Representativeness and Uncertainty in Classification Systems

Paul R. Cohen, Alvah Davis, David Day,
Michael Greenberg, Rick Kjeldsen,
Susan Lander and Cynthia Loiselle
COINS Technical Report 85-26

Department of Computer and Information Science
University of Massachusetts, Amherst

EKSL Report 5

This research is funded by a grant from the National Science Foundation IST 8409623 and DARPA-RADC Contract F30602-85-C-0014.

Representativeness and Uncertainty in ClassificationSystems

Paul Cohen, Alvah Davis, David Day, Michael Greenberg, Rick Kjeldsen, Susan Lander, and Cynthia Loiselle

Experimental Knowledge Systems Laboratory, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts 01003

Abstract

The choice of implication as a representation for empirical associations and for deduction as a mode of inference requires a mechanism extraneous to deduction to manage uncertainty associated with inference. Consequently, the interpretation of representations of uncertainty is unclear. Representativeness, or degree of fit, is proposed as an interpretation of degree of belief for classification tasks. The calculation of representativeness depends on the nature of the associations between evidence and conclusions. Patterns of associations are characterized as endorsements of conclusions. We discuss an expert system that uses endorsements to control the search for the most representative conclusion, given evidence.

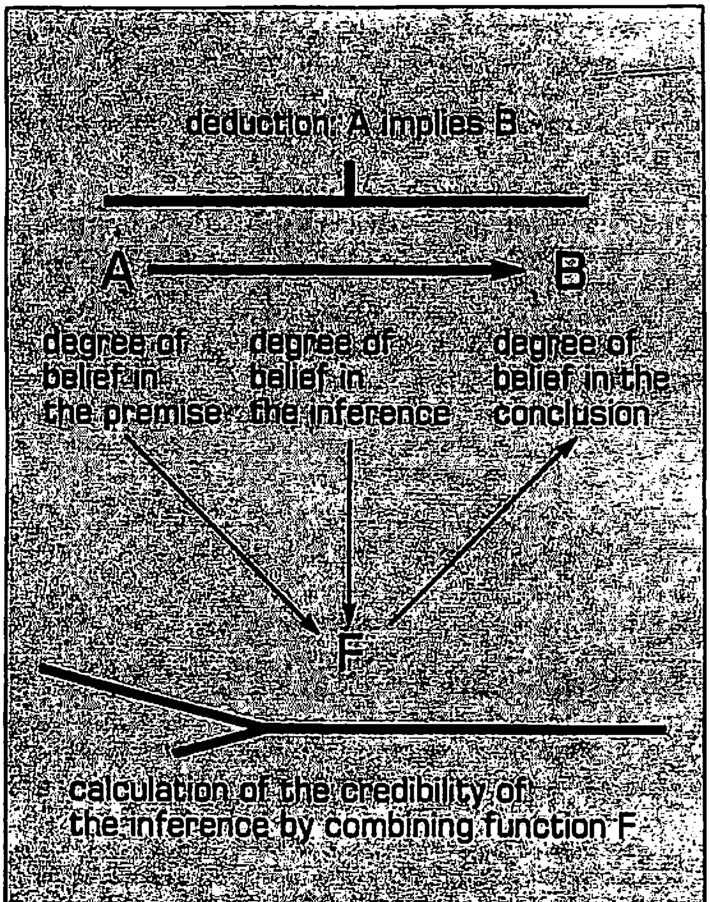
Tasks can be classified by the kinds of uncertainty that characterize them. Planning tasks, for example, are characterized by uncertainty about the interactions of plan steps. Strategic planning is further characterized by uncertainty about the intentions and actions of an opponent. Perception is characterized by too much data too noisy for bottom-up interpretation, and ambiguous with respect to top-down models. The subject of this article is *classification*, an important task for many AI systems. Most expert systems are classification problem solvers. They heuristically associate data with one or more known solutions; the problem is to match data with the solution that explains them best (Clancey, 1984).

Uncertainty in classification problem solving has two major sources. The first is that data may be inaccurate or incomplete, and the second is partial matching. This article is not concerned with the quality of data; we focus instead on uncertainty inherent in the design and behavior of classification systems. The partial matching problem has two forms, easily illustrated by the following common, empirical association: *A person with a queasy stomach, fatigue, aching limbs, and a fever has flu in its early stages.* Now consider a person with a marginal fever, complaining of poor appetite, headache, and a persistent twitch in his left eye. This case seems to exhibit manifestations not stated in the rule for flu and fails to display manifestations that are so stated. We are uncertain whether the person has flu for two distinct reasons: we cannot be certain that the actual symptoms fail to match the stated ones (Does "marginal fever" count as a fever? Does "headache" count as aching limbs?); and we cannot be certain that the rule for flu includes all and only the relevant manifestations of flu.

This article is concerned with the representation of uncertainty due to partial matching in classification systems. Many classification systems represent uncertainty with a number or range, but the interpretation of these numbers is unclear, due in part to the way that empirical associations themselves are represented. Rules, or *productions*, are the representation of choice for empirical associations in classification systems. They are modular, modifiable, and easily understood; and since they are logical implications, *deduction* is the obvious mode of inference for rule-based systems. But against these advantages we must weigh the assumptions that implication is an adequate representation of empirical association and that deduction is an adequate mode of inference for associative reasoning. Subscription to these assumptions leads to a *parallel certainty inference* approach to reasoning under uncertainty (Cohen, 1983). Since deduction is not defined for statements that are neither true nor false, deductive inference

This work is supported by NSF Grant IST-8409623 and DARPA-RADC Contract F30602-85-C-0014. We wish to acknowledge Bruce McCanless and Marg Burggren for their patient assistance and Daniel Corkill for his advice on drafts of this article.

in uncertain domains maintains an uncomfortable, parallel coexistence with calculations of the certainty of deductions (see Figure 1). While deduction preserves the meaning of logical statements, it sheds no light on the meaning of statements qualified by degrees of belief. The information we need to interpret the degree of belief in an empirical association—the *type* of the association, be it causal, correlational, based on physical connectivity, or whatever—is thrown away when the association is cast into the uniform mold of implication.



Parallel Certainty Inference.

Figure 1.

The associations that underlie an inference, if represented explicitly, permit us to reason about the credibility of the inference. For instance, since stress and hypertension are causally associated, we can infer that a person under stress will develop hypertension. Further, a person under a particular kind of stress, say job-related pressure might develop hypertension. The associations KIND-OF and CAUSE are the basis for inferring hypertension from job pressure, and also determine the credibility of that inference. Given this information, we can write rules of plausible inference (Collins, 1978) that are similar to rules of deductive inference:

$$\frac{\begin{array}{c}\text{CAUSE(X,Y)}\\ \text{KIND-OF(Z,X)}\end{array}}{\text{CAUSE(Z,Y)}}$$

If X causes Y and Z is a kind of X, then Z also causes Y. The point is that the credibility of these inferences depends not on X, Y, or Z, but on our everyday understanding of the terms CAUSE and KIND-OF. These terms tell us everything we need to know about the credibility of the inference above, provided we know what they mean. A qualitative approach to reasoning about uncertainty requires a set of terms—the associative basis of inferences—and enough understanding of the meanings of the terms to write rules like the one above. We call the terms *endorsements* (Cohen and Grinberg, 1983; Cohen, 1983, 1984). This article describes a method for reasoning with endorsements that represent uncertainty due to partial matching in classification tasks.

We begin with a review of relevant literature, in which it becomes clear that the semantics of representations of uncertainty are poorly understood. Then we suggest that representations of uncertainty for classification systems should be interpreted in terms of the similarity or fit between data and solution, not in terms of the relative frequency of their co-occurrence. Lastly, we show how these ideas work in an expert system for matching research proposals to the appropriate sources of funding.

