Order Number 9108104

Cross-examination as a model of knowledge elicitation in the design of expert systems

Fulda, Joseph Simcha, Ph.D.

City University of New York, 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

CROSS-EXAMINATION AS A MODEL OF KNOWLEDGE ELICITATION

IN THE DESIGN OF EXPERT SYSTEMS

by

JOSEPH SIMCHA FULDA

A dissertation submitted to the Graduate Faculty in Computer Science in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1990

This manuscript has been read and accepted for the Graduate Faculty in Computer Science in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

5/18/90
Date

Michael Anshel
Chair of Examining Committee

5/18/90
Date

Executive Officer

Michael Levin
Elliott Mendelson
Howard Rubin
Supervisory Committee

The City University of New York

ABSTRACT

CROSS-EXAMINATION AS A MODEL OF KNOWLEDGE ELICITATION IN THE DESIGN OF EXPERT SYSTEMS

by

Joseph Simcha Fulda

Adviser: Professor Michael Anshel

Statement of the problems: (1) The methodologies for eliciting knowledge from domain specialists are "soft" and unreliable—i.e., the variations on the interview scheme, even those which make heavy use of modern technology, do not result in a reproducible knowledge base. Of course, some methodologies are superior to others and advances have been made, but most existing methodologies do not rise to the level of genuine knowledge engineering.

(2) If there are n items of evidence, each of which can take on m values, and based on the evidence D decisions are possible (in general, there is no reason to assume that the decisions are mutually exclusive or that the evidence must be binary), then the knowledge engineer must ask the domain specialist $D(m^n)$ questions, a combinatorially impossible task for even moderate n, before he can build his rule base.

Thesis: Use of the most highly structured question-and-answer paradigm, namely that of the legal process, can result in both a reproducible knowledge

base and in one built without combinatorial explosion, thus resulting in a tractable problem. The most important technique of the legal question-and-answer process and the one requiring the most creativity on the part of the attorney is adversary examination, particularly cross-examination. By using ten principal heuristics, we introduce an algorithm for cross-examination which makes it possible to focus on the hard cases—those where inconsistency is likely to occur—and to effectively cover the remainder of the $D(m^n)$ cases. This is not always possible. But in cases where the evidence is more or less cumulative and independent, it is not only possible, it has been done for this dissertation using the medical domain. The heuristics for cross-examination are used on the hard cases until all inconsistencies are removed from the knowledge base and it proves futile to try to produce additional hard cases or additional inconsistencies.

ACKNOWLEDGMENTS

The original research embodied in this dissertation was performed from September 1988 through December 1989 as part of my duties as Research Assistant Professor of Biomathematical Sciences at the Mount Sinai School of Medicine. I am grateful to the medical school for providing an ideal research environment. I am deeply appreciative of the expertise and professional excellence of Professor Clyde Schechter, M.D., Assistant Professor of Community Medicine and Assistant Professor of Medicine, who served brilliantly, effectively, and graciously as the domain specialist throughout the research, and to whom the medicine in Chapters 4 and 5 is entirely due.

I am deeply grateful to Professor Craig Benham, my Chairman, who found this work valuable enough to grant me 100% released time at full salary so that I could compile this work and turn the research into a fully developed dissertation during Spring 1990 as Associate Professorial Consultant in Biomathematical Sciences.

I appreciate the publication of the majority of the original research embodied in this dissertation, by and with the advice and consent of my mentor, by the International Association of Knowledge Engineers.

I am exceptionally appreciative of the firm guidance of the examining committee if not for whose wisdom in having me frame my proposal narrowly and concisely this work might never have reached completion. I am struck by the fact that almost all of their collective and unanimous recommendations presage Professor Ziolkowski's (Dean of the Graduate School at Princeton University) extensive remarks in the lead article of the Spring 1990 issue of *The American Scholar*, Phi Beta Kappa's journal of arts and letters.

It is my pleasure to acknowledge Professor Michael Levin's valuable and insightful comments on the logical structure discussed in Chapters 5, 6, and 7. I first truly studied applied logic under his guidance well over a decade ago and his critical reviews of my work in the area ever since have always been most helpful. He has again proven most helpful as Vice-Chairman of the committee.

I am most appreciative and grateful to Professor Michael Anshel, my mentor for well over a decade, in whose four classes as an undergraduate at The City College my interest in artificial intelligence and related areas was first sparked. His support, career guidance, and inspiration over the years have proven invaluable to my professional development. Early on he taught me, by example and encouragement, the value of original research and the importance of its dissemination to the academic and scholarly community in archival sources. Professor Anshel has also encouraged me to follow my interdisciplinary interests in law, logic, and their relationship to artificial intelligence. For all these reasons, it is my pleasure and honor to dedicate this work to him with respect and appreciation.

I would never have attained the doctorate without the benefit of the toweringly inspirational figure of my father, Professor Manfred Fulda, whose hard work and great works touched my life and that of many others in so many ways. Nor could I have attained whatever I have, without my mother, Mrs. Naomi Fulda, who taught me to read, write, and study, among so much else, and whose many labors on my behalf collectively sum to love. My sister, Miss Aviva Fulda, who is within as without and the truest person I know, has always stood by me. For the above and much besides, I can never thank my family enough. *Dei Gratia*.

Joseph S. Fulda, New York, March 1990

TABLE OF CONTENTS

CROSS-EXAMINATION AS A MODEL OF KNOWLEDGE ELICITATION IN THE DESIGN OF EXPERT SYSTEMS

1. INTRODUCTION

We present an overview of this work, chapter-by-chapter.

The Field of Knowledge Acquisition. The design and development of expert systems is a relatively new endeavor which has, nevertheless, resulted in an explosion of activity in recent years. Acquiring the knowledge to produce a system that demonstrates expertise in its domain is a task logically prior to the design of such a system. But science does not develop as neatly as all that and, in point of fact, knowledge acquisition—as a field of academic inquiry—is younger and far less mature than that of expert systems design and development. As the literature attests, the first expert systems were designed and developed using ad hoc techniques rather than engineered methodologies. The youth of the field is reflected in the way the literature is scattered throughout AI and cognitive science and cognate disciplines and in the fact that there is only one journal, *The International Journal of Man-machine Studies*, which has covered the topic for any length of time. This year, one of the two major computer science professional organizations, the IEEE Computer Society, introduced the *IEEE Transactions on Knowledge and Data Engineering*, and two even newer journals have just opened their pages. The other major computer science professional organization, the ACM, has recognized the emerging field with a special refereed issue of its AI newsletter, which also contains an extended and invaluable bibliography to the scattered literature mentioned earlier.

In the literature review, the knowledge acquisition field will be taxonomized along two dimensions, source of expertise and method of acquisition.

Knowledge Elicitation. When the source of expertise is an expert in the field and the method of transfer involves the spoken or written word, with or without technological or other aids, we are in the subfield known as knowledge elicitation, the method used almost exclusively by practitioners and we will survey the subfield and its literature. This is despite the so-called "knowledge acquisition bottleneck" which refers to the difficulties inherent in knowledge elicitation, on which the literature will be discussed.

Knowledge Elicitation as a Quasi-legal Process Using an Additive Model: An Illustration from the Medical Domain. The rule-based paradigm which has dominated expert systems design will be presented, with references, and we will discuss why it lends itself to combinatorial explosion. Then our additive model for expert systems design is introduced, using evidentiary weights (which may be expressions, Boolean or arithmetic) and thresholds. The key insight is the replacement of evidentiary items by evidentiary weights, thus vastly reducing the combinatorial explosion, and cumulating them until it can be decided which decisions are indicated. A similar problem reduction by replacing decisions with equivalence classes of decisions is also sometimes possible. The defects of the additive model are discussed, as well as those situations in which it is clearly inapplicable. The assignment of evidentiary weights and thresholds is a very tricky task and one which greatly reduces the model's reliability, on which immediate elaboration.

The legal process requires the cumulation of (often) independent evidence to reach a (or one of several) decision(s). In the Anglophone world, it is accomplished by the most stunning use of the question-and-answer paradigm this author is aware of: adversary examination. As Professor Fletcher of Columbia University Law School writes, in law the most feared outcome "is a trial that ends in error" (Fletcher, 1988). Thus, a robust process is a necessity.

It is the thesis of this dissertation that the ends of knowledge elicitation in the design of expert systems, particularly when using an additive model, are best served when the knowledge engineer takes on the role of counsel for the defense and tries to upset the domain specialist's choices of evidentiary weights and thresholds through the techniques of adversary examination, and, in particular, cross-examination. Thus this chapter draws a very extensive analogy between the question-and-answer process in the law and the question-and-answer process used for knowledge elicitation. Evidence, depositions taken at EBT's, and the rest of the legal process are all given analogues, some of which have stunning applicability to knowledge elicitation, and some of which have virtually no applicability owing to the different aims of scientific discovery and legal inquiry. But the most important contribution is the introduction of ten principal heuristics for the process of cross-examination which collectively bring the art of knowledge elicitation to the level of knowledge engineering. As cross-examination proceeds, the numbers (expressions) used as evidentiary weights and thresholds change and new hard cases are produced. The process does not terminate until the knowledge engineer is unable to generate further inconsistencies or hard cases (candidates for inconsistencies), i.e., until the domain specialist has survived the rigors of cross-examination. When that occurs, the knowledge base is reproducible *and* (compared to a rule base) compact. This has been verified as an experimental finding, on which immediate elaboration.

An extended example of the methodology introduced is presented, one that took many man-months to develop. The case study involves the design of a knowledge base for the diagnosis of the cause(s) of tiredness induced by some malignancies in a particular segment of the patient population. The domain chosen is one in which the methodology just described is particularly suitable. There are three reasons for this, one involving the diagnostic task in general, one involving the sub-domain of tiredness, and one involving metastatic disease in particular. This chapter gives us the opportunity to extend the legal analogy to the knowledge elicitation process in a concrete manner, thus demonstrating how a knowledge base is actually designed using our methodology.

The Logic of the Model and Expert Inconsistency: Illustrations from the Medical Domain. Using the medical domain to reify the discussion once more, we present Schoenmaker's paradigm and extend it to the predicate calculus and show how it applies to rule-based systems. The idea is simple: propositions which are facially consistent are inconsistent when additional propositions assumed in the background by everyone but explicitly stated nowhere in the rule base are considered. In the case of rule-based systems such inconsistencies occur when multiple experts are used (or when a database compiled from multiple sources is used) to produce a knowledge base. However, it neatly turns out that an almost identical paradigm defines the inconsistencies detected by the knowledge engineer working with a single domain specialist using our methodology (an additive model coupled with cross-examination). A variety of medical cases reify the discussion, the medical domain being particularly suited for classical, two-valued logic.

The Contributions of the Present Model. Then follows an overview of the three contributions we have made.

The Conceptual Architecture of the Knowledge Elicitation Model and Formal Heuristics for Cross-Examination. Each phase of the knowledge elicitation process using the present methodology will be discussed and the connection between the parts clarified both verbally and diagrammatically: this is what is meant by a conceptual architecture. In addition, each principal heuristic for cross-examination will be given a partial formalization in SETL pseudocode. SETL is an extremely high-level language excellently suited for expert systems design, knowledge representation, and logic programming; perhaps only its principal designer's modesty has prevented it from becoming a standard AI language.

Future Research Directions. Although well beyond the scope of this thesis it would be interesting to combine cross-examination with non-verbal methods of knowledge acquisition from experts. In particular, a clinical trial comparing an expert system designed using our model with the actual practices of the domain specialist on a sample population should prove interesting. If there are inconsistencies between the reproducible beliefs of the domain specialist and his actual practice, he may not be aware of all the factors he uses to arrive at his clinical decisions. After a thorough cross-examination, such discrepancies are unlikely to be resolved by modification of the evidentiary weights and thresholds. Rather a careful process-trace will likely reveal additional factors (or, rarely, fewer factors) that subliminally affect the expert's clinical practices and those should be subjected to conscious scrutiny.

