CHAPTER

Human-Intensive Techniques

10

Vimla L. Patel and Edward H. Shortliffe

10.1 Introduction

In the chapters of this section, we discuss the various ways that knowledge is derived to serve as a basis for clinical decision support. In this first chapter, we focus on human expertise as a source of knowledge. In subsequent chapters, we explore data-intensive methodologies, and techniques for synthesizing the collected knowledge of the medical literature. We then further explore how advances in genomics can offer new types of knowledge for clinical decision making and, in turn, how large databases and the data-intensive methodologies can be applied to personal decision making.

When we consider what makes human beings excellent at clinical decision making, we generally acknowledge that there are two key determinants: how much the experts know, and how well they apply what they know when devising solutions to problems that may arise. Thus, as we consider the creation of optimal decision support systems, we must similarly consider both the knowledge that they embody and the processes they adopt when applying that knowledge. A system can be “dumb” if the knowledge it needs is lacking or faulty, and it can demonstrate “poor judgment” if it reaches inappropriate conclusions despite a wealth of necessary factual knowledge. We recognize that it means little if we cram huge amounts of knowledge into a system but the program subsequently cannot use it wisely or appropriately.

In this chapter we focus on the acquisition of knowledge so that it can be encoded for use in decision support systems. Our emphasis is on what we can learn through interaction with human beings who are excellent at the same task for which the system is intended. As we have suggested, that means that we need to understand both the factual knowledge that is required to solve the relevant problems and the judgmental knowledge that characterizes a decision maker who gets to the heart of a problem effectively, discards irrelevant information, and demonstrates an ability to be creative rather than to solve problems by rote formula every time they arise.

As noted previously, Chapters 11 and 12 discuss analytical methods for identifying new or relevant knowledge from databases or the literature, such as data mining

Clinical Decision Support.
Copyright © 2014 Elsevier Inc. All rights reserved.

techniques and meta-analysis. Here, rather, we focus on the elicitation of knowledge by interacting with expert human beings – analyzing their behaviors, inferring their beliefs and knowledge, asking them to explain their thought processes and actions, shadowing them while they perform expert tasks, and applying formal or informal methods for extracting from those behaviors and explanations the factual and judgmental knowledge that they appear to be applying. Such interactions can be undertaken by human beings interacting with experts (often called *knowledge elicitation* by cognitive scientists, or *knowledge engineering* by computer scientists) or by computer programs that experts can use to convey what they know for capture in a computer-based representation (often called *interactive transfer of expertise*). This chapter focuses on the former processes, with brief discussion of the history of computer-based transfer of expertise from human being to computer in Section 10.4. Other chapters (Chapters 11–14) discuss current approaches to the inference of new knowledge by analyzing large data sets (data mining).

There are a number of reasons why we want to capture expert knowledge (Crandall et al., 2006). These include:

- **Knowledge preservation.** We want to capture “wisdom,” which develops with expertise. Such knowledge is typically experiential, too often undocumented, and we lose it once the expert retires or otherwise leaves the job.
- **Knowledge sharing.** Captured expert knowledge, meaningfully represented, can be reused in training programs, where trainees can be taught to develop expert strategies and functional efficiency. Such knowledge also can be shared among those who need to use it for a wide variety of decision-making tasks.
- **Knowledge to form the basis for decision aids.** New technology can be created, based on the expert knowledge, to help practitioners make better decisions. The technology, properly implemented, must embody the concepts, principles, and procedures of the work domain.
- **Knowledge that reveals underlying skills.** As the use of expert knowledge is explicated, it also reveals underlying strategies and skills, and how heuristics and intuition are applied in practice.
- **Understandability of human-derived knowledge.** As opposed to the “logic” underlying data-intensive or probabilistic techniques, the rationale for a judgment by a human expert can be explained in a form that makes sense and can thus be more readily accepted by a user.

Although the computer-based representation of knowledge is covered in Section IV of this volume, it is difficult to discuss the acquisition of knowledge without considering the representational issues that motivate and guide the acquisition process. Furthermore, the entire effort to capture and utilize knowledge in computer programs is predicated on the recognition that knowledge has a central role to play in providing tailored guidance through decision support systems. For example, cognitive psychologists have recognized the centrality of domain-specific knowledge in the skilled solving of complex problems (Patel and Groen, 1991a;

Patel and Groen, 1991b). Researchers in artificial intelligence have been noting for decades that "knowledge is power" and that general representations and search strategies, once a primary focus in that field, are limited in their ability to create intelligent behavior in machines (Feigenbaum et al., 1971). Knowledge-dependent computer applications, such as *expert systems* that use expert knowledge to perform complex problem solving and decision-making tasks (Duda and Shortliffe, 1983), are intended for use when the real experts are scarce, expensive, inconsistent, or simply unavailable on a routine basis. This characterization begs the questions "What is an expert?" and "How do we distinguish the knowledge and abilities of experts from those who are novices, or less expert, in a field?" Although we can easily agree that experts are those who have special skills or knowledge derived from extensive experience in their domain of expertise, their ability to achieve accurate and reliable performance also shows flexibility and adaptiveness in their environment that is difficult to explain by factual knowledge alone. We recognize that experts know "how" and "when", not just "what," and any attempt to capture knowledge for computer representation and use must recognize that these two general classes of knowledge are equally important.

Knowledge acquisition is a very general term that may be defined as the process of identifying and eliciting knowledge from existing sources – from domain experts, from documents, or inferred from large datasets – and subsequently encoding that knowledge so that it can be verified, validated, and utilized. This volume discusses the design and implementation of such knowledge-based systems and the evaluation of their performance. However, reproducible methods to acquire such knowledge, and to assure its accuracy, are typically discussed separately, even though they are intimately related to the design and construction of decision support programs. A knowledge base used in a clinical decision support system might contain knowledge structures that represent potential findings and diagnoses and the relationships among them (*conceptual* or *factual* knowledge), a knowledge structure representing guidelines or algorithms used to operate on this knowledge structure (*procedural* knowledge), and possibly also a knowledge structure with application logic used to apply these guidelines and algorithms to the underlying conceptual structure (*strategic* knowledge). All these types of knowledge must be combined to achieve a functioning decision support facility, and the elicitation of knowledge needed for expert performance must address each type of knowledge, not just "facts."

The techniques and theories that enable knowledge elicitation can be viewed within the context of the process illustrated in Figure 10.1. The process begins with methods to "extract" knowledge from human experts (knowledge acquisition [*KA*] or knowledge elicitation [*KE*]), followed by the representation of that knowledge (*KR*) in a computationally tractable form that supports knowledge-based agents or applications (Hoffman et al., 1995). Many people would then include the verification and validation of the output of those knowledge-based agents or applications as part of the complete process, since they provide feedback regarding the quality of the contents of the underlying knowledge structures.

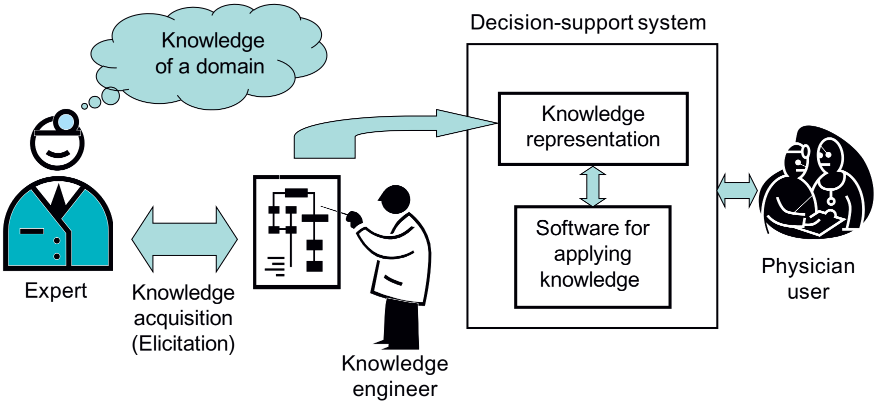

FIGURE 10.1

The classical view of knowledge engineering, in which an individual who knows the technical details of a system's representational conventions also has the skills of interviewing and observation necessary to work closely with an expert (or a group of experts) in order to obtain the needed knowledge and to convert it to a computationally useful form.

