Domain Knowledge and Hypothesis Generation in Diagnostic Reasoning

GUY-MARIE JOSEPH, MA, VIMLA L. PATEL, PhD

The role of domain knowledge in the process of hypothesis generation during diagnostic reasoning was examined. Subjects were given a clinical case presented one segment at a time on a microcomputer. They were prompted to think aloud after presentation of each segment of the clinical case. A combination of discourse and protocol analysis techniques was used to investigate the problem solving process in two groups of experts working on an endocrine problem. The groups consisted of high-domain-knowledge subjects (HDK), endocrinologists, and low-domain-knowledge subjects (LDK), cardiologists. The results showed no significant differences between the groups in terms of selection of relevant and critical cues from the case. However, specific differences were found in the links or relations between the cues, with the HDK subjects using more relations to connect important information. The HDK subjects generated accurate diagnostic hypotheses early in the problem encounter and spent the rest of the time confirming the hypotheses by explaining the given cues. The LDK subjects also generated accurate diagnostic hypotheses but were unable to discriminate between and eliminate alternative hypotheses. A two-stage problem solving process and its relationship to domain specific knowledge are proposed. *Key words:* medical problem solving; hypothesis generation; psychological processes; expert reasoning. **(Med Decis Making 1990;10:31–46)**

Diagnostic reasoning in medicine is considered by many to be a classification task that involves the selection of the most likely hypothesis or hypotheses from a wide range of possibilities.[1] The generation of such hypotheses is dependent upon a set of observations—signs, symptoms, test results, and results of physical examination—that are related to the underlying disease processes. One of the major questions facing researchers in clinical reasoning is how expert physicians gather and use clinical information to generate accurate diagnoses. This interest has been motivated by both theoretical and practical concerns. Studies of expertise in medicine and other domains have led to the finding that the early generation of diagnostic hypotheses is an important strategy in solving problems.

The rapidity with which the initial hypotheses are generated constitutes a striking feature of the behavior of experienced clinicians. Often, with only the age, sex, and presenting complaint of the patient, the clinician unhesitatingly selects a single working hypotheses. Studies have shown that the earlier a good hypothesis set is created, the more predictive it is of the quality of the diagnosis.[2] Differences in the time course of diagnosis generation seem consistent with the speed differences found in the chess-playing abilities of masters versus beginners,[3] as well as the findings from studies comparing experts and novices solving physics problems.[4]

Elstein and colleagues[5] were among the first researchers to apply information processing theory to the study of diagnostic reasoning, proposing the "hypothetico-deductive" model as a model of expert reasoning. In their studies the experts solved the diagnostic problems through a process of hypothesis generation and verification. Hypotheses are consistently generated early in a workup when very little information is available to the physician. They are used to overcome limitations of memory capacity and serve to narrow the size of the problem space, since a hypothesis can account for many possible manifestations. This narrows the scope of possible inferences, and thus uncertainty is controlled. Since it would be impossible to conduct an efficient inquiry without some hypothetical goal that would tell the inquirer when to stop, hypotheses serve to transform an open medical problem (What is the patient's illness?) into a set of closed problems that are much easier to solve (Is the illness X or Y or Z?). If the illness is deemed to be X, then predictions can be made about the patient's data. This

Received January 4, 1988, from Cognitive Studies in Medicine: Centre for Medical Education, McGill University, Montreal, Quebec, Canada. Revision accepted for publication June 19, 1989. Presented in part at the 8th Annual Meeting of the Cognitive Science Society, Amherst, Massachusetts, August 1986, and at the Annual Meeting of the American Education Research Association, Washington, DC, April 1987. Supported in part by grants from the Canadian Medical Research Council (MH-0004) and Canadian Natural Sciences and Engineering Research Council (A2598) to Vimla Patel.

Address correspondence and reprint requests to Dr. Patel: Centre for Medical Education, McGill University, 3655 Drummond Street, Room 529, Montreal, Quebec, Canada H3G 1Y6.

Table 1 • A Case of Hashimoto's Thyroiditis with Myxedema Pre-coma

*Medical history**

S1. A 63-year old woman with a one-week history of increasing drowsiness and shortness of breath was brought to the emergency room by her daughter.
S2. The patient had not been well for over a year.
S3. She had complained of feeling tired all the time, had a loss of appetite, a 30-pound weight gain, and constipation.
S4. A month later she had been diagnosed as having "chronic laryngitis" and had been prescribed a potassium iodide mixture as an expectorant.

*Physical examination**

S5. Physical examination revealed that the patient was pale, drowsy, and obese, with marked periorbital edema.
S6. She had difficulty speaking, and when she did speak her voice was slow and hoarse.
S7. There were patches of vitiligo over both her legs.
S8. Her skin felt rough and scaly.
S9. Her body temperature was 36°C.
S10. Pulse was 60/min and regular.
S11. Blood pressure was 160/95 mm Hg.
S12. Examination of her neck revealed no jugular venous distention.
S13. The thyroid gland was enlarged to approximately twice the normal size.
S14. It felt firm and irregular.
S15. There was grade 1 galactorrhea.
S16. The apex beat could not be palpated.
S17. Chest examination showed decreased movements bilaterally and dullness to percussion.
S18. There was no splenomegaly.

*Results from neurology, x-ray, and laboratory tests**

S19. Neurologic testing revealed symmetrical and normal tendon reflexes, but with a delayed relaxation phase.
S20. Urinalysis was normal.
S21. Chest x-ray showed large pleural effusions bilaterally.
S22. ECG revealed sinus bradycardia, low-voltage complexes, and nonspecific T-wave flattening.
S23. Routine biochemistry (SMA = 16) showed Na = 125, K = 3.8, BUN = 8 mg/100 ml.
S24. Arterial blood gases: Po_2 = 50 mm Hg, Pco_2 = 60 mm Hg.
S25. The patient was admitted to the intensive care unit for further management.

*These headings were not presented to subjects during the on-line case presentation.

line of reasoning is termed predictive reasoning, where the hypothesis is used to predict the cues in the data[6] and further refine the problem space.

The underlying theory of the hypothetico-deductive model is that the early generation of hypotheses is the major strategy used by physicians to direct their search toward accurate hypotheses within the problem space. Hence, the physician concentrates only on the components that are likely to yield the accurate solution. If, as most studies seem to suggest, one of the most important components of diagnostic reasoning is hypothesis generation, then obvious questions to ask are what determines the hypotheses physicians generate and when are they generated?

The study reported in this paper attempts to assess the role of hypothesis generation in diagnostic reasoning. It focuses on how a subject's representations of the problem change over time and the domain knowledge necessary to construct and modify these representations. "Think-aloud" protocols provide rich, complex data that are approximately concurrent with the subject's reasoning[7,8] and thus provide information about the changes in the subject's representation of the problem. However, protocol analysis methods have been limited in their success in providing more than global information about subjects' processing.[6] One of the major concerns raised by the use of protocol analysis is the lack of formal, objective methods for analyzing complex verbal material.[8] Methods of propositional analysis[9-11] have been proposed to meet this concern and have most often been applied to stimulus texts and recall protocols. Although recall protocols provide quite detailed data about representation in memory and inferences based on domain knowledge, they are very limited in the amount of information they can provide about the problem solving process.

Recently, a number of researchers have addressed these problems by combining the two methods of protocol and discourse analysis techniques.[12] This strategy has proven especially useful for studying diagnostic reasoning in medicine.[13,14] As a further refinement, in the present study, subjects were prompted to think aloud at regular intervals. This on-line procedure, where one segment of information is presented at a time, allows finer control of the stimuli the subject is to respond to, and may provide more detailed information about the subject's changing representations of the problem as a function of access to specific portions of the input text.

In the present study we are primarily interested in showing the precise relationship between domain-specific knowledge, hypothesis generation, and problem representation of physicians while solving a problem.

