25, & p = SQR \\
0.00, & \text{else}
\end{cases}
$$

This is obviously a Bayesian mass function since it distributes its mass exclusively over the most primitive elements of the frame of discernment (p. 104). From this information alone the point probabilities of Table 6 follow. Table 7 summarizes the impact of this information in combination with the evidence from the equal-sides and equal-angles knowledge sources. Note that the higher prior probability of parallelogram overrides the preponderance of evidence supporting rectangle, resulting in parallelogram being the more likely.

If this, or any other, Bayesian mass function is combined with another mass function that is fully committed to a single proposition, the resulting support (and plausibility) for each proposition is the Bayesian conditional. In other words, Dempster's rule has the same effect as Bayes' rule of conditioning in this limiting case. Table 8 summarizes such an example where the Bayesian chance function of the previous example has been combined with a mass function that commits all of its mass to the proposition PS=2P (i.e., two pairs of parallel sides). Notice that the results are the Bayesian conditionals $P(p_i|PS=2P)$.

EA=0	[.08, .08]	[-*----------------]
EA=1AP	[.08, .08]	[-*----------------]
EA=1OP	[.05, .05]	[-*----------------]
EA=1P	[.12, .12]	[--*---------------]
EA=2AP	[.10, .10]	[--*---------------]
EA=2OP	[.35, .35]	[-------*----------]
EA=2P	[.45, .45]	[---------*--------]
EA=4	[.35, .35]	[-------*----------]
EA≥1P	[.93, .93]	[----------------*-]
ES=0	[.08, .08]	[-*----------------]
ES=1AP	[.08, .08]	[-*----------------]
ES=1OP	[.10, .10]	[--*---------------]
ES=1P	[.17, .17]	[---*--------------]
ES=2AP	[.05, .05]	[-*----------------]
ES=2OP	[.30, .30]	[------*-----------]
ES=2P	[.35, .35]	[-------*----------]
ES=4	[.40, .40]	[--------*---------]
ES≥1P	[.93, .93]	[----------------*-]
IKT	[.05, .05]	[-*----------------]
ITR	[.10, .10]	[--*---------------]
OKT	[.03, .03]	[*-----------------]
OKT,0,0	[.01, .01]	[*-----------------]
OKT,0,1	[.01, .01]	[*-----------------]
OKT,1,0	[.01, .01]	[*-----------------]
OKT,1,1	[.01, .01]	[*-----------------]
OTR	[.12, .12]	[--*---------------]
OTR,0,0	[.03, .03]	[*-----------------]
OTR,0,1	[.03, .03]	[*-----------------]
OTR,1,0	[.03, .03]	[*-----------------]
OTR,1,1	[.03, .03]	[*-----------------]
PRL	[.20, .20]	[----*-------------]
RCT	[.10, .10]	[--*---------------]
RHM	[.15, .15]	[---*--------------]
SQR	[.25, .25]	[-----*------------]

Table 6. Inference results: $MASS_B$.[7]

---

[7]All entries have been rounded to two decimal places.

EA=0	[.01, .01]	[*------------------]
EA=1AP	[.01, .01]	[*------------------]
EA=1OP	[.02, .02]	[*------------------]
EA=1P	[.03, .03]	[*------------------]
EA=2AP	[.01, .01]	[*------------------]
EA=2OP	[.44, .44]	[--------*----------]
EA=2P	[.45, .45]	[---------*---------]
EA=4	[.51, .51]	[----------*--------]
EA≥1P	[.99, .99]	[------------------*]
ES=0	[.01, .01]	[*------------------]
ES=1AP	[.01, .01]	[*------------------]
ES=1OP	[.01, .01]	[*------------------]
ES=1P	[.02, .02]	[*------------------]
ES=2AP	[.02, .02]	[*------------------]
ES=2OP	[.72, .72]	[-------------*-----]
ES=2P	[.75, .75]	[-------------*-----]
ES=4	[.23, .23]	[----*--------------]
ES≥1P	[.99, .99]	[------------------*]
IKT	[.02, .02]	[*------------------]
ITR	[.01, .01]	[*------------------]
OKT	[.00, .00]	[*------------------]
OKT,0,0	[.00, .00]	[*------------------]
OKT,0,1	[.00, .00]	[*------------------]
OKT,1,0	[.00, .00]	[*------------------]
OKT,1,1	[.00, .00]	[*------------------]
OTR	[.01, .01]	[*------------------]
OTR,0,0	[.00, .00]	[*------------------]
OTR,0,1	[.00, .00]	[*------------------]
OTR,1,0	[.00, .00]	[*------------------]
OTR,1,1	[.00, .00]	[*------------------]
PRL	[.39, .39]	[-------*-----------]
RCT	[.34, .34]	[------*------------]
RHM	[.06, .06]	[-*-----------------]
SQR	[.17, .17]	[---*---------------]

Table 7. Inference results: $MASS_{ES} \oplus MASS_{EA} \oplus MASS_B$.[8]

---

[8]All entries have been rounded to two decimal places.