2. THE FIELD OF KNOWLEDGE ACQUISITION

The design and development of expert systems is a relatively new endeavor which has, nevertheless, resulted in an explosion of activity in recent years. Acquiring the knowledge to produce a system that demonstrates expertise in its domain is a task logically antecedent to its design. But applied science does not develop as neatly as all that and knowledge acquisition—as a field of academic inquiry—is younger and orders of magnitude less mature than that of expert systems design and development. As the literature attests, the first knowledge bases for expert systems were compiled using ad hoc techniques rather than engineered methodologies. This may be partly because it was simply assumed that experts can teach knowledge engineers in an educational rather than professional manner. Indeed, the early literature usually cited as part of knowledge acquisition background is in psychology, cognitive science (in its older meaning), and education and has a distinctly different flavor from the later literature. And, although it is usually cited, we found no instance of its actual use except for personal construct psychology (see, e.g., Boose, 1985). Insights gained from psychology are indeed ubiquitous; but the methodology of knowledge acquisition when the object is teaching does not seem to carry over to the case where the object is designing a knowledge base. It was also presumed, wrongly but understandably, that knowledge acquisition would prove fairly easy. Today, Boose (1985) calls knowledge acquisition not just a "major problem" but also the "central task" in the design of the entire expert system. LaFrance (1989) quotes Duda and Shortliffe, both pioneers in the theory and development of expert systems and technology, to the effect that "expertise elicitation (is) one of the most complex and arduous tasks encountered in the construction of an expert system." On a more positive note, she cites another author as regarding knowledge elicitation as "more an art than a science." Shadbolt (1989) describes the problem simply as "vexing," while Lavrac (1989) finds it "a demanding mental process." Crandall (1989) identifies knowledge elicitation as "a major hurdle," and Hoffman (1989) calls the interviewing process often involved "exhausting." All of this speaks to the problems of the knowledge engineer and those responsible for producing the product; the same is true, however, on the expert's side. There is a fair amount of literature discussing the psychology of experts placed under the gun. Brown (1989) in "The Taming of an Expert: An Anecdotal Report" gives an excellent description of expert-knowledge engineer dynamics. On a more abstract level, Waldron (1989) gives the communications problems involved a brief treatment. However, it is Micciche (1989) who makes the most relevant point: "a primary responsibility of the knowledge engineer is to establish an environment in which the expert can be more able, comfortable, and willing to communicate his knowledge." The expert, in other words, is a resource and it is (and here we go beyond Micciche) squarely the responsibility of the knowledge engineer—like any engineer—to use his resources, including personnel, effectively. Thus, as Micciche concludes "interviewing, despite one's initial impressions, is a challenging task." There is, of course, a technical term for all these problems collectively: "the knowledge acquisition bottleneck." It may be inferred from Hoffman (1989) that the phrase is due to Edward Feigenbaum, a world-class researcher, who coined it back in 1969 when nobody thought it would be the problem it has since been seen to be. There is, here, an implied physical analogy, and Michie (1987) presents several "maps" with corridors which very effectively illustrate the problem (and his particular solution, on which more later). These knowledge engineering maps and the "bottleneck" have now become accepted paradigms and explain the principal reason for massive delays in expert systems projects. I faced the same expectations of ease as most other knowledge engineers and progressively narrowed the medical domain from tiredness to tiredness in older patients to tiredness caused by malignancies in older patients to tiredness caused by malignancies in certain (well defined) populations of older patients. The definition of tiredness was also narrowed and consideration was limited to those cases where it is *the* presenting symptom. Even though a genuinely engineered methodology was used the corridor was but widened somewhat. That it was widened is amply illustrated by the enormous number of data elicited from the domain specialist; that the bottleneck remained a problem is amply illustrated by the successive, almost comical, revisions of the patient population under study. It is entirely due to the project's domain specialist that we found ways to successively narrow the project's scope without diminishing its intellectual or medical value and rewards.

This introduction began by speaking of the youth of the field and to our knowledge, there are but four journals which deal primarily with knowledge engineering. The oldest, *The International Journal of Man-machine Studies*, while not by any means limited to this area treats of it frequently and generally produces two volumes each year. The journal reflects the state of the field and a very new branch of engineering cannot be expected to develop but from humble beginnings. My acquaintance with the pre-Socratic philosophers' scientific conceptions acquired as an undergraduate reinforced the old lesson about seeing further when elevated by your predecessors' shoulders. Today, of course, much of ancient science and technology seems a curious mixture of rejected ideas, speculation without basis, ad hoc techniques, and occasional insight. The same is true of knowledge engineering today, although the pace of change is radically faster, of course. It is no derogation of one's chosen field for a mature scientist to admit this openly, honestly, and with humility. The intention of such an admission is to spur the field to live up to its high expectations and not to result in the general academic disrepute into which, say, educational theory has fallen, despite such high initial expectations.

The remaining three journals have just opened their pages: *Knowledge Acquisition* in Great Britain, *Data and Knowledge Engineering* from Elsevier, and the *IEEE Transactions on Knowledge and Data Engineering*. At the date of this writing, only the last of these was available and in the one volume published, rather surprisingly, there is little in the way of pieces directly concerned with knowledge engineering. In addition, the journal has chosen a somewhat idiosyncratic distinction between data and knowledge, although it may be, at a high level, conceptually equivalent to the standard distinction between data and knowledge (symbols without meaning in contradistinction to symbols or aggregates carrying meaning) accepted in data structures by computer scientists. In addition to the IEEE Computer Society, the other main academic, professional organization in computer science, the ACM, has published a single, very extended, refereed issue of its AI newsletter. That, a major textbook (on which more later), and pieces scattered throughout the AI literature generally or presented at conferences, constitute the bulk of the literature on knowledge acquisition that receives attention. Most of this is not overly different from the articles representative of *The International Journal of Man-machine Studies* and the same comments and remarks apply thereto.

Knowledge acquisition has traditionally been understood to include knowledge elicitation—the extraction of knowledge from experts using verbal means—as well as nonverbal means of acquiring the knowledge to construct a knowledge base for an expert system. In practice, completely nonverbal knowledge acquisition is extremely rare, but we will here review what can, in theory at least, be done without knowledge elicitation from a domain specialist. LaFrance (1989) suggests that "rarely, the expert becomes a knowledge engineer who then relies on introspection to encode the expertise"; however, the rest of the article does not address this. We would speculate, tritely, that two heads are better than one, and that the elicitation process, however slow and frustrating, provides much more insight than introspection. Later when a quasi-legal approach to knowledge elicitation is introduced, we will remark in the same vein that a lawyer who has himself for a client.... Hoffman (1989) who does a thorough survey of the state-of-the-art points out several additional means. First, there is direct observation. While a domain specialist (or several) is needed, there is no elicitation involved. However, both examples cited have the experts thinking through their thoughts out loud, whether by "encouragement" or because they are deliberating as a group. However, it seems that this is just the limiting case of an unstructured interview: no questions, only answers. As one who is promoting the legal process as a model, such observations are seen as entirely unhelpful in constructing a *robust* knowledge base. Second, there are documentary sources—"training manuals, operating manuals, handbooks, etc." Again, the same objections apply, although at least the presentation is (presumably) not ad hoc. Third, and far superior, is the idea of the knowledge engineer selecting a test sample (representative, difficult, borderline, salient, etc.) based on archival data and then observing the expert as he carries out the pre-selected sample of tasks. While the methodology reviewed is entirely different, the types of samples that it is suggested that one might choose are surprisingly similar to the types of situation-combinations chosen for our research as will be detailed when we present our heuristics for cross-examination. Hoffman (1989) makes the best attempt I've seen to transfer psychological methods to knowledge acquisition as here intended but may fail to truly go beyond standard knowledge elicitation techniques (which have not yet been surveyed here) as is shown by his Tables 1 and 2, as understood by this reader.

Before moving on to knowledge elicitation, it would be well to point out that there is much literature on both experts and expertise *in se*. The literature raises troubling questions about the value of expertise in general (Shanteau, 1987) and the conflicts about what kind of models of expertise work (Stewart, 1987). In addition, there are reams of literature on what makes an expert different from a novice and on what precisely characterizes expertise. This literature is interdisciplinary and sometimes interesting; however, since it is not of particular relevance to this study we will not survey it.

3. KNOWLEDGE ELICITATION

Knowledge elicitation remains the principal, almost exclusive, method of acquiring knowledge for knowledge bases used in expert systems. Contrary, however, to initial expectations, a survey of this literature is not quite the task it seems. The reason for this is simple: existing methods largely fall into two categories—automated knowledge elicitation and interviewing. In each category, there are admittedly an endless variety of variants each described in different, often idiosyncratic language. Examining these variants has convinced me that it is well to be doing a survey rather than a catalogue: so much has been done and written. The variants matter less or more, the former for the interview methodology, the latter for the automatic approaches, so we accord the latter a more detailed treatment. Of course, we will treat direct and cross-examination as a method of knowledge elicitation in depth too, since that is the present contribution. It, too, is a variant of the interview method, with the critical difference being that an algorithm and validation method are provided thus raising it to the level of an engineering methodology.

INTERVIEW VARIANTS

Bainbridge (1979) in an early discussion treats the use of questionnaires and their variants thoroughly. Of course, use of these instruments in the social sciences is the subject of much writing. (By "interview" we mean a process proceeding from verbal reporting, not necessarily oral reporting.) He also treats the question of distortion in communication and discusses, *inter alia*, the relationship between observed behavior and verbal reporting, a topic with which this thesis will close. Belkin (1987) discusses observation, as also discussed earlier in citing Hoffman (1989), but adds a discussion on the use of the linguistic method of discourse analysis on transcripts gained from domain experts. (Incidentally, many of these methods sometimes employ audio- or videotape and transcripts as aids to uncertain ends.) Johnson (1987) gives a thorough review of what is known as "protocol analysis" and by which is meant verbalization during performance and is often referred to as the "think-aloud technique"; generally, the method has someone asking for explanations as the task proceeds, but not necessarily. LaFrance (1987) covers the well-known knowledge-acquisition grid. At first glance, it might appear that the terminology does not refer to a knowledge elicitation method. However, the grid is an aid to "encoding" the information *after* it has been elicited. Still, there is a degree of formalization here because the elicitation may be directed by the knowledge engineer's *prior* knowledge of the information required for the grid. An obvious drawback is its tediousness. Boose (1985) also discusses the knowledge-acquisition grid as well as graphs. Indeed, several authors describe the decision-analytic process in the computer science and operations research tradition of trees or graphs, but usually that is a construct that is used, once again, after the knowledge is elicited. To have an expert begin with such a structure would fall between the beginnings of automation and simple introspection. Crandall (1989) describes what she refers to as a "semi-structured interview" which she terms in what seems a misnomer the "critical decision method." While it is more structured than the think-aloud method to which it is compared, it is not otherwise different from other variants. Crandall is not speaking, indeed, about a better interview, but about trained, aware, and astute interviewers with some knowledge of the domain and the critical factors that influence decisions in the domain (hence the method's name). However, that seems to be an entirely—or almost entirely—different set of issues. Hoffman (1989) once again provides the best survey of interview methods discussing "free-flowing dialogue" and "pre-planning of questions and their order" (this is the right track, but the problem with calling such interviews structured, as does Hoffman, is that while planning is suggested, no actual plan is given). He does, however, discuss briefly test cases (see Chapter 2) and as we noted earlier that does provide the beginnings of a plan. Finally, he distinguishes between introspective, retrospective, and event-recall interviews, the latter of which is also sometimes referred to as "case-based reasoning." Shadbolt (1989) shares the general view that think-aloud techniques are not particularly valuable and also found that several knowledge elicitation techniques "did not seem to complement each other... and combinations of techniques did not reveal much more information than any of the techniques on their own (with the exception of protocol analysis)." Shadbolt also discusses "contrived" techniques, particularly those termed "card sorts" and "laddered grids." The latter is a technique where the expert constructs a graph on his own with "a small set of prompts" from the knowledge elicitor to guide him in its construction, while the former is a technique involving cards with domain elements printed on them which are then sorted by the expert along every dimension he thinks relevant to try to cover situation-action pairs. The author's description of such knowledge-elicitation techniques as "vexing" is understandable.

One final point. It has been called into question whether verbal reporting *can* be complete and accurate (Stevenson, 1988). People may not have "conscious access to all their cognitive processes.... When solving a problem, people may use fast, automatic processes which are not available to conscious reflection" (Stevenson, 1988). Stevenson's article is worth reading, especially in light of modern discoveries regarding the role of the brain's right hemisphere. Stevenson's concern is yet another reason for advocating the combined approach suggested for future research at the close of this thesis.

AUTOMATION

Automation holds great promise for knowledge acquisition and is advancing rapidly, so rapidly that as this is being written the techniques referred to may be becoming obsolescent. Machine inference from examples which allow generalization is the preferred method. The expert need teach but a certain amount; thereafter the computer will pick up the concepts, with or without the assistance of a knowledge engineer. Discussing this briefly is Martin (1989) and comprehensively are Michie (1987), Quinlan (1987), and especially Lavrac (1989) who does seem to have developed a genuinely engineered automation methodology. As Michie (1987) explains, the idea of automatic debriefing through induction—rules from examples—suggests itself from classical debriefing in which the domain specialist frequently resorts to tutorial examples rather than the rules or patterns behind them. This technique has proven highly efficient. Michie (1987) states that in one case traditional knowledge engineering took months of work, while machine induction generated an equally accurate knowledge base in thirty-four seconds. Michie (1987) admits, however, that this promising technique, despite its increased accuracy and efficiency, has a major drawback when compared to standard interviewing techniques: the artificially generated rules seem to be totally opaque to human experts examining them. This is of concern since the domain specialist must provide feedback on the accuracy of the rules generated and must therefore be able to follow them. Quinlan (1987) provides an extended case study of this technique, using the domain of thyroid-related diseases and their diagnosis. He also presents a thorough description of an induction algorithm and the various experiments using it. Both papers are well worth reading, Michie's for the general information, Quinlan's for additional detail. The work of Silvestro (1988) like that of many others not cited here uses automation as merely an aid to knowledge elicitation. Other approaches are discussed in Kahn (1985), Narayan (1987), Slator (1989), and Burzesi (1989) to whose work I now turn.*

* It should be noted that although Burzesi is a five-time student whose thesis I officially mentored with a colleague, my role was only to encourage his own highly original research. Burzesi is an independent worker in whose research I played no substantive role, save the final assignment of an A.

