21

Anticipating the Second Decade

Edward H. Shortliffe and William J. Clancey

The research efforts described in this book may paradoxically appear to be both hideously complex and yet ridiculously simplistic—complex in the range of concepts they attempt to capture, encode, and use effectively, but simplistic in the important areas of human knowledge and common sense that they ignore (but that we know can be crucial to excellent clinical decision making). Viewed in this light, the research invites the question whether "ultimate" AIM systems, when they are eventually constructed, will be manageable and amenable to ongoing refinement. Or will they become so large and complex that they will totally outgrow the ability of their developers to cope with their knowledge bases and with the need for ongoing verification and updating?

It is certainly true that the research has raised at least as many new questions as it has answered old ones, but such is the nature of scholarly investigation. It is unlikely that we will ever see the day when all questions have been answered and all the problems solved. However, as the field progresses, we believe that useful (albeit limited) tools will increasingly become available, particularly as the hardware revolution (made possible by large-scale integration) provides the AIM field with cost-effective vehicles for moving advice programs from research laboratories to hospitals and private offices. Hardware and software advances are also beginning to offer us new models of system-building environments, ones in which graphical capabilities and interactive tools provide knowledge engineers with effective methods for dealing with systems that are much too large to be managed using traditional hard-copy listings for reference (Tsuji and Shortliffe, 1983).

Many of the ideas presented in this chapter were previously discussed by E. H. Shortliffe (1982a; 1982b).

In this final chapter, we summarize the trends of the past decade while citing the important research problems that remain to be solved in the years ahead. The discussion is motivated by a summary of the design considerations that have been identified by asking physicians what they would demand from a clinical consultation system before they would be willing to use it routinely. We also identify those kinds of medical problems for which practical systems can be built soon, using the kinds of techniques that have been developed during the 1970s. The ultimate systems are still probably many decades away, but existing techniques help define a subset of problems with which we are already prepared to deal.

21.1 What Physicians Want

Researchers in the field of medical decision making must contend with a great deal of ambivalence on the part of the potential physician users of their systems. On the one hand, there is a "show me" attitude expressed by a profession that has heard the potential of clinical computing extolled for more than ten years but has yet to see a widely accepted decision support system. On the other hand, there are indications that the environment is changing, with an increased acknowledgment that clinical decision-making research can validly contribute to medical practice. For example, we have seen significant clinical changes result from theoretical work in clinical decision analysis (e.g., the recent American Cancer Society recommendations regarding mammography and PAP smear screening) and the development of an ambitious, well-received journal in the field (Lusted, 1981). Studies of physician attitudes (Teach and Shortliffe, 1981) have also shown that there is a growing curiosity about computers and a heightened faith in their potential. This phenomenon has been further demonstrated by the emergence of doctors with home computers and customized office systems, and by the success of educational programs designed to introduce physicians to computers for both business and clinical applications.

The study of physicians' attitudes towards clinical consultation systems (Teach and Shortliffe, 1981) showed that a significant segment of the medical community believes that assistance from computer-based consultation systems will ultimately benefit medical practice. Teach and Shortliffe also studied the physicians' demands regarding desirable features for such systems if they are to be useful and clinically accepted. The resulting design considerations highlight performance capabilities that are a challenge to medical computer scientists. Consider, for example, the six design features that physicians rated *most important* for future consultation systems:

1. they should be able to explain their diagnostic and treatment decisions to physician users;

2. they should be portable and flexible so that the M.D. can access them at any time and place;
3. they should display an understanding of their own medical knowledge;
4. they should improve the cost-efficiency of tests and therapies;
5. they should automatically learn new information when interacting with medical experts; and
6. they should display common sense.

No current consultation system meets all these criteria, but the list does help identify both the research challenges that lie ahead and the criteria for assessing new systems that may be introduced. The first, third, fifth, and sixth of these criteria are central issues being addressed by researchers in the AI field and thereby emphasize the importance of AI as an ingredient in the development of clinically acceptable decision aids.