EA=0	[0.00, 0.00]	[*-----------------]
EA=1AP	[0.00, 0.00]	[*-----------------]
EA=1OP	[0.00, 0.00]	[*-----------------]
EA=1P	[0.00, 0.00]	[*-----------------]
EA=2AP	[0.00, 0.00]	[*-----------------]
EA=2OP	[.50, .50]	[--------*---------]
EA=2P	[.50, .50]	[--------*---------]
EA=4	[.50, .50]	[--------*---------]
EA≥1P	[1.00, 1.00]	[-----------------*]
ES=0	[0.00, 0.00]	[*-----------------]
ES=1AP	[0.00, 0.00]	[*-----------------]
ES=1OP	[0.00, 0.00]	[*-----------------]
ES=1P	[0.00, 0.00]	[*-----------------]
ES=2AP	[0.00, 0.00]	[*-----------------]
ES=2OP	[.43, .43]	[-------*----------]
ES=2P	[.43, .43]	[-------*----------]
ES=4	[.57, .57]	[----------*-------]
ES≥1P	[1.00, 1.00]	[-----------------*]
PS=0	[0.00, 0.00]	[*-----------------]
PS=1P	[0.00, 0.00]	[*-----------------]
PS=2P	[1.00, 1.00]	[-----------------*]
IKT	[0.00, 0.00]	[*-----------------]
ITR	[0.00, 0.00]	[*-----------------]
OKT	[0.00, 0.00]	[*-----------------]
OKT,0,0	[0.00, 0.00]	[*-----------------]
OKT,0,1	[0.00, 0.00]	[*-----------------]
OKT,1,0	[0.00, 0.00]	[*-----------------]
OKT,1,1	[0.00, 0.00]	[*-----------------]
OTR	[0.00, 0.00]	[*-----------------]
OTR,0,0	[0.00, 0.00]	[*-----------------]
OTR,0,1	[0.00, 0.00]	[*-----------------]
OTR,1,0	[0.00, 0.00]	[*-----------------]
OTR,1,1	[0.00, 0.00]	[*-----------------]
PRL	[.29, .29]	[-----*------------]
RCT	[.14, .14]	[--*---------------]
RHM	[.21, .21]	[----*-------------]
SQR	[.36, .36]	[-------*----------]

Table 8. Inference results: $MASS_B \oplus MASS_{PS=2P}$.

This further illustrates that when true Bayesian information is available, it is fully exploited within this framework. Yet when this exacting information is unavailable, which is the usual circumstance within evidential domains, it continues to perform in a reliable and productive way!

Suppose we discover that this Bayesian distribution we have been using $MASS_B$ is only applicable 75% of the time and the chance distribution governing the remaining 25% of the trials is unknown. This situation can be accurately modelled by a mass function that distributes 75% of its mass in accordance with the known chance distribution and attributes its remaining mass to θ.

$$MASS_{B'}[p] = \begin{cases} .0375, & p = IKT \\ .075, & p = ITR \\ .005625, & p = OKT,0,0 \\ .005625, & p = OKT,0,1 \\ .005625, & p = OKT,1,0 \\ .005625, & p = OKT,1,1 \\ .0225, & p = OTR,0,0 \\ .0225, & p = OTR,0,1 \\ .0225, & p = OTR,1,0 \\ .0225, & p = OTR,1,1 \\ .15, & p = PRL \\ .1125, & p = RHM \\ .075, & p = RCT \\ .1875, & p = SQR \\ .25, & p = \theta \\ 0.00, & \text{else} \end{cases}$$

Its impact, summarized by Table 9, reflects what is unknown. When this mass function is combined with the evidential information already presented pertaining to ES and EA, the result (Table 10) reflects these unknowns in addition to those already present in the evidential infor-