There is a variant of Figure 10.1 in which the knowledge is acquired not from a single expert collaborator but rather from a group of experts, perhaps through a consensus development process or by studying several experts and merging what one has learned into a single knowledge base. The field of cognitive science offers several methods for understanding the reasoning processes, mental models, and knowledge used by experts when they solve problems, as well as for dealing with team decision making and consensus development. We shall present some of those notions in the subsection that follows. There are also formal methods by which experts work together, supported by the literature and formal research studies, to reach consensus in formulating knowledge (e.g. the process of evidence-based guideline development (Peleg et al., 2006)). The acquisition and representation of consensus guidelines are further discussed in Chapter 16.

Finally, there has been substantial work to develop computer programs that acquire knowledge directly from experts (see Figure 10.2). Termed *knowledge acquisition systems* or *knowledge authoring systems* (see Section 10.4), these programs are intended to fill the role of knowledge engineer, providing human beings with a computational environment for assessing what knowledge is missing from a system and transferring their knowledge so that it can be encoded for that system's use. Such programs may be tightly coupled with the decision support system itself, allowing the system's decision-making abilities to be assessed and debugged as part of the knowledge acquisition/enhancement process. They always rely on access to the pre-existing knowledge in the system, as is indicated by the arrows going in both directions between the computer and the knowledge base in the figure.

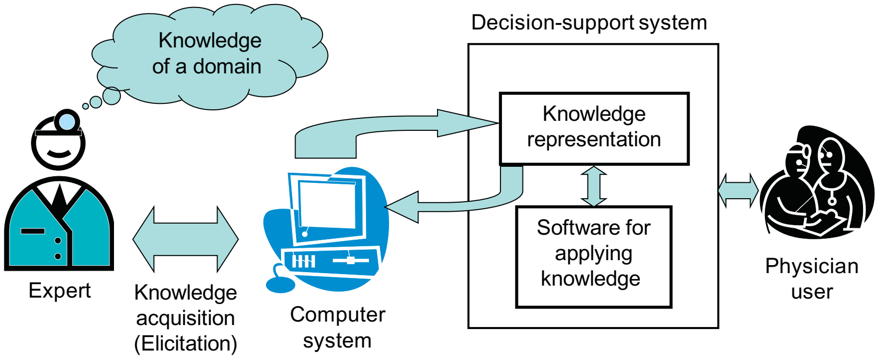

FIGURE 10.2

The interactive transfer of expertise using a computer program for knowledge acquisition. Note that such programs will generally both create new knowledge *and* use preexisting knowledge to guide the knowledge acquisition process. See also Section 10.4.

10.2 Theoretical basis for knowledge acquisition

We now focus on the frequently cited theoretical basis that underlies the numerous methods and techniques that exist to elicit domain knowledge from sources such as relevant experts. The currently accepted psychological basis for KA depends on defining and acknowledging the concept of expertise. Two major goals of expertise research have been to understand what distinguishes outstanding individuals in a domain from less outstanding individuals, and to characterize the development of expertise. This approach originated with the pioneering research of deGroot (deGroot, 1965) in the domain of chess, from which it extended to investigations of expertise in a range of content domains, including physics (Chi et al., 1981; Larkin et al., 1980), music (Sloboda, 1991), sports (Allard and Starkes, 1991), and medicine (Patel and Groen, 1991b). This research has shown that, on average, the achievement of expert levels of performance in any domain requires about ten years of full-time experience. An "expert" is someone who has achieved a high level of proficiency, as indicated by various measures, such as international "Elo" ratings in chess (named for the system's creator, Árpád Élő, a Hungarian-born American physics professor), world rankings in various athletic endeavors, and certification by a sanctioned licensing body, as in medical subspecialties.

10.2.1 The nature of expertise

In medicine, the expert–novice paradigm has contributed to our understanding of the nature of medical expertise and skilled clinical performance. Expert physicians have extensive general knowledge of medicine (acquired through medical school and residency training) and deep, detailed knowledge of their relatively narrow areas of specialization (acquired from both training and clinical experience). Every experienced physician has acquired common wisdom and medical knowledge as well as certain mastery in the application of medical skills; this constitutes generic expertise. Investigators have suggested the following classification of levels of expertise (Patel and Groen, 1991b):

- A *beginner* is a person who has only routine, lay knowledge of a domain; an example is a typical patient.
- A *novice* is someone who has begun to acquire the prerequisite knowledge assumed in the domain, such as a medical student; novices have a basic familiarity with the core concepts, the language, and to a lesser extent, the culture of medicine.
- An *intermediate* is above the beginner level but below the subexpert level and is typically a senior medical student or a junior resident.
- A *subexpert* (e.g. a specialist solving a clinical problem outside his or her domain of expertise) possesses generic knowledge and experience that exceeds that of an intermediate but lacks specialized knowledge of the medical subdomain in question.
- An *expert* (e.g., a cardiologist or an experienced intensive care nurse) has specialized knowledge of the subdomain in addition to broad generic knowledge.

The development of expertise has been shown to follow a somewhat counterintuitive trajectory. It is often assumed that the novice becomes an expert by a steady, gradual accumulation of knowledge and fine-tuning of skills. That is, as a person becomes more familiar with a domain, his or her level of performance (e.g. accuracy and quality) gradually increases. It turns out, however, that one generally can document a degradation in performance as a subject moves from novice to expert. This has been referred to as the *intermediate effect* (Patel and Groen, 1991a). It has been repeatedly demonstrated that superior expert performance is mediated by highly structured and richly interconnected domain-specific knowledge. Experts' knowledge is hierarchical and densely interconnected, which allows new pieces of information to become well integrated. Given that a novice's knowledge base is sparse and an expert's knowledge base is intricately interconnected, an intermediate may have many of the pieces of knowledge in place but lack the extensive connectedness of an expert, leading to the intermediate effect just mentioned. For example, expert cardiologists are routinely called upon to integrate clinical findings at various levels of aggregation, from biochemical abnormalities evidenced in blood tests to perturbations at the system level to clinical manifestations as expressed in the patient's complaints. After the performance degradation phase due to the intermediate effect, practitioners develop the missing connections among concepts in their knowledge base and, as they gain experience in the execution of a task, their performance becomes increasingly smooth, efficient, and automatic.

A great deal of experts' knowledge is finely tuned and highly automated, enabling them to execute a set of procedures in an efficient, yet highly adaptive manner, which is sensitive to shifting contexts. They can readily filter out irrelevant information. Novices, as opposed to intermediates, do not conduct irrelevant searches, simply because they lack knowledge rich enough to generate such searches. Studies demonstrate that expert performance is not a result of generally superior memory skills, but it is a function of a well-organized knowledge base adapted to recognizing familiar configurations of stimuli. The nature of experts' organized knowledge can also account for their superior perceptions of patterns. This is demonstrated compellingly in studies of expert radiologists, where they can be shown to look at the x-ray image at a glance, to develop an immediate impression, and then to search the image for findings that fail to fit or that otherwise modify the initial impression. For more details on the nature of expertise, refer to several of the key papers in the field (Chi et al., 1988; Ericsson, 1996; Ericsson and Smith, 1991; Feltovich et al., 1997).

One of the things that domain experts know about is the procedures they use in their practice. They learn many "heuristics" or rules of thumb (Chapman and Elstein, 2000). These compiled, top-level procedures can lead experts to skip steps when they describe the processes by which they carry out their task. Some such heuristics are shared with other experts, but others are ones they have created on their own (Patel et al., 1994). In addition, experts have meta-cognitive awareness of their own strategies and how they manage their resources (Glaser, 1996). Metacognition refers to the collection of cognitive process and functions that individuals use when thinking about their own cognition (about the way that they think).

