Memory & Cognition
1990, 18 (4), 394-406

Medical expertise as a function of task difficulty

VIMLA L. PATEL, GUY J. GROEN, and JOSÉ F. AROCHA
McGill University, Montreal, Quebec, Canada

This paper is concerned with factors that disrupt the pattern of forward reasoning characteristic of experts with accurate performance. Two experiments are described. In the first, the performances of cardiologists, psychiatrists, and surgeons in diagnostic explanation of a clinical problem in cardiology were examined. In the second, the performances of cardiologists and endocrinologists in diagnostic explanation of clinical problems within and outside their domains of expertise were examined. The performances of researchers and practicing physicians are also compared. The results of Experiment 1 replicated earlier results regarding the relationship between forward reasoning and accurate diagnosis. There were no differences in recall as a function of expertise. Experts did not show any bias toward using specific knowledge from their own areas of expertise. The results of Experiment 2 showed that the breakdown of forward reasoning was related to the structure of the task. In particular, nonsalient cues induced some backward reasoning even in subjects with accurate diagnoses. Some differences were also found between the types of explanation used by researchers and practitioners. The practitioners referred more to clinical components in their explanations, whereas the researchers focused more on the biomedical components.

An important distinction made in rule-based models of problem solving involves the directionality of inferences. *Forward reasoning* (or *forward chaining*) involves the generation of a hypothesis on the basis of data given in the problem. *Backward reasoning* involves the generation of data on the basis of a hypothesis. The distinction between forward and backward reasoning originated not in psychology but in artificial intelligence, where early expert systems tended to use one of these two kinds of inference mechanisms exclusively. However, this distinction motivated two empirical studies (Larkin, McDermott, Simon, & Simon, 1980; Simon & Simon, 1978) that were interpreted as showing that directionality is directly related to expertise. At least on routine problems, experts used forward reasoning, whereas novices used backward reasoning.

It is important to note that this interpretation was based more on theory than on data. The data, which were based on only 3 subjects, simply demonstrated the existence of the two methods of reasoning and suggested a correlation with level of expertise. The interpretation of this correlation was derived from the well-established finding in artificial intelligence that forward reasoning is an optimal inference mechanism for solving well-structured problems in the presence of a large base of relevant knowledge, whereas backward reasoning is effective when a problem is ill-structured or relevant knowledge is absent. The obvious question of the generalizability of the results exists, and the experimenters made no attempt to manipulate the conditions under which the directionality of reasoning might change. There was also a confounding between accuracy and expertise, since no attempt was made to study erroneous expert performance.

However, there is some converging evidence, which indicates that the phenomenon of expert forward reasoning may be quite robust. Patel and Groen (1986) presented 7 cardiologists with a clinical case in cardiology and asked them to explain the underlying pathophysiology. Since diagnoses were invariably included as part of the explanations, it was possible to rate the explanations on a basis of diagnostic accuracy, as well as to analyze the protocols. It was found that the 4 subjects who gave completely accurate diagnoses also generated their explanations by means of a process of forward reasoning. In contrast, the remaining 3 subjects, who gave inaccurate or incomplete diagnoses, used a mixture of forward and backward reasoning. These results indicate the possibility of a strong relationship between diagnostic accuracy and the presence of forward reasoning. However, they also raise the issue of what causes the pattern of forward reasoning to break down. Our purpose in this paper is to consider this question in the context of medical reasoning.

The strong relationship between forward reasoning and accuracy suggests that whatever determines the latter should also determine the former. Greeno and Simon (1988) suggest that the common factor may be the ability of experts to chunk large pieces of information. This notion can be used to explain the greatly enhanced performance of experts, since high-level chunks can encode large quan-

This research was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada (OGP0002598) and the Medical Research Council of Canada (MH0004) to Vimla Patel. We would like to thank the physicians Yogesh Patel and Sheldon Magder for their assistance with the medical aspects of the studies. Special acknowledgments are due to Kevin Dunbar, David Kaufman, and several reviewers for valuable comments on earlier drafts of the paper. We thank Susan St.-Pierre for putting the manuscript together. Correspondence should be addressed to Vimla L. Patel, Cognitive Studies in Medicine, Centre for Medical Education, McGill University, 3655 Drummond Street, Room 529, Montreal, Quebec H3G 1Y6, Canada.

tities of information, which can then result in the firing of more complex production rules than would be possible otherwise. In the case of nonexpert subjects, such reasoning can only be handled by a much larger number of production rules. Under this assumption, the existence of such large chunks makes possible the use of forward chaining, since such chaining tends not to converge to a solution if the problem space is too large (Winston, 1984).

An important factor in substantiating the chunking notion is the assumption that the existence of large chunks of this kind can be demonstrated through the use of free recall tasks. However, there is evidence that, at high levels of expertise, recall is unrelated to accuracy of performance. Charness (1985) has shown that the performance of experts in chess and bridge remains invariant despite age-related decreases in short-term memory. Similarly, Patel and Groen (1986) showed that inaccurate diagnosis by cardiologists (even though it occurred within their own domain of cardiology) was associated with the use of backward reasoning. However, there was no difference in recall of the case between subjects with accurate and those with inaccurate diagnoses. In fact, physicians' recall of clinical cases is quite independent of the domain of specialization. This independence makes it difficult to apply notions of expertise that assume an intimate connection between enhanced recall and efficiency in reasoning, in the sense that large chunks are enabling the use of complex production rules. Both the kind of rules utilized and the strategies used to generate the rules seem quite independent of chunking phenomena.