Burzesi (1989) is especially well written, available to me *in toto*—not in the almost-abstracted form many publications have, and is also an excellent example of an engineered methodology which works, which is especially effective in cases of combinatorial explosion, and which belongs to the automation category. This work deals with dynamic, physical systems and defines a model with sources, sinks, processors, reservoirs, conduits, and port-connections as the components. The system's state is represented by a configuration (a vector) giving the state of each component. Typical states are "inconsistent" (the heater is on, the water flows through, the water comes out cold), a "problem," or "normal." Initially, the knowledge engineer classifies whole configurations as normal, inconsistent, or as a problem. If the configuration is inconsistent or a problem, the system queries to find out which subconfiguration(s) is responsible. Fairly soon, because its acquired knowledge of inconsistent and problem subconfigurations effectively "covers" whole configurations it has never seen before, it is classifying same with complete accuracy. Burzesi's automated tool thus learns from the particular to the general in the inductive tradition.

If we may jump ahead to the present research, the idea is the same at a very high level—for the interview method. Asking the right (as determined by the heuristics) questions of the expert and subjecting the answers to cross-examination effectively "covers" easier combinations (questions consisting of symptom combinations) since the model is additive. This will become clear in the following chapter.

4. KNOWLEDGE ELICITATION AS A QUASI-LEGAL PROCESS USING AN ADDITIVE MODEL: AN EXTENDED ILLUSTRATION FROM THE MEDICAL DOMAIN

This chapter combines a model, its validation, and its application, which, in this particular case, have proven to be inextricably intertwined. As will become clear in detail as we progress, an additive model or a Bayesian approach with the domain specialist reeling off evidentiary values is impeached with only the slightest of effort: experts are simply incapable of quantifying thousands of judgments accurately without a method of validation. Our method, a quasi-legal approach, provides, as we shall show in detail, the necessary robustness, reliability, and reproducibility which give the numbers scientific value. The methodology is thus genuinely engineered—but in the AI sense, with heuristics still requiring a degree of artistic talent in the knowledge engineer serving in the role of cross-examiner. However, no matter how intriguing a model and its validation, without reification on a particular domain illustrating the concepts, the model loses most of its force. It is my express intention to introduce the legal concepts and their medical analogues together, so that all the different things we mean starting with evidence and other basic concepts become crystallized for the reader. The decision to join the description of the model, its validation, and its application has resulted in a chapter of unusual length, but in light of the peer review that the substance of that combination (Fulda, 1989) has already undergone and in view of the natural cohesion inherent in the combination, we have proceeded this way.

Although technically this is probably best described as an additive model, we do frequently refer to it as a matrix-based model. Knowledge engineers accustomed to thinking in terms of grids, programmers charged with implementing expert systems where the knowledge base is a huge matrix, and computer professionals, generally, accustomed as they are to making use of numerous software aids of the grid or matrix type will find this terminology comfortable and as part of the computing community we have on occasion sacrificed precision for comfortable usage.

An additive model, of course, is not the norm. Typical expert systems are constructed as rule-based systems described extensively and reified elegantly in (Winston, 1984) with extensive technical and expository detail also provided by (Buchanan, 1988). See also (Taylor, 1988) and the review thereon. In "Fairytales," an article published by Allen Newell, a founder of artificial intelligence, the essence of intelligence is operationally defined as follows: "This is what it means to apply knowledge to action: it means to let the action taken reflect knowledge of the situation..." This is also what lies at the heart of the most advanced AI software and the most commercially and scientifically attractive: expert systems with rule bases. Indeed, as is amply illustrated in the two references cited above, the "productions" or "if-then" rules are most frequently and easily described as "situation-action rules" with the form "If *situation* then *action*." Newell's concept is the old engineering concept of feedback recast in the language and context of a later time. Given such a history, not to mention commercial viability and scientific achievement, it requires some explanation to account for desertion of such a popular model.

As is described later in much detail, the reason is another well-known concept important in engineering: combinatorial explosion. There are many domains in which the number of situation-action pairs becomes far too large to handle and that is where the additive model with its adding of individual situations to create a global picture and with its grouping of actions into equivalence classes comes into play. An extensive case is made that the medical case study chosen involves such a domain, but it should not be supposed that the important domains of applicability are few and far between. The range of those domains is quite large and varied (Stewart, 1987) and the potential for this method to make genuine contributions is especially enhanced when combined with unobtrusive observation of the domain specialist at work or examination of his records. A specific example of such a potential study in the medical domain is given in the final chapter of this work. What follows, then, is, in essence, an extended description of the original research undertaken for this dissertation, excepting its logic on which more later.

LAW APPLIED TO KNOWLEDGE ENGINEERING AND THE COMBINATION APPLIED TO MEDICINE: THE CORE OF THIS RESEARCH

The Project and the Population. Diagnosing the underlying disorder whose chief presenting symptom is tiredness is a formidable task. Almost all symptomatic disease causes tiredness and the causes of tiredness span the full range of internal medicine: depression, anxiety, sleep disorders, metabolic disorders of various sorts, malignancies, cardiac disease, pulmonary disease, renal disease, a wide range of subacute and acute infectious diseases, and the various anemias, among others. The knowledge engineering undergirding an expert system which would take on the full range of this problem would take many man-years to develop, by which time the diagnostic procedures involved would be long obsolete. We thus sought to constrain the population under study in such a way as to render the problem tractable yet nevertheless to leave it filled with the subtleties and complexities, albeit in less breadth and depth, that make for its basic interest.

Dr. Clyde B. Schechter, previously a mathematics professor and now a physician on the faculties of community medicine and medicine of the Mount Sinai School of Medicine, served as domain specialist on the project. His hospital appointment has been in ambulatory care for many years, a setting in which tiredness is a frequent complaint. As we will see, the usual knowledge representation techniques proved inadequate for the domain considered here (tiredness induced by malignancies), so the object was to design a model of the knowledge base which would greatly simplify the problem without significantly reducing the accuracy of the results. Both the knowledge engineer and the domain specialist have worked in population modelling using a variety of techniques, thus complementing each other to an unusual degree and minimizing the frustrated communication between knowledge engineer and domain specialist that is typical when the expertise of the team is without overlap.

Older patients (males who have attained the age of fifty and post-menopausal females) experience all the maladies of younger patients, although with different prior probabilities, workup risks, and levels of treatability and consequently different orderings for the diagnostic workups, different diagnostic procedures, and the like. Hence, limiting consideration to the population of older patients greatly simplifies the problem, without diminishing its intrinsic interest.

We also limited our consideration to those cases in which *the* presenting symptom is tiredness or fatigue, thus ruling out diagnoses which would cause other symptoms so pronounced and severe as to be almost necessarily causes of complaint. Consistent with this constraint, we limited our consideration to patients whose general appearance is healthy—physicians "eyeball" patients and, almost subliminally, classify them as healthy, acutely ill, toxic, chronically ill, or cachectic in general appearance. This is less a constraint than it might seem at first glance, first because many patients refuse to seek medical help despite the severity of the symptoms (for example, see Table 4.1 and Table 4.3: patients may be coughing up blood, show signs of ataxia, or observe the spontaneous appearance of ecchymoses, yet nevertheless do not seek medical help till tiredness or fatigue simply overwhelms them or renders them dysfunctional). Second, general appearance, while used by all internists in medical decision making, is, of course, a rather crude indication of health. As this project has amply shown, many patients with a healthy general appearance are nevertheless quite ill. Still the constraint does limit the domain to some extent: for example, most acute infectious diseases will send even the least worried patient off to the physician for evaluation.

The Combinatorial Setting. We describe some of the knowledge engineering techniques used in the problem domain, those that prove most useful in a combinatorial setting.

Such a setting is provided when the patient reports not being short of breath, being always tired, and having a normal sleep history. Older patients with a healthy general appearance and this patient profile are sufficiently likely to have their tiredness explained by a malignancy that at this point a history, a physical exam, and a preliminary workup should be undertaken to determine which, if any, of the workups for cancer should be ordered by the internist. The relevant history items are given in Table 4.1, the relevant physical findings are given in Table 4.3, and the cancer workups and when they are indicated by the history, physical exam, and in medias res findings (to be explained later and given in Table 4.5) are given in Table 4.6.

We now consider why the diagnosis of cancer is a serious combinatorial problem. There are three reasons for this, one characteristic of medicine in general, one characteristic of cancer in general, and one specific to the broad domain at hand—the diagnosis of tiredness.

Medical diagnosis, in general, is based on a large number of history items and physical findings supplemented by laboratory workups, many of which are neither very sensitive nor very specific with respect to a particular disease. Medical evidence is said to be sensitive when the presence of the disease is a very good indication of the presence of the evidence in question as well. Medical evidence is said to be specific when the absence of the disease in question is a very good indication of the absence of the evidence in question as well. Sensitivity and specificity do not necessarily go hand-in-hand. Thus, a physical finding of enlarged lymph nodes is a sensitive but not specific indication of lymphoma, while a physical finding of a neck mass is a specific but not sensitive indication of throat cancer. In the absence of sensitive and specific evidence of disease, the physician must rely on a large aggregate of not-very-sensitive and not-very-specific evidence, posing a combinatorial problem.

Cancer is almost sui generis in pathology: it can spread both rapidly and discontinuously. Discontinuous spread is known as metastasis, and the history items and physical findings given in Table 4.1 and Table 4.3 as evidence of malignancies may be either evidence of a primary cancer site, evidence of spread, or both. Metastatic disease can thus show symptoms almost anywhere—without a clear way of tracing those symptoms to the primary site. Indeed, it is not uncommon for a diagnosis of cancer to be clear, with the primary unknown. Thus, there are many more history items and physical findings to consider in the diagnosis of cancer than in, for example, the diagnosis of all cardiac, pulmonary, and renal disease within the base population (older patients with a healthy general appearance) combined.

Furthermore, when these are indications of spread, they tend to be even less sensitive and less specific than is normally the case in medicine.

Finally, we chose to study tiredness, which can be an early manifestation of a malignancy in a great many cancers. There are a few other symptoms of this nature: bone pain, weight loss, and chronic pain in the territory of a peripheral nerve, for example. Yet, unlike tiredness, the causes of these do not even approach spanning the entire range of internal medicine. No less than fourteen cancers can cause tiredness in a patient in the base population with the profile given above, without necessarily causing symptoms so severe and pronounced as to make them, rather than tiredness, the cause of complaint: leukemia, lymphoma, breast cancer, colon cancer, lung cancer, stomach cancer, pancreatic cancer, cancer of the gall bladder, primary hepatic cancer, throat cancer, prostate cancer, ovarian cancer, uterine cancer, and cervical cancer.

Since each evidentiary item can be either present or absent and there are a total of 75 such items given in Table 4.1, Table 4.3, and Table 4.5, the number of combinations that have to be considered is 14×2^{75}, a staggering number (5.29×10^{23}) which results in a problem that is computationally intractable. Furthermore, even if it were computationally tractable for a supercomputer, it would remain computationally intractable for the knowledge engineer and domain specialist whose task it is to consider each subset of the possible evidence and decide which of the cancer workups are indicated. (The matter is made even worse since some types of cancer require multiple workups depending on the strength of the evidence suggestive of the malignancy in question.)

Now, it is certainly true that we have presented a huge combinatorial problem as even larger than it is. This is so for several reasons: (1) For some of the fourteen cancers, many of the 75 items of evidence are irrelevant—totally insensitive and totally inspecific; (2) Some combinations of symptoms, although not many, are virtually physically impossible and therefore do not have to be considered—for example, the combination of spontaneous ecchymoses, a neck mass finding, a prostate abnormality, and ataxia; (3) Some combinations of symptoms given the patient population and the patient profile will have resulted in a diagnosis before reaching the cancer workup in question—for an example, see Table 4.6; (4) It turns out that all supersets of evidentiary items requiring a workup also require a workup and all subsets of evidentiary items not requiring a workup also do not require a workup. It should be emphasized that this is a medical finding, not a mathematical necessity, a point to which we shall recur in the next section.

Regardless of just how combinatorially intractable the problem is, that it is so cannot be questioned. Hence, the usual technique of rule-based knowledge representation cannot be used. That would require a rule for each combination of evidentiary items for each cancer, subject to the exceptions in the last paragraph. The number of rules in the knowledge base would be huge and the number of antecedents in many of the rules would be extremely large. Hence, it proved necessary to design a model to render this apparently exponential problem tractable without sacrificing much in the way of diagnostic accuracy.

The Model. The fundamental insight is to replace an item of evidence with a measure of its evidentiary strength. Thus, for the purposes of diagnosing leukemia we may regard a history of blood in the stool, blood in the urine, vaginal bleeding (recall that the women in the population are post-menopausal), and coughing up blood as equivalent: within the context of a decision as to whether a leukemia workup is indicated or not, these distinct evidentiary items have equal evidentiary value. This alone would vastly reduce the combinatorial problem.

But to really reduce the complexity of the problem from its apparent exponentiality, it is necessary to do more, namely to find a method of combining these measures of evidentiary strength which yields a simple measure of combined evidentiary strength which can then, in turn, be checked against a threshold. If the numbers are verified properly, the subject of the next section, simple arithmetic addition will suffice. Table 4.2, Table 4.4, and Table 4.5 give the measures of evidentiary strength for each history item, physical finding, and in medias res finding, while Table 4.6 gives the thresholds (and other conditions that have to be met) for the workups to be indicated.