Thus, when such experts work with knowledge engineers or KA programs, their goal is to transfer their existing knowledge to the computer so that it is able to replicate human expert performance in the task for which they have specialized expertise. Given the complexity of the types of knowledge and perceptual issues that characterize human expertise, it is challenging to capture such knowledge and to encode it for computer use so that expert performance by a decision support system can be achieved.

10.2.2 Role of mental models

One of the significant challenges of knowledge representation, especially from a cognitive perspective, is to devise mechanisms for capturing and representing the products of human clinical comprehension. For example, during a clinical diagnosis task, clinicians perceive, focus on, comprehend and create solutions using available patient information. The summary results of this process (e.g. a diagnosis or an assessment and plan) are often documented for future users (e.g. another physician during a later shift). However, the intermediate thought processes (i.e. how the diagnosis was reached, or what relationships among the available data were considered) are quite difficult to capture and use for the future. Such a representation would be useful, not only for characterizing the nature of knowledge that is used for clinical diagnosis, but also for developing intelligent applications that can support decision making. While much has been written about rule-based, probabilistic and knowledge-based solutions, there has been very little research on how to capture and distinguish among the corresponding mental models of clinicians, especially given the variability in clinicians' expertise in a given disease (e.g. the differing mental models of a disease between a specialist in the relevant discipline and a less experienced individual such as a house officer or even a primary-care physician; the patient's mental model of his or her own disease is of course even more rudimentary).

Such mental models are especially useful for documenting the basis for the diagnosis and for sharing information regarding patients and their care transitions. These models are designed to answer questions such as "how does this work?" or "what will happen if I take the following action?" or "why is the patient's blood glucose level so high given the medications he is currently receiving?" Running a model corresponds to a process of mental simulation for generating possible future states of a system from an observed or hypothetical state. An individual's mental models provide predictive and explanatory capabilities of the function of a given system (Patel and Kaufman, 2006).

10.2.3 Team-based decisions and shared knowledge

The development of clinical guidelines and other decision support tools involves multiple team players, such as attending physicians, consultants, clinical trainees, computer scientists, and psychologists with a range of expertise, unique vocabularies, and specific mental models. Shared mental models are an extension of the mental model concept and reflect the shared and collective knowledge of a team. They provide reciprocal expectations, which enable teams to coordinate and make predictions about the behavior and needs of their colleagues (Cannon-Bowers et al., 1993). Individual mental models can be studied through a wide range of experimental tasks that involve prediction and explanation. Shared or team mental models are best captured using naturalistic[1] or quasi-naturalistic methods that characterize communication and collective expertise. The study of teams necessitates a convergence of methods that focus on both individual and collective performance.

10.3 Cognitive task analysis

What, then, are the approaches that have allowed nonexperts to analyze, understand, and encode the ways that individual experts make decisions? The general approach that cognitive scientists use in analyzing the basis for human performance is known as *cognitive task analysis* (CTA). Its purpose is to capture the way the mind works – to capture *cognition*. CTA should describe the basis for skilled performance that is being studied. The methods in this field are varied, and a detailed exposition is beyond the scope of this book, but in using CTA, cognitive scientists try to capture what people are thinking about, what they are paying attention to, the

[1]Naturalistic methods are those that involve direct observation of individuals performing tasks in the real world rather than in controlled experimental situations.

strategies they are using in making decisions, what they are trying to accomplish, what information they discard, and what they know about the way a process works (Crandall et al., 2006). The three key aspects of CTA are 1) knowledge elicitation, 2) data analysis, and 3) "knowledge representation," where in the generic case the representation of knowledge conforms to formal criteria and methods that may not be inherently computational, even though they might provide insight when one is constructing a computer system's knowledge base in the same domain. Cognitive scientists will utilize one of a variety of knowledge representation schemes to describe and capture what they have learned and to compare the expertise and reasoning processes of individuals (for example, novices versus experts when presented with identical problems). In the following sections, we briefly describe each of these three key aspects of CTA.

10.3.1 Knowledge elicitation (KE) methods

Conducting KE studies is often complex and resource-intensive. As a result, it is important to select the appropriate KE methods and tools at the outset of such projects in order to ensure that the end product is amenable to the planned application domain. One of the key issues to consider when planning a KE study is the source of the knowledge to be elicited. The use of domain experts is probably the most common and simultaneously problematic source of knowledge (Scott et al., 1991). The use of domain experts presupposes the selection of individuals with sufficient domain knowledge, interest in participating in the KE process, and minimal bias – a combination of attributes not always easily attained.

Further complicating the use of domain experts is the frequent need to collect knowledge from multiple experts. Groups of experts are often needed to mitigate the problems associated with using single experts, as described later in this chapter (Liou, 1990), which may lead to knowledge elicitation with incomplete or potentially ineffective contents. However, though the use of multiple experts has the potential benefit of utilizing group synergies to generate consensus knowledge that is greater than the sum of the knowledge contributed individually (Boy, 1997), it is also not without its potential pitfalls, most notably the difficulties surrounding the merging of multiple experts' knowledge and the potential for the resulting knowledge to represent a single expert's opinion or input, rather than a true group consensus (Liou, 1990). Despite these potential concerns, the benefits of using multiple experts in the knowledge elicitation process generally outweigh the disadvantages.

Straightforward interview techniques are often used because they require a minimum level of resources, can be performed in a relatively short time frame, and can yield a significant amount of qualitative knowledge. The disadvantages of interview techniques include a frequent lack of quantitative data, which are needed for the input into the next step in the process. Furthermore, the results can often be biased due to the framing or presentation of questions or the selection of topics that are of interest only to researchers (Boy, 1997; Hawkins, 1983). But, perhaps most importantly, interviews simply lead to introspective opinions of the collaborating experts, and the knowledge elicited may *not* correspond to what they actually do when solving problems in the domain. For this reason, most knowledge engineers and psychologists who perform knowledge elicitation would prefer to observe the experts as they carry out tasks, either in simulated or "real world" environments. In order to gain insight into their mental processes, the experts may be asked to talk aloud about what they are doing and thinking *while they are performing the task*. In the world of cognitive science, such responses generated during problem solving are known as *think-aloud protocols* (Ericsson and Simon, 1993).

In contrast, ethnographic evaluations of expert performance are naturalistic observational studies conducted in context, with a minimum of knowledge engineer or psychologist involvement in the workflow or situation under consideration ("fly on the wall" observations). Such studies also implicitly evaluate the knowledge used by those experts and have been have been used in a variety of domains, ranging from air traffic control systems to complex healthcare delivery applications (Adria et al., 2003; Cohen et al., 2006; John et al., 1995; Laxmisan et al., 2005). One of the primary benefits of contemporary ethnographic research methods is that they are specifically tailored to minimize potential observational or researcher-induced biases (e.g. the Hawthorne effect[2]), while maximizing the role of collecting information in context, providing situation-specific knowledge. The resulting qualitative data generated by observational studies are often characterized as being "rich" or "concrete" (Rahat et al., 2005). The advantages of observational techniques are similar to interviews in that they require a minimum of resources, and further, provide for the capture of generally unbiased and contextual information. The disadvantages of observational techniques are again similar to interviews, in that they are time-intensive and do not easily yield large amounts of quantitative data. When quantitative data are generated from the observational studies, it is often a time- and resource-intensive task to code generated transcripts in order to extract data. Furthermore, in the absence of think-aloud protocols, it is left to the researchers to infer thought processes and knowledge structures from the behaviors that they have observed. However, one could debrief the subjects after the observations, using specific probes (questions) to get their interpretations to check for accuracy.