The independence of chunking and reasoning indicates that the common factor is not to be found in phenomena related to comprehension. A number of possibilities remain, however. One is that the presence or absence of forward reasoning is simply a matter of cognitive style. Another is that the method of reasoning is determined by the degree of relevant prior knowledge. This can be examined by varying the closeness of match between the subjects' expertise and the expertise demanded by the task. In most domains, there are degrees of expertise. For example, a chess master may be expert in a particular, limited class of openings. The chess master will be likely to perform somewhat differently if forced to play an opening outside that class. In fact, he or she will be more likely to reason inaccurately and hence lose the game. An analogous situation exists in medicine. Cardiologists and endocrinologists may both be expert physicians, but it is clear that diagnoses are less likely to be accurate when an endocrinologist diagnoses a case in cardiology and vice-versa. This kind of performance is not equivalent to novice or intermediate performance, and hence we assign it a special term in this paper, namely, *subexpertise*.

Experiment 1 was designed to examine the possibility that method of reasoning is a function of the degree of relevant prior knowledge, by systematically comparing the performance of subexperts with that of experts. A quite different possibility is that, whereas diagnostic accuracy is determined by prior knowledge, the breakdown of forward reasoning is determined by coherence of explanation. This will be examined in connection with Experiment 2.

EXPERIMENT 1

This experiment was designed to compare the results obtained by Patel and Groen (1986) with diagnostic explanations of subexperts (specifically, specialists in domains of medicine other than cardiology). The experimental design and the methods of analysis were identical to those used in the original experiment.

Subexperts should be expected to diagnose a case less accurately than experts. The main issue is to examine how closely the absence of forward reasoning is related to inaccurate performance. It is, however, also possible to examine the extent to which the diagnostic explanation is determined by the subject's knowledge base. A considerable amount of research on expert problem solving has argued that the solution process is determined by the ability to represent specific aspects of content. This position, which essentially reflects conventional thinking on expert systems (e.g., Hayes-Roth, Waterman, & Lenat, 1983), assumes that both accuracy of diagnosis and directionality of reasoning are determined by prior knowledge. Forward reasoning occurs because the rules stored in long-term memory are naturally chained together in a forward direction when the knowledge base is sufficient to induce accurate results. Backward reasoning occurs because the subject is forced to resort to weak methods. If this is the case, the explanations of subexperts should exhibit a bias toward the use of rules more typical of their own specialized areas.

Method

The subjects were 6 surgeons and 6 psychiatrists who volunteered to participate in the study. These physicians were practitioners with MD degrees and Board Certification in their respective specialities. They were affiliated with at least one of the teaching hospitals of McGill's Faculty of Medicine.

The problem text used in this study was the same as the one used in Patel and Groen (1986). It describes a patient with a diagnosis of *acute bacterial endocarditis*, which is a bacterial infection affecting heart valves. The case has four main components: (1) *infection*—the patient contracted the infection from a contaminated needle, possibly from intravenous drug use; (2) *aortic valve insufficiency* (AVI)—the aortic valve became infected and the blood gushed back through the valve, causing heart murmurs; (3) *emboli*—the patient manifested a blood-shot eye, due to a capillary's bursting from the lodging of a particle from the infected valve; and (4) *acuteness*—the patient was seriously ill and needed immediate attention. All the components in the text are necessary to generate a complete and accurate diagnosis. Problems of *acute bacterial endocarditis* are reasonably familiar to cardiologists but not to other specialists such as psychiatrists or surgeons.

The basic experimental procedure was identical to that used by Patel and Groen (1986). The subjects were shown a written description of the case. Each subject was tested individually and allowed

2.5 min to read the case description, which was then removed. The subjects were then asked to write down as much of the text as they remembered, and then explain (again in writing) the underlying pathophysiology of the case without reference to either the text or the previous recall response. Finally, the subjects were asked to provide a diagnosis.

For data analysis, the diagnoses were coded as *accurate*, *partially accurate*, or *inaccurate*. A diagnosis was characterized as accurate when it was completely correct (i.e., all of its components were included); partially accurate, when some components of the diagnosis were included; and inaccurate when no diagnostic components were included.

Using techniques of propositional analysis based on Kintsch (1974) and Frederiksen (1975), both the clinical text and the response protocols were analyzed. The protocols were scored for recalls against the stimulus texts according to the method described in Patel, Groen, and Frederiksen (1986). The stimulus texts were initially segmented into clauses and then propositionally represented. Next, the subjects' protocols were segmented and matched against the original text propositions. The scoring involved marking every item in the subjects' protocols that corresponded to the text base as defined in the original text. On the basis of experts' classifications, the propositions in both clinical cases were classified as *disease relevant* or *disease irrelevant* propositions, and a record was made of the number of relevant and irrelevant propositions each subject recalled.