Representation of Uncertainty in AI Programs

Szolovits and Pauker (1978) contrast *categorical* reasoning, characterized by judgments "made without significant reservations," with *probabilistic* reasoning. They examine a Bayesian model of judgment and conclude that it is unrealistic for medical decision-making due to its "voracious demand for data" and the failure of the assumptions that reduce that demand. They describe four probabilistic approaches, but recognize that the numbers manipulated by these techniques have different interpretations. Two approaches (those of the PIP and INTERNIST systems) calculate degrees of fit to prototypes; a third bases its "probabilities" on causal arguments (the CASNET system); the fourth (MYCIN) uses numbers that are related to conditional probabilities. For all the commonality between the so-called probabilistic approaches, Szolovits and Pauker might have called their paper "categorical and non-categorical reasoning in medical diagnosis" and avoided the word probability altogether. We will adopt the term non-categorical in recognition of the lack of consensus in AI about what we mean by probability.

A second approach to non-categorical reasoning has been to design control structures that minimize the effects of uncertainty. Opportunistic control structures such as "island driving" are one manifestation of this general approach (Hayes-Roth and Lesser, 1977). Another is *diversification*. A system may, in recognition of unresolvable uncertainty, "cover" as many alternative hypotheses as possible. MYCIN took this approach to therapy recommendation (Shortliffe, 1974); FOLIO (Cohen and Lieberman, 1983) recommended portfolios that were diversified due to uncertainty about investors' goals. In all these cases, uncertainty is acknowledged but not given a probabilistic representation. Non-categorical reasoning is not necessarily probabilistic. But since probability is closely tied to uncertainty in many AI systems, several interpretations of the concept are worth examining.

If categorical reasoning is judgment without significant reservations, then non-categorical reasoning suggests qualified judgment. This is apparently the sense in which the term "probability" was originally used:

> What we now call the mathematical theory of probability was originally called the theory of games of chance. Probability was an entirely different topic; something was probable when there was good argument or good authority for it. When James Bernoulli and others began to use the word probability in connection with the theory of games of chance, they were expressing the ambition that this theory might provide a general framework for evaluating evidence and weighing arguments. (Shafer, 1984, p. 7)

Recently, we have seen two proposals that noncategorical reasoning should be probabilistic in the *original* sense of being supported by good argument or authority. Cohen and Grinberg (1983) and Cohen (1983) outline a theory of *endorsements*, which are reasons to believe and disbelieve propositions and Doyle delineates theories of *reasoned assumptions* (Doyle, 1983a, 1983b).

In contrast to these approaches, AI has developed several techniques that are numeric but not probabilistic in the sense of games of chance. These include certainty factors (Shortliffe and Buchanan, 1984), fuzzy set theory (Zadeh, 1975), and the theory of belief functions (Shafer, 1976; Lowrance and Garvey, 1982). It is unclear how to interpret the numbers used in these techniques. They are subjective, in the sense that they reflect aspects of beliefs about events, not aspects of events themselves. They are constructed by psychological effort and are probably an amalgam of several considerations, including utility and relative frequency (Buchanan and Shortliffe, 1984, p. 217; Bar Hillel, 1982). The numbers are probably not accurate, in the sense that an expert may assess different numbers for the same situation. Studies conducted by the MYCIN group suggest that the numbers can be modified by as much as 20% without significantly changing the certainty-ranking of hypotheses (Buchanan and Shortliffe, 1984, p. 219). The numbers are used most often in conditioning designs (Shafer and Tversky, 1985), where degrees of belief in hypotheses are conditioned on evidence. There is a close syntactic affinity between a conditional probability and a degree of belief in a conditional statement such as an inference rule. Thus numbers are sometimes combined as if they were conditional probabilities, although they are not. Any interpretation of the numbers may or may not be preserved by the functions that combine them. This concern is especially germane since humans (even statistically skilled ones) produce "probabilities" conditioned on evidence that vary sharply from those produced by bayesian conditioning rules (*e.g.*, Tversky and Kahneman, 1982, p. 35; Eddy, 1982).

These considerations are not intended to condemn the use of numbers in programs that reason under uncertainty. Provided the meaning of the numbers is clear and preserved by their rules of combination, we have no concern. But numbers are a representation of last resort in AI: when we know what they mean, we abandon them for more symbolic representations. We use numbers only because we do not know what they represent. The remainder of this article proposes a symbolic representation—and its interpretation—for uncertainty in classification systems.

Representativeness

We suggest that the interpretation of probability in classification systems should be in terms of similarity, not in terms of games of chance. This interpretation has precedent in some frame-based expert systems (*e.g.*, PIP and INTERNIST) and in psychological literature, where it is called the *representativeness* heuristic:

> Many of the probabilistic questions with which people are concerned belong to one of the following types: What is the probability that object A belongs to class B? What is the probability that event A originates from process B? What is the probability that process B will generate event A? In answering these questions, people typically rely on the representativeness heuristic, in which probabilities are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B. (Tversky and Kahneman, 1982, p. 4)

Assessments of subjective probability in classification situations are insensitive to factors that affect probability (such as prior probability distributions) and sensitive to the resemblance between data and their classification. For example, Kahneman and Tversky asked subjects to classify individuals as librarians or truck drivers on the basis of personality sketches. They found that the classification was insensitive to the prior distribution of librarians and truck drivers in the population. An individual described as "neat, methodical, and shy" was classified as a librarian even if the prior probability of being a librarian was low. Remarkably, subjects ignored prior probability even when the personality sketches were completely uninformative, assessing a probability of 0.5 for each alternative instead. Translating these results to the expert systems literature, we would expect degrees of belief in heuristic associations between data and solutions—often represented as conditional probabilities—to be interpreted not in terms of relative frequency, but in terms of the degree to which data are representative of a solution. We might hope that experts would use probabilistic information more efficiently than novices, but evidence suggests that experts are as prone to judgment by representativeness as the rest of us (Kahneman and Tversky, 1982, p. 35).

Intuitively, the degree to which evidence is representative of a conclusion determines the credibility of the conclusion given the evidence. But if representativeness is to be useful as an interpretation of uncertainty in AI programs, we need a way to measure it. Kahneman and Tversky do not specify its determinants (though see Bar Hillel, 1982, for a domain-specific attempt). Our approach is to represent propositions as structured objects and to measure representativeness in terms of the nature of the associations between structures. This is illustrated in Figure 2, which shows four propositions in a network representation:

P-1: *tobacco CAUSE cancer*

P-2: *cigarettes CAUSE cancer*

P-3: *tobacco PART-OF cigarettes*

P-4: *cigarettes HAS-PART tobacco*

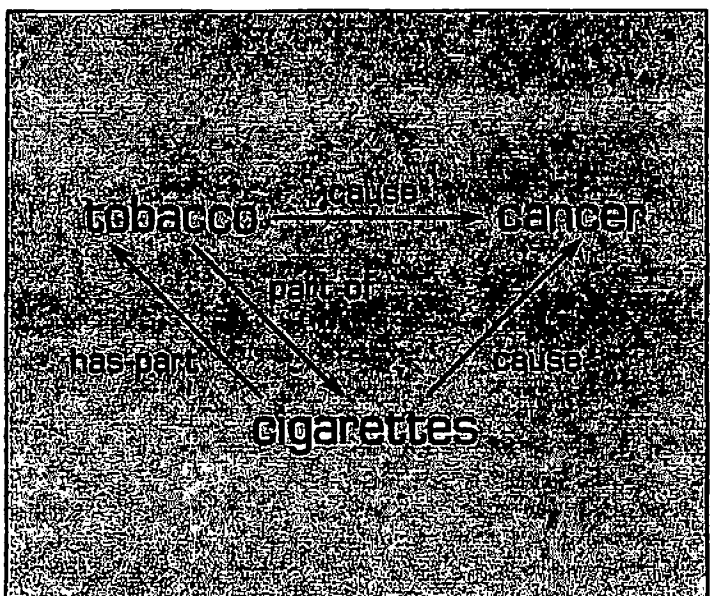


Figure 2

Given a common interpretation of PART-OF and its inverse HAS-PART, P-1 is credible evidence for P-2; but P-2 is not credible evidence for P-1, since some other PART-OF of cigarettes (besides tobacco) could be responsible for cancer. The PART-OF association that holds between tobacco and cigarettes determines how representative P-1 is of P-2; similarly, the HAS-PART relationship determines how representative P-2 is of P-1. We say that if PART-OF is the basis of an inference (*e.g.*, P-1 → P-2), then the inference is credible to the extent that PART-OF *preserves representativeness*. If we know a priori that PART-OF preserves representativeness better than HAS-PART, then P-3 and P-4 tell us all we need to know to rank by credibility the inferences P-1 → P-2 and P-2 → P1.