Addition works because in assigning the measures of evidentiary strength, it is constantly kept in mind that two 1's must equal a 2, three 5's must equal a 15, and so on. A more sophisticated mathematical justification for using an additive scale, communicated to me by Clyde Schechter, follows:

From a Bayesian decision-making perspective, a workup is indicated if and only if the probability of a given disease exceeds a certain threshold. When a series of history items and physical findings has been obtained, the current disease probability differs from its prior probability (i.e., its frequency in that part of the base population meeting the profile) in accordance with Bayes' Theorem. Thus, if the evidentiary items are independent, the posterior odds of disease is the product of the likelihood ratios and the prior odds. Taking logarithms, the likelihood ratios' logarithms are additive measures of evidentiary weight.

It turns out that there are not examples of more-than-marginally sensitive and specific evidentiary items that suggest that a malignancy of a certain class is absent (in this patient population with this patient profile), but if there were such evidence it would be assigned a negative number. In that case, it would no longer follow that all supersets of evidence requiring a workup also require a workup since additional evidence might actually weaken the case for a workup. Likewise, it would no longer follow that all subsets of evidence not requiring a workup also do not require a workup since the removal of evidence might actually strengthen the case for a workup.

Despite the simple elegance of the additive model, it is not without defects. That, of course, is inherent in the nature of models. One problem is that it can be difficult for the domain specialist to assign measures of evidentiary strength to two items, one of which is sensitive but not specific, the other of which is specific but not sensitive. The uneasiness felt by the domain specialist is explained by the fact that the model constrains him to assign commensurable quantities to somewhat incommensurable entities. The next section will discuss how the knowledge engineer can be of assistance in this task.

A second problem is that clusters of symptoms surrounding the same primary may be subadditive (the independence condition is violated), while bizarre—but not impossible—combinations of symptoms may seem to defy additivity altogether. Again, the next section will discuss how the knowledge engineer can assist the domain specialist in selecting appropriate values.

Other problems can be dealt with by simply expanding the model's concept of evidentiary strength to include algebraic and Boolean expressions as well as raw scores. For example, the number of fully excised benign breast masses in a woman's history can be a deciding factor as to whether a workup is indicated by the evidence suggestive of breast cancer. The score for each such mass is 4, and the evidentiary strength of the item is 4x, where x is an integer from 0 to 3 (12 points already exceeds the relevant threshold, so x need never exceed 3) signifying the number of such masses. Likewise, for lung cancer—but not for leukemia or lymphoma—a history of chronic obstructive pulmonary disease is just as indicative of the need for a workup as a chronic cough, but the presence of both is no more indicative than the presence of either. Hence, the score assigned to history item 2 for lung cancer is 11 less the score assigned to history item 1 for lung cancer. An examination of Table 4.2 and Table 4.4 will show other examples of the use of algebraic expressions to refine the model. Similarly, Boolean expressions can be used to refine the model. Heavy alcohol consumption is a risk factor in throat cancer if and only if the patient is a smoker. Likewise, a host of neurological symptoms is relevant to a diagnosis of lymphoma if and only if the patient has tested HIV+, in which case it is highly relevant. An examination of Table 4.2 and Table 4.4 will show other examples of the use of Boolean expressions to refine the model.

In some cases, namely those where the same evidentiary strength arises because of the same or nearly the same evidentiary findings, the description of the evidence can simply be written inclusively. Examples include the neurological findings just referred to, stool appearance (size, shape, consistency), and the finding of an abdominal mass or abdominal distention, the latter simply indicating an undetected abdominal mass.

Another way in which the model simplifies reality without a reduction in diagnostic accuracy is by grouping cancers for which the evidence is almost identical under one heading. Thus we were able to treat ovarian, uterine, and cervical cancer under the single heading of pelvic cancer. Likewise, we were able to treat stomach, pancreatic, gall bladder, and primary hepatic cancer under the single heading of abdominal cancer. For the patient population and profile in question, the differential sensitivity and specificity of the evidence for these cancers, even collectively, is so marginal as to be below the threshold of discrimination, as psychologists say. Furthermore, items specific to only some of the cancers grouped under a single rubric approach 0 in evidentiary value, thus allowing a large reduction in problem size without a comparable tradeoff in diagnostic accuracy. Of course, once the cancer workups are performed, differential diagnosis between the cancers subsumed here under a single heading is usually made.

THE LEGAL ANALOGY

Thus far, we have presented what might be termed "a knowledge engineer's appreciation of the subtleties and complexities of the medical domain—a case history." The remainder of this chapter will make a contribution to knowledge engineering itself, by drawing an extensive legal analogy. The heart of the analogy can be summed up in a single sentence: evidence is used to reach a judgment using the skills of direct examination and cross-examination. The highly structured discourse of the law provides a far better mechanism than any looser interview process for reasons that will become increasingly vivid as the analogy is set up and drawn in detail in the next three sections. In the final section of the chapter, we will discuss the contrast between science and law, in general, and why we consider knowledge engineering in a combinatorial domain only a quasi-legal process and why the development in the following sections is an analogy, not a metaphor.

Evidence. Evidence is classified in several ways: exhibits, material testimony, and expert testimony; direct evidence and circumstantial evidence; admissible evidence and inadmissible evidence. All of these classifications have medical analogues. Many of the physical findings, a neck mass for example, are exhibits. All of the history items are material testimony. Many of the physical findings, a pleural rub, for example, are expert testimony. The in medias res findings, on which we have maintained silence so far, present an interesting and subtle variation on expert testimony. *In medias res* is Latin for "in the middle of things." Often during the preliminary workup or a cancer workup, evidence is found that is suggestive of a cancer not being currently worked up. Such evidence is not ignored; instead it is scored and added to the remainder of the evidence for the malignancy in question and then checked against the threshold. The legal analogue of this is the testimony of an expert witness about some matter not covered in the scope of his *voir dire*. Such testimony would normally be inadmissible (even if responsive); however, if the opposing counsel opens the door, as attorneys say, to acceptance of this line of questioning and expertise, most judges will permit the testimony to be received in evidence. Of course, going back to the medical domain, the patient no doubt considers the physician's board certification in internal medicine to be an all-inclusive *voir dire*, is not an adversary but one who gladly opens the door to helpful evidence, and the like (see the final section for more on this contrast as a generality).

The physical finding of an abdominal mass is direct evidence; the physical finding of abdominal distention is (high-quality) circumstantial evidence. Likewise, all evidence of metastasis is circumstantial, while evidence of a primary tumor is direct. The differential value placed on direct and circumstantial evidence is reflected in the far lower values assigned to evidence of metastatic disease as compared with the values assigned to evidence of primary disease.

In the final section of this chapter, we will discuss why no evidence is inadmissible, but the medical analogue of inadmissible evidence is downgraded in scoring. All history items are—to the computer—hearsay evidence. Hearsay is excluded from evidence in the legal setting for two reasons: it cannot be confronted directly and thereby challenged and it may be heavily interpreted. Both of these problems are present with history items. There is no way to challenge the patient's assertion that he saw blood, rather than beets, in his stool, and the physician may often use his estimation of the patient's credibility and reliability in assessing histories. (Indeed, he must do so.) There are a great many evidentiary items that have analogues in both the listing of history items (Table 4.1) and the listing of physical findings (Table 4.3). Examples include: blood in stool, hoarseness, abdominal masses, ecchymoses, enlarged lymph nodes, and weight loss. In almost every instance, for almost every cancer, the physical finding is weighted more heavily than the corresponding history item. The computer weights what the physician witnessed more than what it is told the patient has witnessed. The few exceptions to this rule have good medical reasons. (Incidentally, the hearsay rule has exceptions, too.)

While the medical concepts of sensitivity and specificity do have legal analogues, they are of little use. As we shall see, however, the legal analogues of positive predictive value and negative predictive value are of considerable utility. Evidence has positive predictive value if its presence is a good indication of the presence of the disease in question. Evidence has negative predictive value if its absence is a good indication of the absence of the disease in question. The reason for the lack of utility of the legal analogues of the medical concepts of sensitivity and specificity is that virtually all legal evidence is both not sensitive and highly specific. The prevalence of a particular class of crime is not much reason to expect any particular evidence for any instance of it. The absence of a certain class of criminal acts is every reason to expect no evidence of it. True, these are not universal rules: For those who want to be apprehended and leave as many clues as they can, the evidence will be sensitive; for those being framed or subject to bizarre coincidences that circumstantially combine to effectively frame them, the evidence will not be specific.

It should be pointed out that it is because the presence or absence of disease is the cause and presence or absence of evidence is the effect (usually) that sensitivity and specificity, rather than positive predictive value and negative predictive value, are considered more medically *fundamental*. Yet even in the medical domain, it is the cumulative predictive value, positive and negative of the evidence that is most *useful* in reaching a diagnosis. Furthermore, a few minutes with a paper and pencil should suffice to convince the reader that any three of the following five variables suffice to compute the remaining two: sensitivity, specificity, positive predictive value, negative predictive value, and frequency in the population in question.

Turning to the legal analogues of predictive values, confessions have high positive predictive value of crime. "Missing witnesses" have high negative predictive value of crime. This requires some elaboration for those not familiar with the law. The prosecution has an affirmative duty to call all witnesses who would be expected to give evidence favorable to the prosecution. If it is established that someone is in that category and the prosecutor fails to call him to the stand, the judge may decide, especially if so petitioned by defense counsel, to instruct the jury that the prosecution has failed to adequately explain the witness' absence and that the jury may infer that such witness would have testified in a manner contrary to the prosecution's case. This instruction is known as "the missing witness charge."

Direct Examination. The direct examination of the domain specialist is relatively straightforward and its purpose is to elicit a descriptive model of a physician's (best) practices. A truly descriptive model of when workups are to be performed is impossible in this context, since physicians perform workups for screening and treatment as well as and interspersed with those for diagnosis. Since the domain specialist for this research is board certified in preventive medicine, it proved particularly necessary during cross-examination to make sure that at least part of the intent behind a mammography or a colonoscopy was diagnostic and based on medical indications of disease.

In a civil suit, both sides directly examine the parties, the plaintiff at discovery (a deposition is taken), the defense at trial. More is uncovered during the discovery process than will be admissible at trial. But that part of the deposition that is admissible at trial had better be consistent with trial testimony, since the former can be read into the record at trial. Then the witness will be asked the fateful question: "Do you remember being asked those questions and making those answers?" and he will be in for a tough cross-examination. (Recall that the attorney who took the deposition does the cross-examination at trial.) I used an analogue of this technique in eliciting all the relevant history items and physical findings. Early on, a listing of all evidence that might indicate any cancer was requested. Later on, separate lists for each cancer were requested. The results were impressive. The "deposition" had many more items, notably including the circumstantial evidence of metastasis. Yet, the more-focused "trial testimony" included some rare but textbook cases of evidence of primary disease such as Cushing's Syndrome, leukemic, retinal infiltrates, and chloroma. Since the function of the attorney who took the deposition is, at trial, cross-examination, all items given in the "deposition" were reviewed by the knowledge engineer to see if the domain specialist wished to stand by his earlier "testimony."

Another technique in direct examination is to use earlier testimony to see if evidence is missing. Earlier in the project, patients with a different profile (short-of-breath) had history items such as exposure to asbestos and other lung toxins listed as factors in pulmonary disease. After putting smoking on the list of history items relevant to lung cancer, we immediately asked about exposure to asbestos. It was added. We then asked about exposure to other common lung toxins. They were not added, inasmuch as they cause lung diseases other than cancer.

Even the most knowledgeable domain specialist will not recall everything at once. It is the function of the knowledge engineer during direct examination to jog the expert's memory as much as possible.

One final point: during cross-examination, some values will be changed. In that case, it may be necessary to do a re-direct examination by asking, say, whether the item upped from, say, a 4 to a 5, has the same evidentiary value as some other 5. I have found that re-direct examination does not play a major role in the knowledge engineering process, although sometimes necessary.

Cross-Examination. The most challenging task for an attorney and by analogy for a knowledge engineer is cross-examination. An attorney is not often in a position to choose his witnesses, with the exception of his experts which he spends time and money in doing. The knowledge engineer is well-advised to expend considerable effort before beginning his project to choose an appropriate domain specialist. Clinicians for this project were interviewed by the knowledge engineer for over a month! It is not that expertise is hard to come by; rather it is witnesses who can withstand rigorous cross-examination both emotionally and intellectually that are a rarity. Experts with a weak knowledge of modelling or such concepts as sensitivity and specificity will produce matrices like those of Table 4.2 and Table 4.4 and Table 4.5, which repeatedly dissolve upon cross-examination. Since the knowledge engineer is not himself a domain specialist (and even if he were, the interaction is crucial: a lawyer who has himself for a client...), all he can do is ask. If the answers fall apart session after session, no progress will be made. Furthermore, an experienced knowledge engineer who knows how to cross-examine will, at first, get an incorrect response to every third question or so. This arises only in part from human fallibility or carelessness; in greater measure it arises from the slight defects in the additive model around which cross-examination will center (see below). After a three-hour knowledge engineering session consisting entirely of abstruse or borderline questions, an expert without the right temperament and disposition for such a project—and I am not being jocular here—will either break down and cry or fly into a mad rage, reciting his credentials at light speed, and the like, when, in reality, 70% is about the best that can be hoped for.