21.2 Two Decades of Research

Medical decision-making research in the 1960s emphasized the use of the computer to deal with probabilistic information, to recognize patterns using numerical techniques, to model physiological processes that were amenable to mathematical simulation, or to encode algorithmic approaches to routine clinical chores. The field was then in its first decade as an identifiable area of research, and the emphasis was on how to get machines to make decisions that were both accurate and reliable. Formal statistical approaches that had been impractical before computers became available were, quite naturally, the first techniques to be tried as physicians and engineers began to appreciate the computer's potential as a clinical tool.

In the 1970s, however, there was a shift in research direction. As was outlined in Chapter 3, investigators increasingly realized that there are several key problems that escape attention if the research focuses solely on the development of techniques for reaching good decisions. These include:

1. the problem of *data acquisition*—how to acquire, encode, and control for variations in the descriptors that define patients and populations;
2. the problems of *knowledge acquisition and representation*—how to acquire and encode the kinds of judgmental perceptions and the commonsense approach that characterize expertise in the clinical decision-making areas being modeled;
3. the problem of *explanation*—how to build decision support programs that not only give advice but are able to defend their decisions in terms physicians can understand; and

4. the *logistics of integration*—how to design and implement computer-based decision aids that fit smoothly into the daily routine of physicians' practices, that acknowledge their hectic schedules, and that seek to demystify and simplify the mechanics of the human-computer interface.

Several early approaches to these problems were developed during the last decade. Large patient data bases have been constructed and used to aid in defining prognoses for new cases (Feinstein et al., 1972; Fries, 1972; Rosati et al., 1975). Investigators who depend on valid statistics to support their decision-making systems have begun to look at geographical variations in populations in order to assess the transferability of programs (de Dombal, 1979). Hospital information systems have become increasingly common and provide promising early models for the way in which relevant data will eventually be routinely acquired (Lindberg, 1977). There has also been complementary work in the development of large computer-based text documents designed to bring up-to-date knowledge of a domain to the practicing physician (Bernstein et al., 1980).

During the same period, AI approaches have become prominent and have suggested several methods for encoding uncertainty, representing expert knowledge, and modeling the reasoning processes of accomplished clinicians. The symbolic reasoning techniques described in this book have suggested ways decision-making programs can explain their reasoning to physicians, thereby allowing the user to decide whether to follow the system's recommendations. Interactive techniques have been developed that also allow experimental systems to interview experts and to acquire new knowledge directly from them (Davis, 1979).

Finally, there have been several notable experiments that have sought new ways to encourage physicians to interact with computer programs. These have included systems using light pens (Watson, 1974) or touch screens (Schultz and Davis, 1979) and decision support programs integrated into large-scale hospital information systems (Pryor et al., 1982). These efforts and others have demonstrated that physicians will learn to use computers and accept their role if the benefits of the technology outweigh the costs of learning how to use the device and integrating it into one's normal routine.

21.3 The Challenges Remaining

A litany of recent accomplishments partly serves to emphasize the significant problems still remaining, however. Many of the experiments we have cited are only first steps toward the development of clinically useful tools. Some of the major barriers are practical ones relating to the logistics of interfacing patient data bases with expert systems, issues of legal liability (Brannigan, 1981), and the problem of training system users and knowledge engineers. At a more basic level, as is true with any emerging science, the development of short-term solutions tends to lead to a new understanding of the nature of the remaining problems and helps define the fundamental research directions for the future. Current results suggest that the following areas are among those requiring attention in the decade ahead:

1. additional *psychological studies,* similar in motivation to some of the pioneering studies of the 1970s (Elstein et al., 1978; Kassirer and Gorry, 1978), that will provide new insights into optimal methods for simulating expert decision-making performance and may suggest novel approaches to the organization of knowledge and its interaction with probabilistic information;
2. improved *techniques for representing and using causal and mechanistic relationships* (because expert decision-making behavior sometimes depends on an ability to reason from "first principles" rather than relying on empirical associations between observations and hypotheses);
3. improved methods for *acquiring expert knowledge, encoding it, and checking it for inconsistencies or incompleteness* (Davis, 1979; Suwa et al., 1982; Politakis and Weiss, 1984), thereby helping avoid the problems of knowledge base construction that have been major impediments to the development of expert systems;
4. enhanced *explanation capabilities,* ideally guided by an improved understanding of how human beings explain things to one another and, in particular, how they adapt their explanations to the knowledge and experience of the individual requesting advice;
5. experimentation with *new machine architectures* (e.g., parallel processing or networking of multiple coordinated processors) that may permit an optimal assignment of languages and interfaces for the individual subtasks required by high-performance decision-making programs;
6. experiments that seek to provide an *optimal melding of symbolic techniques* drawn from artificial intelligence research *and the analytic techniques* of formal statistics, pattern recognition, and decision theory; and
7. research into novel ways that developing *technologies for personal computing and graphics* might heighten both the acceptability and cost-effectiveness of systems to aid physicians with their decision-making tasks.

21.4 Steps in Demonstrating the Effectiveness of a Consultation System

With significant fundamental problems such as those above requiring solutions, can *anything* of practical use for decision support be implemented soon? Can we define clinical problems that are amenable to short-term solutions and that will allow AIM researchers to undertake validating experiments in active clinical environments rather than in hypothetical experimental settings such as those used for the evaluation of MYCIN (Yu et al., 1979a; 1979b) and INTERNIST-1 (Chapter 8)? We believe that the answer to both of these questions is "yes." Short-term clinical implementation is inherently intertwined with evaluation issues, however, and we have accordingly found it useful to define a series of steps through which an advice system must pass as it moves from a research environment to ongoing clinical use.

Diagnostic programs have tended to be assessed on the basis of their *decision-making accuracy*—the issue that is usually central to the system's design and to the motivation of the system's developers. Yet there are several additional components to the evaluation process when it is performed optimally. In order to control for confounding variables, we have suggested (Shortliffe and Davis, 1975) that system evaluations should be undertaken in a series of steps as follows:

1. *Demonstrate a need for the system.* Are there data indicating that physicians need help with the task for which the consultation system is designed to assist, and if so, is a computer necessary to provide that assistance?
2. *Demonstrate that the system performs at the level of an expert.* Can it be formally shown that the system reaches the same decisions as experts who are presented with the same clinical decision tasks? If there are frequent disagreements, can it be shown that the system is correct at least as often as the experts are? Note that the determination of correctness thereby requires some "gold standard" against which the performance of both experts and the consultation system can be measured.
3. *Demonstrate the system's useability.* Can physicians easily learn to handle the mechanics of interacting with the consultation system? Is the response time adequate? Is the system's performance sufficiently transparent so that the clinician can obtain the information he or she needs in an efficient and straightforward manner?
4. *Demonstrate acceptance of the system by physicians.* Can it be shown that clinicians offered the decision tool will in fact return to use it, even when access to it is entirely optional?
5. *Demonstrate an impact on the management of patients.* If physicians use the system, can it be shown that they follow the advice it offers? If not, has it favorably changed their behavior in some other way?
6. *Demonstrate an impact on the well-being of patients.* If physicians are following the recommendations of the consultation system, can it be shown that patients are benefiting from its use? Are there objective measurements of patient-care quality that can be assessed before and after the decision aid has been introduced?

7. *Demonstrate cost-effectiveness of the tool.* If all the other validation criteria have been satisfied, can it be shown that there is a version of the consultation system that is cost-effective when both costs and benefits are assessed using some generally accepted criterion?

These seven steps for demonstrating the effectiveness of a medical consultation system are idealized and difficult to traverse. We know of no medical decision-making system that has rigorously been shown to meet formal validation criteria at all seven steps of development. In fact, most systems have been assessed only at step 2, and remarkably few have met even the criterion of need specified in step 1.