EA=0	[.06, .31]	[-*****--------------]
EA=1AP	[.06, .31]	[-*****--------------]
EA=1OP	[.04, .29]	[******--------------]
EA=1P	[.09, .34]	[-******-------------]
EA=2AP	[.08, .33]	[-******-------------]
EA=2OP	[.26, .51]	[-----*****----------]
EA=2P	[.34, .59]	[------******--------]
EA=4	[.26, .51]	[-----*****----------]
EA≥1P	[.69, .94]	[-------------******-]
ES=0	[.06, .31]	[-*****--------------]
ES=1AP	[.06, .31]	[-*****--------------]
ES=1OP	[.08, .33]	[-******-------------]
ES=1P	[.13, .38]	[--******------------]
ES=2AP	[.04, .29]	[******--------------]
ES=2OP	[.23, .48]	[----******----------]
ES=2P	[.26, .51]	[-----*****----------]
ES=4	[.30, .55]	[------*****---------]
ES≥1P	[.69, .94]	[-------------******-]
IKT	[.04, .29]	[*******-------------]
ITR	[.08, .33]	[-******-------------]
OKT	[.02, .27]	[*****---------------]
OKT,0,0	[.01, .26]	[*****---------------]
OKT,0,1	[.01, .26]	[*****---------------]
OKT,1,0	[.01, .26]	[*****---------------]
OKT,1,1	[.01, .26]	[*****---------------]
OTR	[.09, .34]	[-******-------------]
OTR,0,0	[.02, .27]	[*****---------------]
OTR,0,1	[.02, .27]	[*****---------------]
OTR,1,0	[.02, .27]	[*****---------------]
OTR,1,1	[.02, .27]	[*****---------------]
PRL	[.15, .40]	[---*****------------]
RCT	[.08, .33]	[-******-------------]
RHM	[.11, .36]	[--*****-------------]
SQR	[.19, .44]	[---******-----------]

Table 9. Inference results: $MASS_B$.[9]

---

[9] All entries have been rounded to two decimal places.

EA=0	[.00, .06]	[*--------------------]
EA=1AP	[.00, .06]	[*--------------------]
EA=1OP	[.01, .07]	[*--------------------]
EA=1P	[.01, .07]	[*--------------------]
EA=2AP	[.00, .02]	[*--------------------]
EA=2OP	[.36, .42]	[-------*-------------]
EA=2P	[.36, .42]	[-------*-------------]
EA=4	[.56, .63]	[-----------**--------]
EA≥1P	[.94, 1.00]	[-------------------**]
ES=0	[.00, .13]	[***------------------]
ES=1AP	[.00, .13]	[***------------------]
ES=1OP	[.00, .02]	[*--------------------]
ES=1P	[.01, .13]	[***------------------]
ES=2AP	[.01, .07]	[*--------------------]
ES=2OP	[.72, .89]	[---------------****--]
ES=2P	[.78, .90]	[----------------***--]
ES=4	[.09, .21]	[-***-----------------]
ES≥1P	[.87, 1.00]	[------------------***]
IKT	[.01, .07]	[*--------------------]
ITR	[.00, .02]	[*--------------------]
OKT	[.00, .01]	[*--------------------]
OKT,0,0	[.00, .01]	[*--------------------]
OKT,0,1	[.00, .01]	[*--------------------]
OKT,1,0	[.00, .01]	[*--------------------]
OKT,1,1	[.00, .01]	[*--------------------]
OTR	[.00, .02]	[*--------------------]
OTR,0,0	[.00, .01]	[*--------------------]
OTR,0,1	[.00, .01]	[*--------------------]
OTR,1,0	[.00, .01]	[*--------------------]
OTR,1,1	[.00, .01]	[*--------------------]
PRL	[.30, .39]	[-----***-------------]
RCT	[.43, .56]	[--------***----------]
RHM	[.02, .07]	[*--------------------]
SQR	[.06, .15]	[-**------------------]

Table 10. Inference results: $MASS_{ES} \oplus MASS_{EA} \oplus MASS_{B}$..[10]

[10] All entries have been rounded to two decimal places.

mation from the knowledge sources. Such varying degrees of ignorance cannot be properly captured by Bayesian point probabilities.