Method

SUBJECTS. Nine senior physicians associated with the Faculty of Medicine at McGill University volunteered as subjects for the study. Given that the task was the diagnosis of an endocrine disorder, the high-domain-knowledge (HDK) group consisted of four endocrinologists and the low-domain-knowledge (LDK) group consisted of five cardiologists. The physicians were all board-certified, practicing physicians with five to ten years of experience (practice) in their fields.

CLINICAL PROBLEM. The stimulus was a text describing an endocrinology problem based on a real case and modified by an endocrinologist for the purposes of this study. The clinical information in the case was arranged in the typical order of the patient's medical history, findings from the physical examination, and x-ray and laboratory test results, as shown in table 1.

The problem describes the case of an elderly woman who was brought to the emergency room by her daughter in a state of drowsiness and confusion heading towards a coma. Prior to her admission into the emergency room, the patient had consulted a physician for speaking difficulty and throat irritation. The diagnosis had been chronic laryngitis, for which a potassium iodide mixture had been prescribed as an expectorant. This had precipitated an acute hypothroid crisis, leading to pre-coma.

The accurate diagnosis of the case is Hashimoto's hypothyroidism with myxedema pre-coma. The diagnosis can be divided into three components varying in specificity, from general to specific, hypothyroidism, myxedema pre-coma, and an autoimmune condition called Hashimoto's thyroiditis.

PROCEDURE. Each subject was tested individually in his or her office at the hospital. Before the presentation of the case, subjects were given a short practice session to familiarize themselves with the experimental apparatus and procedure.

The stimulus material was presented to subjects on a Macintosh microcomputer, one segment at a time. Subjects controlled the rate of presentation of each segment by pressing the mouse button, and then each segment was replaced by the next on the display. It was not possible to access information presented in prior segments at anytime during the presentation of the case. Subjects were instructed to verbalize their thoughts about the role and importance of the information in each segment in reaching the correct diagnosis. Subjects' verbalizations were recorded for subsequent transcription and analysis. After presentation of the entire case, subjects were asked to provide a summary of the case and then to offer their final diagnoses.

ANALYSES

Since the major aim of our study was to look at how experts process information for solving clinical problems, we performed an in-depth analysis of each protocol. The nature of our study is qualitative and descriptive, and the methods selected for data analysis are ones that are sensitive to subtle changes in representation. This analysis would also lead potentially to variables that can discriminate between various levels of expertise.

The methods of analysis used in this study include the use of techniques from both discourse and protocol analyses. Four levels of analysis were carried out on each subject's protocol: 1) segmentation of protocol; 2) selection and organization of clinical case information; 3) hypothesis generation; 4) problem representation.

A combination of techniques from protocol and discourse analyses are used in order to study the selection and organization of information for hypothesis generation and the relationship of problem representation to hypothesis generation. Protocol analysis allows one to study the problem solving moves in relation to transitions in knowledge states, especially where the variables are easily specified. Discourse analysis allows one to study detailed semantic description that captures complex relations in the protocols.

Subjects' transcribed protocols were divided into syntactic units or "segments" using the method described by Dillinger.[15] This division facilitates identification and further analysis of the parts of the text and of the protocols. The method is based on Winograd's system of clausal analysis.[16] An example of segmentation is given in appendix A.

Next, the protocols were coded for the selection and organization of case information. Specifically, the protocols were analyzed for cue selection and links generated from cues selected. Parts of subjects' verbalization were coded as *repetitions* if they reproduced the stimulus text verbatim and did not explicitly associate the cue(s) with any aspect(s) of the correct diagnostic components. Subjects' comments were coded as *interpretations* if the comments explicitly linked information in the text segment to any aspect(s) of the diagnostic components. Examples of coding are given in appendix B.

For a measure of organization of the incoming information, the interpretation statements were further coded by type of cue (critical, relevant, and irrelevant) selected and type of cue linked to the accurate diagnostic components. The method of determining the critical and relevant information in the clinical case was based on the work of Patel and colleagues,[9,17] and consisted of asking four expert endocrinologists to categorize the cues in the case (signs, symptoms and results from physical examination and laboratory tests) as critical, relevant, or irrelevant to the diagnosis. A critical cue is defined as one viewed by the expert consultants as *necessary* for successful diagnosis of the case. A relevant cue is defined as a cue that provides *potentially important* information for reaching an accurate diagnosis. An irrelevant cue is one that is not considered directly relevant to the accurate diagnosis. The inter-rater reliability between experts was 0.92 for scoring the critical cues and 0.85 for scoring the relevant cues.

In clinical reasoning, the term "diagnostic hypothesis" has been used to refer to any ideas, diagnoses, or guesses that label the phenomena observed, or to proposed explanations that will guide the investigation of the patient's problem.[18] These hypotheses can refer to a variety of concepts such as syndromes, specific disease entities, disorders, pathophysiologic processes, and anatomic or biochemical disturbances. Examples of boding for the number of hypotheses generated are given in appendix C. Hypotheses were coded only the first time they occurred. In example 2a, *cardiac* and *chest disorder* were coded as two hypotheses generated. The criteria for identifying hypotheses were not defined by a set of explicit rules but were based on the other methods used in literature.[5,18,19] Hypotheses were analyzed with the assistance of the consulting endocrinologist.

The reference model was developed with the aid of protocols and interviews with two expert endocrinologists (who did not participate as subjects), as well as information from standard textbooks in endocrinology. The experts were given the whole clinical case and then asked to explain the underlying pathophysiology. Specific questions were used to probe for details. This reference model is taken to represent the knowledge required for generating the accurate diagnosis (fig. 1).

The reference model itself is divided into three main sections, which represent the underlying pathophysiologic processes related to the three diagnostic components. The first refers to hypothyroidism and is found on the upper right side of figure 1. For example, hypothyroidism causes *decreased metabolic functioning*, which in turn causes the symptoms of *tiredness* and *loss of appetite*. The decrease in metabolic functioning also causes an inability to generate body heat, relating to the findings of blood pressure of 160/95 mm Hg and body temperature of 36°C. Hypothyroidism also causes fluid retention, which in turn causes the weight gain, paleness, obesity, and periorbital edema.

The second section represents the pathophysiologic processes related to myxedema, and is represented on the left side of the reference model. For example, the excessive secretion of vasopressin leads to the patient's drowsiness and to dilutional hyponatremia, which is indicated by the laboratory result "Na = 125." The increased production of prolactin leads to galactorrhea, pericardial effusion to the EKG results and the apex beat's not being palpated, and so on.

The third section of the reference model represents the specific nature of the hypothyroidism, Hashimoto's thyroiditis, and appears on the lower right. This relates to vitiligo and a firm, irregular, enlarged thyroid gland. This disorder is also more common among elderly patients.

The nodes in the figure represent cues (critical and relevant), the components of the diagnosis, and some of the pathophysiologic, information linking them. The links represent relations between the nodes; most of which are causal relations (CAU:) with the exception of an occasional association (ASSOC:), category relation (CAT:), or conditional dependency relation (COND:). Then nodes and links are arranged to reflect the chain of ramifications that tie the disease to its manifestations. The central determinants of the patient's condition (in the middle of the diagram) cause particular conditions that eventually lead to the observed signs, symptoms, or laboratory findings (at the periphery of the diagram). The reference model is used here to evaluate the representations of the problem that subjects generate after the presentation of each text segment.

Each subject's interpretation of the given segment is mapped onto the reference model by doing an overlay of the subject's representation on the reference model. In the figures that follow, cues and links selected by the subjects as new information are highlighted with thick lines, and the diagnostic hypotheses generated by the subjects are set in black circles. The pathophysiologic mechanisms used by the subjects are underlined and set in bold type.

This method of analysis is designed to capture changes in representation in terms of generation of new hypotheses, refinement of old ones, and generally, restructuring of information. It is also intended to reflect the continuities in the problem representation, as each new segment appears, including confirming prior hypotheses by linking them with new text information, as well as the elaboration of prior hypotheses through the addition of new nodes and links.