Some observers of the field may argue that the theoretical issues in the development of high-performance consultation systems are still so great that it is folly to focus attention on steps 3 through 7 at this time. Yet many significant theoretical barriers to the successful implementation of consultation systems do not arise at step 2 and will not be met until the subsequent steps are encountered.

21.5 Characteristics of an Optimal Application Domain

Attitude surveys such as that of Teach and Shortliffe (1981) help delineate some of the issues that must be addressed by system builders if clinically acceptable decision tools are to be developed. However, since most of these issues are best studied and assessed at the later stages of system implementation, scientists who wish to address them in their *current* research must select an appropriate clinical problem area. The following criteria for that selection seem to be particularly pertinent:

1. As indicated above, there must be a *demonstrated need for help* in the domain. A program that deals with an "interesting" problem, but one with which physicians already do rather well, will generate little interest.
2. Equally as important, there must be a *recognized need for help* by the physicians themselves. Data showing poor performance by the overall population of physicians will not necessarily convince individual practitioners that they are among those needing help. Demand will come only from perceived need on the part of the intended users.
3. The domain should ideally provide a *core of formalized and readily available knowledge*. We have learned that knowledge base development can be an arduous and time-consuming aspect of consultation system re-

search. Theoretical issues regarding knowledge completeness, consistency, and acquisition must inevitably be faced when a complex system is built for a domain in which expert knowledge is poorly formalized.

4. The domain must provide a *straightforward mechanism for introducing a computer-based tool into the daily routine* of the physicians who use it. This point has several corollaries. First, use of the computer should ideally replace a task that is already being performed; this helps guarantee that the system will require a minimal additional time commitment. Second, the mechanical interface must be rapid, congenial, and easy to learn to use. And third, the decision tool's design must demonstrate a respect for the physician's hectic schedule.
5. The program should *maintain the physician's role as ultimate decision maker* (e.g., by giving explanations for recommendations and allowing the user to override any advice that is offered).
6. The system developers must be able to identify *highly motivated collaborators* from the domain of expertise.
7. The problem area should allow the initial prototype system to *avoid major theoretical barriers* (e.g., the domain should not require solutions to problems such as the development of approaches to the management of inexact inference, generalized methods for the management of temporal reasoning, encoding of strategic knowledge for domain-specific problem solving, or generation of highly customized explanations that demonstrate "first principle" understanding of the clinical area).

Criteria such as these have guided the development and progress of one of the newer AIM research activities. That project, known as ONCOCIN, is an expert system designed to aid physicians in the management of patients receiving cancer chemotherapy. The program is based on AI representation and control techniques similar to those described in this book (Shortliffe et al., 1981), but much of the effort has focused on getting the program implemented for use by physicians. Experience with the program and its users, both before and after its clinical introduction in May 1981, has recently been described (Bischoff et al., 1983). In order to provide a congenial high-speed interface, the system required a novel system architecture that separated the reasoning and interactive components (Gerring et al., 1982). ONCOCIN has also provided a productive environment for research on methods to ensure knowledge base completeness and consistency (Suwa et al., 1982) and on specialized explanation techniques (Langlotz and Shortliffe, 1983). Because of its initial promising success, plans have been made to convert the program to run on professional workstations and to use them as a vehicle for disseminating the technology to nonacademic settings (Tsuji and Shortliffe, 1983). More detailed discussions of ONCOCIN may be found in a recent book by Buchanan and Shortliffe (1984).