Even when chance information is totally unavailable, useful results can be obtained. In this last example, we assume that four knowledge sources have examined a presented polygon, and have returned with the following information:

$$MASS_{ED}[p] = \begin{cases} .80, & p = ED{=}N \\ .20, & p = \theta \\ 0.00, & \text{else} \end{cases}$$

$$MASS_{RD}[p] = \begin{cases} .60, & p = RD{=}Y \\ .20, & p = RD{=}N \\ .20, & p = \theta \\ 0.00, & \text{else} \end{cases}$$

$$MASS_{BD}[p] = \begin{cases} .40, & p = BD{\geq}1 \\ .30, & p = BD{=}2 \\ .30, & p = \theta \\ 0.00, & \text{else} \end{cases}$$

$$MASS_{ES}[p] = \begin{cases} .80, & p = ES{=}2P \\ .20, & p = \theta \\ 0.00, & \text{else} \end{cases}$$

Combining these bodies of evidence has the effect of distinguishing isoceles kite as the most likely identification (Table 11). This evidence also can be fed forward to predict unexamined features (Table 12), or fed back to make better predictions about examined ones (e.g., compare the ES portions of Tables 2 and 11). This flexibility follows directly from the internal consistency of the model. Inferencing is unconstrained; convergence is guaranteed!

All of the results in this section were produced by machine. A general graph-theoretic database facility, GRASPER 1.0 [Lowrance 1978; Lowrance and Corkill 1979], supports dependency graph construction, editing, and retrieval. A LISP implementation of Dempster's rule and an accompanying inference engine realize the reasoning component.

BD=0	[.00, .06]	[*-----------------]
BD=1	[.57, .63]	[-----------**------]
BD=2	[.37, .43]	[-------**----------]
BD≥1	[.94, 1.00]	[-----------------**]
ED=Y	[.07, .12]	[-*----------------]
ED=N	[.88, .93]	[---------------**-]
RD=Y	[.65, .70]	[------------*------]
RD=N	[.30, .35]	[------*------------]
ES=0	[0.00, .07]	[*-----------------]
ES=1AP	[0.00, .07]	[*-----------------]
ES=1OP	[0.00, .01]	[*-----------------]
ES=1P	[0.00, .07]	[*-----------------]
ES=2AP	[.55, .61]	[-----------*-------]
ES=2OP	[.28, .34]	[-----**-----------]
ES=2P	[.85, .89]	[---------------**--]
ES=4	[.09, .13]	[-*----------------]
ES≥1P	[.93, 1.00]	[-----------------**]
IKT	[.55, .61]	[-----------*-------]
ITR	[0.00, .01]	[*-----------------]
OKT	[.01, .04]	[*-----------------]
OKT,0,0	[0.00, .04]	[*-----------------]
OKT,0,1	[0.00, .04]	[*-----------------]
OKT,1,0	[0.00, .04]	[*-----------------]
OKT,1,1	[0.00, .04]	[*-----------------]
OTR	[.00, .02]	[*-----------------]
OTR,0,0	[0.00, .02]	[*-----------------]
OTR,0,1	[0.00, .02]	[*-----------------]
OTR,1,0	[0.00, .02]	[*-----------------]
OTR,1,1	[0.00, .02]	[*-----------------]
PRL	[.22, .27]	[----*-------------]
RCT	[.06, .10]	[-*----------------]
RHM	[.07, .12]	[-*----------------]
SQR	[.01, .04]	[*-----------------]

Table 11. Inference results: $MASS_{BD} \oplus MASS_{ED} \oplus MASS_{RD} \oplus MASS_{ES}$.[11]

---

[11] All entries have been rounded to two decimal places.

EA=0	[0.00, .09]	[**------------------]
EA=1AP	[0.00, .09]	[**------------------]
EA=1OP	[.55, .61]	[-----------*--------]
EA=1P	[.55, .64]	[-----------**-------]
EA=2AP	[0.00, .01]	[*-------------------]
EA=2OP	[.29, .36]	[-----**-------------]
EA=2P	[.29, .36]	[-----**-------------]
EA=4	[.07, .14]	[-**-----------------]
EA≥1P	[.91, 1.00]	[------------------**]

Table 12. Inference results: $MASS_{BD} \oplus MASS_{ED} \oplus MASS_{RD} \oplus MASS_{ES}$ on EA.[12]

---

[12] All entries have been rounded to two decimal places.

CHAPTER VII

CONCLUSIONS

Summary

This thesis is both a description of a general representation of dependency information and its use as a basis for inferential reasoning, as well as a description of a specific representation of evidential support and its use as a basis for evidential reasoning.

The utility of graphical representations of dependency information has been previously demonstrated as a basis for mechanized inferential reasoning. Dependency graphs are a generalization and refinement of these representational ideas. They are more general, because they are capable of representing dependency relationships of arbitrary order and specificity; they are more refined, because they embody consistency conditions that guarantee the integrity of any inferences based upon them.