Results and Discussion

SELECTION AND ORGANIZATION OF CASE INFORMATION

Overall, there were no differences in the total numbers of textual cues selected for use by HDK and LDK subjects. When the cues were separated by critical, relevant, and irrelevant categories, for both groups of subjects, the percentages of total cues selected were higher for relevant and critical categories than for irrelevant categories. Significant differences were found between the HDK and LDK subjects with respect to the categories of cues selected, with HDK subjects focusing more on the critical and relevant information than LDK subjects. The results were statistically significant at $p = 0.05$. The patterns of cues selected by the two groups of subjects are presented in figure 2.

Overall, the mean percentage of links generated to relate cues selected to arrive at the diagnostic com-

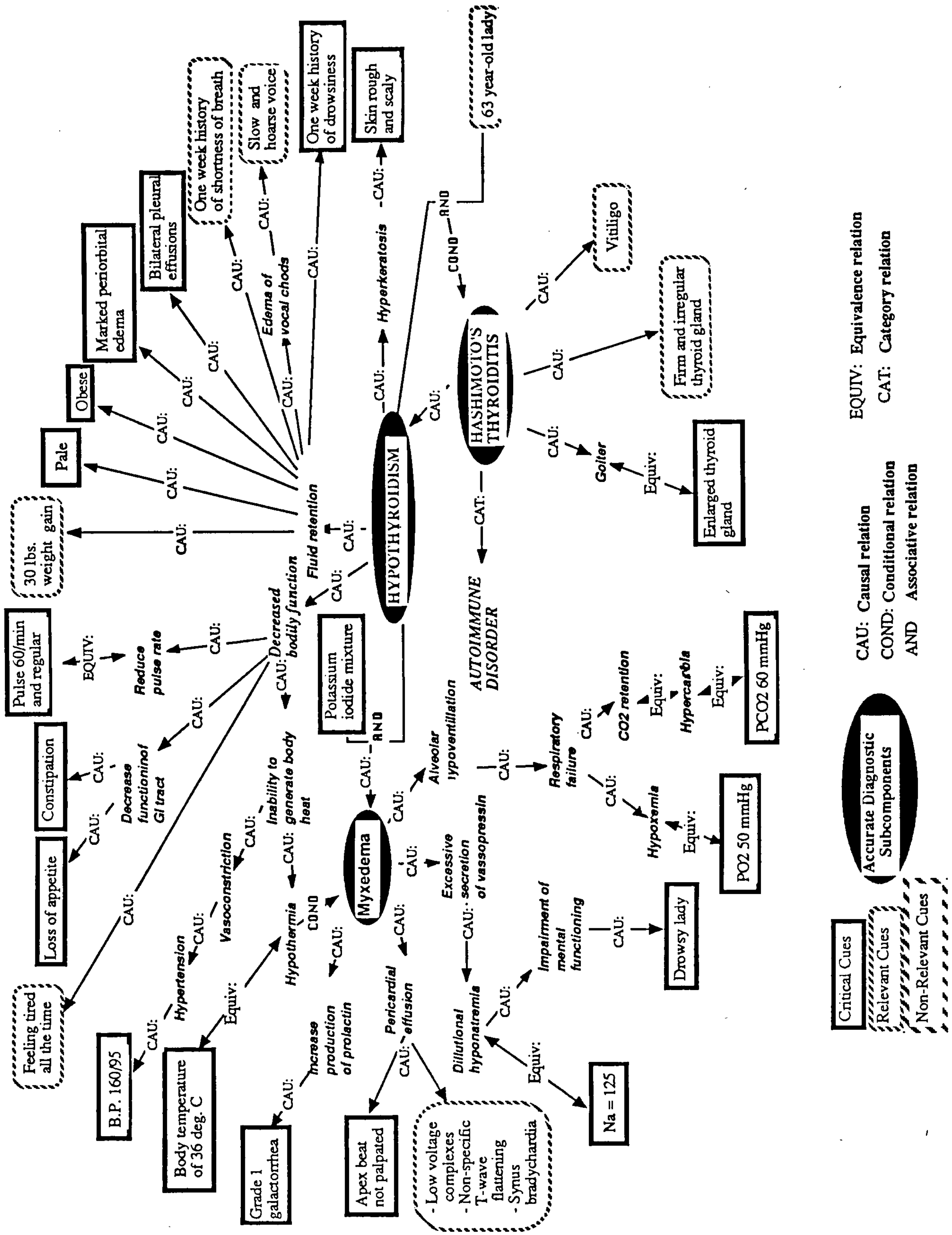

FIGURE 1. A reference model for the Hashimoto's thyroiditis with myxedema pre-coma.

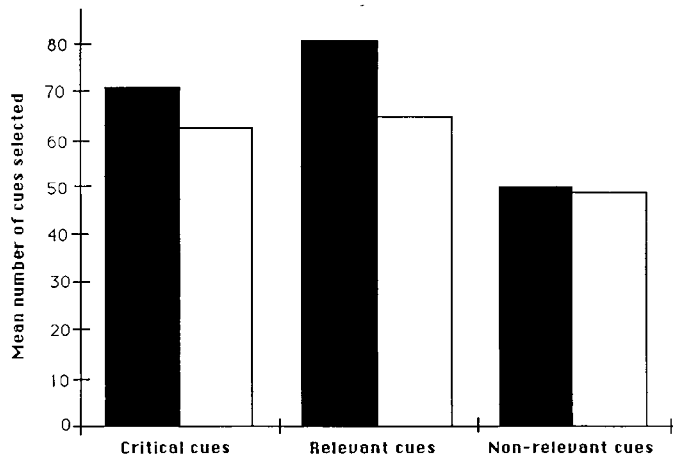

FIGURE 2. Mean numbers of cues of each type selected by high-domain-knowledge (HDK) (black bars) and low-domain-knowledge (LDK) (white bars) subjects.

ponents was higher for the HDK subjects than for the LDK subjects. This result was significant $p = 0.01$. The difference was especially attributable to the links generated between critical cues. However, it should be noted that the deviation from the mean was higher for the LDK subjects than for the HDK subjects.

The results (fig. 3) indicate that the HDK subjects selected the important input information and generated links to organize the information during the interpretation of the text segments. The LDK subjects were also able to select the important information, but were unable to organize this information during interpretation. The results show not only the differences in problem representation by the two groups of subjects, but also the nature of such differences, which reflects the HDK subjects' organization of the domain-specific knowledge.

THE TIME COURSE OF HYPOTHESIS GENERATION

Analysis of the time course production of diagnostic hypotheses focused on differences between HDK and LDK subjects in 1) the relationship between the number of hypotheses generated in the course of presentation of the case description, and 2) the time and the order of production of accurate diagnostic components.

Figure 4 gives the pattern of the cumulative sum of mean new hypotheses produced with each new segment of information for HDK and LDK subjects. This pattern of results was true for those HDK and LDK subjects who accurately identified the diagnostic components. Except for one LDK subject whose pattern of reasoning was like that of an HDK subject, there was no deviation from the mean for the generation of the diagnostic components. However, alternative hypotheses were being considered by the LDK subjects.

Those LDK subjects who produced accurate diagnostic components produced them later than did the HDK subjects. The HDK subjects produced the hypothyroidism component after segment 3 and the LDK subjects produced it after segment 4 (*A* in fig. 4). There was a difference of only one segment in diagnosing the general component of the diagnosis. However, the second, more specific, component of myxedema was diagnosed after segment 4 by HDK subjects and after segment 6 by LDK subjects (*B* in fig. 4). Finally, the identification of Hashimoto's thyroiditis occurred after segment 7 for the HDK subjects but only after segment 11 for the LDK subjects (*C* in fig. 4). The lag in the generation of accurate diagnostic components is shown to be a function of organization of relevant knowledge required for the diagnosis, since there were no differences between the two groups in their selections of relevant information.

The time course of hypothesis generation for each group is shown in figure 4. Each point on the figure represents the total-to-date (for the segment number indicated) of new hypotheses generated. The slope of the lines represents the pattern of hypothesis generation: the larger the slope, the greater the number of new hypotheses being generated; zero slope indicates no change, i.e., no generation of new hypotheses. There is a clear difference in the patterns of hypothesis generation for the two groups.