Representativeness, defined in terms of the associations between structures, is an appropriate representation for uncertainty caused by partial matches. We say that P-1—*tobacco causes cancer*—partly matches P-2, the proposition *cigarettes cause cancer*. The degree of match is determined by the PART-OF association between tobacco and cigarettes, and any uncertainty introduced by using P-1 as evidence for P-2 is captured fully by the PART-OF association.

Given this, we can interpret the subjective probability of a conclusion given evidence in terms of the representativeness, or match, between them. For example, the subjective likelihood that a grant proposal (P) will be funded by an agency (A) depends on the degree of overlap between the research interests of P and A, the match between the desired level of funding for P and the typical award size of A, the degree to which P meets any geographic or demographic requirements of A, and so on[1]. In section 4 we discuss an expert system for matching research proposals to the agencies most likely to fund them. But for now, we will use the words *proposal* and *agency* more loosely: a proposal is an arrangement of evidence, and an agency is one of the conclusions that can be drawn from the evidence. We might equally well talk about symptoms and diseases or core samples and lithographic strata. The theme is the same: the most likely of several conclusions is the one that best matches the available evidence.

We have adopted a network formalism to facilitate judgments of the match between evidence and conclusions. Figure 3 shows a proposal (P2) to study the effectiveness of methadone as a treatment (called an INTERVENTION in our system) for heroin addiction. An agency (A2) wants to support work on psychological counseling for drug ad-

[1] The likelihood of an agency funding a proposal also depends on the prior probability of funding. Failure to account for this factor leads to the same error as in Kahneman and Tversky's Librarian example. We thank Dan Corkill for this observation. The GRANT system, discussed below, is based on the assumption that all agencies have the same prior probability of funding a proposal. GRANT can be thought of as updating this probability based on information about the research topic of the proposal.

dicts. Is the match between P2 and A2 a good one? If A2 is interpreted as the proposition, "Someone wants to fund the study of psychological counseling for drug addicts," is P2 evidence for this proposition, and how strong is the evidence? The mismatches between P2 and A2 are two: P2 mentions methadone while A2 focuses on psychological counseling, and P2 is interested in heroin addicts in contrast to A2's drug addicts. The credibility of P2 as evidence for A2 clearly depends on the relationships that hold between the components of P2 and A2.

Since heroin addiction ISA drug addiction, an interest in the latter is credible evidence of an interest in the former. Methadone treatment often has psychological counseling as a component, so the evidence (methadone treatment) is related to the conclusion (psychological counseling) by HAS-PART. Earlier, in the context of Figure 2, we said that the HAS-PART relationship preserved representativeness relatively poorly. But in this case, it seems reasonable to infer an interest in psychological counseling from an interest in methadone treatment, given the HAS-PART relation that holds between them. This apparent contradiction illustrates that the credibility of an inference actually depends on one factor in addition to the nature of the associations on which it is based: the nature of the inference itself. We can infer the *existence* of an object given that it is PART-OF another object that exists, but we cannot infer a *property* of an object given that it is PART-OF another object with that property. If cigarettes are carcinogenic, we cannot infer that tobacco is carcinogenic given only that cigarettes HAS-PART tobacco. Similarly, if methadone treatment is addictive, we cannot infer psychological counseling is addictive, given only that methadone treatment HAS-PART psychological counseling. This said, we confine ourselves for the rest of this article to inferences about the existence of things, or rather, to inferences about the existence of an *interest* in things. Although HAS-PART does not preserve representativeness for inferences about properties of things, we may credibly infer that a funding agency is interested in methadone treatment if it is interested in psychological counseling and methadone treatment HAS-PART psychological counseling.

On the basis of these considerations, we conclude that P2 is representative of A2. Given P2, we may infer A2, and any uncertainty in the inference is captured by the associations ISA and HAS-PART. These "pathways" of associations between evidence and a conclusion determine the credibility of the conclusion given the evidence. They play an important role in our approach; we call them ***path endorsements***. Endorsements are reasons to believe and disbelieve propositions and are the basis of explanations and control decisions in uncertain reasoning (Cohen, 1983). Path endorsements, too, summarize reasons to believe and disbelieve. They tell us whether a proposition is acceptable evidence for another, if not why not, and they provide a basis for control of reasoning. We will see that they are derived directly from the representation language of a domain, so they characterize uncertainty in a domain in the same language as is used to describe conclusions. This is an improvement over previous manifestations of the theory of endorsements, in which endorsements were statements in a language extraneous to the domain and justified on intuitive grounds.

Path endorsements sometimes specify that an associative path is *not* representative and cannot be the basis of a credible inference. Consider two propositions. A3 is "Someone wants to study the effectiveness of counseling for treating eating disorders" P3 is "Someone wants to study the effectiveness of counseling for psychological disorders." These propositions may be illustrated as network diagrams as in Figures 2 and 3; equivalently, they may be represented as frames:

```
(P3 ISA instance of INTERVENTION with
        TYPE    =counseling
        MANAGE=psychological-disorder)

(A3 ISA instance of INTERVENTION with
        TYPE    =counseling
        MANAGE=eating-disorder)
```

Is P3 evidence for A3? As before, the answer depends on the relationships that hold between the components of P3 and A3. Both name *counseling* as the TYPE of INTERVENTION, but they differ on the phenomenon the counseling is intended to MANAGE. P3 mentions psychological disorders; A3 is more specifically interested in eating disorders. In Figure 3 we noted that heroin addiction was an instance of, and thus evidence for, drug addiction. In contrast, psychological disorders are *not* an instance of eating disorders; the opposite is true. The difference between the cases is easily summarized:

```
MANAGE(P2)  =  heroin-addiction
MANAGE(A2)  =  drug-addiction
----------------------------------
thus, ISA(MANAGE(P2),MANAGE(A2))

MANAGE(P3)  =  psychological-disorder
MANAGE(A3)  =  eating-disorder
----------------------------------
thus, ISA-INVERSE(MANAGE(P3),MANAGE(A3)).
```

An ISA link was the basis of the inference P2 $\rightarrow$ A2, but an ISA-INVERSE link is the basis of the inference P3 $\rightarrow$ A3. The one corresponds to inferring a superclass given evidence of the subclass, the other corresponds to the opposite direction of inference. In general, one can credibly infer the existence of a superset (psychological disorders) from a subset (eating disorders) but not vice versa. ISA preserves representativeness but ISA-INVERSE does not. Consider this example:

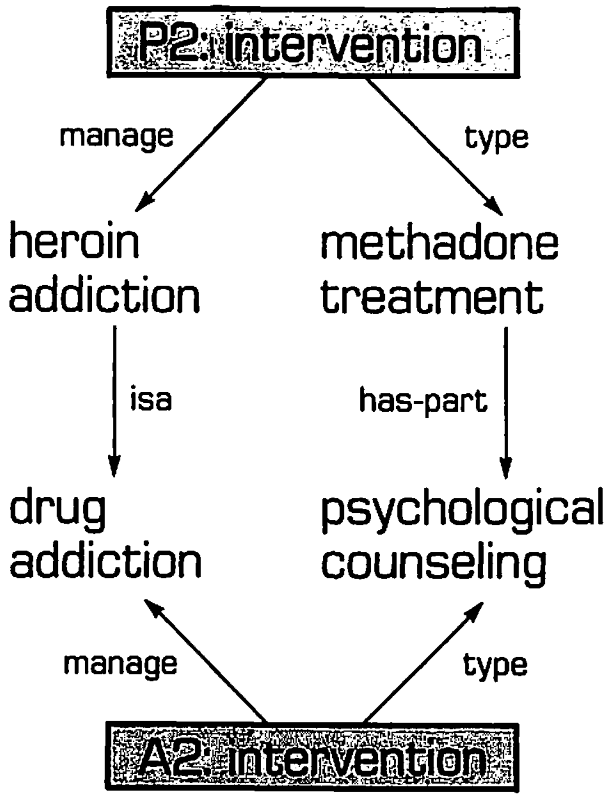


Figure 3.