Dependency graphs represent dependency relations. Four different classes of dependency relations have been defined, giving rise to four different classes of dependency graphs. This taxonomy has allowed these concepts to be incrementally introduced on the basis of their order and specificity, from the most restrictive to the least restrictive. Dependency relations have been defined as coordinated sets of dependency relationships. Their consistency conditions play the coordinating role, guaranteeing that all redundantly expressed information is compatible.

Each class of dependency graph gives rise to a distinct inference rule and corresponding inference engine. A dependency-graph inference-engine extrapolates from partial confidence information, expressed as a dependency-graph covering, towards more complete information. It makes these predictions based on the dependency information represented by a dependency graph. The initial confidence information can be expressed relative to any subset of propositions in the graph. Propositions are not predefined to serve as either the stimulus or response of some pre-selected inferential steps, but can serve in either role, at any time. Inferencing is unconstrained; feedback and feedforward can freely occur without fear of contradiction; reasoning loops are guaranteed to converge. This reasoning operation is based on whatever information is available, be it pratial or total. One is not forced to estimate information that is truely unavailable for informative inferences to be made.

A dependency-graph model consists of a dependency graph and an accompanying inference engine. The dependency graph reflects the perceived dependencies among a set of propositions relative to an environment being modeled. The inference engine is capable of making predictions based on these perceived dependencies, taking incomplete information about the confidences of these propositions and extending it through inferential reasoning. If the dependency graph and the initial confidence information accurately reflect the environment, then so do the predictions. The internal consistency of these models is guaranteed. Inaccuracies must be attributed to external, and not internal, inconsistencies.

Frequently, the environmental situations of interest in artificial intelligence domains are evidential. Propositions are not known to be true or false, but are attributed subjective degrees of belief based on bodies of evidence extracted from the environment by unreliable sources of knowledge. The freedom to express partial information within dependency-graph models makes them a suitable host for Shafer's mathematical theory of evidence. The adoption of Shafer's theory leads to the adoption of a subset of dependency-graph models as appropriate models of evidential support. A number of evidential relationships, derived from Shafer's theory, constitute the base components from which dependency-graph models of evidential support are constructed.

Dependency-graph models of evidential support take single bodies of evidential information, expressed in terms of support/plausibility or mass, and extend them. Evidential reasoning is used to extrapolate evidence from those propositions that it directly bears upon to those it indirectly bears upon. The only other information required is that contained in the dependency-graph, which represents the range of possibilities. If the evidential information is expressed in terms of mass, model breakdown is precluded, even in the case of a knowledge source and a model with incompatible views of the environment. This condition is guaranteed because evidence expressed in terms of mass does not carry any dependency information; therefore, it cannot refute the dependency information in a model.

When a source of evidential information is unreliable, so are the predictions based upon the information it provides. This is the typical situation in artificial intelligence domains, where knowledge sources are prone to error. Since the consensus of several independent opinions is generally more reliable than any single opinion, more reliable predictions should be possible if they are based on the combined opinions of several independent knowledge sources. The same dependency-graph that provides the appropriate information for the extrapolation of single bodies of evidence, also provides the necessary information for the combination of multiple bodies of evidence. Dempster's rule of combination, an integral part of Shafer's theory, provides the appropriate theoretical foundation for this process. It is order independent; it treats Boolean, Bayesian, and evidential beliefs in a uniform manner; and it does not require a priori chance density information, though this can be fully exploited when it is available.

The feasibility of these techniques, as a foundation for automated evidential reasoning, has been demonstrated. A general system embodying these techniques has been implemented and an initial application explored. The results are promising and suggest that these techniques are applicable in more ambitious domains.

Innovations

Dependency-graph models of evidential support offer some significant advantages over the previous approaches to evidential reasoning in artificially intelligent systems. Many of these follow from the rejection of the probabilistic, rule-based approach, in favor of a possibilistic, relational approach.

The axiom base in a rule-based system consists of a predetermined set of directed inferential steps. In addition each step has some associated probabilistic information, describing the a priori probability of that rule being a valid inferential step. The problem with this approach is that the required probabilistic information is generally unavailable and cannot be accurately estimated. This leads to internal inconsistencies and thereby to contradictory inferential paths. And even when this probabilistic information is consistent, the rules of inference employed typically are not, again leading to contradictory results. In order to prevent divergent behavior, these systems impose arbitrary restrictions on their reasoning processes. For example, both PROSPECTOR and MYCIN are forced to eliminate reasoning loops i.e., chains of inferences across several rules, beginning and ending with the same proposition.