Before the presentation of segment 7, the two groups of subjects had generated approximately the same number of hypotheses, although the HDK subjects generated the first hypothesis earlier than did the LDK subjects. The HDK subjects used more of the infor-

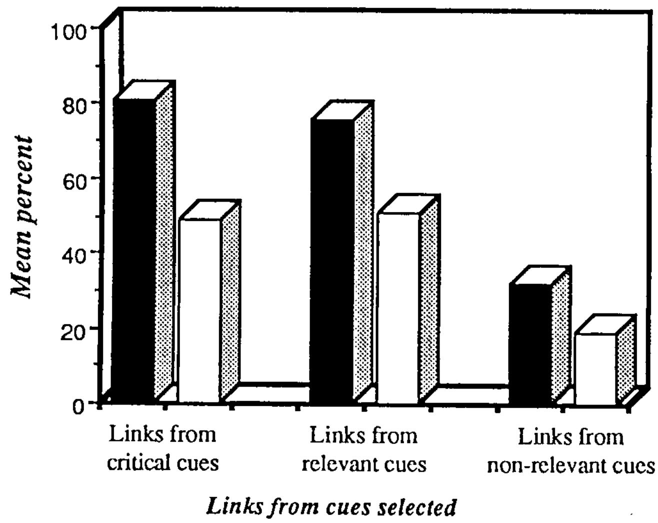

FIGURE 3. Mean percentages of links generated from cues of each type by high-domain-knowledge (HDK) (black bars) and low-domain-knowledge (LDK) (white bars) subjects.

mation contained in the medical history (1-4) in generating diagnostic hypotheses than did the LDK subjects.

For example, both groups reacted similarly to the cues in segment 2 ("The patient had not been well for over a year") and generated few hypotheses because this information is non-specific. However, the information presented in the next segment, 3 ("She complained of feeling tired all the time, had a loss of appetite, a 30-pound weight gain and constipation") led to generation of more hypotheses, since the occurrence of these symptoms together is strongly suggestive of hypothyroidism. The HDK subjects responded to these cues with many more hypotheses than did the LDK subjects since the symptoms are common to many other endocrine and metabolic disorders. This is indicated by the sharp difference in the numbers of hypotheses generated for this segment.

After the presentation of segment 4 ("A month later she had been diagnosed as having 'chronic laryngitis' and had been prescribed a potassium iodide mixture as an expectorant"), there was little quantitative difference between the groups. The information presented in this segment provided the case of the myxedema, and was the last segment of the patient's history.

The two groups responded differently to the presentation of segment 5, the first segment giving the result of the physical examination ("Physical examination revealed that the patient was pale, drowsy, and obese, with marked periorbital edema"). The HDK subjects generated very few hypotheses, whereas the LDK subjects generated the largest number to that point. The two groups responded similarly to the next two segments, generating few hypotheses in response to segment 6 and many more in response to segment 7 ("There were patches of vitiligo over both her legs"). This cue strongly suggests the component of Hashimoto's thyroiditis.

The analyses of hypothesis generation suggest that although the resulting diagnoses were very similar, the two groups produced them by different means. Unlike the LDK subjects, the HDK subjects distinguished two phases in producing the diagnosis. In the first stage (before presentation of segment 7), the HDK subjects used the information from the medical history and a few findings from the physical examination to generate the accurate diagnostic components.

After producing the accurate diagnosis, the HDK subjects generated few new diagnostic hypotheses, whereas the LDK subjects continued to generate many hypotheses. The HDK subjects sporadically ruled out some of the hypotheses generated earlier, and most often used the new findings from the physical examination to confirm the diagnosis and determine secondary problems. The LDK subjects, on the other

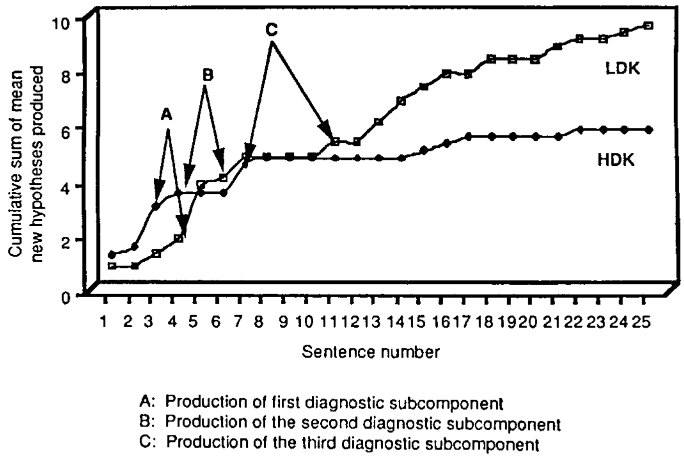

FIGURE 4. Hypothesis generation over the time course of information presented by high-domain-knowledge (HDK) and low-domain-knowledge (LDK) subjects.

hand, focused primarily on associating the findings from the physical examination and the laboratory tests results with new possibilities, as well as hypothyroidism and myxedema. They generated very few secondary problems, and did not rule out the diagnostic hypotheses that they had generated earlier, even when the new hypotheses were contradictory to some of the earlier ones. In general, the HDK subjects narrowed uncertainty while LDK subjects increased it.

DETAILED ANALYSIS OF TWO SAMPLE PROTOCOLS

The changing representations of the problem by one representative subject's protocol from each group is discussed. These protocols were selected such that they illustrate specific differences between the two groups.

The HDK and LDK subjects' representations of the case after the presentation of segment 1 are illustrated in figure 5. The HDK subject organized the information by grouping the findings together and identifying a disease that would explain the combined findings. In the process, two diagnostic hypotheses (cardiac or chest disorder) were generated as tentative explanations. The LDK subject, on the other hand, generated many disease hypotheses from individual cues. For example, the "cardiac and respiratory" problem was generated from "one-week history of shortness of breath"; medication overdose and metabolic or endocrine disorder were generated from "drowsiness." There was one overlapping hypothesis generated by both subjects, the hypothesis of cardiac disorder.

Neither of the subjects added much new information to the representation of the problem after the presentation of segment 2. The HDK subject retained the same two diagnostic hypotheses. The LDK subject, however, generated a new diagnostic hypothesis,

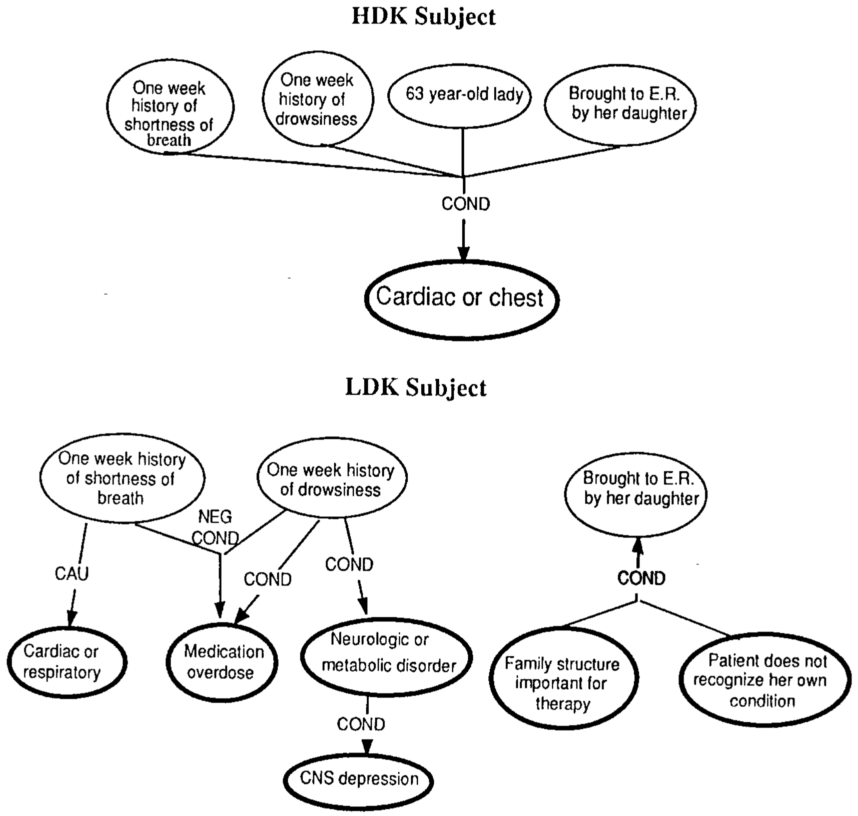

FIGURE 5. Semantic representations of protocols produced by the high-domain-knowledge (HDK) and the low-domain-knowledge (LDK) subjects in response to sentence 1 of the endocrine problem.