$$\text{autism} \xrightarrow{\text{ISA}} \text{mental disorder} \xrightarrow{\text{ISA-INV}} \text{schizophrenia}$$

Schizophrenia and autism are both instances of psychological disorders. It is reasonable to infer an interest in psychological disorders given an interest in autism, but we must be careful not to proceed further to infer an interest in schizophrenia. We know that an interest in autism does not credibly imply an interest in schizophrenia because the path between them includes an ISA-INVERSE link. The path ISA::ISA-INVERSE is common enough to warrant a name—we call it a SIBLING endorsement. Any conclusion with a SIBLING endorsement is not credible. It is useful to know which chains of associations are the basis of credible inferences and which are not, since this information can be used to control search for credible conclusions in an associative network. We will return to this point shortly.

So far we have considered the degree of representativeness between individual components of a proposition. An analysis of Figure 3 showed that methadone treatment was credible evidence of psychological counseling, and heroin addiction was credible evidence of drug addiction; we did not say how to "sum" these pieces of evidence to compute an overall match between P2 and A2. We are currently implementing such a scheme, though it is not part of the program described in Section 4. We will sketch our approach here.

The overall degree of fit between P2 and A2 (Figure 3) depends on whether the components of P2 are acceptable evidence for A2 and on the importance of that evidence. We use path endorsements to determine the first criterion, and numerical weights for the second. For instance, if the emphasis of A2 is *primarily* on drug addiction, we might represent this emphasis with numbers:

(A2 instance of INTERVENTION with
    TYPE (= psychological counseling) (weight .2)
    MANAGE(= drug addiction) (weight .8))

The path endorsements of P2 (HAS-PART for the inference of psychological counseling given methadone treatment and ISA for the inference of drug addiction from heroin addiction) are both acceptable, so the overall degree of fit in the inference P2 → A2 is 1.0. The importance of the weights above is manifest when a proposal does *not* provide adequate evidence for all parts of an agency. Consider another proposal, P4, to study psychological counseling as a treatment for schizophrenia (Figure 4). The path from schizophrenia to drug addiction has the SIBLING endorsement, discussed above, so P4 is not adequate evidence of an interest in drug addiction. The path from psychological counseling (in P4) to psychological counseling (in A2) has the EQUAL endorsement, since they are the same concept. The degree of fit between P4 and A2, then, is

$$(.8 \times (\text{weight of SIBLING})) + (.2 \times (\text{weight of EQUAL}))$$

Currently, path endorsements have weights of 0 or 1, indicating that they support credible inference or don't. Given this, the overall credibility of the inference P4 → A2 is .2.[2]

Although endorsements can have arbitrary numeric weights, all our work is based on weights of 0 or 1. This is unrealistic—some credible associative paths are clearly more credible than others—but our research program is currently angled toward discovering reasons for credibility in the associations of a domain. We are more concerned with why associative inferences are credible than with the degree of their credibility.

Given that endorsements determine the acceptability of evidence, we can use them to control search in an associative network. Figure 5 shows a proposal (P5) to study management of hypertension by regulating dietary

[2]Although we have yet to explore the possibilities, statements about the weights of combinations of slots are also possible; for example, evidence for slots X and Y may individually be inadequate evidence for a frame, but adequate in conjunction; or, a slot may be "criterial," so that the fit between evidence and conclusion is zero lacking evidence for that slot, irrespective of the evidence for the other slots.

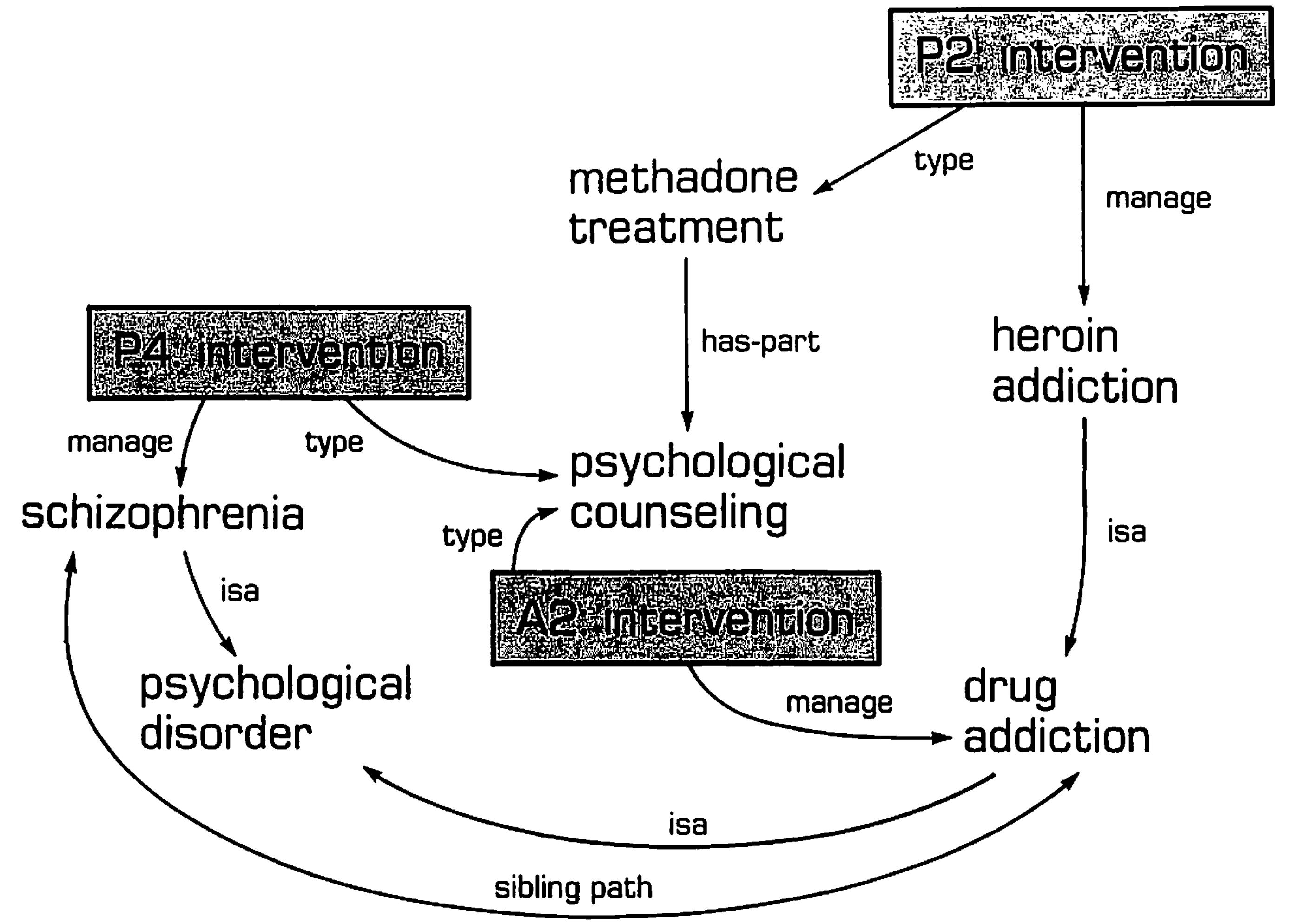


Figure 4.

sodium, and three agencies (A4, A5, A6) interested in nutrition, cardiovascular disorders, and psychological counseling to reduce stress, respectively. Four boundaries are drawn around these structures. The boundary around P5 is called the *evidence boundary* since it circumscribes the structure that represents evidence. The others are called *conclusion boundaries* and are drawn around A4, A5, and A6. If the evidence boundary can be extended by following pathways that have acceptable endorsements, until it includes all nodes in a conclusion boundary, then the degree of belief in the conclusion given the evidence is 1.0. If the evidence boundary cannot be extended to include *all* the nodes in the conclusion boundary, then the degree of belief in the conclusion given the evidence is a function of the relative weights of the nodes that are included and excluded, as described above.