We now discuss the process of cross-examination itself. We have already remarked at length that the problem domain is combinatorially intractable. Therefore, any set of matrices that are produced at first will be fairly inaccurate. Hence, the best way to begin is with a global cross. Add up the scores for each cancer and rank them. Ask the domain specialist to rank them as well. If there are significant disparities, the domain specialist should adjust the scores to preserve comparability. Second, add up the scores for each evidentiary item and rank them. Again, discrepancies in the overall strength of an evidentiary item as an indication of some malignancy or other should be adjusted as appropriate. Finally, a more subtle global cross should be applied. Some of the items in Table 4.1 and Table 4.3 are sex-specific and are so indicated. Again, the totals for each cancer should be computed, this time separately for males and females. Then the domain specialist should be asked for which sex there is potentially stronger evidence for a particular malignancy. And, again, incorrect responses may require adjustment.

In addition to the matrices, the domain specialist is responsible for selecting the thresholds for the various workups. It is the knowledge engineer's responsibility to ensure that if the sum of various matrix values compared against the threshold disagree with the domain specialist's assessment of what the appropriate diagnostic workup is (or is not) that the appropriate values are adjusted. This is done by asking several questions involving the matrix values and several involving the threshold and seeing which is more often off. (If neither is ever off, it may be that the domain specialist has answered the original question incorrectly and it should be posed again.) These questions might be termed local cross-examination (since they deal with specific subsets for specific cancers), and the best cross-examinations are those which mix a variety of techniques, ten of which we give here.

(1) Choose matrix values that sum exactly to the threshold.

(2) Choose matrix values whose sum falls just below the threshold.

(3) Choose a series of matrix values whose sum is at or near the threshold, varying exactly one matrix item at a time—with the matrix item being varied always having the same value.

(4) Choose a selection of quite unrelated matrix items.

(5) Choose a selection of very highly correlated matrix items.

(6) Choose matrix values whose sum falls right below the threshold and upon obtaining the *correct* response ("no workup indicated") add a matrix item whose value is 0.

(7) If you observe a false positive for a combination of matrix values 1 point below the threshold, try a combination 2 points below.

(8) If you observe a false negative for a combination of matrix values at the threshold, try a combination 1 point above.

(9) If the thresholds differ by sex as a percentage of the total possible score, ask in which direction.

(10) The November 15, 1988 New York *Times* (science section) suggests that different answers may result if the same items are presented in a different order; this is an occasionally helpful technique.

Of course, every so often throw in a question for no special reason, so that the above heuristics do not become too familiar to the domain specialist; furthermore remember that "magicians never reveal their tricks."

At this point the reader should review the defects associated with the additive model discussed earlier and see how some of the above techniques assist the knowledge engineer in alleviating them.

It is recommended that after the local cross-examination of matrix values and thresholds, the global cross-examinations be retried. This time, however, the model should be fairly robust and changes should be made at this point with great reluctance. One point of note is that it is unwise to truncate values at the highest threshold for a particular cancer's most extensive workup, since that precludes effective global cross-examination at the end, both vertically (per cancer) and horizontally (per evidentiary item). (See (Fulda, 1988) for a full discussion of a related point in a similar context.)

We have now discussed the who, when, what, and how of cross-examination. The most important question, however, "Why?", is yet to be discussed. We began this chapter by pointing out the necessity of simplifying an exponential problem using a model. Therefore, neither the matrix values nor the thresholds matter *in se*. Their combination matters because it must reflect the 2^n subsets. The only way to verify that the model is a perfect representation (it is not) is by trying out all possible sums of matrix values and comparing them with the thresholds. Of course, however, if that were tractable, the model would be unnecessary in the first place. The key, therefore, is to choose those subsets most likely to be modelled incorrectly. That is precisely what the various local cross-examination techniques are designed to do: focus on the hard cases and rely on the subset-superset rule to take care of the easier cases.

The technique works eminently well. One knows he is finished cross-examining when the number of false positives (unindicated suggested workups) and false negatives (indicated unsuggested workups) becomes vanishingly small. And cancer-by-cancer it does. Cross-examination may require several thousand questions, but what is that compared to 5.29×10^{23}.

Before closing our discussion of cross-examination, it is important to point out that every time matrix values are changed new "hard" cases may be created, thus requiring re-cross. Unlike re-direct, re-cross plays a central role in assuring the reliability of the model.

Science and Law. At this point, the reader probably regards cross-examination as a legal technique which when suitably adapted can be of scientific use. This is not our conclusion. Reread the answer to the "Why?"

question of the preceding section and summarize the answer in *one* word. The word is the most important word of the scientific method: **reproducibility**. Since cross-examination all but ensures reproducibility, it can be regarded as a scientific technique as well as a legal technique. A loose interviewing process performed ten times may produce ten different models. A formal, legalistic process of examination and cross-examination performed ten times will result in the same or equivalent (scaled) models. This is an experimental observation made on portions of the domain and is confirmed by the eventual inability of the knowledge engineer to elicit incorrect responses from the domain specialist, even firing off two questions at a time as attorneys sometimes do to fluster a witness, and even waiting till after the weekend when the domain specialist has forgotten his answer to a close call. Since the model is descriptive of the expert's thought processes, it is as reproducible—and as scientific—as they are.

Nevertheless, and although we clearly regard the law as a technical field, it is important to distinguish between the ends and means of science, on the one hand, and those of the law, on the other. In science, we begin with evidence working as a team and seek the truth as we can best understand it. To ensure reproducibility we are at liberty to use any methodology that seems to ensure that the same evidence will yield the same conclusion. In law, we start with perspectives on the truth—the client's—and working as adversaries seek to adduce evidence to support our predetermined perspectives on the truth. A good lawyer does not seek reproducibility. Quite the contrary—regardless of the evidence, he must make a case for his client.

As a consequence of this fundamental difference in purpose and perspective, any scientific evidence that has probative value, even mere correlations that fly in the face of good philosophy of science, is given some weight. Since it is the truth that is sought, any path to it is allowable, provided only that no steps along the path are taken that are larger than warranted by their probative value. Legal evidence, in contrast, which falls below a certain threshold of probative value is inadmissible, because the law must protect the rights of the parties and their right to present their perspectives on the truth based on principles consistent with the very idea of conflicting perspectives on the truth. To give but one example, hearsay allows for no effective rebuttal; it must therefore be, generally, inadmissible. The values implicit in the scientific approach are superior for scientific inquiry, while the values implicit in the legal approach are superior for legal inquiry.

TABLE 4.1. HISTORY ITEMS.

1. Chronic cough?
2. Chronic obstructive pulmonary disease?
3. Coughing up blood?
4. Hoarseness of at least two weeks duration?
5. Blood in stool?
6. Abdominal masses?
7. Fully excised benign breast mass? (F)
8. Biopsied benign breast mass? (F)
9. Never-biopsied breast mass? (F)
10. Weakened urinary stream? (M)[1]
11. Difficulty initiating urination?
12. Blood in urine? (observed or history of 4+ on urinalysis)
13. Incomplete voiding?
14. Pelvic pain? (F)
15. Vaginal bleeding? (F)
16. Spontaneous ecchymoses?
17. Pneumonia, more than once in last five years?
18. Masses in the axillae?
19. Enlarged lymph nodes in the neck?
20. Change in bowel habits?
21. Change in size, shape, or consistency of stool?
22. Post-prandial ($^1/_2$-hour) pain?
23. Regurgitation?
24. Early satiety? (in the absence of causative drugs)
25. Appetite loss? (in the absence of causative drugs)
26. Vomiting?
27. Pain during intercourse? (F)
28. Difficulty swallowing?
29. Weight loss? (history or observed against history)
30. Bone pain?
31. Chronic pain in the territory of a peripheral nerve?
32. Smoker?
33. Heavy consumer of alcohol?
34. History of exposure to asbestos?
35. History of exposure to benzene?
36. Family history of this cancer?
37. History of HIV+?

TABLE 4.2. HISTORY ITEMS SCORE SHEET.

	Leuk	Lymph	Brst	Colon	Lung	Abdm	Thrt	Pros	Pelv
1.	1	2	0	0	11	0	0	0	0
2.	0	0	0	0	11-x(1)	0	0	0	0
3.	2	1	0	0	17	0	0	0	0
4.	0	.5	0	0	7	0	0	0	0
5.	2	1	0	5	0	5	0	0	0
6.	0	2	0	5	0	7	0	0	0
7.	0	0	4x, x=0..3	0	0	0	0	0	0
8.	0	0	11	0	0	0	0	0	0
9.	0	0	11	0	0	0	0	0	0
10.	0	0	0	1	0	0	0	[15]	2
11.	0	0	0	1	0	0	0	[15]	2
12.	2	1	0	1	0	0	0	[15]	5
13.	0	1	0	1	0	0	0	[15]	2
14.	0	1	0	1	0	0	0	[15]	6
15.	2	1	0	1	0	0	0	0	6
16.	9.5	1	0	0	0	0	0	0	0
17.	5	5	3	1	7	1	2	0	0
18.	1	$(10)^2$	11	0	2	0	0	0	0
19.	1	$(7)^2$	6	0	2	3	5	0	0
20.	1	2	0	5	0	3	0	(4)	3
21.	1	2	0	7	0	1	0	(4)	3
22.	0	1	1	0	.5	9	0	0	0
23.	0	0	1	0	.5	5	0	0	0
24.	0	2	1	0	0	9	0	0	0
25.	1	2	1	7	.5	9	4	(4)	.5
26.	1	2	0	7	1	9	0	0	1
27.	0	1	0	1	0	0	0	0	6
28.	0	2	2	0	7	5	13	0	0
29.	3	3	3.5	4	4	9	4	(4)	5
30.	5	5	5	9	10	5	5	(4)	5
31.	1	1	5	5	11	5	7	(4)	5
32.	0	0	1	0	2	0	—	0	0
33.	0	0	0	0	0	0	2 iff 32+	0	0
34.	0	0	0	0	1	0	0	0	0
35.	.5	0	0	0	0	0	0	0	0
36.	0	0	10	0	0	0	0	0	0
37.	0	1	0	0	0	0	0	0	0

TABLE 4.3. PHYSICAL FINDINGS.

1. Enlarged lymph nodes?
2. Hoarseness? (observation or confirmation)
3. Breast mass? (F)
4. Abdominal mass or distention?
5. Enlarged bladder?
6. Enlarged liver?
7. Enlarged spleen?
8. Ecchymoses, not confined to distal portions of arms and legs?
9. Epigastric tenderness?
10. Pelvic—cervical ulcerations? (F)
11. —intravaginal masses? (F)
12. —uterine masses? (F)
13. —adnexal mass? (F)
14. Chest—fixed wheeze? (H indicates high central location)
15. —isolated patch of rales?
16. —signs of pleural effusion?
17. —pleural rub?
18. Weight loss? (Observation or equivalent evidence) against (records or equivalent evidence)
19. Neck mass?
20. Tongue ulcer or *single* enlarged parotid gland?
21. Horner's syndrome?
22. Myositis?
23. Cushing's syndrome?
24. Any of the following neurological symptoms not otherwise explained by a CAT scan of brain: Ataxia, cerebellar tremor, dysdiadokokinesis, monoparesis, or hemiparesis?
25. SVC syndrome?
26. Leukemic infiltrates in retina?
27. Chloroma?
28. DRE—rectal abnormality?
29. —prostate abnormality? (M)
30. —positive stool guiac?
31. —black, maroon, or red stool color?

TABLE 4.4. PHYSICAL FINDINGS SCORE SHEET.

	Leuk	Lymph	Brst	Colon	Lung	Abdm	Thrt	Pros	Pelv
1.	1	—	11	5	—[3]	6	—[3]	(4)	4
2.	0	.5	1	0	3	0	17	0	0
3.	0	7	—	0	0	0	0	0	0
4.	0	15	1	10	1	9	0	(4)	5
5.	0	2	0	2	0	0	0	[15]	6
6.	5	20	11	10	5	7	0	(4)	0
7.	5	20	2	5	2	6	1	0	0
8.	11	5	0	0	0	0	0	0	0
9.	0	1	.5	0	1	7	0	0	0
10.	0	2	0	0	0	0	0	0	6
11.	0	2	0	0	0	0	0	0	6
12.	0	2	0	1	0	0	0	0	6
13.	0	2	0	1	0	0	0	0	6
14.	0	2	1	0	17	0	10 iff H	0	0
15.	0	2	1	0	(5)[4]	1	0	0	0
16.	1	4	1	1	17	2	0	(4)	.5
17.	1	2	1	1	17	1	0	(4)	.5
18.	5	5	10	9	10	9	7	(4)	6
19.	0	10	2	0	6	5	17	0	0
20.	0	10-s(19)	0	0	0	0	17-s(19)	0	0
21.	0	0	0	0	5	0	0	0	0
22.	0	0	0	0	5	0	0	0	0
23.	0	0	0	0	5	0	0	0	0
24.	0	22 iff H37+	11	10	5	0	0	0	0
25.	0	22	11	0	17	0	0	0	0
26.	11	0	0	0	0	0	0	0	0
27.	11	0	0	0	0	0	0	0	0
28.	0	2	0	10	0	0	0	(4)	6
29.	0	.5	0	10	0	0	0	[15]	0
30.	5	3	1	10	0	5	4	(4)	3
31.	4 iff 30+	1 iff 30+	0	0	0	2 iff 30+	0	0	1 iff 30+

TABLE 4.5. IN MEDIAS RES FINDINGS and SCORE SHEET.