"chronic disorder." This hypothesis was generated from the activation of a text cue that was irrelevant for the accurate diagnosis.

The HDK subject's representation of the case after presentation of segment 3 is illustrated in figure 6. The findings presented in the segment led to the generation of the general diagnostic component (hypothroidism). After that, the subject elaborated on the patient's general condition, generating causal links from pathophysiologic processes to other text cues (e.g., respiratory distress causing the history of shortness of breath and hypothroidism causing fluid retention, which in turn caused the patient's weight gain). In this case a hypothesis is generated from textual cues and used to explain the other given cues not used to generate the hypothesis. The hypothesis of hypothroidism is confirmed.

In the LDK subject's representation (fig. 7), the evaluation of individual findings led to the generation of more diagnostic hypotheses (psychiatric problem and depression), together with a general description of the first diagnostic component (endocrine disorder or one of decreased glandular function). None of the diagnoses was specific enough to explain all the findings.

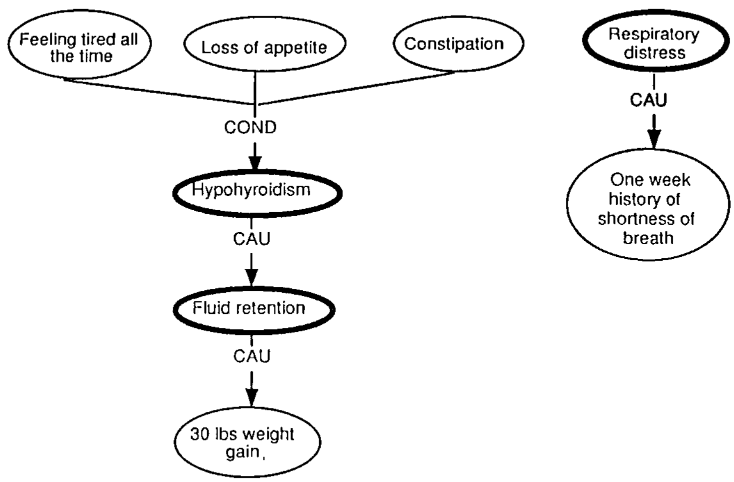

FIGURE 6. Semantic representation of protocol produced by the high-domain-knowledge (HDK) subject in response to sentence 3 of the endocrine problem.

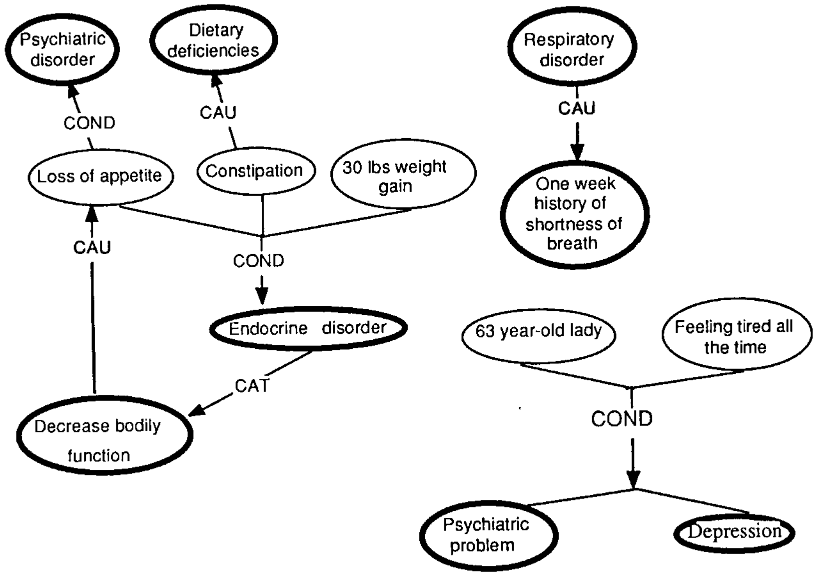

FIGURE 7. Semantic representation of protocol produced by the low-domain-knowledge (LDK) subjects in response to sentence 3 of the endocrine problem.

The subjects' representations of the case after the presentation of segment 4 are illustrated in figure 8. The HDK subject used causal explanation as in the reference model to generate the second diagnostic component (myxedema) and to elaborate on related problems (renal failure) and other pathophysiologic processes (fluid retention and iodine-induced goiter). This subject used a minimal number of cues to generate the second diagnostic component. The LDK subject, on the other hand, used more of the cues presented in the segment to generate a general statement of the same component ("advanced hypothyroidism"). The links used to generate the accurate diagnostic component were mostly conditional relations rather than causal, suggesting more uncertainty.

The difference between causal and conditional links can be interpreted according to their strengths of implication. Causal links imply cause-and-effect rela-

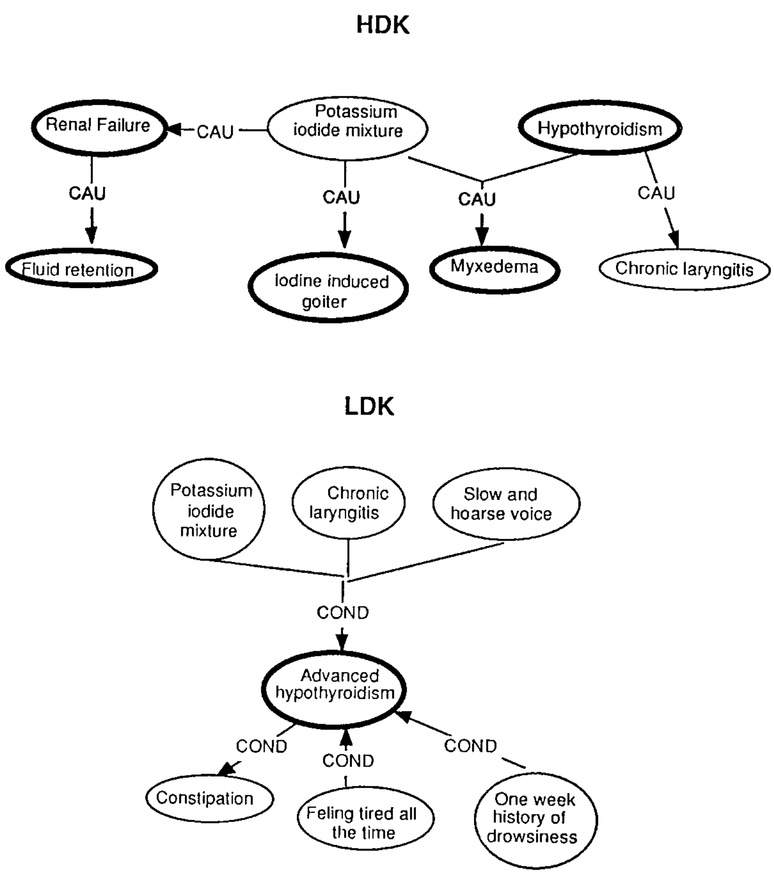

FIGURE 8. Semantic representation of protocols produced by the high-domain-knowledge (HDK) and low-domain-knowledge (LDK) subjects in response to sentence 4 of the endocrine problem.

tionship. The statement "potassium iodide mixture given to a patient, who is already hypothyroid, will cause myxedema" is an example of a causal relation.