If these previous systems were recast as dependency-graph models, the dependency-graph consistency conditions would not be satisfied. It is these conditions that guarantee the integrity of dependency-graph models. When these conditions are satisfied, it has been shown that a dependency-graph model is a sound inferential system. Errors would have to be attributed to external, and not internal, inconsistencies. When these conditions are not satisfied, errors may be attributable to either internal or external problems.

In place of directed inference rules, dependency graphs are composed of coordinated sets of dependency relationships. These relationships are undirected, providing the appropriate basis for inferential reasoning in any direction. These relationships are not limited to those describing total dependence or independence, as in the probabilistic, rule-based formalism. When precise information about one proposition only partially constrains another proposition, it can be so represented. Partial information about the probability of any proposition can always be directly incorporated into the reasoning. One need not overstate the available information to fit the formalism.

This ability to represent and reason from partial information is critical in evidential domains. By its nature, evidence is partial, characterized by varying degrees of ignorance. Boolean and Bayesian based formalisms do not properly capture this aspect of evidential information. They force evidential information into a form belying its precision. Dependency-graph models of evidential support, based on Shafer's theory of evidence, do not require such overstatements. If partial information is all that is available, useful predictions can be made from it. If total Boolean or Bayesian information is available, it too can be exploited. Further, this reasoning takes place with or without the use of a priori chance densities, whose precise estimation is required in the Bayesian approach. The axiom base in our approach need only describe the possibilities, clearly making it easier to construct than one requiring precise a priori probabilities, particularly since these are typically unavailable and difficult (or impossible) to estimate accurately.

Two distinct reasoning processes have been defined in terms of these graphs. One extrapolates from those propositions that a body of evidence directly bears upon, to those propositions that it indirectly bears upon. The other combines distinct bodies of evidence, pooling the information. The previously developed approaches to inexact reasoning do not always properly distinguish these two types of reasoning, and this is a source of their consistency problems. Dependency-graph models of evidential support carefully avoid this confusion. The result is a system that freely reasons without the need of ad hoc ordering constraints. The conclusions are invariant with respect to the order of the inferential steps. Feedback and feedforward occur freely.

Dependency-graph models of evidential support provide a common framework for the combination and extrapolation of evidential information provided by disparate sources of knowledge. They do so in a formally consistent way, guaranteeing the integrity of their predictions.

Dependency-graph models of evidential support, though not a panacea, do offer some significant advantages over the previously developed systems for inexact reasoning.

Areas for Further Investigation

This final section suggests some areas for further investigation, prompted by this thesis. One of the most obvious shortcomings of this work is the absence of a decision rule. Presumably, one who models an evidential domain is ultimately interested in making a decision based on evidential information. As was recently observed [Barnett 1981], a Bayesian approach to evidential reasoning initially suppresses ignorance before any decisions are made, by requiring that precise point probabilities be specified from the onset. On the other hand, a Dempster-Shafer appraoch preserves varying degrees of ignorance, complicating the decision process. Although a decision rule has not been developed specifically for this formalism, some closely related work is being performed on decision making in the context of fuzzy sets [e.g., Zadeh 1976; Orlovsky 1978; Watson, Weiss, and Donnell 1980; Adamo 1980; Takeda and Nishida 1980; Tong and Bonissone 1980; Freeling 1980]. It appears that a support-based decision rule might eventually be defined as a special case of a fuzzy decision rule.

Dempster's rule of combination makes some assumptions that are not always justified. In particular, Dempster's rule assumes that bodies of evidence are uniformly reliable and independent. The former is more easily addressed than the latter. If some bodies of evidence are from more reliable sources than others, the mass functions representing them might be maniuplated, prior to any evidential reasoning, to reflect their relative reliabilities; mass functions representing less reliable information could be renormalized, with proportionally less of their mass attributed to propositions other than Θ. Thus, when they are combined, information from more reliable sources would take precedence over information from less reliable sources.

The other problem, evidential dependence, does not seem to lend itself to any simple solution. However, this is not surprising since the same problem is evident in the competing theories. This remains an open question.

Although a general scheme has been outlined for moving from an evidential domain to its representation as a dependency graph, considerably more work needs to be done. When constructing a dependency-graph model, there are a large number of representational alternatives to choose among. No criteria for making these decisions has been identified, other than internal consistency. This is a problem when constructing any axiomatic system, but the problem is particularly acute within this new context.