10.3.1.1 Group techniques

A number of group techniques for expert KE have been reported, including brainstorming (Osborn, 1953), nominal group studies (Delbecq et al., 1986; Jones and Hunter, 1999), presentation discovery (Payne and Starren, 2005), Delphi studies (Adelman, 1989), consensus decision making (McGraw and Seale, 1988), and computer-aided group sessions (Adams et al., 1999). All of these techniques focus on the elicitation of consensus-based knowledge. It has been argued that

[2]The alteration of behavior by the subjects of a study, resulting because they know that they are being observed. The term was coined in 1950 by Henry A. Landsberger (when analyzing older experiments from 1924–1932 at the Hawthorne Works (a Western Electric factory outside Chicago)). Ref: Henry A. Landsberger, *Hawthorne Revisited*, Ithaca, 1958.

such consensus-based knowledge is superior to the knowledge that can be gained from a single expert, since the group techniques used to generate such knowledge can reduce individual biases, increase the potential for the incorporation of multiple lines of reasoning, and account for potentially incomplete domain knowledge on the part of individuals (McGraw and Seale, 1988). Besides gaining consensus-based knowledge from a team of experts, such as expert physicians, there is also the potential to gain insight about shared interactions from teams who represent multiple areas of expertise (see the earlier discussion on "Mental Models"). However, conducting such group-technique KE studies can be difficult; it may be difficult to recruit appropriate experts to participate or to schedule mutually agreeable times and locations for such groups to meet. Furthermore, a forceful or coercive minority of experts or single experts might exert a disproportionate influence over the contents of the resulting knowledge collection (Liou, 1990).

10.3.1.2 Biases in logical and probabilistic reasoning

In clinical medicine, much of what experts report during knowledge elicitation is inherently uncertain. Although physicians, including experts in specific clinical subdomains, have been shown to be poor at the formal estimation of probabilities associated with relationships (Berwick et al., 1981; Leaper et al., 1972), they will frequently use terms that show that they are managing uncertainty in their approach to problems (e.g. "suggests," "supports," "goes against," "often," "evokes the possibility"). Despite the challenges, many knowledge engineers and psychologists have sought to obtain true probabilities from experts as part of their knowledge elicitation activities. In addition to poor estimation of probabilities by human beings, bias in their probabilistic reasoning has also been well documented (Kahneman and Tversky, 1982; Lichtenstein and Fischoff, 1980; Tversky and Kahneman, 1983), and types of bias have been categorized (Fraser et al., 1992). These bias types include tendencies (a) to allow undue influence of cognitive availability (recency) of information, mistaking this characteristic for frequency, (b) to anchor judgments on initial estimates, (c) to assess the likelihood of an event based on familiarity or stereotypic representativeness rather than objective frequency, and (d) to overestimate the frequency of rare events.

Following the demonstrations of Tversky and Kahneman, some researchers speculated that various biases might also be manifest in experts (Fischhoff, 1989), and they suggested that knowledge engineers should avoid the use of probabilistic or statistical judgments in knowledge elicitation altogether (Hink and Woods, 1987). The work on probabilistic reasoning bias became a red flag, because the notion of uncertainty is crucial in many expert systems (Fox, 1986; Kuipers et al., 1988; Zadeh and Kacprzuk, 1992). For example, in diagnostic problems one may need to formulate such rules as: "If the patient has spots, then the patient has measles with certainty X" (see, for example, the *certainty factor* uncertainty model used in the MYCIN expert system (Shortliffe and Buchanan, 1975) and subsequently applied in many other domains, both within and outside medicine). If experts provide biased probability estimates, there could be substantial problems for those building expert systems containing rules that are triggered when particular probability values are in effect for specific variables.

In many applications, statistical judgment and the sorts of judgments involved in decision analysis are contrived, in that they can take experts away from their usual way of thinking about problems. However, some investigators have argued that people have little trouble in giving probabilities, and that decision analysis can be used in knowledge elicitation (Fischhoff, 1989), where the focus is on improving judgment by making the decision processes and judgment criteria explicit. Some researchers have expressed doubt that the biases in probabilistic reasoning that have been observed in laboratory research occur with the same frequency and magnitude in any real-world problem-solving situations (Beyth-Marom and Arkes, 1983; Christensen-Szalanski and Beach, 1984).

Bias in logical reasoning also has been observed in the laboratory, where many problems have been observed (Evans, 1989; Fischhoff, 1989; Fraser et al., 1992; Johnson-Laird, 1983):

- A tendency to assign undue weight to the first evidence obtained
- Overreliance on variables that have taken on extreme values
- The tendency to seek evidence that confirms the current hypothesis
- The tendency to reason about only one or two hypotheses at a time
- The tendency to be overconfident
- The desire to maintain consistency with prior hypotheses even if that means devaluing, distorting, or ignoring important information
- Belief in illusory correlations
- The tendency to be overly conservative
- Basing conclusions on hindsight

In their studies of medical decision making, Schwartz and Griffin (1986) cited over 20 relevant papers supposedly demonstrating that experts rely on heuristics. In fact, they argued that experts do not seem to be prone to biases to such an extent that the concern should have practical import in knowledge elicitation work. However, most workers believe that such biasing tendencies are sufficiently common that they must be considered as confounders during the knowledge elicitation process.

10.3.2 Data analysis methods

10.3.2.1 Protocol and discourse analysis

The techniques of protocol and discourse analysis are very closely related, and concern themselves with the elicitation of knowledge from individuals while they are engaged in problem solving or reasoning tasks (i.e. *think-aloud* studies, as mentioned earlier). Such analyses may be performed in order to determine the conceptual entities and relationships between those entities used by individuals while they reason about a problem domain. The basic premises of these techniques are derived from the domains of psychology and cognitive science (Groen and Patel, 1988;

W. Kintsch and Greeno, 1985; Patel et al., 2001). In this approach, not only are a job's task activities charted, but also problem solvers are instructed to explain what they are doing and thinking while they are performing the task. The think-aloud procedure generates a response protocol, which is a recording of the deliberations that is subsequently transcribed and analyzed for propositional content and semantic content. The process of verbalization typically does not significantly affect the normal course of cognitive processes (Ericsson and Simon, 1993), and it can yield information about the reasoning sequences and goal structures in experts' problem solving (Patel and Groen, 1991b; Patel and Ramoni, 1997).

The think-aloud problem-solving/protocol-analysis technique has been used extensively in cognitive research on medical expertise (Johnson et al., 1981; Kuipers et al., 1988; Kuipers and Kassirer, 1984; Patel and Groen, 1986). For example, two early researchers (Kuipers and Kassirer, 1984) found that in a routine case, experts tended to produce very sparse protocols that did not provide much basis for characterizing reasoning patterns. The authors suggested that expert knowledge is so compiled that it is difficult to articulate intermediate steps. This led to using clinical probes to elicit constrained information within the think-aloud paradigm (Groen and Patel, 1988). Patel et al. (1994) showed that experts interpret clinical data from the first few segments of the patient problem evaluation process in terms of high-level hypotheses, which they later evaluate. This serves to partition the problem into manageable units, thus reducing the load on working memory. In contrast, experts out of their domain of expertise (subexperts) generate hypotheses mostly at lower levels, and they keep generating new hypotheses instead of evaluating and discarding some of them.

During such protocol analysis studies, the recorded explanations by subjects are codified for analysis at varying levels of granularity (Feltovich et al., 1989; Patel and Groen, 1991a; Patel and Groen 1991b; Polson et al., 1992). Discourse analysis is the process by which an individual's intended meaning within a body of text or some other form of narrative discourse is analyzed into discrete units of thought (propositions). These units are then analyzed according to the context in which those units appear (propositional relations in semantic structures) as well as the quantification and description of the relationships existing among those same units (Alvarez, 2002). The advantage of this approach to conceptual knowledge elicitation is that it situates the overall elicitation process within the broader distributed sociocognitive context in which individuals perform real-world reasoning and problem solving (Patel et al., 2001; Patel et al., 2002).