Starting at the edges of the evidence boundary— hypertension and dietary-sodium—we consider which associated nodes might be included in an expanded evidence boundary. By the reasoning of earlier examples, ISA is considered an adequate path endorsement, and so the evidence boundary can be extended to include cardiovascular-disorder. As it happens, this is the sole interest of the agency A5, so the match between P5 and A5 is perfect. Stress is one CAUSE of hypertension, but since hypertension has several causes (one of which is dietary sodium), an interest in hypertension is not evidence for an interest in any single cause. (This is similar to the ISA-INVERSE case, discussed above.) Thus, stress is excluded from the evidence boundary. This suggests that agency A6 will be excluded also unless the evidence boundary can be extended by another path to stress, and/or the boundary can be extended to include psychological counseling. Given the network as shown, neither is possible.

The boundary can be extended, however, from the node dietary-sodium to the node nutrient by the acceptable path ISA. But before we can extend the evidence boundary further to include the node nutrition, we have to

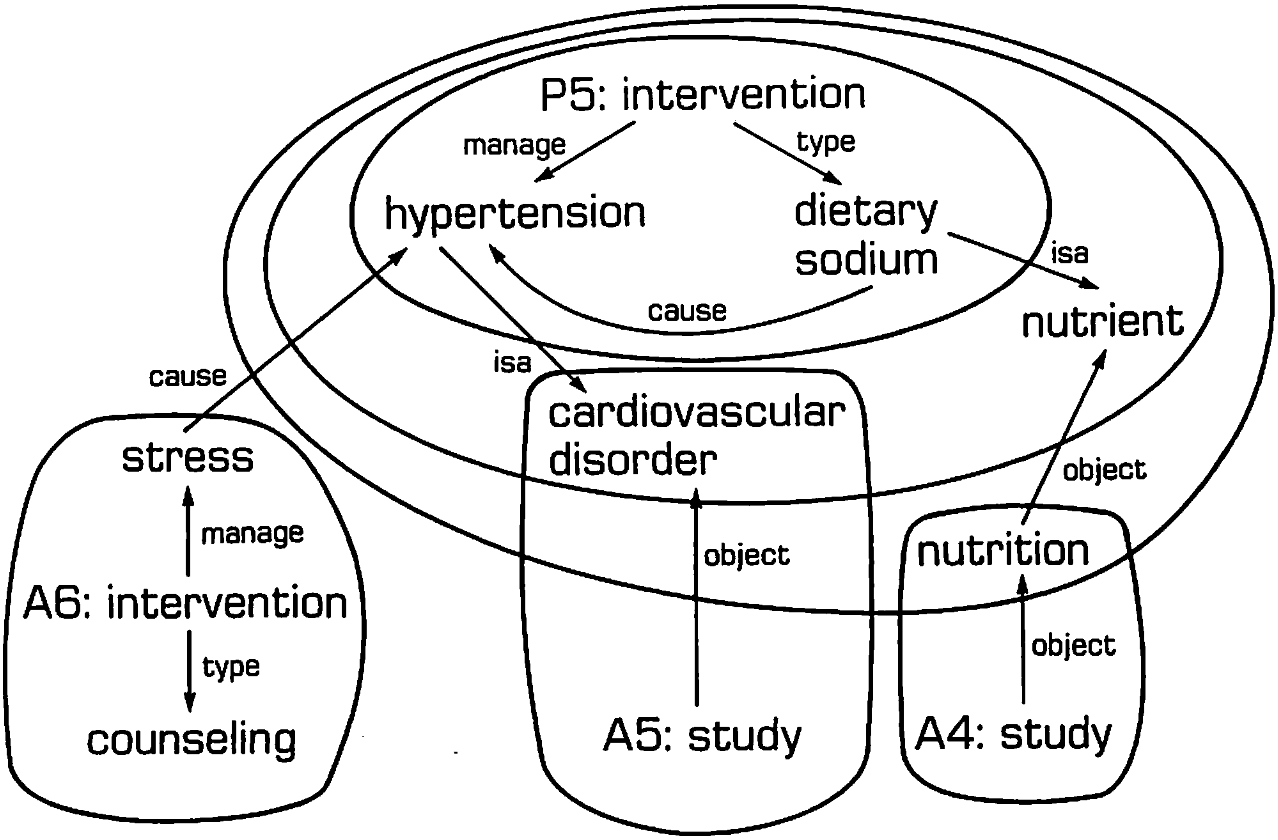


Figure 5.

consider whether the combination of paths ISA followed by OBJECT is acceptable. Would a funding agency consider an interest in nutrients as adequate evidence for an interest in nutrition? And what about an interest in one particular nutrient? Other, formally identical, questions can be posed to an expert in the domain of funding sources. If an agency is interested in drug addiction, is it likely to be interested in a particular drug? That is, if heroin ISA drug, and drug is the OBJECT of drug addiction, is an interest in heroin evidence for an interest in drug addiction? After posing this and similar questions to our expert, we believe that the path endorsement ISA:OBJECT, which denotes a sequence of inferences based on the ISA and OBJECT links, respectively, is adequate: we can infer an interest in nutrition from an interest in dietary sodium. This admits agency A4 into the evidence boundary.

This example illustrated that associative paths between evidence and conclusions can include multiple associations and that not all associative paths are the basis for credible inferences. A complex network of associations can support a great many inferences, and so the rules for extending the evidence boundary to include all and only credible conclusions are important. The next section describes an expert system that, by merit of 19 such rules, infers which federal agencies are most likely to fund a proposal.

GRANT: Expert Reasoning by Constrained Association

GRANT is an expert system that recommends funding sources for research projects. The system is being developed in collaboration with the Office of Research Affairs at the University of Massachusetts.[3] At the beginning of a consultation with GRANT, an investigator states his or her proposal, for example, "to study the effects of stress

[3]The director, Bruce McCandless, and his assistant, Marg Burggren, are our experts.

on hypertension in animal populations." Then GRANT finds and ranks all agencies that fund this and related research. Ideal performance involves finding one or two well-endowed funding sources that wish to fund exactly the proposed research. More often, no such agencies exist, but GRANT finds several that fit the proposal adequately. For example, an agency that wants to study "the causes of hypertension" is a good fit to the proposal just mentioned, even though it doesn't specify "stress" as the cause or "animal populations" as the experimental matrix. A poorer fit is to an agency that funds research in "cardiovascular disease." An agency that wants to fund research distantly related to stress (*e.g.*, "the stressful effects of rotating shifts") is a poor fit.

GRANT finds funding sources in two phases, called *proposal-directed* and *matching*, respectively. Proposal-directed search expands the evidence boundary around a proposal until it cannot be expanded further. During the matching phase, the agencies that fall within the evidence boundary are ranked by their degree of fit to the proposal, using a weighted sum measure as discussed above. This has been implemented in prototype only, so we limit our discussion to proposal-directed search. Proposal-directed search is constrained by the rules that expand the evidence boundary to find topics that are representative of the original proposal. Thus, GRANT finds agencies that fit the proposal as well as possible—agencies that are most *likely* to fund the proposal, given our interpretation of subjective probability in terms of representativeness.

Knowledge Representation

GRANT has two kinds of knowledge, a semantic network of topics in science (called the *topic network*) and a set of heuristic rules for searching this network. The representation of heuristics is described below. Nodes in the topic network are the concepts needed to express the research interests of funding agencies. Every agency is indexed to one or more nodes in the topic network. The topic network consists of about 800 nodes, sufficient to describe the research aims of the 50 agencies that fund the most research at the University of Massachusetts. The average branching factor of the network is about 4. A program called Buildnet helps with the addition of new agencies by prompting for typical agency information and keeping a stack of any concept used to define an agency that is absent from the topic network. A synonym facility allows the user to refer to the same concept in different ways (*e.g.*, salt, sodium, and dietary-sodium are currently synonyms).