This entire exposition has assumed that the domains of interest can be modeled within a propositional framework. In general, this is not the case. When a domain is sufficiently large and complex, a propositional framework is, at best, awkward. Large, complex domains are better modeled in a predicate framework, where truths can be expressed relative to portions of the environment, and general statements made about the implications of these relative truths. Quantified variables and functions make this possible. Extending this work into a predicate framework would certainly extend its range of applicability, though its current range of applicability has yet to be determined.

Finally, these techniques need to be applied to some more ambitious tasks. Work has already begun on applications to sensory-based situation assessment [Garvey, Lowrance, and Fischler 1981] and visual scene analysis [Wesley 1982]. Both of these have shown promising initial results, but it is too early to draw any definite conclusions. The success of these applications will most heavily depend on the ability of their designers to construct the large, complex, dependency-graphs that are needed to capture the domain knowledge. As these graphs become larger and more complex, it becomes increasingly difficult to guarantee their logical consistency. Of course, the same problem exists when constructing any large axiom base. Is the next axiom that is to be added to the base consistent with the axioms already in the base? The question is decidable, but the computational load is significant. However, any initial investment to ensure consistency is repayed many times over in terms of increased system reliability and flexibility.

BIBLIOGRAPHY

Adamo, J. M. (1980). "Fuzzy Decision Trees." Fuzzy Sets and Systems, 4, 207-219, North-Holland Publishing Company.

Barnett, J. A. (1981). "Computational Methods for a Mathematical Theory of Evidence." In IJCAI81.

Bayes, T. (1763). "An Essay Towards Solving a Problem in the Doctrine of Chances," Philosophical Transactions of the Royal Society, 53, 370-418.

Boole, G. (1854). An Investigation of the Laws of Thought, London. Reprinted in 1958 by Dover, New York.

Carnap, R. (1950). "The two concepts of probability." In Logical Foundations of Probability. 19-51. University of Chicago Press, Chicago, Ill. (1950).

Dempster, A. P. (1967). "Upper and lower probabilities induced by a multivalued mapping." Annals of Mathematical Statistics, 38, 325-339 (1967).

Dempster, A. P. (1968). "A generalization of Bayesian Inference." Journal of the Royal Statistical Society, Series B, 30, 205-247 (1968).

Duda, R. O., Hart, P. E., and Nilsson, N. J. (1976). "Subjective Bayesian methods for rule-based inference systems." In Proc. 1976 National Computer Conference (AFIPS Conference Proceedings), 45, 1075-1082.

Duda, R. O., Hart, P. E., Nilsson, N. J., and Sutherland, G. L. (1977). "Semantic Network Representations in Rule-Based Inference Systems." In Pattern-Directed Inference Systems (D. A. Waterman and F. Hayes-Roth, eds.), 203-221, Academic Press (1977).

Erman, L. D. and Lesser, V. R. (1975). "A multi-level organization for problem solving using many diverse cooperating sources of knowledge." In Proc. 4th Int. Jt. Conf. Artificial Intelligence, Tbilisi, USSR, 483-490.

Fine, T. L. (1973). Theories of Probability. Academic Press, New York.

Fisher, R. A. (1956). Statistical Methods and Scientific Inference, Hafner.

Freeling, A. N. S. (1980). "Fuzzy Sets and Decision Analysis," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-10, no. 7.

Friedman, L. (1980). "Reasoning by Plausible Inference." Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California.

Garvey, T. D., Lowrance, J. D., and Fischler, M. A. (1981). "An Inference Technique for Integrating Knowledge from Disparate Sources." In IJCAI81.

Goguen, J. A. (1968). "The logic of inexact concepts." Synthese, 19, 325-373 (1968-1969).

Gorry, G. A. and Barnett, G. O. (1968). "Experience with a model of sequential diagnosis." Comput. Biomed. Res., 1, 490-507 (1968).

Gorry, G. A. (1973). "Computer-assisted clinical decision making." Meth. Inform. Med., 12, 45-51 (1973).

Hempel, C. G. (1945). "Studies in the logic of confirmation." In Aspects of Scientific Explanation and Other Essays in the Philosophy of Science, 3-51, The Free Press, N.Y. (1965).

Konolige, K. (1979). "Bayesian Methods for Updating Probabilities." In "A Computer-Based Consultant for Mineral Exploration," Final Report, SRI Project 6415, (R. O. Duda, P. E. Hart, K. Konolige, and R. Reboh), Artificial Intelligence Center, SRI International, Menlo Park, California, 83-146.

Lowrance, J. D. (1976). "GRASPER 1.0 Reference Manual." COINS Technical Report 78-20, University of Massachusetts, Amherst, MA.