Finding	Lymph	Brst	Lung	Pros
1. Blastic lesion on scan	0	11	0	[15]
2. Lung nodule (possibly w/ satellites) on X-ray	0	0	17	0
3. Hilar adenopathy on X-ray	0	0	17	0
4. Hypokalemia on blood test	0	0	5	0
5. Hypercalcemia on blood test	0	0	5	0
6. Pancytopenia	22	0	0	0
7. Small cell cancer in lymph node	0	0	17	0

Note: An in medias res finding is a finding indicative of some cancer during the workup for some earlier cancer or during the preliminary workup. For more on in medias res findings, see Table 4.6.

TABLE 4.6. CANCER WORKUPS.

Number and name.	Threshold and indications.	Potential in medias res findings.[5]
0. **Preliminary.** SMA-18; CBC; urinalysis—protein, sugar.[6]	**Always performed.**	**4,5,6.**
1. **Leukemia I.** CBC with differential leukocyte count; platelet count.	**4.**	
2. **Prostate I.** Acid phosphatase blood test.	**4 after subtraction of all bracketed points.**	
3. **Lymphoma I.** Lymph node biopsy.	**Lymph nodes to biopsy.**	**7.**
4. **Breast I.** Mammogram.	**11.**	
5. **Abdominal.** EGD and Upper GI Series if endoscopy is contraindicated. Contraindications: difficulty swallowing or chronic bleeding external to the GI tract.	**9.**	**1,2,3.**[7]
6. **Leukemia II.** Bone marrow aspirate and referral to hematologist.[8]	**11 or Leukemia I is positive.**	
7. **Pelvic.** Pelvic sonogram and referral to gynecologist.	**6.**	
8. **Lung I.** (1) Chest X-ray and if positive then (2) tomogram of lesion and referral to pulmonologist; while if negative then (2) CT scan and if positive for lung nodules or infiltrates then (3) if still needed, tomogram of lesion and referral to pulmonologist.	**5.**	**1.**
9. **Breast II.** Breast biopsy.	**Breast mass is found (on exam or Breast I).**	

10. **Colon.** 10. 1.[9]
Colonoscopy followed by barium enema if colonoscopy cannot be completed.

11. **Lymphoma II.** **Lymphoma I is negative or impossible and 22 after subtraction of all parenthesized points.**
Needle biopsy of liver and referral to appropriate specialist, unless contraindicated. Contraindications: liver not enlarged, platelet count abnormal, prothrombin time abnormal, partial thromboplastin time abnormal, sonogram indicates hemangioma or biliary obstruction, or serum protein electrophoresis is consistent with amyloidosis.[10]

12. **Lymphoma III.** **Lymphoma I is negative or impossible and Lymphoma II is not indicated or is contraindicated and 10 after subtraction of all parenthesized points.**
Bone marrow biopsy and referral to appropriate specialist.

13. **Prostate II.** **Prostate I is positive or 15 after subtraction of all parenthesized points.**
Referral to urologist for appropriate procedure (e.g., IVP, Cysto, Bx).[11]

14. **Throat I.** 17.
Laryngoscopy with lesional biopsy, if possible.

15. **Lung II.** **Lung I shows no lung nodules or infiltrates and 17 after subtraction of parenthesized points.**
If there is a pleural effusion then (1) do a pleural biopsy and if positive then (2) diagnose; while if negative or non-specific then (2) do a bronchoscopy if possible and not contraindicated (contraindication: chronic extrapulmonary bleeding) and if positive then (3) diagnose; while if negative or non-specific or impossible or contraindicated then (3) do a thoracotomy unless prior workup findings or in medias res finding 1 show metastatic disease.

If there is no pleural effusion then (1) do a bronchoscopy if possible and not contraindicated (contraindication: chronic extrapulmonary bleeding) and if positive and specific then (3) diagnose.

16. **Throat II.** **17 and Throat I fails to yield a diagnosis.**
Referral to ENT for blind biopsy or other appropriate procedure.

NOTES TO TABLES

[1]Present in women only rarely, although when present it is scored. However, for the purposes of by-sex cross-examination, it is treated as a male symptom only.

[2]The parentheses derive from the fact that in the absence of corroborating physical findings, these history items are regarded as inaccurate. See Table 4.6, workup 11.

[3]These are given no score because they will never be reached: Physical finding 1 will be fully explained by the prior lymphoma workup. See Table 4.6, workup 3.

[4]This item has low specificity: many healthy persons will show this. Thus, in the absence of corroboration, the evidence is regarded as uninformative. See Table 4.6, workup 15.

[5]In medias res findings are to be differentiated from external findings, that is from findings of disease other than cancer in medias res. The domain specialist estimates that roughly 50% of workups will, in fact, be external findings. In addition, the domain specialist estimates that roughly 50% of workups will result in no findings at all.

[6]As an example of the above, if there is an external finding that urinary protein exceeds 2+ or urinary sugar exceeds a trace, workups for proteinuria and diabetes, respectively, will be ordered.

[7]Upper GI Series only.

[8]Whenever "referral" is used in Table 4.6, it (a) excludes referral for mere performance of a workup, and (b) excludes referral for treatment. All listed referrals are for specialized diagnosis beyond the scope of the typical internist.

[9]Barium enema only.

[10]From the higher threshold and all the potential contraindications, it might seem surprising that Lymphoma II precedes Lymphoma III. This is because, as noted above, external findings are always possible, and a needle biopsy of the liver is a rich source of these, while a bone marrow biopsy is not.

[11]Prostate cancer has two workups: one for direct evidence, one for circumstantial evidence. The evidence is, in all cases, not additive. Hence in this one instance—where combinatorial explosion is absent—the additive model is effectively bypassed by the device of parenthesizing and bracketing all values.

Note: This chapter was first published by the IAKE; however, it has been considerably revised.

5. THE LOGIC OF THE MODEL AND EXPERT INCONSISTENCY: ILLUSTRATIONS FROM THE MEDICAL DOMAIN

In his seminal paper, Schoenmakers (1986) presents a logical paradigm that neatly captures the problems inherent in expert systems with knowledge bases constructed from multiple experts (or from a database compiled from multiple sources). The problem, in a word, is inconsistency. Schoenmakers illustrates in a simple but compelling way how inconsistencies can arise which are implicit and hence not capable of being detected.

In this chapter, we apply Schoenmakers' insight to both rule-based systems built from multiple experts and to the single expert matrix-based systems described earlier. It turns out that both types of expert systems are vulnerable to these implicit inconsistencies, although it is necessary to move from the propositional calculus of Schoenmakers' paradigm to the predicate calculus, a development not surprising given that the same development was needed for earlier but parallel work on the legal domain (Fulda, 1988, 1989).

Schoenmakers' Paradigm. In Schoenmakers' paradigm, the first expert claims P, the second expert claims P—>Q, and we may therefore infer Q. But as Schoenmakers observes, *both* may actually believe -Q. Moreover, the knowledge base is internally consistent, as well as consistent with the erroneous conclusion drawn. Furthermore, each expert's position is consistent

($P \;\&\; -Q$ and $P \longrightarrow Q \;\&\; -Q$)! In fact, the two experts disagree on something on which one of them has taken no position, the truth value of P.

The Logical Structure of Rule-based Systems. Rule-based systems without Winstonian censors are generally collections of universally general propositions, i.e. formulations in the predicate calculus such as $(\forall x)(Ax \longrightarrow Bx)$ and often read "All A's are B's."

The Logical Structure of Matrix-based Systems. As we discussed earlier, the basic idea animating the construction of matrix-based systems is to accumulate evidence, each bit of which is assigned numerical evidentiary weight, and then to measure this accumulation against a threshold. If the cumulative weight of the evidence exceeds the threshold, then appropriate action is taken, such as the performance of a diagnostic procedure. In the predicate calculus, this can be represented as $(\forall x)(Ax \;\&\; Bx \;\&\; Cx \longrightarrow Tx)$ or $(\forall x)(Ax \;\&\; Bx \;\&\; Cx \longrightarrow -Tx)$ depending on whether x being an A, B, and C is or is not a sufficient indication for taking the action in question.

Medical Illustrations of Implicit Inconsistencies. The bulk of this section will consist of medical illustrations of implicit inconsistencies of the sort first described by Schoenmakers applied to a real-world domain of importance.

We will present each illustration by table and formulae supplemented by verbal exposition. These illustrations do assume the rule-based paradigm (which, after all, is the dominant paradigm for expert systems today); however, we will discuss matrix-based systems as well and draw the appropriate parallels.

Following are seven examples which reify the problem of using multiple experts in constructing a knowledge base for an expert system—because in each case, the experts differ. In presenting these examples, we both explain the medicine and medical terminology involved and pay close attention to the logical structure involved.

The medical illustrations are: (a) AZT treatment for HIV-infected but asymptomatic patients, (b) Consumption of alcohol when receiving anti-coagulants, (c) Experimental treatment for patients with incurable diseases, (d) Valve replacement in patients with aortic stenosis, (e) The conditions under which an emergency EGD should be performed on patients with hematemesis, (f) Referral to a psychiatrist of patients with recurrent syncope and no evidence of heart or CNS disease after a thorough workup, and (g) Administration of isoniazid to patients with inactive TB.

(a) AZT treatment for HIV-infected but asymptomatic patients. AZT is a drug which has shown promise in treating people with HIV, the AIDS virus. Unfortunately, long-term use of AZT often becomes intolerable to patients and even valueless as the virus mutates and becomes resistant to the drug. Some experts understandably recommend immediate treatment upon detection of the virus in the hope of slowing it down considerably. Other experts, pointing to the difficulties associated with long-term use, would rather save the most potent weapon presently available for later stages of the disease, viz. when it becomes symptomatic. The following table shows the logical formulation of each expert's position, the logical formulation of positions they hold in common, and a definition of the predicates involved.

TABLE 5.1 **Expert 1: $(\forall x)(Ix \rightarrow Ax)$**
Expert 2: $(\forall x)(Ix \rightarrow -Ax)$
Both: $(\exists x)Ix$

Ix: x is HIV-infected but asymptomatic.
Ax: x should be treated now with AZT.

Placing both experts' views in the knowledge base, in addition to what they jointly believe, creates a contradiction. To see this, it is only necessary to instantiate what they both grant and use that parameter, say a, to produce Aa & -Aa. Note that expert 1's positions are consistent, expert 2's positions are consistent, and the knowledge base resulting from both of their beliefs will be consistent, unless the designer is astute enough to include the existentially general proposition which everyone takes for granted. If the designer is not that astute, the computer, regardless of its deductive capabilities, will be unable to detect the implicit inconsistency.

(b) Consumption of alcohol when receiving anticoagulants. While medical experts agree that consumption of alcohol while taking anticoagulants must be limited, some would eliminate it altogether, while others permit up to an ounce daily. The following table illustrates.

TABLE 5.2 **Expert 1: $(\forall x)(Ax \rightarrow Nx)$**
Expert 2: $(\forall x)(Ax \rightarrow -Nx)$
Both: $(\exists x)Ax$

Ax: x is taking anti-coagulants.
Nx: x should not imbibe *at all*.

The logic here is identical to that of (a) above. Yet this is a more difficult scenario to deal with, because what is jointly believed is *so* obvious that it is very hard to imagine someone remembering to place it in the knowledge base.

(c) Experimental treatment for patients with incurable diseases. Experts disagree, philosophically, on when experimental treatments should be tried on human subjects. Some feel that such measures are justifiable only when the patient is terminally ill, because he has nothing to lose. Others argue that the contrary is true and that treatment should be made available when the patient has something to gain and that, in any case, patients in the end-stages of a terminal disease do not produce a study of much scientific value. The following table illustrates. (Note: the domain of discourse is limited to diseases without any established treatment.)

TABLE 5.3. **Expert 1: $(\forall x)(Tx \rightarrow -Lx)$ & $(\forall x)(-Lx \rightarrow Ex)$**
Expert 2: $(\forall x)(-Tx \rightarrow Gx)$ & $(\forall x)(Gx \rightarrow Ex)$
Both: $(\exists x)-Ex$

Tx: x is terminally ill.
Gx: x (and science) have something to gain.
Lx: x has something to lose.
Ex: x should be enrolled in the experimental protocol.

Consider that expert 1's position, after several logical manipulations, amounts to $(\forall x)(\text{-}Ex \longrightarrow \text{-}Tx)$, while expert 2's position amounts to $(\forall x)(\text{-}Ex \longrightarrow Tx)$. Thus, several logical manipulations lead us back to the paradigm of (a) and (b) above, and, again, unless the existentially quantified statement, which is fairly obvious, is expressly placed in the knowledge base, the contradiction will be hidden.