The topic network contains the kinds of objects that populate the domain of funding sources and the associations that hold between them. Specifying the classes of objects and their interrelationships is tantamount to giving a case semantics for the language of funding sources. Figure 6 shows in tree form the objects we reason about. All objects are phenomena, and the ones that funding sources are most concerned with are states, processes, and things. States are typically the goals of an agency, such as safety and nourishment. Other states, such as illness and urban blight, although not strictly goals, are represented similarly. Processes have two forms, intentional and physical. An intentional process is done by someone with some purpose in mind. *Study* and *intervention* are important intentional processes. Studies include art history, biology, ..., zoology. Interventions include various therapies and legislative acts. Other intentional processes include cases where the intention is not very clear, such as smoking. *Covariance* is a physical process, and denotes the many situations in which one thing changes as the consequence of another. All grant proposals are represented as instances of *study, intervention* or *covariance.* For example, the proposal to study the effects of crowding-induced stress on hypertension in animal populations is represented this way:

```
(topic =
   (instance of covariance with
            (dependent-variable = hypertension)
            (independent-variable =
                (instance of stress with
                  (cause = crowding))
            (experimental-matrix = animals))))
```

Other physical processes include various disorders, such as anorexia, bulimia, cancer, and so on. The world of things is organized in a shallow but branchy tree. Living and non-living things are distinguished, as are plants, animals, and people.

These objects are not enough to represent the interests of funding sources or research proposals. In addition, and to provide us with a basis for endorsements, we need to know how these objects associate with each other. Figure 7 shows Figure 6 redrawn with superimposed labeled arcs between the objects. These are the associations known to GRANT. For example, every process has a SETTING, which must be a thing. Every intentional process has a WHO (the person who does it) and a WHO-FOR (the beneficiary). These associations map between intentional processes and people. Some associations are transitive in the sense that their domain and range are the same kind of object. For example, studies have foci, which are themselves studies: molecular biology is an instance of biology (a study) with a FOCUS of chemistry (another study). The granddaddy transitive association is ISA, which holds for any phenomenon. Others include CAUSE: one process causes another and so on. The case frame for each object is just the set of associations unique to the object, plus those associations that characterize the ancestors—in the tree of Figure 7—of the object. For example, the case frame for *study* includes FOCUS and FOCUS-INVERSE, SUBFIELD and SUBFIELD-INVERSE, and EXPERIMENTAL-MATRIX. It also includes

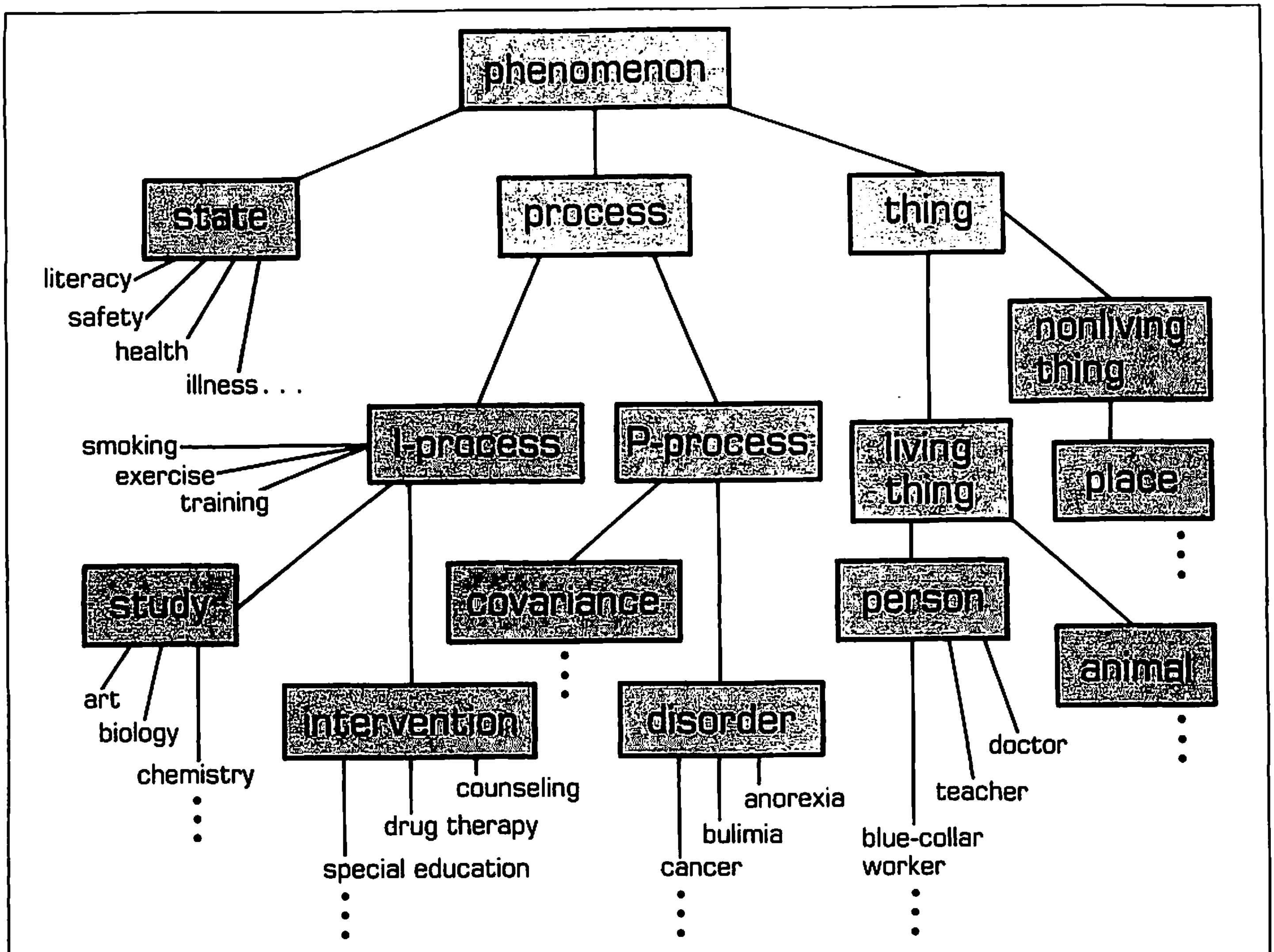


Figure 6.

OBJECT, PURPOSE, WHO, and other case relations inherited from *intentional-process*, as well as other relations inherited from *process* and *phenomenon*.

Path Endorsements

Path endorsements are the associative paths that GRANT might follow to expand its evidence boundary. Some path endorsements are positive, meaning that an inference based on such a path is credible. Some are negative. The SIBLING path endorsement, discussed above, is negative. A related negative path endorsement accrues to the path EFFECT followed by EFFECT-INVERSE. For example, if a researcher is interested in Type-A behavior, then she may be interested in one of its EFFECTs, such as hypertension. But she probably is not interested in something other than Type-A behavior that has hypertension as its EFFECT, such as dietary sodium:

Type-A behavior $\xrightarrow{\text{EFFECT}}$ hypertension $\xrightarrow{\text{EFFECT-INV}}$ dietary sodium

One cannot credibly infer that someone interested in Type-A behavior is interested in dietary sodium. The evidence boundary can be expanded from Type-A behavior to hypertension, but it should not be further expanded to dietary sodium. The endorsement on the path from Type-A behavior to dietary sodium is used to control the expansion of the evidence boundary.