For each pathophysiological explanation protocol, a semantic network representation of the causal and conditional rules was obtained, using the method described in Patel and Groen (1986). Our basic approach in analyzing these data is to begin by representing the propositional structure of a protocol as a semantic network. Within this structure, the propositions that describe attribute information are called *nodes* and those that describe relational information are called *links*. We will call the attributes that are given in the original clinical text as *facts*, and those that are not given, we will call *hypotheses*. The causal (CAU:) and conditional (COND:) relationships that form a major part of the semantic network arising from our data closely resemble the rules in an expert system. The directionality of causal and conditional rules is important since causal rules generally lead away from a diagnosis (diagnosis → explanation of facts), whereas conditional rules lead toward one (observable facts → diagnosis). This system thus consists of hypothesis-directed and data-directed or fact-directed rules, which can be viewed as equivalent to forward and backward reasoning.

In order to be more precise about our analysis, some notions from graph theory are used to define properties of the semantic networks in terms of components (Groen & Patel, 1988). A *graph* is a nonempty set of *nodes* and a set of *arcs*, each leading from a node N to a node N'. A *walk* through a graph is a sequence of nodes such that if a_{i+1} is the immediate successor of a_i in the sequence, then the two nodes are connected by an arc. A *path* is a walk in which all nodes are distinct. In a *directed graph*, an arc connecting N to N' is viewed as distinct from an arc connecting N' to N. To emphasize the distinction, arcs of directed graphs are usually denoted by arrows. If e is an arrow connecting N to N' then N is called the *source* of e, and N' is the *target* of e. A path is said to be *oriented* or *directed* if every node is the source of an arrow connecting it to its immediate successor. In other words, it is a path that follows the direction of the arrows. A *subpath* of an oriented path P is a path that is a subsequence of P. Note that undirected paths are possible in directed graphs if one simply ignores the order of the arrows. A graph is *connected* if a path, directed or undirected, exists between any two nodes. If it is not connected, then it breaks down into *disjoint components*, each of which is connected, but none of which has a path linking it to any other component.

Backward reasoning corresponds to an oriented path or subpath from a hypothesis to a fact, whereas forward reasoning corresponds to a path or subpath from a fact to a hypothesis. The presence of subpaths in these definitions is important, because it gives criteria for forward or backward reasoning between facts. Pure forward reasoning corresponds to a graph in which every oriented path satisfies the forward reasoning criterion. Pure backward reasoning corresponds to a graph in which every oriented path satisfies the backward reasoning criterion.

Wherever possible, statistical analyses were performed on the data to test for significance of results. Because of the small sample size, a series of nonparametric Mann-Whitney tests was performed on each of the dependent variables separately.

Results

Because of the use of procedure and stimulus materials identical to those in Patel and Groen (1986), the specialists in cardiology studied by Patel and Groen were used as a comparison group. Four out of 7 cardiologists arrived at a completely correct diagnosis, whereas the other 3 cardiologists gave partially correct diagnoses. One of the 6 surgeons in the present study was able to provide a completely accurate diagnosis. The rest of the surgeons were only partially accurate with their diagnoses. Psychiatrists, by contrast, provided either partially correct or incorrect diagnoses. None of the psychiatrists provided completely accurate diagnoses of the case.

The protocols were analyzed as described for the presence of pure forward reasoning. The only subject who exhibited this pattern was the surgeon who made the accurate diagnosis. No subject with partially accurate or inaccurate diagnoses showed this pattern of pure forward reasoning.

Analysis of a single subject. To make our analysis clear, we will concentrate in some detail on the results of the single subject who made a completely accurate diagnosis. A discussion of the results with the other subjects will follow the analysis of this single subject.

The explanation of the underlying pathophysiology given by Surgeon 1 does not have a close relation to the original text. The protocol generated by the surgeon and the propositional representation of the protocol are given in the Appendix. The propositions are represented in terms of a relational structure, which is given in Figure 1. In this diagram, the numbers denote propositions numbered in accordance with the Appendix. The arrows correspond to the linking propositions and indicate directionality. These are referred to as links and the other propositions as nodes. The diagram also indicates the nodes that correspond to propositions appearing in the text (text cues) and those that are part of the expert canonical knowledge (intermediate components).

Figure 1 defines a process that leads to a diagnosis similar to the one described in Patel and Groen (1986). To investigate the extent to which the other subjects are using a similar process, we make use of a reference model for the disease. This network contains the knowledge base of an expert cardiologist and the rules necessary to explain the diagnosis. Altogether, there are 15 rules, all of which are in a forward direction toward the diagnosis with no instance of backward chaining. These rules are given in Table 1.