(d) **Valve replacement in patients with aortic stenosis.** In this example, the domain of discourse is limited to patients with hemodynamically significant, treatable aortic stenosis. The narrowing of the aortic valve becomes significant to the flow of blood if its cross-sectional area is less than $^1/_2 cm^2$ [illegible] may be treated in patients with mild symptoms or no symptoms. Experts agree that valve replacement (either a prosthetic valve or a porcine valve graft) is indicated in patient with mild symptoms: Class I (NYHA classification) heart failure (heart failure upon severe exertion), stable angina, or syncope (fainting) without any other symptoms. Some experts, however, believe that it is also indicated in patients with no symptoms (minimal or no myocardial damage) for whom it is a less risky operation and promises better results. Despite this, the experts are divided because placement of a new valve (or a porcine graft) often leads to thrombotic complications and the necessity of taking anticoagulants or, especially in the case of porcine grafts, several operations when the graft or prosthesis later fails and must be replaced. The following table illustrates.

TABLE 5.4. **Expert 1: $(\forall x)(Vx \longrightarrow Mx)$**
Expert 2: $(\forall x)((Mx \vee Nx) \longrightarrow Vx)$
Both: $(\forall x)(Nx \oplus Mx)$ & $(\exists x)Nx$

Nx: x has no symptoms.
Mx: x has mild symptoms.
Vx: x's valve should be replaced.

This is a substantial variation on previous examples. Because of the existentially general proposition, we know that some patient is asymptomatic. We also know from the universally general proposition accepted by all that having mild symptoms and having no symptoms are mutually exclusive. Given that, there is a case where Expert 1 will not order a valve replacement and Expert 2 will. A contradiction is made express, however, only if the system designer thought to include not only the existential as in previous examples but also the obvious mutual exclusivity condition, which is not obvious at all to the computer! The contradiction will ultimately have the form Ma & -Ma, rather than an inconsistency regarding the predicate V. Of course, it can be transformed into such an inconsistency.

(e) **The conditions under which an emergency EGD should be performed on patients with hematemesis.** Here the domain of discourse is limited to patients with hematemesis (vomiting up blood). Some experts would order an emergency EGD (endoscopy of the esophagus, stomach, and duodenum) in order to find the source of the bleeding no matter what state that patient's liver is in, while others would shy away from doing so in patients with decompensated cirrhosis or acute hepatitis. The reason for the caution is that an EGD requires sedatives such as Valium and analgesics such as Demerol, medications which are dangerous if the patient's liver is not functioning. The following table illustrates.

TABLE 5.5. **Expert 1: $(\forall x)((Lx \vee -Lx) \longrightarrow Ex)$**
Expert 2: $(\forall x)(Ex \longrightarrow -Lx)$
Both: $(\exists x)Lx$

Lx: x has decompensated cirrhosis or acute hepatitis.
Ex: x should get an EGD.

The logic of this example is almost identical to that of (d), the only difference being that the mutual exclusivity of Lx and -Lx is a logical, rather than an empirical, truth, making the contradiction more likely to surface depending on the deductive ability of the system. Of course, the knowledge base still must include the existentially general proposition.

(f) Referral to a psychiatrist of patients with recurrent syncope and without any evidence of heart disease or CNS disease after a thorough workup. The domain of discourse is limited to patients without any evidence of heart disease or central nervous system disease after a thorough workup and who would not be referred to a psychiatrist for reasons other than recurrent syncope. While everyone agrees that either depression or a psychiatric disorder such as somatization, hypochondriasis, or schizophrenia are possible causes of recurrent syncope, some experts believe that the patient may instead have an undiagnosable CNS disease which will first be manifest later and will therefore not make a psychiatric referral when not otherwise indicated. Other experts believe that such an undiagnosable, early disease of the CNS is such a rarity as not to be a consideration in treatment. The following table illustrates.

TABLE 5.6. **Expert 1:** $(\forall x)(Sx \longrightarrow -Rx)\ \&\ (\forall x)(Sx \longleftrightarrow (Dx \vee Px \vee Ux))$
Expert 2: $(\forall x)(Sx \longrightarrow Rx)\ \&\ (\forall x)(Sx \longleftrightarrow (Dx \vee Px))$
Both: $(\exists x)Sx$

Sx: x has recurrent syncope.
Rx: x should be referred to a psychiatrist because of fainting spells.
Dx: x has depression.
Px: x has a psychiatric disorder like those listed above.
Ux: x has an undiagnosable disease of the CNS.

The logic of this example is slightly more complicated than that of (a) and (b), since the core of the difference of opinion is over the cause of Sx and if Ux is a significant possible cause, then -Rx. In other words, those who take Ux as a real possibility prefer that diagnosis to Dx or Px. (Recall that there are no other symptoms which would cause referral to a psychiatrist as we defined the domain of discourse.)

(g) Administration of isoniazid to patients with inactive TB. This is yet another variation on the theme presented in (a) and (b), as the table following will show. The minority view is that inactive TB (as discovered by a skin test) should be treated even in patients over 35, while the majority view is that isoniazid is sufficiently dangerous to preclude its use except in younger patients, symptomatic patients, and in special high-risk circumstances.

TABLE 5.7. **Expert 1: $(\forall x)((Tx \ \& \ 35x) \longrightarrow Ix)$**
Expert 2: $(\forall x)((Tx \ \& \ 35x) \longrightarrow -Ix)$
Both: $(\exists x)(Tx \ \& \ 35x)$

Tx: x has inactive TB.
35x: x is at least 35 years old.
Ix: x should be treated with isoniazid.

The logic here is identical to that of (a) and (b), although here the protases (I use the term loosely) of the quantified conditionals (again, loose usage) are two-membered, not one-membered.

Matrix-based Systems. Referring back to our discussion of the logical structure of matrix-based systems, we can see that implicit inconsistencies in matrix-based systems are an extension of case (g), immediately above, the only difference being that, as shown earlier, the protasis of the quantified conditional is often many-membered.

In a previous paper (Fulda, 1988, 1989), we show how implicit inconsistencies can arise even when only one expert is involved and continue with a discussion of the logic involved (propositional attitudes). Likewise a main concern of this work is the elicitation of consistent information from a single domain specialist: the example with A, B, C, and T can thus be understood as arising from two experts or from one at different times. In the latter case, and as described earlier in this work, it is one of the functions of the knowledge engineer to produce an integritous knowledge base, our method being to approach the task as a quasi-legal process and to use the tools of adversary examination, particularly cross-examination.

6. THE CONTRIBUTIONS OF THE PRESENT MODEL

LaFrance (1989) sums up the state-of-the-art in knowledge elicitation most aptly as follows: "In the usual case, the knowledge engineer engages in intense, generally unstructured interviewing of the expert.... (T)he use of the unstructured interview, is by far the most common method. Indeed some developers have taken it for granted that an unstructured interview is essentially the only way to extract expert knowledge...." As we have discussed at length in opening Chapter 4, the use of an additive model, originally suggested by the domain specialist, reduces combinatorial explosion. However, if ordinary ruled-based systems contain errors, a matrix of evidentiary weights is, *in se,* **totally** unreliable. By using the legal process as a model, especially cross-examination, the additive model becomes practicable. This is the first intended contribution of this work. Beyond that, as we have discussed in the literature survey, there have, to my knowledge, been no variants of the interview (as opposed to the automated) techniques that are truly structured. The legal process, in contrast, is highly structured and its use produces a genuinely engineered methodology for knowledge elicitation. This is the second intended contribution of this work.

There is also, perhaps surprisingly, a third contribution. The standard textbook is Karen L. McGraw and Karan Harbison-Briggs' *Knowledge Acquisition: Principles and Guidelines* and is written rather in the style of a management textbook. Chapter 9 entitled "Acquiring Knowledge from Multiple Experts" is completely devoid of any degree of formalism. Thus, the section "Multiple Experts—Pros and Cons" speaks of "ease of access," "weakness of single lines of reasoning," "equality" of experts, "upward-ripple paranoia," "confidentiality," "consensus versus diversity," and the like. Yet the text thoroughly surveys all the literature that we had also mapped out for examination on the subject of multiple experts. Hence, it appears that the preceding chapter formalizes what has remained an informal if recognized problem in knowledge acquisition: expert disagreement.

In this regard, it is well worth noting that the predicate logic, as simple as it may seem, is a powerful tool for treating questions of consistency and can get quite subtle and complex as precision in meaning is forced to the fore. This was first made clear during my work on the logic of expert inconsistency with the domain specialist and was underscored further when Professor Levin, in reviewing the material, uncovered additional ambiguities, subtleties, and, indeed, outright errors. While what we have presented here in the way of a formalism does not partake of the incredible richness of modern logics, it seems to be a substantial advance over management-like discussions on consensus and diversity.

7. THE CONCEPTUAL ARCHITECTURE OF THE KNOWLEDGE ELICITATION MODEL and FORMAL HEURISTICS FOR CROSS-EXAMINATION

As shown in Figure 7.1, the knowledge elicitation model contains one primary component, the knowledge base, and inputs and outputs based on the legal analogy developed earlier. In turn the knowledge base is the basis of the expert system, here shown as a black box, since our sole concern with it is its interaction with the knowledge base constructed using the methods described earlier.

Figure 7.2 presents a partial formalization, using SETL pseudocode, of most of the ten principal heuristics used for cross-examination.

The two figures together formalize, to a certain extent, the art of cross-examination within the context of an expert systems design project. "To a certain extent," because debriefing an expert is as much an interpersonal task—requiring the knowledge engineer to "read" him—as an intellectual task and the heuristics used (and, perhaps, a good part of the paradigm) may have to undergo slight-to-moderate changes to accommodate the personality of the domain specialist. That is why cross-examination is termed an "art": it is heavily dependent on the subject being examined.

Figure 7.1 is self-explanatory, given the previous chapters, except for one apparent oddity. After re-cross, local or global, the matrix is changed without using re-direct. This is because by this time the matrix is fairly robust and the knowledge engineer can adjust the evidentiary matrix with only cursory assistance from the domain specialist. This is also the explanation for our earlier remark that re-cross plays a greater role than re-direct, and is confirmed, experimentally, on the medical domain for which this model was built. In addition to the robustness of the matrix, it has been our experience that the knowledge engineer begins to acquire a fairly good understanding of the domain involved, particularly if it is well-circumscribed. In other words, the domain specialist teaches both knowledge engineer and computer simultaneously.

Figure 7.2 is also rather direct, but a knowledge of SETL (or similar mathematical) notation would be very helpful (Schwartz, 1986). Two of the ten heuristics are not given (highly correlated and unrelated items), because such heuristics are domain-specific and cannot be automated, unless the computer is programmed as very much more than a conventional knowledge base (even with deductive capabilities).

FIGURE 7.1. THE CONCEPTUAL ARCHITECTURE OF THE

KNOWLEDGE ELICITATION MODEL

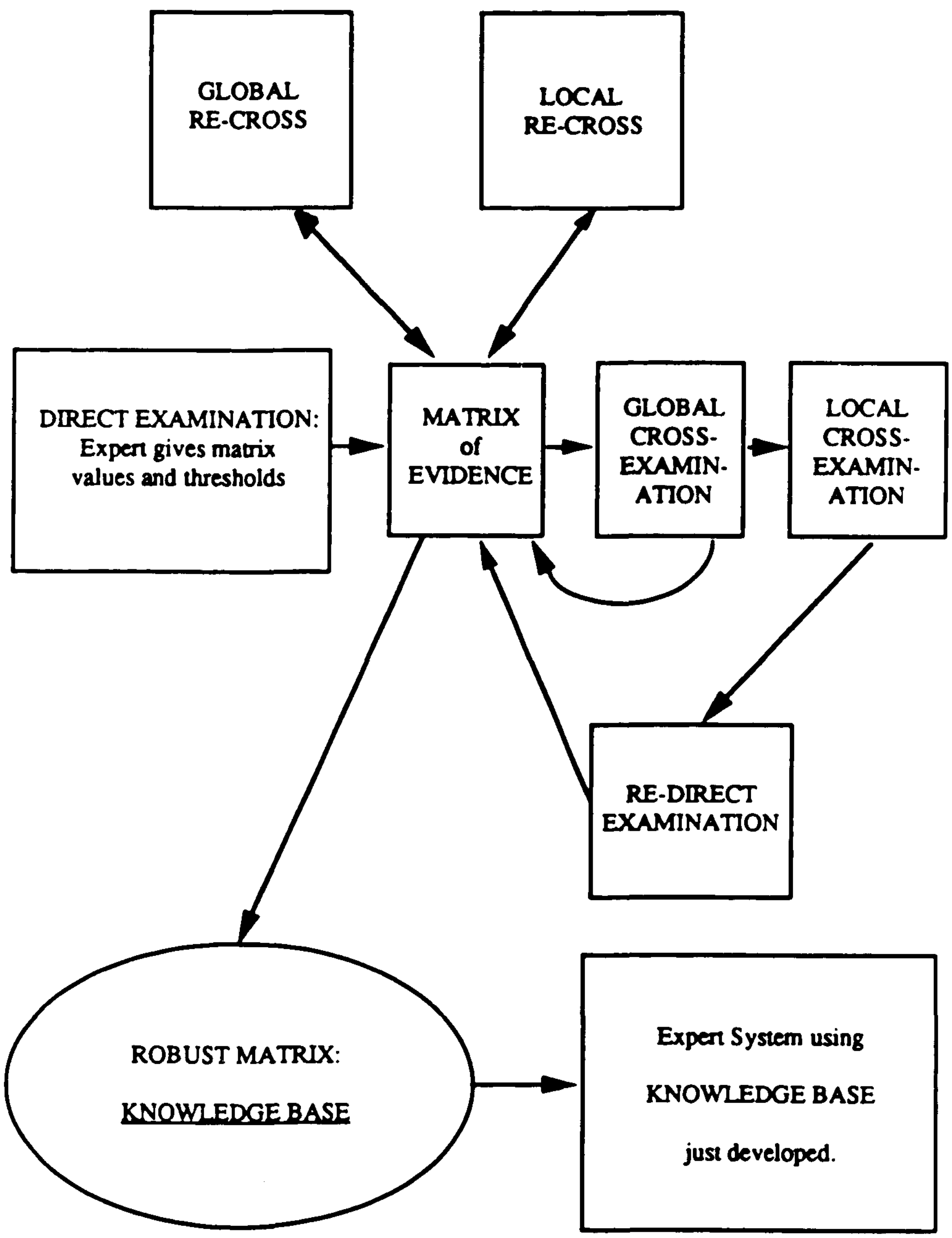

FIGURE 7.2. PARTIAL FORMALIZATION OF HEURISTICS FOR CROSS-EXAMINATION USING SETL PSEUDOCODE

(forall of the (binary) decisions, =: D)

Read T * Threshold for the decision.