10.3.2.2 Concept analysis

In recent years, some CTA researchers have adopted a technique called concept mapping as a method of both eliciting and representing knowledge (Crandall et al., 2006; Novak, 1990). The modern idea of a concept map can be interpreted as a "user-friendly" expression of meaning in a text. Concept maps have been used in many studies of the psychology of expertise, and this work has shown that these maps can support the formation of consensus among experts (Gordon et al., 1993).

Concept maps constructed by domain experts clarify what they wish to express, and they eventually show high levels of agreement (Gordon, 1992). In concept mapping knowledge elicitation, the researchers help the domain practitioners to build a representation based on their domain knowledge, merging the activities of knowledge elicitation and representation. This technique has also proved to be useful as a tool for creating knowledge-based performance-support systems (Dorsey et al., 1999). Concept maps are labeled node-link structures, like the semantic networks described later in the chapter, but are less formal than the networks based on formal propositional representations.

10.3.2.3 Verification and validation of knowledge acquisition

As mentioned earlier, the process of verification and validation of knowledge is ideally and most effectively applied throughout the entire knowledge engineering spectrum. Therefore, it is important to understand the types of verification and validation metrics and techniques available for use within the specific context of KA. *Verification* is the evaluation of a knowledge-based system to ensure that it satisfies the end-user or domain-specific requirements used to define the design of that system (logical consistency, general notions of completeness, avoidance of redundancy, and the like). *Validation* is the evaluation of a knowledge-based system to ensure that it satisfies an external criterion of correctness, e.g. the end-user or domain-specific requirements that are intended to be realized upon implementation and refinement of that system. An example of a validation measurement for a knowledge-based system would be the concordance between the system's reasoning concerning a given set of "real world" input data in comparison to the reasoning that would be used by a domain expert assessing the same input data within the same real-world context. The MYCIN system (Shortliffe, 1976) pioneered these kinds of knowledge-base validation experiments (Yu et al., 1979a; 1979b).

To summarize the distinction, verification is the evaluation of whether a knowledge-based system meets the perceived requirements of the end users or application domain, and validation is the evaluation of whether that system meets the realized (e.g. real-world) requirements of the end users or application domain. However, in both instances, similar evaluation metrics may be used. A number of critical verification and validation criteria exist, such as multiple-source or expert agreement, degree of interrelatedness of the knowledge, and consistency of the generated knowledge.

10.3.2.4 Heuristic methods

The most commonly used approach to evaluating knowledge is the use of heuristic evaluation criteria (Neilsen, 1994). The advantage of this approach is the obvious simplicity of the evaluation method (e.g. knowledge engineers or experts may manually review the knowledge generated and determine if its contents are consistent with the heuristics actually used during expert performance of the related tasks). However, methods for doing this are limited in their tractability when applied to large knowledge sets, since they are difficult, if not impossible, to automate. Furthermore, they make comparison of knowledge "quality" across multiple sets infeasible, because of the qualitative nature of the evaluation results being generated.

10.3.3 Representational methods

Cognitive task analysis also speaks to the representation of interpreted data, rather than just the collection of primary data. CTA techniques generally provide abstract frameworks that assume particular types of knowledge structures as well as underlying reasoning processes.

In the representation of verbal data, investigators have made use of two kinds of representational formalisms: propositional representations and semantic networks. Intuitively, a proposition is an idea underlying the surface structure of a text. The notion's usefulness arises from the recognition that a given piece of discourse can have many related ideas embedded within it. A propositional representation provides a means for representing these ideas, and the relationships among them, in an explicit fashion. In addition, it provides a way of classifying and labeling these ideas. Systems of propositional analysis (Frederiksen, 1975; Kintsch, 1974) are essentially languages that provide a uniform notation and classification for propositional representations. In all these approaches, as in case grammars, a proposition is denoted as a relation (predicate) over a set of arguments (concepts). Sowa's system of conceptual graphs provides another example of a language of this type (Sowa, 1984). Although there are notational differences in the formalisms, the underlying assumption is that propositions correspond to the basic units of the representation of discourse and form manageable units of knowledge representation.

The primary challenge is to represent the structure of verbal or written data arising from observations and interviews as well as from think-aloud protocols. The first stage of analysis involves generating a propositional representation of the acquired text. This is then transformed into a semantic network representation. The network consists of propositions that describe attribute characteristics, which form the nodes of the network, and propositions that describe relational information, which form the links.

The primary relations of interest in these networks are binary dependency relations, specifically, *causal, conditional*, and Boolean connectives (*and*, *inclusive or*, and *exclusive or* relations). In addition, algebraic relations (e.g. *greater than*), identifying relations, and categorical relations (i.e. category membership, part-whole relations) can be expressed. One can also distinguish between the *source* of a process and the *result* of a process. Uncertainty in relations can be represented by modal qualifiers (e.g. *can*), and truth values can be indicated when they deviate from the default value (truth with certainty).

A semantic network is a directed graph formed by nodes and by labeled connecting paths. Nodes may represent either clinical findings or hypotheses, whereas the paths represent directed connections among such nodes. These networks also

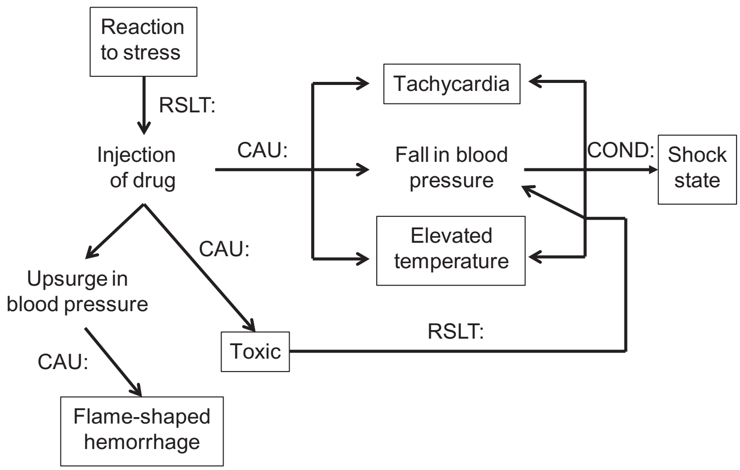

FIGURE 10.3

Semantic analysis of a clinical text. In the diagram, solid rectangles indicate cues from the text, broken lines indicate diagnostic hypotheses, and arrows indicate directionality of relations. **COND:** = *conditional* relation, **CAU:** = *causal* relation, **RSLT:** = *resultive* relation. In this case, the text is taken from an explanation protocol provided by a psychiatrist who had been challenged by a case from the field of cardiology: "*The patient has been reacting to stress, likely by his injecting a drug (or drugs), which has resulted in tachycardia, a fall in blood pressure, and elevated temperature. These findings are due to the toxic reaction caused by the injected drugs. He is in or near shock. The flame-shaped hemorrhage may represent a sequel of an upsurge in blood pressure possibly as a result of his injection of drugs.*"

provide a relatively precise means for characterizing the directionality of reasoning (Groen and Patel, 1988).

Figure 10.3 shows a semantic structure generated using discourse analysis to understand the implied and explicit knowledge contained in a specific text taken from a think-aloud protocol. The example is based on a diagnostic explanation offered by a psychiatrist when presented with a case from cardiology (Patel et al., 1990). The case is not within the subject's domain of specialization, and the diagnosis of a *shock state* is inaccurate. Because of this, the representation is lacking in coherence and contains one possible inconsistency; that is, a patient cannot have both high and low blood pressure at the same time. Furthermore, the underlying mechanism that explains the signs and symptoms in this patient is attributed to toxicity of drugs that the patient has injected in an effort to respond to external psychological stress. This is an inaccurate description of the patient's problem.