The reference model can be decomposed into a number of disease components. To make a diagnosis, one has to understand the disease process in each component and

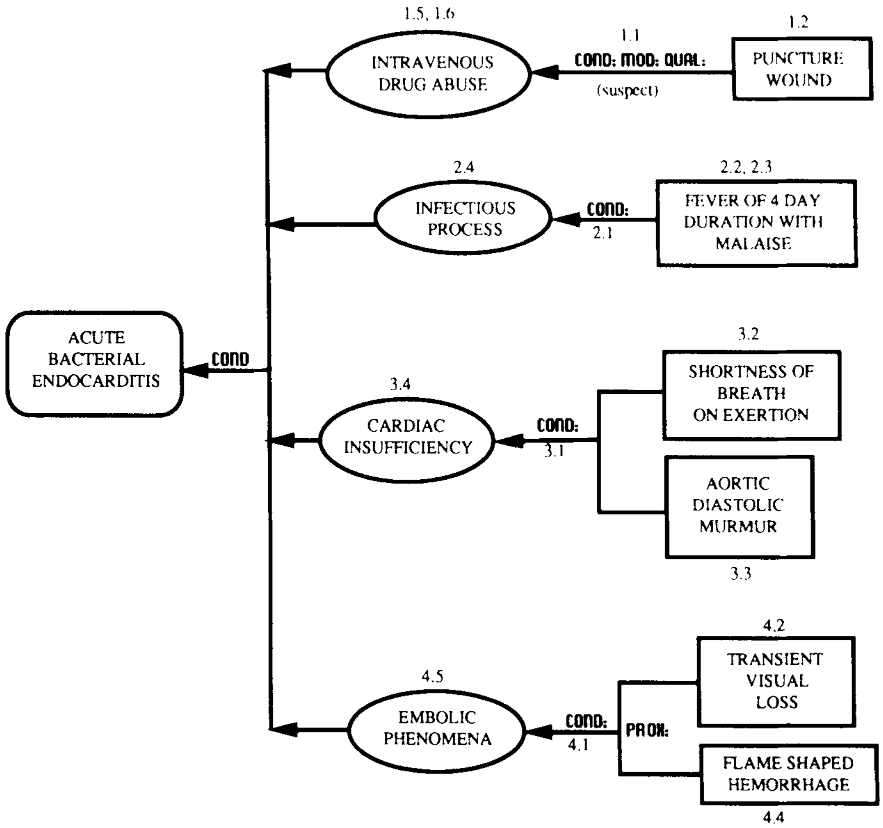

Figure 1. Representation of the relational structure of the patient problem from Surgeon 1. The proposition numbers refer to those in the Appendix and the arrows indicate directionality. Textual cues: ☐ COND: = conditional relation. Intermediate components: ◯ MOD: QUAL: = modality:qualify. Diagnosis: ▢ PROX: = proximal.

also be able to connect these various components. From a rule-based point of view, each disease component can be viewed as a set of rules that lead to an independent part of the diagnosis. The *connecting rules* are (Rule 7 and Rule 15 in Table 1) those that connect component diagnoses. *Critical rules* (Rule 2 and Rule 9 in Table 1) are those that are considered by expert consultants, who did not participate in the study, as the minimal number of rules that would elicit the production of an accurate diagnosis by an expert.

A total of five rules from the reference model, including critical and connecting rules, were used by our single subject. A comparison of this subject with the single subject (a cardiologist) described in Patel and Groen indicates a similar pattern for both subjects, with the relevant propositions from the text being used as input information in the network derived from the explanation protocol. All the rules are in the forward direction, pointing toward the diagnosis. This implies that the subject is using pure forward reasoning, like the cardiologist with the accurate diagnosis. Both the cardiologist and the surgeon make use of critical and connecting rules for accurate diagnosis. The use of two disease components, intravenous drug abuse and aortic insufficiency, was also common to the two subjects with accurate diagnoses.

Recall and diagnostic accuracy. There were no differences in recall of propositions in each of the four disease components (infection, emboli, AVI, and acuteness), by surgeons, psychiatrists, and cardiologists.

The propositions were separated into relevant and irrelevant ones. The overall recall of relevant propositions was high for all three groups of subjects. The propositions corresponding to the infection component were recalled the most and propositions of the AVI component recalled the least, although the differences were not statistically significant. Relevant propositions were separated into accurate, partially accurate, and inaccurate diagnoses pooled over the three groups of experts. Relevant propositions constituted 64% of the total propositions recalled by subjects making accurate diagnoses. Subjects making partially accurate diagnoses recalled 61% of the relevant propositions. Subjects making inaccurate diagnoses recalled 65% of the relevant propositions. There were no significant recall differences between these categories.

Explanation of clinical case. The total number of rules used in the explanation of the clinical case was calculated for accurate, partially accurate, and inaccurate diagnostic categories by the three groups of experts. A total of 38 rules were used by the experts with accurate diagnoses, 31 rules were used by experts with partially accurate di-

Table 1
Production Rules That Yield an Accurate Diagnosis

Production Rule	Antecedent (if)	Consequent (then)
1	puncture wounds and young unemployed male	intravenous drug use
2	intravenous drug use	bacterial infection
3	bacterial infection and emboli	bacterial endocarditis
4	fever	bacterial infection
5	transient blindness	emboli
6	red blood cells in urine	emboli
7	intravenous drug use and bacterial endocarditis	bacterial endocarditis from intravenous drug use
8	low diastolic pressure and normal systolic pressure	aortic valve insufficiency
9	early diastolic murmur	aortic valve insufficiency
10	shaking chills and bacterial endocarditis	acuteness of bacterial endocarditis
11	rigor and bacterial endocarditis	acuteness of bacterial endocarditis
12	normal spleen and bacterial endocarditis	acuteness of bacterial endocarditis
13	short duration of illness and bacterial endocarditis	acuteness of bacterial endocarditis
14	normal heart size and bacterial endocarditis	acuteness of bacterial endocarditis
15	bacterial endocarditis from intravenous drug use, aortic valve insufficiency, and acuteness of bacterial endocarditis	acute bacterial endocarditis with aortic insufficiency from drug use