21.6 On Artificial Intelligence and Medical Computer Science

Those who work in the AIM field are uniformly enthused about the field's potential to do social good but are also aware of the common misinterpretation of their goals and of the frequent failure to acknowledge the fundamental research barriers that remain to be conquered. We have already discussed the problems that lie ahead, and we hope that the reader will share the cautious optimism that we feel about the future. Misinterpretations of the goals of AI research, however, at least partly relate to the phrase *artificial intelligence* itself. For example, the eminent essayist Lewis Thomas recently wrote in a "Notes of a Biology Watcher" column in the *New England Journal of Medicine* (Thomas, 1980):

> The most profoundly depressing of all ideas about the future of the human species is the concept of artificial intelligence. The ambition that human beings will ultimately cap their success as evolutionary overachievers by manufacturing computers of such complexity and ingenuity as to be smarter than they are, and that these devices will take over and run the place for human betterment or perhaps, later on, for machine betterment, strikes me as wrong in a deep sense, maybe even evil. Until now, computers have had the look of useful, often indispensable tools . . . [But] this is what the artificial intelligence people are talking about: a mechanical brain with the capacity to look back over the past and make accurate predictions about the future, then to lay out flawless plans for changing that future any way it feels like, and, most appalling of all, capable of feeling like doing one thing or another. Machines like this would be connected to each other in a network all around the earth, doing all the thinking, maybe even worrying nervously. But being right all the time. Leaving us with time for leisure . . .

We are not sure where Thomas obtained his information about the field, but we hope that this volume has demonstrated his misinterpretation of the nature of AI—both regarding the motives of the researchers and regarding expectations of what can and will be accomplished. One is reminded of a recent book by Weizenbaum that questioned not so much what *could* be accomplished by the AI field but what *should* be accomplished (Weizenbaum, 1976).

In response to Thomas's essay, Shortliffe and Buchanan sent a letter to the editor of *New England Journal of Medicine,* a portion of which was published with other letters on the subject (Shortliffe and Buchanan, 1980). We reproduce the entire original letter here:

> Lewis Thomas' polemic against artificial intelligence responds more to the emotional content of the phrase than to the realities of the techniques

and goals associated with this subfield of computer science. It is ironic that his opinion piece should appear at a time when computing techniques drawn from AI are being increasingly applied in clinical domains.

It is commonly accepted that computers can offer the medical professions significant relief from the complexities of routine information handling and data analysis (e.g., office billing systems, CT scanners). Because of the frequently cited explosion of medical knowledge, much research has also focused on computer-based tools to assist physicians with clinical decision making. Medical computing researchers are being drawn to AI largely because they see in the field techniques that will make programs for physicians more congenial, acceptable, and clinically useful. One of the goals of AI is to construct intelligent assistants that reason symbolically using empirical associations, accepted theory, and experts' judgmental knowledge. Although a textbook is a well-accepted tool, it is static and inflexible in the sense that it fails to customize its knowledge to the consideration of specific patients. By reasoning with general knowledge to suggest an individual approach to a patient's management, a program that can function as an intelligent assistant may further enhance the physician's effectiveness.

Thomas would have us believe that AI research purports to create a network of machines "doing all the thinking . . . leaving us with time for leisure." Yet in its medical applications, AI research is seeking ways to overcome the tendency to estrangement between man and machine, a frequent complaint that has tended to limit the utility of clinical computing. AI workers are attempting to provide us with computer-based tools that will make doctors more effective thinkers and clinical decision makers (Shortliffe, 1980). In his fervor for pursuing the philosophical correlates of a phrase like *artificial intelligence,* Thomas loses sight of the fact that "intelligent" knowledge-based machines may continue to serve as the "useful, often indispensable tools" which he admits he has come to appreciate.

The preceding interchange brings us naturally to a further definition of goals that will guide "the second decade" of AIM research that lies ahead. In addition to the research areas previously outlined, it is clear that two issues stand foremost on the medical computing agenda for the 1980s (Shortliffe, 1983): (1) there must be improved education of medical students and practicing physicians regarding computers and decision making, and (2) there must be an enhanced acceptance of *medical computer science* as an intrinsic component of the modern academic medical environment. The financial and academic support necessary for tackling difficult tasks such as those we have outlined will be made available only if there is improved recognition of the field's potential and of the fundamental research questions that exist for the medical computing community.