Conditional links (if–then implication) between two events imply occurrence of one when the other is presented. These links are not as precise and explicit as the causal links. For example, the statement "I suspect the patient has advanced hypothroidism, given the chronic laryngitis, slow and hoarse voice, and potassium iodide mixture that was prescribed" is an example of a conditional rule.

While both subjects generated the second accurate diagnostic component in response to this segment, they seem to have arrived at them in different ways.

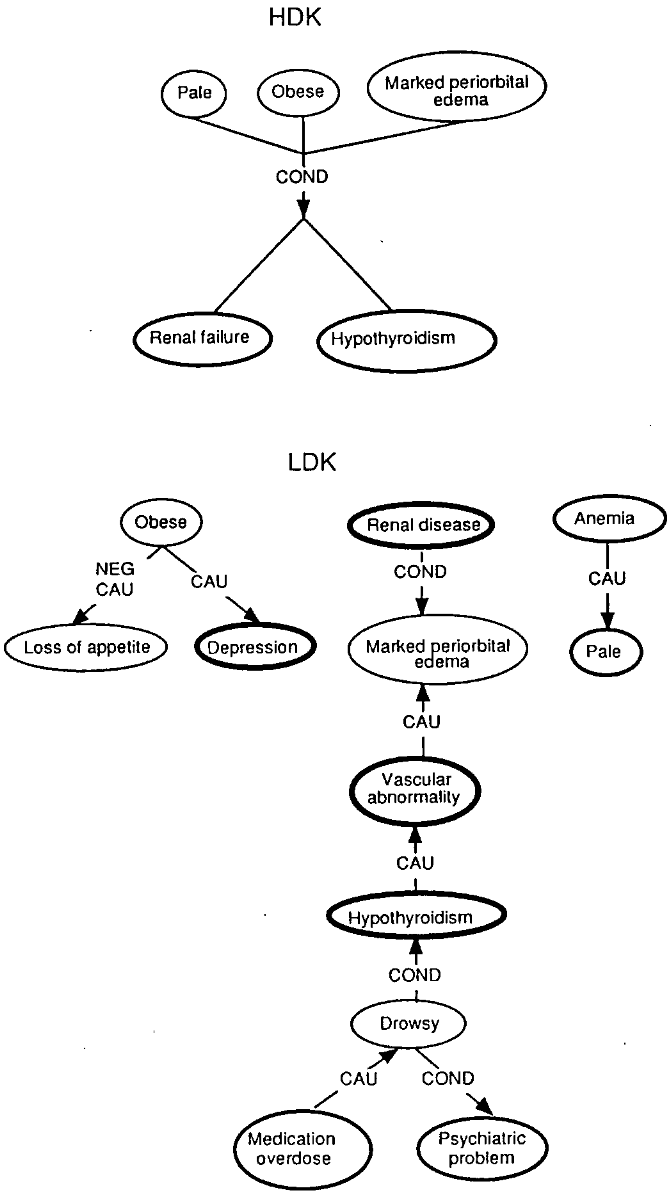

FIGURE 9. Semantic representation of protocols produced by the high-domain-knowledge (HDK) and low-domain-knowledge (LDK) subjects in response to sentence 5 of the endocrine problem.

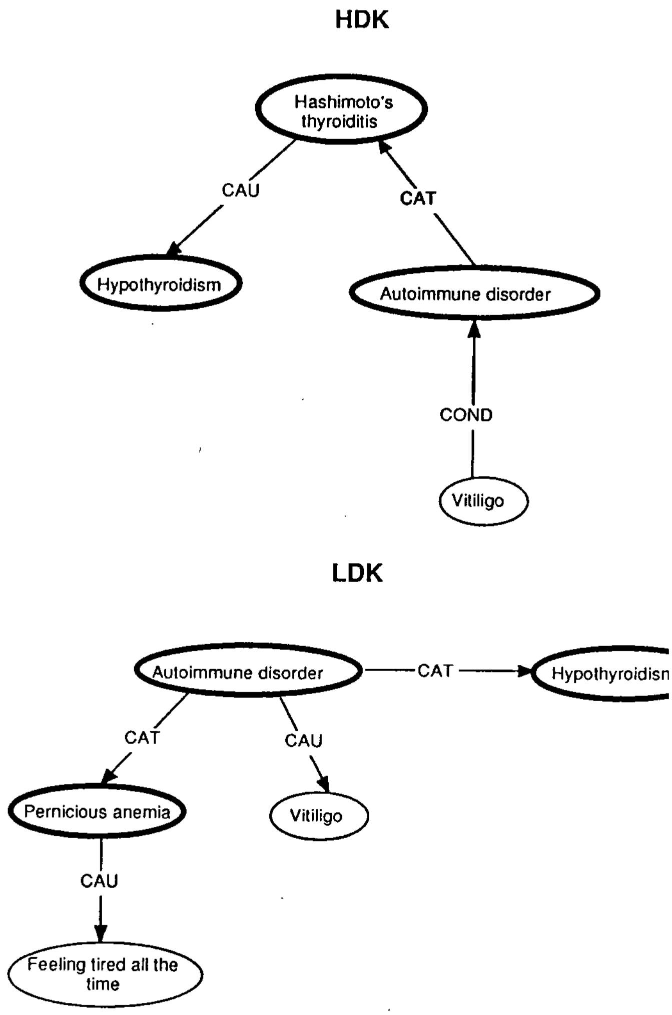

FIGURE 10. Semantic representation of protocols produced by the high-domain-knowledge (HDK) and low-domain-knowledge (LDK) subjects in response to sentence 7 of the endocrine problem.

After the presentation of segment 5, the HDK subject added very little new information to his representation of the problem (fig. 9). Instead, he interpreted the cues as confirming both the renal failure and hypothyroidism, which had been generated previously. The LDK subject generated a general description of the second diagnostic component (i.e., advanced hypothroidism), as well as several other hypotheses. Most of the new hypotheses involving causal links were unrelated to the accurate diagnosis (e.g., anemia, depression, and medication overdose). The subject appeared not to have the detailed knowledge of the disorder to explain the cues generated. This resulted in the generation of new hypotheses (unlikely) in relation to the diagnostic hypotheses. The most striking difference between the representations generated after this segment was in terms of coherence: the HDK subject interpreted the cues as confirming previously generated hypotheses, whereas the LDK subject used them to generate new hypotheses, most of which were unrelated to the accurate diagnosis.

Neither of the subjects generated any new information in their representations of the problem after being presented with segment 6. Both the HDK and the LDK subjects' representations of the case after segment 7 are illustrated in figure 10. The HDK subject generated the third component (Hashimoto's thyroiditis) of the accurate diagnosis from the finding "vitiligo" presented in this segment. The LDK subject recognized that some autoimmune process was involved. However, the subject identified the wrong category of autoimmune process (pernicious anemia) and attempted to justify it by linking it to the previous finding of the patient's tiredness. There were very few other cues that could confirm this diagnosis. He generated the Hashimoto's thyroiditis component of diagnosis several segments later.

The representations generated by the two subjects in response to this segment suggest that they were using very different organizations of information, leading to different specific diagnoses. The HDK subject used the domain-specific knowledge to produce the third (more specific) diagnostic component: He generated the accurate category of autoimmune disorder (Hashimoto's thyroditis) by direct association with one critical cue. The LDK subject, in the absence of well-developed domain knowledge, was unable to generate the specific component and instead focused on another hypothesis that explained the clinical cue of "feeling tired." The problem representations generated by the HDK and LDK subjects and the discussion above suggest that while two of the diagnostic components were generated by both groups of subjects, there were qualitative differences in the hypotheses and representations the subjects generated.