From "Knowledge-based solution strategies in medical reasoning," by V. L. Patel and G. J. Groen, 1986, *Cognitive Science*, **10**, 91-116 (Table 4, p. 104). Copyright 1986 by Ablex Publishing Corp. Reprinted by permission.

agnoses, and 18 rules by experts with inaccurate diagnoses. There was a significant difference between the number of rules used in the two diagnostic components AVI and acuteness, in terms of the accuracy in diagnosis. Overall, more rules were used in these two components by subjects making accurate and partially accurate diagnoses than by subjects making inaccurate diagnoses. Experts with accurate diagnoses used 9 rules in the AVI component and 6 rules in the acuteness component of the disease. The subjects with partially accurate diagnoses used 6 rules in the AVI component and 4 rules in the acuteness component. The subjects who did not diagnose the case accurately did not use any rules in the AVI component and used only 1 rule in the acuteness component overall. These results were statistically significant at the $p < .01$ level for the use of rules in the AVI component and the $p < .05$ level for the acuteness component.

The pattern of rules used by cardiologists with accurate and inaccurate diagnoses shows that two critical rules and some of the connecting rules were used by the subjects with accurate diagnoses, but that they were not used by any of the subjects with inaccurate diagnoses. The rules used by psychiatrists and surgeons were pooled according to partially accurate and inaccurate diagnostic categories. As shown earlier, the subjects who made accurate diagnoses used both a critical rule (one rule) and the connecting rules (two rules). The subjects with partially accurate diagnoses used critical rules but did not use any connecting rules. None of the subjects with inaccurate diagnoses used either of these two types of rules.

For the subjects with partially accurate diagnoses, 80% (24 out of 30) of the rules used were relevant to the diagnosis as compared with 21% (4 out of 19) for subjects with inaccurate diagnoses.

Discussion

In general, these results are highly consistent with the results of Patel and Groen (1986). There are no differences in recall, either as a function of domain specialization or as a function of diagnostic accuracy. The subject with an accurate diagnosis solved the problem by pure forward reasoning, using input information from the text. The other subjects showed mixed directionality in their reasoning, forward and backward. In fact, the only major difference lay in the marked decrease in diagnostic accuracy within the two groups of subexperts.

The implication is that performance on this task is determined by whether or not the subjects possess the relevant knowledge. There is, however, no support for the hypothesis that the area of specialization leads to explanatory bias based on specific areas of interest. This may, however, be due to the small number of subjects utilized.

The most significant aspect of this experiment is the way it reinforces the findings of Patel and Groen regarding the relationship between forward reasoning and diagnostic accuracy. It would seem from these results that forward reasoning is a necessary and sufficient condition for diagnostic accuracy, regardless of a subject's knowledge base. However, it is possible that the homogeneity of these results is due to the fact that an extremely well-structured clinical case was used. Experiment 2 was designed to examine whether this strong relationship between forward reasoning and diagnostic accuracy might break down when more ill-structured clinical cases were used.

EXPERIMENT 2

Since prior knowledge does not seem to determine the presence or absence of forward reasoning, once diagnostic accuracy has been factored out, it is necessary to seek alternative explanations. In Experiment 2, we considered two assumptions (which essentially constituted a dual-process theory). The first was that all subjects will begin by using forward reasoning in generating a diagnostic explanation. The second was that, although diagnostic accuracy is determined by prior knowledge, the breakdown of forward reasoning is determined by the coherence of the explanation. The successful application of pure forward reasoning yields a distinctive pattern in which all rules converge to a unique solution. If it is further assumed that the goal of the subject is to generate a complete explanation that accounts for all the important facts in the clinical case, then an incomplete diagnosis will result in facts that are not connected to the main diagnosis. If the subject is aware of such facts, then the breakdown of forward reasoning may reflect an attempt to explain why these facts are unconnected to the diagnosis. If this hypothesis is correct, the pattern of forward reasoning might also be expected to break down independently of accuracy when experts attempt to explain cases that include symptoms unrelated to the main diagnosis. In Experiment 2, we compared the performance of experts and subexperts on cases with two classes of cues that might be expected to yield a breakdown of forward reasoning. The first consisted of cues indicating an alternative diagnosis that had to be ruled out. The second consisted of cues related to a secondary disease that the patient had, but that was unrelated to the main diagnosis.

Experiment 2 differed from Experiment 1 in the use of cases containing cues that might be expected to yield a breakdown of forward reasoning. It also differed in the use of a within-subject design rather than a between-subject design, to control more closely the possible effects of individual differences. In addition, in an effort to further explore the possibility of response biases, some of the specialists were researchers rather than clinical practitioners.