Read N * Number of evidentiary items.

Read M * (Maximum) number of values per evidentiary item.

Read E * A tuple of evidentiary values of length N.

*Objective: treat the hard cases of the (up to) m^n configurations of the

* evidence, with the intent of subsuming the remainder, avoiding

* combinatorial explosion.

*In each of the heuristics below, upon an incorrect (or, to be more precise, an

* inconsistent) response, it may be necessary to change the value

* of T or some of the values in E.

*Heuristic 1.

(forall subtuples of E which sum to T exactly)

Ask the domain specialist if D is indicated. If he says "no," then it is necessary to adjust some values, as discussed earlier.

end.

*Heuristic 2.

(forall subtuples of E which sum to T-1)

Ask the domain specialist if D is indicated. If he says "yes," then it is necessary to adjust some values.

end.

*Heuristic 3A.

(forall subtuples of E which sum to T exactly)

(for each value in the present subtuple which equals 1)

Delete the evidentiary item (by valuing it 0) from the subtuple of E which summed to T and now sums to T-1. Ask the domain specialist if D is indicated. If he says "yes," then it is necessary to adjust some values.

end.

end.

*Heuristic 3B.

(forall subtuples of E which sum to T-1)

(for each value in E which is exactly 1)

Add the value of the evidentiary item to the tuple (say, at the end). Ask the domain specialist if D is indicated. If he says "no," then it is necessary to adjust some values.

end.

end.

*Heuristics 4 and 5 require extensive knowledge of the domain, more than
*practical to include in a stand-alone expert system.

*Heuristic 6.

(forall subtuples of E which sum to T-1)

(for each value in E which is weighted D (for this decision))

Add the value of the evidentiary item to the tuple (say, at the end).

Ask the domain specialist if D is indicated. If he says "yes," then it is necessary to adjust some values.

end.

end.

*Heuristic 7.

If the domain specialist gives an inconsistent response (i.e., one which contradicts his numbers on direct examination) when being cross-examined using heuristics 2, 3B, or 6, try these heuristics again substituting T-2 for T-1.

*Heuristic 8.

If the domain specialist gives an inconsistent response (as explained earlier) when being cross-examined using heuristics 1 or 3A, try these heuristics again, substituting T+1 for T.

*Heuristic 9.

If T differs among groups (subtuples; in medicine, often by sex) as a percentage of the sum of all values in E, check the direction and magnitude of the difference and make certain—with the domain specialist's assistance—that the differences withstand scrutiny.

*Heuristic 10.

After allowing for sufficient elapsed time, pose the same subset of evidentiary items to the domain specialist, but in a different order, and then observe whether the expert has been consistent or had been influenced by the presentation, in which case, again, some values may have to be adjusted.

end. *main loop

SETL's principal advantage is its allowing quantification over sets and tuples (ordered lists). In the above pseudocode, universal quantification is treated almost identically as in SETL. Existential quantification in SETL is treated, as here, by posing a question: "If there exists an x in S such that P(x) then...." Likewise, our querying of the domain specialist is in essence an erotetic existential quantifier. If the answer is affirmative, the existential quantification holds and, as in SETL which binds a value to x, a (possibly definite) description follows. If the answer is negative, a universally quantified statement results. The principal point is that by setting the quantificational structure, SETL provides a very-high level, formal framework for the problem.

8. FUTURE RESEARCH DIRECTIONS

Although well beyond the scope of this thesis, it would be interesting to combine cross-examination with non-verbal methods of knowledge acquisition from experts. In particular a clinical trial comparing an expert system designed using our model with the actual practices of the domain specialist on a sample population should be interesting.

If there are inconsistencies between the reproducible beliefs of the domain specialist and his actual practice, he may not be aware of all the factors he uses to arrive at his decision or, perhaps less likely, omits some of the practices he claims to use from those he actually uses. After a thorough cross-examination, such discrepancies are unlikely to be resolved by modification of the evidentiary weights and thresholds. Rather a careful process-trace will likely reveal additional factors (or, occasionally, fewer factors, or both) that subliminally affect the expert's practices which should be subject to conscious scrutiny.

In particular, it is well known that some psychiatrists do little or no good, and perhaps harm, while others have a measure of success. It is not well understood why and facile answers relating to "competence" should be rejected in favor of the more sophisticated view, entirely in line with psychoanalytical thought, that the successful psychiatrists themselves are unaware, at least consciously, of many of the factors behind their successes. A model such as proposed here may bring such reasons to light, and enable their use by more practitioners. We suggest psychiatry, in particular, because it is the least-developed branch of medicine with the greatest variation in diagnosis and treatment among specialists. Also, psychiatrists are themselves trained in knowledge elicitation and may find the role of domain specialist comes to them more easily and naturally than to other physicians to whom rigorous cross-examination may seem somewhat demeaning.

A first project would, given the state-of-the-art in psychiatry, almost necessarily focus on psychopharmacology, an increasingly well-defined research area which is helping bring psychiatry into mainstream medicine. One might, for example, debrief an expert on the benzodiazepenes and then, with names deleted, examine his records to see whether his clinical practices match his statements regarding them.

It should be pointed out that the benefits of comparing cross-examination with actual practice can be gained without any computer implementation. Particularly if the domain is very circumscribed—e.g., when to prescribe benzodiazepenes and which to prescribe—the discrepancies between practice and verbal (however structured) methods of knowledge acquisition will be noted simply by examining the matrix arising from the latter and the matrix (or other data structure) constructed from the former. An expert system is designed to save time and lives, but the framework developed here has the same life-saving possibilities (although not with the benzodiazepenes) without implementation. The discrepancies between the two methods of knowledge acquisition are noticeable even if neither is computerized; it is the methodology that matters. Of course, the more ambitious the project in scope the more computerization becomes much to be desired.

Furthermore, although we have presented cross-examination as a variant of the interview method, it is clear from the preceding chapter that, with sufficient development work, the same methodology could be a variant of the automation techniques. We have taken the traditional approach to AI, canonized in the ACM's classification system, that the entire field—including knowledge engineering—comes under the heading of *Computing Methodologies* (category I, with knowledge engineering falling under both I.2.1 and I.2.6).

REFERENCES

Bainbridge, L., "Verbal Reports as Evidence of the Process Operator's Knowledge," *The International Journal of Man-machine Studies* 11(1979):411-436.

Belkin N., Brooks, H., and Daniels, P., "Knowledge Elicitation Using Discourse Analysis," *The International Journal of Man-machine Studies* 27(1987):127-144.

Boose, J., "A Knowledge Acquisition Program for Expert Systems Based on Personal Construct Psychology," *The International Journal of Man-machine Studies* 23(1985):495-525.

Brown, B., "The Taming of an Expert: An Anecdotal Report," *SIGART Newsletter* 108(1989):133-135.

* Buchanan, B. and Smith, R., "Fundamentals of Expert Systems," *Annual Review of Computer Science* 3(1988):23-58.

Burzesi, T., *TECREK: A Software Tool That Helps Ensure the Completeness of Rule-based Expert-system Rule Sets*, M.S. Thesis, 1989, Hofstra University.

Crandall, B., "A Comparative Study of Think-Aloud and Critical Decision Knowledge Elicitation Methods," *SIGART Newsletter* 108(1989):144-146.

Fletcher, G., *A Crime of Self-Defense: Bernhard Goetz and the Law on Trial* (The Free Press, 1988).

Fulda, J., "A Case Study in Computational Science," *Simuletter* 19(1988):(1)23-25.

Fulda, J., "The Logic of Expert Judging Systems and the Rights of the Accused," *AI & Society: The Journal of Human and Machine Intelligence* 2(1988):266-269 and in expanded form in *Computers and Law: The Journal of the Society for Computers and Law* 60(1989):14-16.

Fulda, J., Reviews in *Computing Reviews* 29(1988):383-384; 384; 665; 665-666; 30(1989):430; 492; 492-493.

Fulda, J., "Knowledge Engineering in a Combinatorial Setting as a Quasi-legal Process: A Medical Case Study—Tiredness Induced by Malignancies in Certain Populations of Patients," *Proceedings of the Annual Conference of the International Association of Knowledge Engineers* (1989):79-104.

Hoffman, R., "A Brief Survey of Methods for Extracting the Knowledge of Experts," *SIGART Newsletter* 108(1989):19-27.

Johnson, P., Zualkernan, I., and Garber, S., "Specification of Expertise: Knowledge Acquisition for Expert Systems," *The International Journal of Man-machine Studies* 26(1987):161-181.

Kahn, G., Nowlan, S., and McDermott, J., "Strategies for Knowledge Acquisition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 7(1985):511-522.

LaFrance, M., "The Knowledge Acquisition Grid: A Method for Training Knowledge Engineers," *The International Journal of Man-machine Studies* 26(1987):245-255.

LaFrance, M., "The Quality of Expertise: Understanding the Differences between Experts and Novices," *SIGART Newsletter* 108(1989):6-14.

Lavrac, N., "Methods for Knowledge Acquisition and Refinement in Second Generation Expert Systems," *SIGART Newsletter* 108(1989):63-69.

Martin, J. and Redmond, M., "Acquiring Knowledge by Explaining Observed Problem Solving," *SIGART Newsletter* 108(1989):77-83.

McGraw, K. and Harbison-Briggs, K., *Knowledge Acquisition: Principles and Guidelines* (Prentice-Hall, 1989).

Micciche, P. and Lancaster, J., "Applications of Neurolinguistic Techniques to Knowledge Acquisition," *SIGART Newsletter* 108(1989):28-33.

* Michie, D., "Current Developments in Expert Systems" in Quinlan, J. (ed.), *Applications of Expert Systems* (Addison-Wesley, 1987).

Narayan, N. and Viswanadham, N., "A Methodology for Knowledge Acquisition and Reasoning in Failure Analysis of Systems," *IEEE Transactions on Systems, Man, and Cybernetics* 17(1987):274-288.

Newell, A., "Fairytales," *Viewpoints* 3, n.d., Carnegie Mellon University (unpublished).

* Quinlan, J., Compton, P., Horn, K., and Lazarus, L., "Inductive Knowledge Acquisition: A Case Study," in Quinlan, J. (ed.), *Applications of Expert Systems* (Addison-Wesley, 1987).

Schoenmakers, W., "A Problem in Knowledge Acquisition," *SIGART Newsletter* 95(1986):56-57.

Schwartz, J., Dewar, R., Dubinsky, E., and Schonberg, E., *Programming with Sets: An Introduction to SETL* (Springer-Verlag, 1986).

Shadbolt, N. and Burton, M., "The Empirical Study of Knowledge Elicitation Techniques," *SIGART Newsletter* 108(1989):15-18.

Silvestro, K., "Using Explanations for Knowledge-base Acquisition," *The International Journal of Man-machine Studies* 29(1988):159-169.

* Shanteau, J., "Psychological Characteristics of Expert Decision Makers" in Mumpower, J., et al. (eds.), *Expert Judgment and Expert Systems* (Springer-Verlag, 1987).

Slator, B., "Extracting Lexical Knowledge from Dictionary Text," *SIGART Newsletter* 108(1989):173-174.

* Stevenson, R., Manktelow, K., and Howard, M., "Knowledge Elicitation: Dissociating Conscious Reflections from Automatic Processes" in *Proceedings of the Fourth Conference of the British Computer Society* (1988):565-579.

* Stewart, T. and McMillan, C. "Descriptive and Prescriptive Models for Judgment and Decision Making: Implications for Knowledge Engineering" in Mumpower, J., et al. (eds.) *Expert Judgment and Expert Systems* (Springer-Verlag, 1987).

* Taylor, J., "Expert Systems: Where Do We Go from Here?" in Cotterell, A. (ed.), *Advanced Information Technology in the New Industrial Society: The Kingston Seminars* (Oxford University Press, 1988).

Waldron, V., "Investigating the Communication Problems Encountered in Knowledge Acquisition," *SIGART Newsletter* 108(1989):143-144.

Winston, P., *Artificial Intelligence* (2nd Edition) (Addison-Wesley, 1984).

* These pieces have been reviewed by the present author in *Computing Reviews*.