Whereas the diagnoses of the HDK subject were more precise, the LDK subject provided general descriptions of the more specific disease components. For example, the HDK subject generated the second diagnostic component (myxedema) by associating the hypothesis that the patient was hypothyroid with the fact that she had been given a potassium iodide mixture (fig. 11). On the other hand, the LDK subject generated a general description of the same component by associating the prescription of potassium iodide with the finding of slow and hoarse voice (fig. 12). The HDK subject's problem representation was more focused than that of the LDK subject; consequently, it maps onto the reference model more directly than does the LDK subject's representation. Only the first hypothesis that the HDK subject generated was inaccurate; those of the LDK subject were inaccurate throughout the first seven segments.

This pattern of results was true for all nine subjects studied. The individual variations in the lines of reasoning were due to different diagnostic hypotheses considered after exposure to different segments of the problem. However, the nature of the representations generated by all four HDK experts was different from that of those generated by all five LDK experts, in that the former were coherent, with causally related links, whereas the latter consisted of associative and conditional relations.

General Discussion

The focus of the present study was on diagnostic problem solving by experts with high and low levels of knowledge of the domain. Specifically, the study focussed on the processes by which the cues in the case description are selected and organized into a coherent situation in order to solve the problem. No differences were found between the HDK and LDK subjects' abilities to select relevant information from irrelevant data. The differences were attributed to the selective organization of this information for use in solving the problem. This finding is similar to the results reported elsewhere.[14]

The results of the present study describe the precise nature of the knowledge organization, in that the HDK subjects organized information in a coherent form with strong causal relations between various selective data, whereas the LDK subjects linked information with a greater use of weak conditional and associative relations. This pattern of results is unlike that of novices, who are unable to select the relevant information from the case.[9]

However, the present study goes further to explicate the nature of the problem solving processes used by subjects with high and low levels of knowledge of the domain. Given the partial information, the HDK subjects selected the relevant cues to generate a hypothesis. This was done very early in the clinical encounter, as found by other researchers. Furthermore, patient information was used to evaluate the selected hypothesis in order either to confirm the hypothesis or to generate an alternative one. Since the HDK subjects usually generated accurate initial hypotheses, the rest of the time was spent in confirming and establishing the specific nature of the given hypothesis, (refining the initial hypothesis) rather than in generating alternatives.

Similarly, the LDK subjects also selected relevant cues to generate a hypothesis. However, not having the detailed characterization of the disease in question, the evaluation of new information often led to the generation of alternative diagnostic hypotheses rather than confirming the existing one. The process repeated itself until a cluster of cues was obtained that confirmed the original hypothesis. This process resulted in the generation of a number of alternative

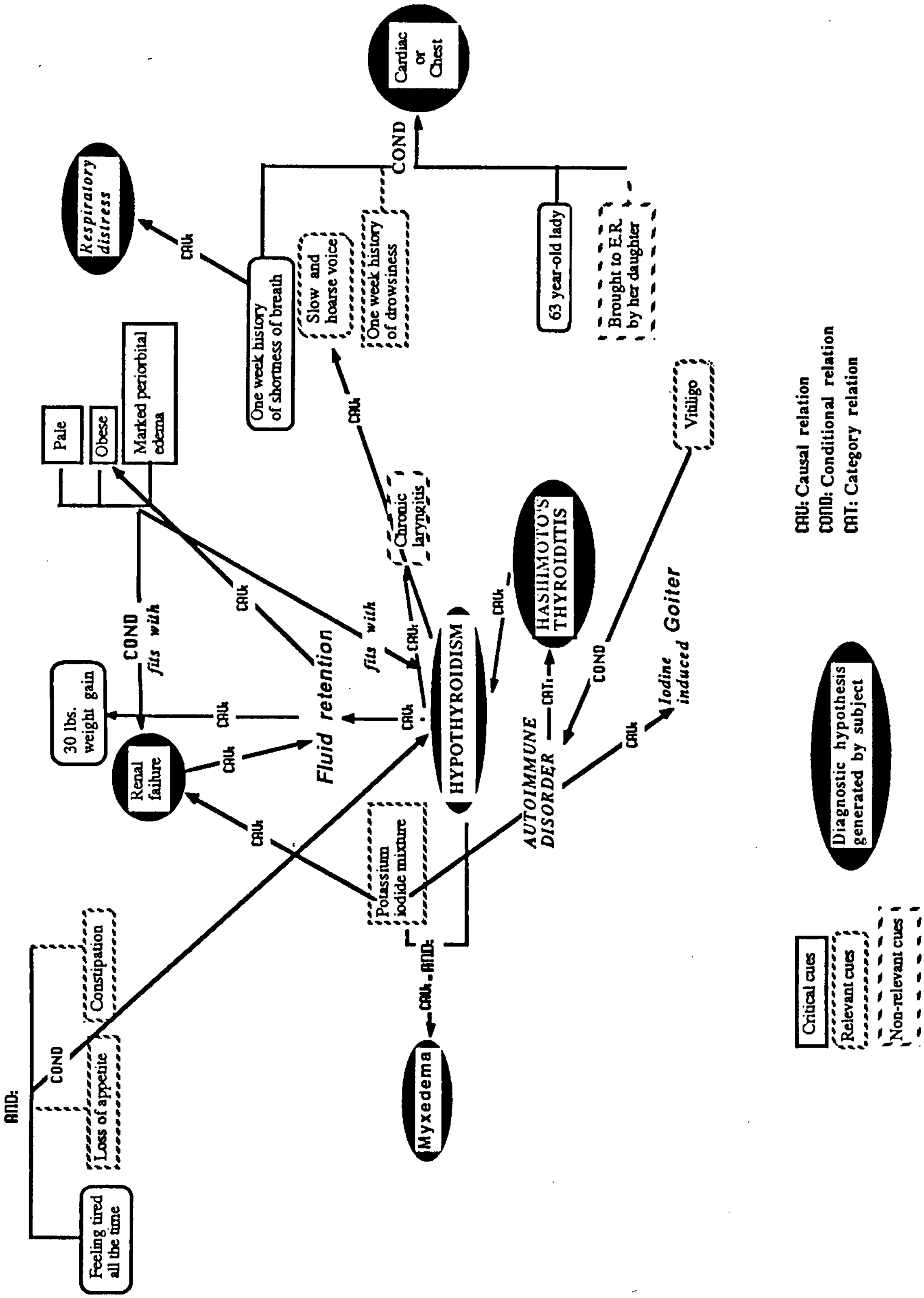

FIGURE 11. Part of the problem representation generated by the high-domain-knowledge (HDK) subject in response to the first seven sentences (sentences 1 through 7).