The diagram consists of nodes linked by arrows. The arrows have labels indicating the relationship between nodes. The two most important are **CAU:** and **COND:**. The arrows labeled **CAU:** represent causal relations, and those labeled

COND: represent conditional relations. **CAU:** means that the source node causes the target (e.g. *upsurge in blood pressure* causes *flame-shaped hemorrhage*), and **COND:** means that the source node is an indicator of the target (e.g. *tachycardia* indicates *shock state*). A difference between the two relates to the strength of implication: **COND:** expresses a directional conditionality, $P1 \rightarrow P2$, which implies if proposition $P1$ is true then $P2$ is true. **CAU:**$P1 \rightarrow P2$, is a stronger relation indicating that one variable, $P2$, is a functional result of another, $P1$.

10.4 History and current status of computer-based knowledge acquisition

The knowledge contained in any large-scale decision support system is so extensive and complex that it has become unreasonable to consider managing such knowledge bases manually. As a result, specialized environments have been constructed that allow trained individuals to enter new knowledge, and maintain or "curate" what is already there. Such systems often require structural knowledge of a domain over which the inferential knowledge is overlaid. Today, that structural knowledge, which defines the concepts in a domain and some aspects of the hierarchical relationships among them, is known as an *ontology* of that domain. Knowledge base developers and maintainers typically begin with the creation of a basic ontology for a field and then build inferential structures and relationships that allow a knowledge system to draw conclusions and generate advice. These knowledge representation issues are discussed in several chapters in Section IV.

We mention this topic here because there is a continuum in the development of computer systems for knowledge acquisition between those that are used for entering knowledge acquired through another means and those that actually interact with experts to extract, encode, and maintain that knowledge. Today, systems in the former category dominate, among which the well-known Protégé system is an important example[3]. Protégé supports the creation of ontologies and the encoding of related complex knowledge in a domain (Musen, 1992; Tudorache et al., 2013). But it would be rare to identify clinical experts who would be able to sit down with Protégé and "teach" it what they know about their domains of expertise. Protégé is for programmers and knowledge engineers to use after they have identified the knowledge that needs to be encoded.

The notion of obtaining knowledge directly from experts using an interactive dialog had its roots in the field of artificial intelligence in the early 1970s. For example, Carbonell pioneered the notion of computer-based mixed-initiative dialogs, focusing on educational uses but recognizing that an ability to interview and interact with a knowledgeable user had broad implications for computational extraction of knowledge, as well as its conveyance in the educational setting

[3] http://protege.stanford.edu/ (Accessed August 1, 2013).

(Carbonell, 1970). Heavily influenced by Carbonell's work, Shortliffe later experimented with a mixed-initiative dialog system that would allow physicians to teach the MYCIN program new knowledge (rules) in the domain of infectious disease therapy (Shortliffe, 1976). The basic notion was that an expert ought to be able to challenge MYCIN with a new patient case, use its explanation facilities to determine what faulty or missing knowledge explained any errors in performance (see Chapter 18 in Buchanan and Shortliffe (1984)), and then to enter new or corrected rules for MYCIN to incorporate into its knowledge base. Consider, for example, Figure 10.4, in which an expert attempts to enter a new rule using natural language methods. In this example, the program has incorrectly interpreted the expert's statement regarding the gastrointestinal (GI) track. The physician has therefore indicated that the program's understanding of the rule is not yet correct. The completion of the session, leading to correct interpretation of the rule, is shown in Figure 10.5.

MYCIN's early foray into knowledge acquisition was later extensively expanded and enhanced by Davis (1979). It was he who coined the phrase "interactive transfer of expertise" to describe the notion of an expert interacting with an intuitive, natural interface implemented in a computer program to "teach" the

****NR***(the expert physician indicates that he would like to enter a New Rule)*
The new rule will be called RULE200

IF: 1** **ORGANISM IS A GRAM NEGATIVE ROD**
and 2** **IT IS ANAEROBIC**
and 3** **IT ENTERED THE BLOOD FROM THE GI TRACK**
and 4** **<CR>***(the physician enters a carriage return to indicate completion)*
THEN: 1** **IT IS PROBABLY A BACTEROIDES**
On a scale of 1 to 10, how much certainty would you affix to this conclusion?
** **9**
And 2** **<CR>**

This is my understanding of your rule:

<u>RULE200</u>
IF: 1) The site of the culture is blood, and
2) The site of the culture is GI, and
3) The stain of the organism is gramneg, and
4) The morphology of the organism is rod, and
5) The aerobicity of the organism is anaerobic
THEN: There is strongly suggestive evidence (.9) that the identity of the organism is bacteroides

Okay? (YES or NO)
** **NO**

FIGURE 10.4

A mixed-initiative knowledge-acquisition dialog between MYCIN and an infectious disease expert. The physician's entries are indicated by bolded text in capital letters, with MYCIN's responses in mixed upper- and lower-case. Comments are italicized.

```
Change PREMISE? (YES or NO)
** YES
Want to see the current version? (YES or NO)
** NO
Change (C), Delete (D), or Add (A) a component:
** C
Which component? (enter number):
** 2
Enter the new component:
** PORTAL OF ENTRY IS THE GI TRACK
More changes to this PREMISE? (YES or NO)
** NO
Change rule's conclusion? (YES or NO)
** NO

This is my understanding of your rule:
RULE200… etc.
```

FIGURE 10.5

In this continuation of the session from Figure 10.4, the expert indicates what corrections need to be made and is able to verify that the revised rule is what was intended.

machine about his or her knowledge of a domain. Davis's system, which generalized the methods beyond the MYCIN domain, was known as Teiresias. The program's power came in part from the creation of a notion known as *rule models*, meta-level descriptions of the knowledge categories and relationships that characterized the domain in question and guided the interactive knowledge acquisition process. Depicting a full dialog with Teiresias would be too detailed for this volume, but several examples can be found in Chapter 9 of Buchanan and Shortliffe (1984) and in Davis's summary article, which shows how the approach could be used in a totally different domain (Davis, 1979). There was also a journal titled *Knowledge Acquisition* that published six volumes between 1989 and 1994.[4]

Although the performance of these early programs was promising, the complexity of their creation, maintenance, and use made it difficult to get experts to work with them directly. They much preferred to work with knowledge engineers and psychologists who used the knowledge elicitation techniques we have previously described. Thus, in the 1980s, there was a gradual move toward creating powerful knowledge authoring and editing tools that could be used by knowledge engineers *after* they had elicited the pertinent knowledge from human experts. Graphical user interfaces, unavailable in the 1970s when MYCIN and Teiresias were created, encouraged the adaptation of visual programming concepts for use in knowledge base construction and maintenance. One of the earliest efforts was Musen's creation of OPAL, a graphical authoring environment for entering and maintaining cancer

[4] http://www.sciencedirect.com/science/journal/10428143 (Accessed August 3, 2013)

chemotherapy research protocols that were used by ONCOCIN to guide oncologists in the treatment of cancer patients (Musen et al., 1988). OPAL later was generalized to be used for knowledge entry and editing in any domain, and this led to the creation of Protégé, which is today heavily used for ontology construction and maintenance (Gennari et al., 2003).

Today, although experiments continue, it is a rare knowledge-elicitation tool that is designed and successfully implemented for use directly by physicians or other clinical experts. We instead see continued emphasis on the specialized skills of individuals who know the computational systems but who also have the interpersonal skills, and ability to learn about what is often a new domain to them, in order to work closely with experts, and groups of individuals, in order to elicit the knowledge that is needed for medical decision support. In addition, there is a great deal of work that seeks to derive new knowledge from large datasets, especially in the modern era of "big data." Many of these approaches are discussed in the remaining chapters in Section III.