Method

Expert physicians associated with the Faculty of Medicine volunteered for the study. The subjects were 8 cardiologists and 8 endocrinologists. Four cardiologists and 4 endocrinologists were researchers in medicine, with both MD and PhD degrees, who spent about 70% of their time in biomedical research. The rest of the subjects were practitioners, with MD degrees, who spent about 70% of their time seeing patients, and with little or no research activity.

Two texts were constructed, which described the history, the physical examination, and the laboratory tests of two patients. The first text, which described the case of a 63-year-old woman suffering from an endocrine disorder called *Hashimoto's Thyroiditis*, precipitated to myxedema pre-coma, is as follows:

> A 63-year-old woman with 1-week history of increasing drowsiness and shortness of breath was brought to the Emergency Room by her daughter. The patient had not been well for over a year. She complained of feeling tired all the time, had a loss of appetite, a 30-lb weight gain, and constipation. A month later she had been diagnosed as having "chronic laryngitis" and was prescribed a potassium iodide mixture as an expectorant.
>
> Physical examination revealed a pale, drowsy, obese lady with marked periorbital edema. She had difficulty speaking, and when she did speak her voice was noted to be slow and hoarse. There were patches of vitiligo over both her legs. Her skin felt rough and scaly. Her body temperature was 36° C. Pulse was 60/min and regular. B.P. was 160/95. Examination of her neck revealed no jugular venous distention. The thyroid gland was enlarged to approximately twice the normal size. It felt firm and irregular. There was grade 1 galactorrhea. The apex beat could not be palpated. Chest examination showed decreased movements bilaterally and dullness to percussion. There was no splenomegaly. Neurological testing revealed symmetrical and normal tendon reflexes, but with a delayed relaxation phase. Urinalysis was normal. Chest x-ray: low voltage complexes and nonspecific T-wave flattening. Routine biochemistry (SMA = 16) showed Na = 125, BUN = 8 mg/100 ml. Arterial blood gases: PO2 = 500 mm Hg, PCO2 = 60 mm Hg. The patient was admitted to the Intensive Care Unit for further management.

The diagnosis can be decomposed into three components. The most general and prototypical component is hypothyroidism. This is indicated by the textual cues suggesting fluid accumulation and decreased thyroid function. The second component, myxedema, indicates that the patient is in an advanced state of hypothyroidism. The clinical cues that constitute the third component suggest a very specific origin for the disease process: the autoimmune process known as Hashimoto's thyroiditis. Cues from each of the three components need to be recognized to accurately diagnose the problem.

The second clinical text, which described the case of a 62-year-old man who was diagnosed as having *cardiac tamponade with pleural effusion*, is as follows:

> This 62-year-old retired Air Force mechanic was apparently well until 5 months before presenting to the hospital. He then noted he was "winded" after walking about 40 ft. He was increasingly breathless lying down, tired, using four pillows to sleep and most recently is sleeping sitting up. He has occasionally awoken extremely short of breath. He has a mild nonproductive cough and agrees that his voice is a little hoarse. During this time, his legs ha[illegible] been swelling. His appetite has decreased, yet his abdomen has increase[illegible] and he has gained weight. He says "no food tastes good" and he has constant mild nausea but has not vomited. He has had no chest or abdominal pain. He does not smoke, drinks alcohol socially but less lately. His only admission to the hospital was for a heart attack 12 years ago. He recovered completely and was walking 6 miles a day a year ago. He is taking no medication.
>
> On examination: H.R. 80/min and regular. B.P. 120/98 mm Hg. Pulsus paradoxicus 12 mm Hg. No cyanosis. Pronounced peripheral edema of legs and presacrum. Some edema over abdominal wall and scrotum. Abdomen was large with shifting dullness and a fluid wave was demonstrated. Liver edge was smooth, 3 cm below the right costal margin. Spleen was not palpated. No masses. Jugular veins distended to the angle of the jaw at 45°; apex not palpable, heart sound faint, no S3, no S4, no murmurs. Some dullness to percussion at right lung base. Breath sounds diminished at both lung bases with decreased chest expansion. Fine end inspiratory crepitutions noted. Remainder of examination was normal.
>
> Hb = 13.5 gm%, WBC = 5,500 with a normal differential. Prothrombin time 12.5 (control 11.8), P.T.T. 34 (control 34), T4 = 7.5 (normal 4.5-10.5). Urinalysis was normal except urobilinogen 4.0 (normal 0.1-1.0); SMAC 16 normal except: Albumin 3.5 (N = 3.7-4.9), total bilirubin 1.7 (N = 0.2-1.0); alkaline phosphotase 169 (N = 30-105). Chest x-ray: "Enlarged cardiac silhouette; no evidence of pulmonary edema, right pleural effusion, partial atalectasis in right lower lobe." ECG: Remote inferior myocardial infarction. Diffuse ST sagging with T-wave inversion. Generally low voltage QRs with voltage fluctuation.
>
> This patient has been referred from an outlying hospital for definitive management.

This case is more difficult than the endocrinology case, because there are more commonalities between knowledge for various causal patterns leading to alternative diagnoses. To diagnose the case, the physician must decide whether the problem is caused by a left- or right-sided failure, and then identify the presence of pericardial effusion and cardiac tamponade. Determining the actual causal process (right-sided heart failure) is a difficult task, because many of the diagnostic features are common to different diagnostic possibilities. There are, however, a few cues that either serve to rule out alternative diagnoses, or are not related to the main diagnosis.