Lowrance, J. D. and Corkill, D. D. (1979). "The Design of GRASPER 1.0: A Programming Language Extension for Graph Processing." COINS Technical Report 79-6, University of Massachusetts, Amherst, MA.

Luce, R. D. and Suppes, P. (1965). "Preference, utility, and subjective probability," In Handbook of Mathematical Psychology (R. D. Luce, R. R. Bush, and E. Galanter, eds.), 249-410, John Wiley & Sons, N.Y. (1965).

McSkimin, J. (1976). The Use of Semantic Information in Deductive Question-Answering Systems. Ph.D. Thesis, Department of Computer Science, Univ. of Md., College Park, MD.

Nilsson, N. J. (1980). Principles of Artificial Intelligence. Tioga Publishing Company, Palo Alto, CA.

Orlovsky, S. A. (1978). "Decision-Making with a Fuzzy Prefernece Relation." Fuzzy Sets and Systems, 1, 155-167, North-Holland Publishing Company.

Rosenfeld, A., Hummel, A., and Zucker, S. (1976). "Scene Labelling by Relaxation Operations." IEEE Transactions on Computers, pp. 562-569.

Shackle, G. L. S. (1952). Expectation in Economics. Cambridge Univ. Press, Cambridge, U.K. (1952).

Shackle, G. L. S. (1955). Uncertainty in Economics and Other Reflections. Cambridge Univ. Press, Cambridge, U.K. (1955).

Shafer, G. (1973). Allocations of Probability. Princeton doctoral dissertation, available from University Microfilms, Ann Arbor.

Shafer, G. (1975). "A Theory of Statistical Evidence." In Foundations and Philosophy of Statistical Theories in the Physical Sciences, Vol. II. (W. L. Harper and C. A. Hooker, eds.), Reidel.

Shafer, G. (1976). A Mathematical Theory of Evidence. Princeton University Press, Princeton, New Jersey.

Shortliffe, E. H. (1974). "MYCIN: A rule-based computer program for advising physicians regarding antimicrobial therapy selection." Memo AIM-251, Artificial Intelligence Lab., Stanford University, Stanford, CA.

Shortliffe, E. H. (1976). Computer-Based Medical Consultations: MYCIN. Elsevier, New York.

Shortliffe, E. H. and Buchanan, B. G. (1975). "A model of inexact reasoning in medicine." Math. Biosci., 23, 351-379 (1975).

Takeda, E. and Nishida, T. (1980). "Multiple Criteria Decision Problems with Fuzzy Domination Structures." Fuzzy Sets and Systems, 3, 123-136, North-Holland Publishing Company.

Tong, R. M. and Bonissone, P. P. (1980). "A Linguistic Approach to Decisionmaking with Fuzzy Sets," IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-10, no. 11.

Tornebohm, H. (1966). "Two measures of evidential strength." In Aspects of Inductive Logic (J. Hintikka and P. Suppes, eds.), 81-95, North-Holland Publishing Co., Amsterdam (1966).

Trigoboff, M. (1976). "Propagation of information in a semantic net." In Proc. AISB Summer Conference, Edinburgh, Scotland, 334-343.

Tversky, A. (1972). "Elimination by aspects: a theory of choice." Psychol. Rev., 79, 281-299 (1972).

Waltz, D. G. (1972). "Generating Semantic Descriptions from Drawings of Scenes with Shadows." AI TR-271, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

Warner, H. R., Toronto, A. F., and Veasy, L. G. (1964). "Experience with Bayes' Theorem for computer diagnosis of congential heart disease." Anns. N.Y. Acad. Sci., 115, 558-567 (1964).

Watson, S. R., Weiss, J. J., and Donnell, M. L. (1979). "Fuzzy Decision Analysis." IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-0, no. 1.

Wesley, L. P. (1982). "The Use of an Evidential-Based Model for Representing Knowledge and Reasoning about Images in the VISIONS System." In Proc. Workshop on Computer Vision: Representation and Control, Franklin Pierce College, Rindge, New Hampshire.

Zadeh, L. A. (1965). "Fuzzy Sets." Information and Control, 8, 338-353 (1965).

Zadeh, L. A. (1976). "The Linguistic Approach to Decision Analysis." Directions in Large-Scale Systems, Y. C. Ho and S. K. Mitter, eds., Plenum Press.

Zadeh, L. A. (1979). "On the Validity of Dempster's Rule of Combination of Evidence." Memorandum No. UCB/ERL M79/24, Electronics Research Laboratory, College of Engineering. University of California, Berkeley, CA.