This knowledge has both declarative and procedural representations. The declarative form of a path endorsement is

```
(defpath endorsement-name endorsement-class
    ([path endorsement-name]
     (step link-predicate node-predicate)
     [path endorsement-name] . . .
  . . .
```

One specifies the name of the endorsement and its

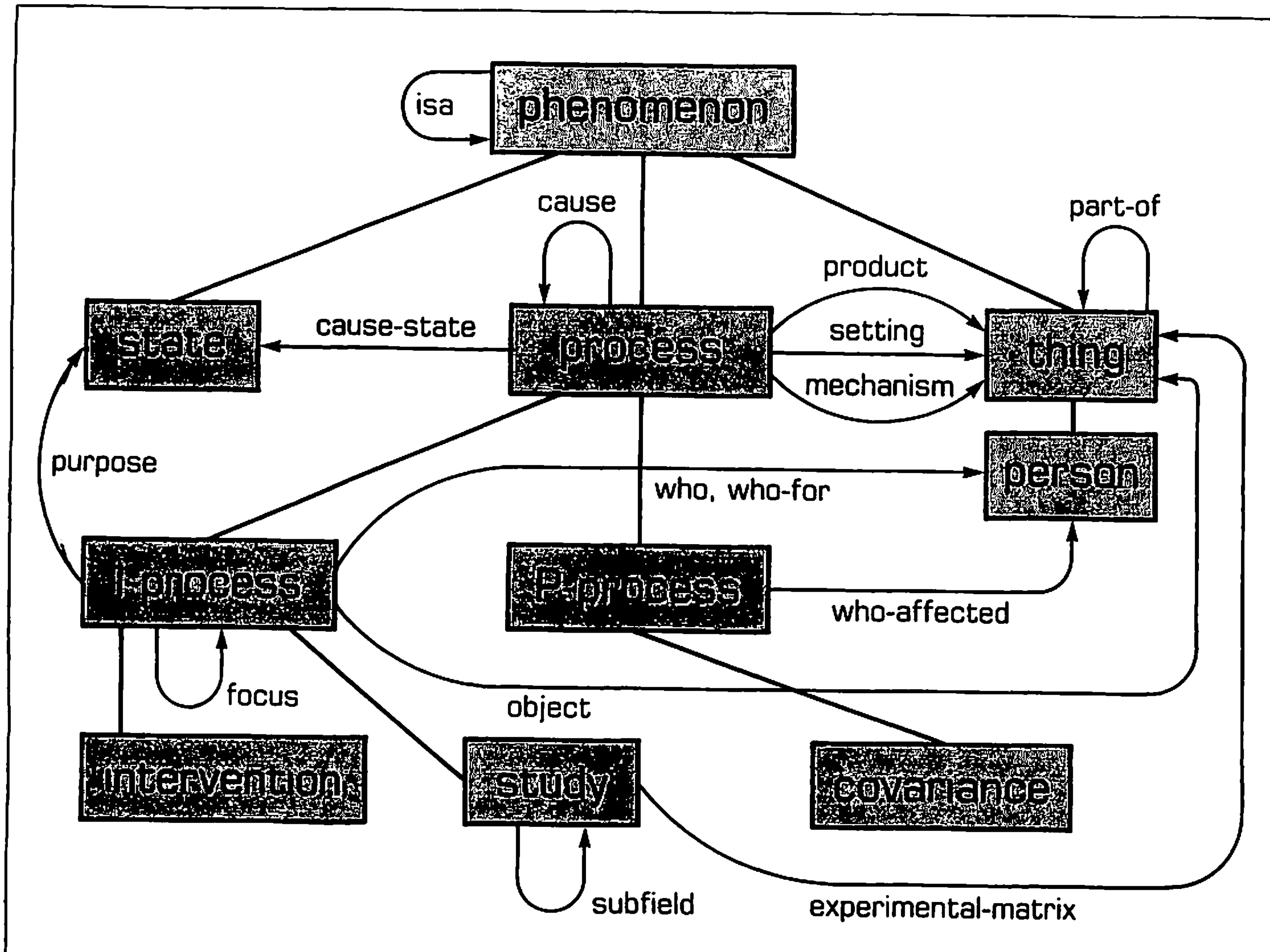


Figure 7.

class. In different versions of GRANT we have used the class variable both to specify the qualitative type of an endorsement (*e.g.*, endorsements based on transitive associations) and to specify the weight of the endorsement. The path itself is composed of an indefinite number of components. Each may include another, previously named, path endorsement at any point in its own specification. Each specifies a STEP from the current node over a link that satisfies LINK-PREDICATE to a node that satisfies NODE-PREDICATE. And each includes an optional repeat factor. For example, the EFFECT:EFFECT-INVERSE endorsement is represented this way:

```
(defpath sibling-effect trash
         (step effect all)
         (step effect-inverse all))
```

The sibling-effect endorsement belongs to the class trash and is composed of a step over an effect link to any node followed by a step over an effect-inverse link to any node. The classification "trash" means, in the current version of GRANT, that the path cannot be used to expand the evidence boundary. Another example is used to avoid including in the evidence boundary very general concepts:

```
(defpath general trash
         (step all (predicate general-nodep)))
```

That is, any link to a node that satisfies the predicate general-nodep will not be allowed to extend the evidence boundary. The nodes that satisfy this predicate in GRANT are thing, behavior, state, covariance, person, illness, and some others. It is vacuous to infer that because a researcher is interested in, say, anorexia, he is interested in illness. Moreover, since the fan-out of illness is very high, including the node in the evidence boundary makes GRANT work too hard to consider—and reject—the many extensions to the evidence boundary that might be made from the illness node.

The procedural form of this knowledge is a compiled

Rete net (Forgy, 1982) of all paths for which endorsements have been defined. Expanding a node is equivalent to adding a link to an extant path; this new path is discriminated through the Rete net of paths and its endorsement is returned.

Control Structure

GRANT first expands its evidence boundary as far as possible, then determines the best fit between the evidence and all solutions captured in the evidence boundary. These are the *proposal-directed* and *matching* phases, respectively. The matching phase is accomplished by a weighted-sum measure as described above, but since it has been implemented in prototype form only, we confine our discussion to proposal-directed search. Proposal-directed search has two versions. In one, we used an agenda to control the order in which nodes within the evidence boundary were expanded. The idea was to expand the nodes with good path endorsements before those with poor ones. Six classes of endorsements were used, of which "trash" was the worst. A general sibling (any link followed by its inverse) was also a poor endorsement. Good endorsements included those based on transitive relationships such as CAUSE and ISA. This form of proposal-directed search is essentially a best-first search in which the evaluation function returns the ordinal rank of the class to which an endorsement belongs. However, it requires an *a priori* ranking of endorsements. The second form of proposal-directed search requires just two classes of endorsements—acceptable and unacceptable. It expands all nodes with acceptable endorsements. The results described in the next section are based on this version of proposal-directed search.

Results

This section discusses a test of endorsement-constrained reasoning. Twenty-three proposals were selected from the files of the Office of Research Affairs at the University of Massachusetts. Proposal-directed search was run on these proposals in two modes, called *minimum-distance* and *endorsement-constrained*. In the minimum-distance (MD) mode, GRANT blindly expands the evidence boundary around a proposal. Each node in the proposal is expanded to all its neighbors, which are said to have a *radius* of 2. Each node at radius 2 is expanded to all its neighbors at radius 3, and so on. GRANT reports the agencies it finds at each radius. It continues expanding the evidence boundary until it finds at least 10 agencies. In fact, it finds an average of 15 agencies per proposal, since it reports all the agencies within the radius at which it finds its 10th agency. The minimum-distance method usually found its 10th agency at a radius of 4. It went beyond that radius 4 times for 23 proposals. In endorsement-constrained (EC) mode, GRANT expands the evidence boundary in accordance with path-endorsements. It reports agencies as it finds them, and it quits expanding the boundary when no node within the boundary can be expanded without incurring a negative path endorsement.

For each proposal, we asked our expert to rank the agencies found by MD search by the judged likelihood that each would fund the proposal. He preferred to classify every agency as good or bad. We asked for any agencies that *should* have been included in the list produced by MD search, but only once did our expert exercise this option. Thus, a blind MD search finds the agencies that the expert wants: its *hit rate* is 100%. But it also finds a large number of agencies that he doesn't want. That is, it eventually finds all the agencies that are judged representative of a proposal, as well as many unrepresentative agencies. As noted above, the average number of agencies found by MD search is 15. But on average, only 2 of these were judged good by our expert. Thus, the overall *false positive rate* is (15 - 2) / 15 = 87%.[4] These figures provide a standard against which to compare EC search. EC search preserves representativeness better than MD search to the extent that its hit rate is higher and its false positive rate is lower.