The procedure used was the same as the one used in Experiment 1. Each subject was required to read, recall, and explain the underlying pathophysiology of the clinical case before providing the final diagnosis for both the cardiology and the endocrinology texts. The order of text presentation was controlled such that all subjects worked on the problem in their own domain first, followed by the problem outside their domain of expertise. The order was not counterbalanced, because the specialists are not normally presented with problems outside their area of specialty. Because of this, presentation in reverse order might lead to results influenced by motivational factors, which are irrelevant with respect to the issues of concern in the experiments.

Results

Diagnostic accuracy. Because there were no clear criteria for partially accurate diagnoses for the two clinical problems, we used only two diagnostic categories, accurate and inaccurate. If the accurate diagnosis was given as a part of a differential list of diagnoses, it was considered accurate. Three out of 4 endocrinology practitioners and 4 out of 4 endocrinology researchers made completely accurate diagnoses on the clinical problem in endocrinology. However, all the endocrinologists made inaccurate diagnoses on the cardiology case. No alternative diagnoses were made on the endocrinology case, as compared with the cardiology case, where alternative diagnoses were provided.

Three out of 4 cardiologist practitioners and 2 out of 4 cardiologist researchers provided completely accurate diagnoses for the clinical problem in cardiology. However, in contrast with the endocrinologists, the cardiologists provided other alternatives beside the correct diagnosis. Seven out of 8 cardiologists provided incomplete diagnoses of the clinical problem in endocrinology, with only the most general aspects of the problem being identified. There were alternative diagnoses provided.

Overall, the diagnoses of the researchers showed more use of a pathophysiological level of explanation together with the clinical aspect of the problem explanation. In contrast, the diagnoses of the practitioners were restricted to the clinical manifestations only.

Relationship between recall and explanation. All subjects were able to recall the information in the text, irrespective of their specific areas of expertise. The results show that both cardiologists and endocrinologists were able to select the relevant propositions and provide a summary of the most important features of the text, even when an inaccurate diagnosis was ultimately given. This result is similar to the findings in Experiment 1, even though a far more difficult text was used. However, the issue of interest lay in the relation of the pathophysiology protocol to the recall protocol. In this case, the propositions in the recall protocols were matched against the propositions in the pathophysiological explanation protocols to determine the overlapping information.

The researchers repeated a high percentage of the recall propositions in the pathophysiology when solving the patient problems in their own speciality, with 48% of the propositions repeated for the endocrinology text and 45% for the cardiology text. However, this was not true for solving the case outside their area of specialization, where 5% of the propositions recalled were used in the explanation of the problem in endocrinology and 25% of the propositions in the cardiology problem. Practitioners, however, showed the opposite effect: They used fewer cues from the summary to explain the pathophysiology of the problem in their own domain, with 14% overlap between recall and explanation for the endocrinology problem and 7% overall for the cardiology problem. However, they used more cues to explain the problem outside their domain of expertise, with 39% for endocrinology and 59% for cardiology.

When the subjects are in their own domain, the practitioners use very little of the propositions in the original case to explain the patient problem. In contrast, the researchers use more of the original propositions. When out of their domain, this relationship is reversed. Practitioners use many more propositions from the clinical case, and researchers use far less propositions. There is an efficient use of textual propositions in the case of practitioners in their own domain, in that a few relevant propositions from the text are used in the explanation. These relevant propositions are necessary for the diagnosis.

Pattern of reasoning in diagnostic explanation. All the subjects with accurate diagnoses used forward reasoning on one large component leading to a diagnosis in their explanation protocols. This was accompanied by one to two small components with backward reasoning to explain the textual cues not accounted for in the major component. This is true for 3 out of 4 endocrinologists who made the accurate diagnosis.

This was also true for 3 out of 4 cardiologist practitioners who made partially accurate diagnoses on the cardiology problem. The subjects used forward reasoning for the major component in the protocol and backward reasoning to tie up the nonsalient cues. However, all 4 cardiologists used forward reasoning at the beginning to rule out the possible alternative diagnosis. These cardiologist and endocrinologist practitioners solving the problem out of their domain of expertise used a mixture of forward and backward reasoning.

The researchers, all 4 in endocrinology and 3 out of 4 in cardiology, used forward chaining as a major pattern of reasoning and backward chaining to tie up loose ends. This is the same pattern of reasoning shown by the practitioners in their own domain of expertise. However, the researchers made greater use of biomedically oriented information, as compared with the more clinical or patient-oriented information used by the practitioners. When these researchers solved problems outside their domain of expertise, all subjects displayed a greater use of backward reasoning. The diagnoses were inaccurate and incomplete in this case.

We will concentrate on the pathophysiological explanations contained in four protocols that indicate the pattern of forward and backward chaining clearly. The semantic networks were generated by the method given in Patel and Groen (1986). We begin with practitioners explaining the clinical case in their own domain. Figure 2 gives the semantic network representation of the protocol of an endocrinology practitioner explaining the endocrinology problem.