10.5 Conclusions

In the modern world, knowledge management has become a major focus of activity in diverse businesses, including health care. Because of the effort required to develop and validate such knowledge, there is growing recognition of the need to share knowledge components when they are developed and optimally to involve experts in providing, assessing, and maintaining the knowledge that is needed. Although we are creating large institutional, local, regional, and national databases, only some of the knowledge that we require to inform practice and policy can be derived solely by analyzing those data or the literature (see Chapters 11 and 12). Many areas of clinical endeavor still depend heavily on the kind of judgmental knowledge and experience that is difficult to acquire from anyone other than those who have the wisdom and efficiency that comes with experience and lifelong learning. Thus, despite the formal analytical methods that are appropriately being used to make sure that we learn as much as we can from our accumulated experience stored in pooled databases and in the literature, knowledge elicitation from experts, and groups of experts, will continue to be a crucial component of knowledge creation and management for clinical decision support. The early promise of computer-based transfer of expertise to knowledge systems has not been borne out, although significant research opportunities and potential continue to exist. The re-emergence of such systems may be facilitated by our increasing knowledge of human problem-solving methods and by enabling improvements in technology. For now, however, it is the direct interaction among experts, and between experts and knowledge engineers, that will serve a crucial role in assuring the development of high quality and accepted knowledge bases that in turn enable the development and effective use of decision support systems.

References

Adams, L., Toomey, L., Churchill, E., 1999. Distributed research teams: meeting asynchronously in virtual space. J. Comput. Mediated Commun. 4 (4).

Adelman, L., 1989. Management issues in knowledge elicitation. IEEE Trans. Syst. Man Cybern. 19, 448–461.

Adria, H.L., William, A.S., Stephen, D.K., 2003. GMS: Preserving Multiple Expert Voices in Scientific Knowledge Management. ACM Press, San Francisco, California.

Allard, F., Starkes, J.L., 1991. Motor skill experts in sports, dance and other domains. In: Ericsson, A., Smith, J. (Eds.), Toward a General Theory of Expertise: Prospects and Limits. Cambridge University Press, New York, pp. 126–152.

Alvarez, R., 2002. Discourse Analysis of Requirements and Knowledge Elicitation Interviews. IEEE Computer Society.

Berwick, D.M., Fineberg, H.V., Weinstein, M.C., 1981. When doctors meet numbers. Am. J. Med. 71, 991–996.

Beyth-Marom, R., Arkes, H.R., 1983. Being accurate but not necessarily Bayesian: Comments on Christensen-Szalanski and Beach. Organ. Behav. Hum. Decis. Processes 31, 255–257.

Boy, G.A., 1997. The group elicitation method for participatory design and usability testing. interactions 4 (2), 27–33.

Buchanan, B.G., Shortliffe, E.H. (Eds.), 1984. Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Addison-Wesley, Reading, MA.

Cannon-Bowers, J.A., Salas, E., Converse, S., 1993. Shared mental models in expert team decision making. In: Castellan Jr., J. (Ed.), Current Issues in Individual and Group Decision Making. Lawrence Erlbaum, Hillsdale, NJ.

Carbonell, J.R., 1970. AI in CAI: an articial intelligence approach to computer-assisted instruction. IEEE Trans. on Man-Machine Systems, MMS 11, 190–202.

Chapman, G.B., Elstein, A.S., 2000. Cognitive processes and biases in medical decision making. In: Chapman, G.B., Sonnenberg, F.A. (Eds.), Decision Making in Health Care: Theory, Psychology, and Applications. Cambridge University Press, New York, pp. 183–210.

Chi, M.T.H., Feltovich, P.J., Glaser, R., 1981. Categorization and representation of physics problems by experts and novices. Cogn. Sci. 5, 121–152.

Chi, M.T.H., Glaser, R., Farr, M.J., 1988. The Nature of Expertise. Lawrence Erlbaum, Hillsdale, NJ.

Christensen-Szalanski, J.J., Beach, L.R., 1984. The citation bias: fad and fashion in the judgment and decision making literature. Am. Psychologist 39, 75–78.

Cohen, T., Blatter, B., Almeida, C., Shortliffe, E., Patel, V., 2006. Distributed cognition in the Psychiatric Emergency Department: A cognitive blueprint of a collaboration in context. Artif. Intell. Med. 37 (2), 73–83.

Crandall, B., Klein, G., Hoffman, R.R., 2006. Working minds: a practitioners guide to cognitive task analysis. MIT Press, Cambridge, MA.

Davis, R., 1979. Interactive transfer of expertise: acquisition of new inference rules. Artif. Intell. 12, 121–158.

DeGroot, A.D., 1965. Thought and Choice in Chess. The Hague. Mouton, The Netherlands.

Delbecq, A.L., Van de Ven, A.H., Gustafon, D.H., 1986. Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes. Green Briar Press, Middleton, Wisconsin.

Dorsey, D.W., Campbell, G.E., Foster, L.L., Miles, D.E., 1999. Assessing knowledge structures: relations with experience and post-training performance. Hum. Perform. 12, 33–57.
Duda, R.O., Shortliffe, E.H., 1983. Expert systems research. Science 220, 261–268.
Ericsson, A., 1996. The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games. Lawrence Erlbaum Publishers, Hillsdale, NJ.
Ericsson, K.A., Simon, H.A., 1993. Protocol Analysis: Verbal Reports as Data. MIT Press, Cambridge, MA.
Ericsson, Y.A., Smith, J., 1991. Toward a General Theory of Expertise: Prospects and Limits. Cambridge University Press, New York.
Evans, J., 1989. Bias in Human Reasoning: Causes and Consequences. Erlbaum, Hillsdale, NJ.
Feigenbaum, E.A., Buchanan, B.G., Lederberg, J., 1971. On generality and problem solving: a case study using the DENDRAL program. In: Meltzer, B., Michie, D. (Eds.), Machine Intelligence. Edinburgh University Press, Edinburgh, pp. 165–190.
Feltovich, P.J., Spiro, R.J., Coulson, R.L., 1989. The nature of conceptual understanding in biomedicine: the deep structure of complex ideas and the development of misconceptions. In: Evans, D., Patel, V. (Eds.), Cognitive Science in Medicine MIT Press, Cambridge, MA.
Feltovich, P.J., Ford, K.M., Hoffman, R.R., 1997. Expertise in Context. AAAI/MIT Press, Menlo Park, CA.
Fischhoff, B., 1989. Eliciting knowledge for analytical representation. IEEE Trans. Syst., Man, Cybern. 19, 448–461.
Fox, J., 1986. Knowledge, decision making, and uncertainty. In: Gale, W.A. (Ed.), Artificial Intelligence and Statistics. Addison-Wesley, Reading, MA, pp. 57–76.
Fraser, J.M., Smith, P.J., Smith, J.W., 1992. A catalog of errors. Int. J. Man Mach. Stud. 37, 265–307.
Frederiksen, C.H., 1975. Representing logical and semantic structure of knowledge acquired from discourse. Cognit. Psychol. 7, 371–458.
Gennari, J.H., Musen, M.A., Fergerson, R.W., Grosso, W., Crubézy, M., Eriksson, H., et al., 2003. The evolution of Protégé: an environment for knowledge-based systems development. Int. J. Hum-Comp. Stud. 58 (1), 89–123.
Glaser, R., 1996. Changing the agency for learning: acquiring expert performance. In: Ericsson, K.A. (Ed.), The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games. Lawrence Erlbaum, Hillsdale, NJ.
Gordon, S.E., 1992. Implications of cognitive theory for knowledge acquisition. In: Hoffman, R.R. (Ed.), The Psychology of Expertise: Cognitive Research and Empirical AI. Springer Verlag, New York, pp. 99–120.
Gordon, S.E., Schmierer, K.A., Jill, R.T., 1993. Conceptual graph analysis: knowledge acquisition for instructional system design. J. Hum. Factors 35 (3), 459–481.
Groen, G.J., Patel, V.L., 1988. The relationship between comprehension and reasoning in medical expertise. In: Chi, M.T.H., Glaser, R., Farr, M.J. (Eds.), The Nature of Expertise. Lawrence Erlbaum, Hillsdale, NJ, pp. 287–310.
Hawkins, D., 1983. An analysis of expert thinking. Int. J. Man.-Mach. Stud. 18 (1), 1–47.
Hink, R.F., Woods, D.L., 1987. How humans process uncertain knowledge. AI Mag. 8, 41–53.
Hoffman, R.R., Shadbolt, N., Burton, A.M., Klein, G.A., 1995. Eliciting knowledge from experts: a methodological analysis. Organ. Behav. Hum. Decis. Processes 62, 129–158.
John, H., Val, K., Tom, R., Hans, A., 1995. The role of ethnography in interactive systems design. interactions 2 (2), 56–65.