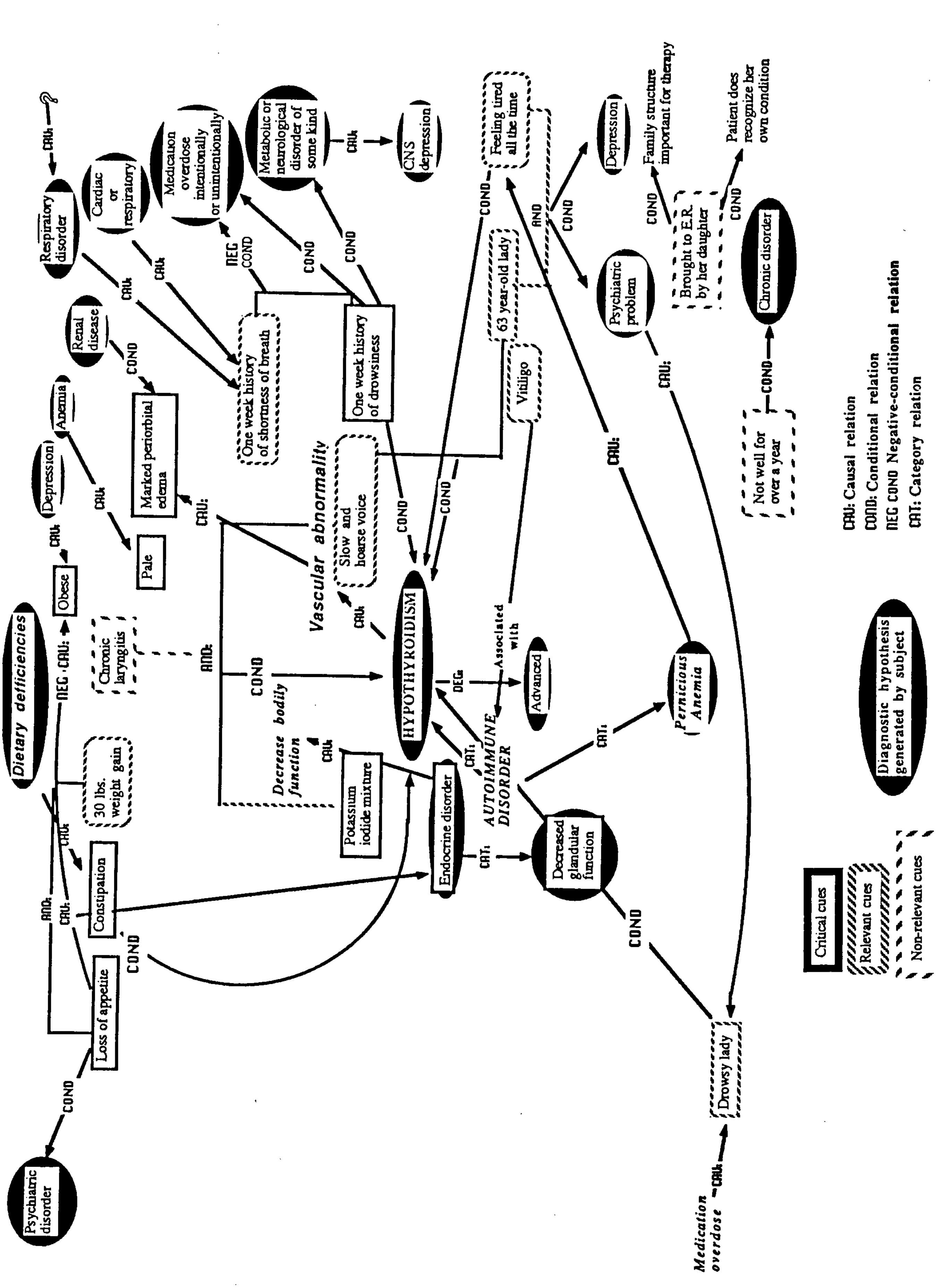

FIGURE 12. Part of the problem representation generated by the low-domain-knowledge (LDK) subject in response to the first seven sentences (sentences 1 through 7).

hypotheses for which specific knowledge was not available for discrimination.

Two phases are identified in solving clinical problems. The first phase is equivalent to the three components of cue acquisition, hypothesis generation, and cue interpretation referred to the literature.[15] However, these three components are considered to be dependent on domain-specific knowledge. In our study, the fact that the time courses of hypothesis generation (times to accurate identification of two of the diagnostic subcomponents) were somewhat similar for the two groups of subjects in the organizational phase suggests a limited role of domain-specific knowledge (at least at this level of expertise)—instead, the time course may be determined by general problem solving skill.

The second phase involves the process of evaluating the problem representation(s) produced. This is equivalent to the fourth component of hypothesis evaluation reported in the literature.[15] It appears that, given the general knowledge of the domain, an accurate diagnostic hypothesis can be generated. But the certainty with which such a hypothesis is accepted as "true" depends on very specific knowledge of the domain. Presumably, problem representation would have to be evaluated with respect to accounting for the cues presented, internal coherence, and relative certainty with which one hypothesis can be proposed over another.

Evidence for a two-step framework comes from the clear change in hypothesis-generating activity after segment 7 (for the HDK subjects), and from the fact that the LDK subjects continued to generate new hypotheses even after they had produced all of the components of the correct diagnosis. This suggests a more important role of domain-specific knowledge in evaluating hypotheses than in generating them.

The next phase of this study is to take the specific variables identified in this research that differentiate problem solving behaviors of subjects at different levels of expertise and test to see whether these variables are meaningful within the context of a larger group of subjects. These data would then also be amenable to statistical analysis.

One limitation of this study was that the physicians did not have the opportunity to request patient information but rather were given information in passive form (a text). An obvious extension of this research will involve the use of an interactive on-line task environment. Such an environment is more flexible and will allow subjects to explore different hypotheses freely, ask questions at any time, and backtrack to review information presented earlier. The protocols and time data thus produced, when analyzed with more refined versions of the methods used here, promise to yield much more detailed information about the intricate processes of medical decision-making.

The authors thank Drs. Y. C. Patel and D. St. Vil for their assistance with the medical aspects of the study. They acknowledge the support of the physicians at McGill University who generously volunteered their time as subjects. Special thanks go to M. Dillinger, A. Breuleux, Robert Bracewell, L. Coughlin, and D. Kaufman for their helpful discussions at various phases of the study, and Susan Young-St. Pierre for preparing the manuscript. Finally, the authors acknowledge the valuable comments and suggestions from Michael Ravitch and the reviewers who read the earlier versions of this paper.

References

1. Shortliffe EH. Reasoning methods in medical consultation systems: artificial intelligence approaches. Comp Progr. Biomed. 1984;18:5–14.
2. Barrows HS, Freghtner JW, Neufeld VR, et al. Analysis of the clinical methods of medical students and physicians. Final report, Ontario Department of Health. Grants ODH-PR-273 and ODH-DM-226. McMaster University, Hamilton, Ontario, 1978.
3. Chase WG, Simon HA. Perception in chess. Cog Psychol. 1973;1:55–81.
4. Larkin J, McDermott J, Simon DP. Expert and novice performance in solving physics problems. Science. 1980;208:1335–42.
5. Elstein AS, Shulman LS, Sprafka SA. Medical problem solving. An analysis of clinical reasoning. Cambridge, MA: Harvard University Press, 1978.
6. Patel VL, Evans DA, Chawla AS. Domain specificity and knowledge utilization in diagnostic explanation. In: Proceedings of the Ninth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Lawrence Erlbaum, 1987;195–202.
7. Ericsson KA, Simon HA. Protocol analysis: verbal reports as data. Cambridge, MA: MIT Press, 1984.
8. Greeno JG, Simon HA. Problem solving and reasoning. In: Atkins RC, Hernstein RJ, Lindzey G, Luce RD, eds. Stevens' handbook of experimental psychology (revised edition). New York: John Wiley and Sons, 1985.
9. Patel VL, Groen GJ, Frederiksen CH. Differences between students and physicians in memory for clinical cases. Med Educ. 1986;20:3–9.
10. Frederiksen CH. Cognitive models and discourse analysis. In: Cooper CR, Greenbaum S, eds. Written communication annual. Volume 1: Linguistic approaches to the study of written discourse. Beverly Hills, CA: Sage,1986.
11. Kintsch W. The representation of meaning in memory. Hillsdale, NJ: Lawrence Erlbaum. 1974.
12. Kintsch W, Greeno JG. Understanding and solving word arithmetic problems. Psychol Rev. 1985;92:109–29.
13. Patel VL, Groen GJ. Knowledge based solution strategies in medical reasoning. Cog Sci. 1986;10:91–116.
14. Patel VL, Groen GJ, Arocha JF. Medical expertise as a function of task difficulty J Memory Cognition. (in press).
15. Dillinger M. Segmentation and clause analysis. Technical report #3, McGill University Discourse Processing Laboratory, Montreal, Quebec, 1987.
16. Winograd T. Understanding natural language. New York: Academic Press, 1972.
17. Coughlin LD, Patel VL. Text comprehension and the effect of expertise in the domain of medicine. J Med Educ. 1987;62:818–28.
18. Barrows HS, Tamblyn RM. Problem-based learning: an application to medical education. New York: Springer-Verlag, 1980.
19. Johnson PE, Duran AS, Hassebrock R, et al. Expertise and error in diagnostic reasoning. Cog Sci. 1981;5:235–83.