The hit rates for MD and EC search are 100% and 80%, respectively (see Table 1). MD search finds all the good agencies, EC search finds 4 out of 5. The false positive rates are 87% and 32%, respectively. Most of the agencies found by MD search are judged bad; a third of those found by EC search are bad. Note that the false positive rate for EC search *overall* is less than for MD search at radius 2. Thus, EC search is better able to discriminate representative from non-representative conclusions, even when the associative path between evidence (proposals) and conclusions (agencies) is very short.

Table 2 gives a clearer picture of the false positive rates. It shows the average number of agencies returned by MD search at each radius and the number of those agencies that were judged good. The ratios of these numbers are found in the *hits/try* column. These numbers are incremental, not cumulative. For example, at radius 2 MD search finds an average of 2.26 agencies of which 1.0 is good. At radius 3, it finds, on average, 2.44 *additional* agencies, of which an average of .57 is good. Thus the percentage of hits per try at radius 2 is 48%; at radius 3 the percentage is 23%; at radius 4 the percentage drops to 3%. Clearly, most good agencies (those judged representative of a proposal by our expert) are "near" the proposal, but not all nearby agencies are representative. Moreover, rep-

[4] The formulae for hit rate, false positive rate, and miss rate are:

hit rate =

$$\frac{\text{(number of agencies judged good by GRANT and the expert)}}{\text{(number of agencies found by the expert)}}$$

false positive =

$$\frac{\text{(number of agencies judged good GRANT and bad by the expert)}}{\text{(number of agencies judged good by GRANT)}}$$

miss rate = 1 - hit rate.

Minimum-distance	Agencies found	Good agencies	Bad agencies	Hit rate	FP rate	Miss rate
radius = 2	*2.26	1.00	*1.26	*52%	39%	48%
radius = 3	*4.70	1.57	*3.13	*83%	58%	17%
radius = 4	12.48	1.83	10.65	*94%	80%	*6%
radius = 4	15.09	2.00	13.09	100%	87%	*0%
Endorsement Constrained	2.78	1.48	1.30	80%	32%	20%

Average Hit and False Positive Rates for 23 Proposals.

Table 1.

resentative agencies *can* be found relatively far from the proposal. The "good agencies" column of Table 1 shows that 50% of the representative agencies are at radius 2, 28% are at radius 3, 13% are at radius 4, and 9% are at distances greater than 4. To the extent that endorsement-constrained search can find these more distant agencies and still maintain a relatively low false positive rate, it is superior to minimum distance search. In fact, for 23 proposals, EC search found a total of 64 agencies, 18 at a radius of 3 or more. EC search thus finds 28% of its agencies at radii where MD search makes most of its false positive errors.

Endorsement-constrained search is nonetheless imperfect. Its false positive rate is 32%, and it fails to find 20% of the agencies judged good by the expert. The reasons for these errors are of two kinds. First, EC search rarely misses a good agency within radius 2, but many of its false positive errors occur there. Conversely, most of its misses are at radii greater than 2, but false positives are rare. A path of length 2 between a proposal and an agency implies that they share a common node; for example, the node *psychological-counseling* is shared by proposal P4 and agency A2 in Figure 4. P4 would be judged representative of A2 by endorsement-constrained search, even if they have nothing else in common! This is because, as we said above, we have no mechanism in GRANT to sum the degree of fit between *all* components of a proposal and an agency. Provided *any* component of an agency is within the evidence boundary of a proposal, it is judged good by EC search. Once we solve this problem, the false positive rate for EC search should drop. Second, the reason that EC search misses agencies at more distant radii is that its path endorsements tightly restrict the expansion of the evidence boundary. We can relax these endorsements at the expense of more false positives. This should prove a good strategy once we can rule out candidate agencies based on their total degree of fit to the proposal.

Conclusion

Endorsement-constrained search finds a sizeable portion of the agencies considered representative by our expert and does not find many agencies that the expert considered unrepresentative. EC search thus operationalizes representativeness. It finds agencies judged likely to fund a proposal, based on the nature of the associations between the proposal and agencies. Viewed as a classification problem-solver, EC search finds the best classes (agencies) given evidence (proposals). Uncertainty in this task arises from partial matches between components of evidence and conclusions. Path endorsements are an explicit representation

	Agencies found	Good agencies	Hits/Try
radius = 2	2.26	1.00	.44
radius = 3	2.44	0.57	.23
radius = 4	7.78	0.26	.03

Average Hits/Try at Incremental Radii for 23 proposals.

Table 2.

of this uncertainty and are successfully used to control search for those conclusions that minimize uncertainty.

References

Bar Hillel, M. (1982) Studies of representativeness. In D. Kahneman, P. Slovic, & A. Tversky, *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press.

Buchanan, B. G. & Shortliffe, E. H. (1984) *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project.* Addison-Wesley.

Clancey, W. J. (1984) *Classification problem solving.* AAAI-84, 49-55.

Cohen, R. (1983) Heuristic reasoning about uncertainty: An artificial intelligence approach. Doctoral dissertation, Stanford University. *STAN-CS-83-986.* Also published by Pitman Advanced Publishing Program. London: 1985.

Cohen, R. (1984) *Progress report on the theory of endorsements.* Proceedings of the IEEE Workshop on Principles of Knowledge-based Systems. Denver.

Cohen, R., & Grinberg, M. R. (1983) A theory of heuristic reasoning about uncertainty. *AI Magazine,* Vol. 4, No. 2.

Cohen, R., & Lieberman, M. D. (1983) *FOLIO: An expert assistant for portfolio managers.* IJCAI-8.

Collins, A. M. (1978) Fragments of a theory of human plausible reasoning. *TINLAP-2,* 194-201.

Doyle, J. (1983a) Methodological simplicity in expert system construction: The case of judgments and reasoned assumptions. *AI Magazine.* Vol. 4, No. 2.

Doyle, J. (1983b) Some theories of reasoned assumptions: Essays in rational psychology. *CMU-CS-83-125,* Carnegie-Mellon University.

Eddy, D. M. Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky, *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press.

Hayes-Roth, F. & Lesser, V. R. (1977) *Focus of attention in the Hearsay-II speech understanding system.* Tech. Rep. , Computer Science Department, Carnegie-Mellon University,

Kahneman, D., & Tversky, A. (1982) Subjective probability: A judgment of representativeness. In D. Kahneman, P. Slovic, & A. Tversky, *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press.

Lowrance, J. D. & Garvey, T. D. (1982) *Evidential reasoning: A developing concept.* Proceedings of the International Conference on Cybernetics and Society, 6-9.

Shafer, G. (1976) *A mathematical theory of evidence.* Princeton University Press.

Shafer, G. (1984) Probability judgment in artificial intelligence and expert systems. *Working paper 165.* School of Business, The University of Kansas.

Shafer, G. & Tversky, A. (1985) Languages and designs for probability judgment. To appear in *Cognitive Science,* 1985.

Shortliffe, E. H. (1974) MYCIN: A rule-based computer program for advising physicians regarding antimicrobial therapy selection. *STAN-CS-74-465,* Stanford University.

Shortliffe, E. H. & Buchanan, B. G., (1984) A model of inexact reasoning in medicine. In Buchanan & Shortliffe, (Eds.), *Rule based expert systems.* The MYCIN experiments of the Stanford Heuristic Programming Project. Chapter 11. Addison-Wesley.

Szolovits, P. & Pauker, S. G. (1978) Categorical and probabilistic reasoning in medical diagnosis. *Artificial Intelligence,* 11. 115-144.

Tversky, A. & Kahneman, D. (1982) Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky, *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press.

Zadeh, L. A. (1975) Fuzzy logic and approximate reasoning. *Synthese,* 30. 407-428