Johnson, P.E., Duran, A.S., Hassebrock, F., Moller, J., Prietula, M., Feltovich, P., et al., 1981. Expertise and error in diagnostic reasoning. Cogn. Sci. 5, 235–283.

Johnson-Laird, P.N., 1983. Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness. Harvard University Press, Cambridge, MA.

Jones, J., Hunter, D., 1999. Using the Delphi and nominal group technique in health services research. In: Mays, N., Pope, C. (Eds.), Qualitative Research in Health Care. BMJ Books, London.

Kahneman, D., Tversky, A., 1982. On the study of statistical intuitions. Cognition 11, 123–141.

Kintsch, W., 1974. The Representation of Meaning in Memory. Lawrence Erlbaum, Hillsdale, NJ.

Kintsch, W., Greeno, J.G., 1985. Understanding and solving word arithmetic problems. Psychol. Rev. 92 (1), 109–129.

Kuipers, B., Moskowitz, A.J., Kassirer, J.P., 1988. Critical decisions under uncertainty. Cogn. Sci. 12, 177–210.

Kuipers, B.J., Kassirer, J.P., 1984. Causal reasoning in medicine: analysis of a protocol. Cogn. Sci. 8, 363–385.

Larkin, J.H., McDermott, J., Simon, H.A., Simon, D.P., 1980. Expert and novice performances in solving physics problems. Science 208, 1335–1342.

Laxmisan, A., Malhotra, S., Keselman, A., Johnson, T.R., Patel, V.L., 2005. Decisions about critical events in device-related scenarios as a function of expertise. J. Biomed. Inf. 38 (3), 200–212.

Leaper, D.J., Horrocks, J.C., Stanliand, J.R., deDombal, F.T., 1972. Computer-assisted diagnosis of abdominal pain using estimates provided by clinicians. Br. Med. J. 4, 350–354.

Lichtenstein, S., Fischoff, B., 1980. Training for calibration. Organ. Behav. Hum. Perform 26, 49–171.

Liou, Y.I., 1990. Knowledge Acquisition: Issues, Techniques, and Methodology. ACM Press, Orlando, Florida, United States.

McGraw, K.L., Seale, M.R., 1988. Knowledge elicitation with multiple experts: Considerations and techniques. Artif. Intell. Rev. 2, 31–44.

Musen, M.A., 1992. Dimensions of knowledge sharing and reuse. Comput. Biomed. Res. 25 (5), 435–467.

Musen, M.A., Combs, C.M., Walton, J.D., Shortliffe, E.H., Fagan, L.M., 1988. OPAL: toward the computer-aided design of oncology advice systems. In: Miller, P.L. (Ed.), Selected Topics in Medical Artificial Intelligence. Springer-Verlag, New York, pp. 167–180.

Neilsen, J., 1994. Usability Engineering. Academic Press Professional, Boston.

Novak, J.D., 1990. Concept maps and vee diagrams: two metacognitive tools for mathematics and science education. Instr. Sci. 19, 29–52.

Osborn, A., 1953. Applied Imagination: Principles and Procedures of Creative Thinking. Scribner, New York.

Patel, V.L., Groen, G.J., 1986. Knowledge-based solution strategies in medical reasoning. Cogn. Sci. 10, 91–116.

Patel, V.L., Groen, G.J., 1991a. Developmental accounts of the transition from medical student to doctor: some problems and suggestions. Med. Edu. 25 (6), 527–535.

Patel, V.L., Groen, G.J., 1991b. The general and specific nature of medical expertise: a critical look. In: Ericsson, A., Smith, J. (Eds.), Toward a General Theory of Expertise: Prospects and Limits. Cambridge University Press, New York, pp. 93–125.

Patel, V.L., Kaufman, D.R., 2006. Cognitive science and biomedical informatics. In: Shortliffe, E.H., Cimino, J.J. (Eds.), Biomedical Informatics: Computer Applications in Healthcare and Biomedicine, third ed. Springer-Verlag, New York.

Patel, V.L., Ramoni, M., 1997. Cognitive models of directional inference in expert medical reasoning. In: Ford, K., Feltovich, P., Hoffman, R. (Eds.), Human & Machine Cognition. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 67–99.

Patel, V.L., Groen, G.J., Arocha, J.F., 1990. Medical expertise as a function of task difficulty. Mem. Cogn. 18 (4), 394–406.

Patel, V.L., Arocha, J.F., Kaufman, D.R., 1994. Diagnostic reasoning and expertise. Psychology of Learn. and Motivation: Adv. Res. Theory 31, 137–252.

Patel, V.L., Arocha, J.F., Kaufman, D.R., 2001. A primer on aspects of cognition for medical informatics. J. Am. Med. Inform. Assoc. 8 (4), 324–343.

Patel, V.L., Kaufman, D.R., Arocha, J.F., 2002. Emerging paradigms of cognition in medical decision-making. J. Biomed. Inform. 35 (1), 52–75.

Payne, P.R., Starren, J.B., 2005. Quantifying visual similarity in clinical iconic graphics. J. Am. Med. Inform. Assoc. 12 (3), 338–345.

Peleg, M., Gutnik, L.A., Snow, V., Patel, V.L., 2006. Interpreting procedures from descriptive guidelines. J. Biomed. Inform. 39 (2), 184–195.

Polson, P.G., Lewis, C., Rieman, J., Wharton, C., 1992. Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. Int. J. Man-Mach. Stud. 36 (5), 741–773.

Rahat, I., Richard, G., Anne, J., 2005. A General Approach to Ethnographic Analysis for Systems Design. ACM Press, Coventry, United Kingdom.

Schwartz, S., Griffin, T., 1986. Medical Thinking: The Psychology of Medical Judgment and Decision Making. Springer Verlag, New York.

Scott, A.C., Clayton, J.E., Gibson, E.L., 1991. A Practical Guide to Knowledge Acquisition. Addison-Wesley, Reading, MA.

Shortliffe, E.H., 1976. Computer-Based Medical Consultations: MYCIN, Vol. 2. Elsevier/North Holland, New York.

Shortliffe, E.H., Buchanan, B.G., 1975. A model of inexact reasoning in medicine. Math. Biosci. 23, 351–379.

Sloboda, J., 1991. Musical expertise. In: Ericsson, K.A., Smith, J. (Eds.), Toward a General Theory of Expertise: Prospects and Limits. Cambridge University Press, New York, pp. 153–171.

Sowa, J.F., 1984. Conceptual Structures. Addison Wesley, Reading, MA.

Tudorache, T., Nyulas, C., Noy, N.F., Musen, M.A., 2013. WebProtégé: a collaborative ontology editor and knowledge acquisition tool for the Web. Semantic Web 4 (1), 89–99.

Tversky, A., Kahneman, D., 1983. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. Psychol. Rev. 4, 293–315.

Yu, V., Buchanan, B.G., Shortliffe, E.H., Wraith, S.M., Davis, R., Scott, A.C., et al., 1979a. Evaluating the performance of a computer-based consultant. Comput. Prog. Biomed. 9, 95–102.

Yu, V., Fagan, L.M., Wraith, S., Clancey, W.J., Scott, A.C., Hannigan, J.F., et al., 1979b. Antimicrobial selection by a computer: a blinded evaluation by infectious disease experts. JAMA 242, 1279–1282.

Zadeh, L.A., Kacprzuk, J., 1992. Fuzzy Logic for the Management of Uncertainty. Wiley, New York.