The explanation presents very little text information, and it is constructed to justify a diagnosis. The order of reasoning is completely forward except at the very end of the protocol, where a condition related to low general metabolism (hypometabolic state) is explained in terms of a possible outcome (respiratory failures), and where the existence of a low serum level of sodium ions (hyponatremia) is explained in terms of impaired water excretion. There is one large component with forward chaining leading to a diagnosis, and there are two small components with backward reasoning to explain the textual cues.

Because it was a more difficult case, nobody obtained a completely accurate diagnosis for the cardiology problem. The reasoning of the 4 cardiology practitioners followed the pattern in the protocol of the endocrine practitioners explaining the endocrine problem. The subjects used pure forward reasoning, except to account for factors not critical for the diagnosis. However, most of the rules were not used to generate a causal explanation, but rather to rule out possible alternative disorders. This was accomplished before proceeding to the explanation of the actual process. Figure 3 gives the semantic representation of the protocol of a cardiology practitioner explaining the cardiology problem. This protocol illustrates the pattern of reasoning described. The strategy used is to rule out the alternatives or interfering information such as textual cues that rule out pericardial constriction, tamponade, right heart failure, anemia, and so forth, as seen in the top half of Figure 3, before attempting to diagnose the actual problem, which is seen in the bottom half of Figure 3. This was a common strategy in all of the four protocols generated by the practitioners. The use of ad-

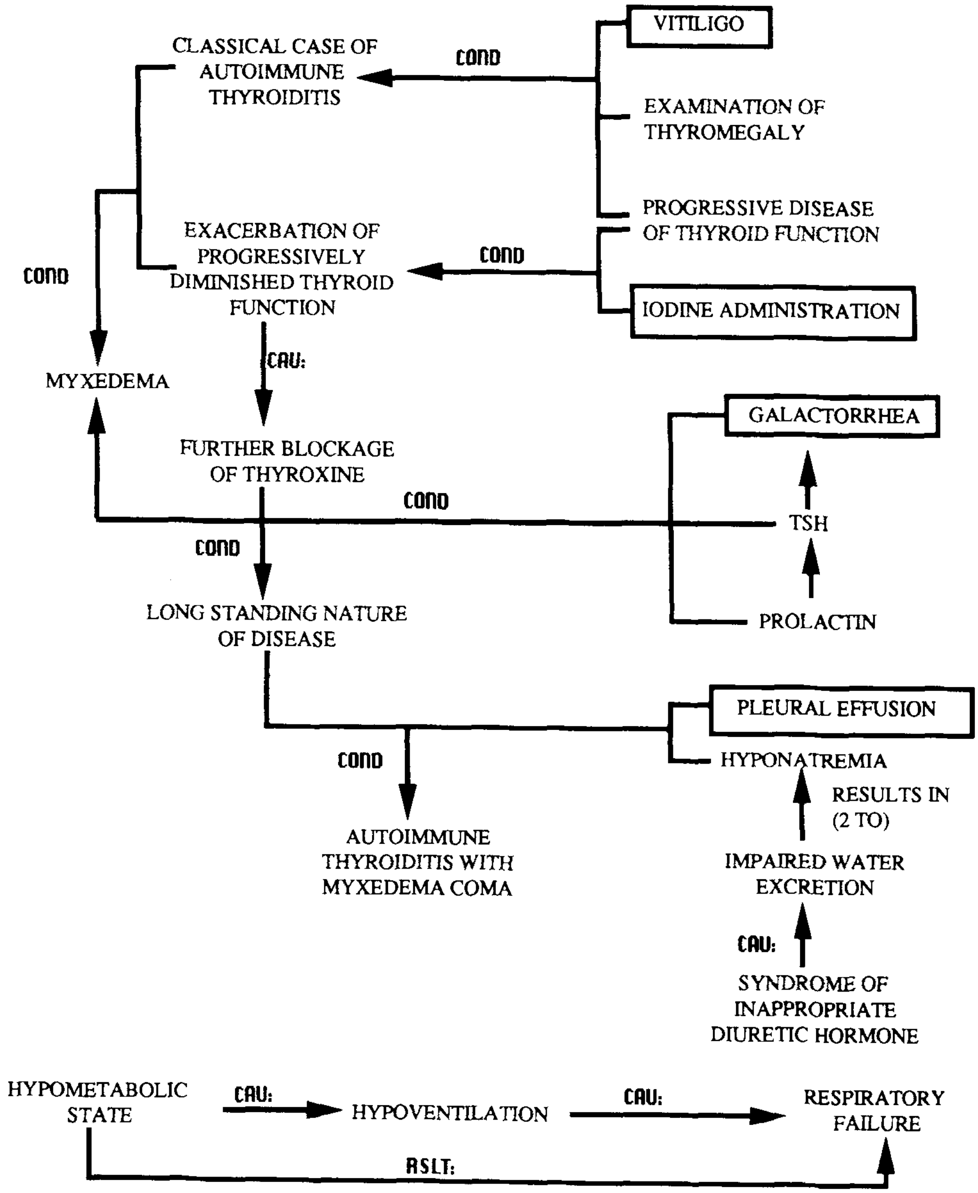

Figure 2. Representation of the relational structure of the patient problem in endocrinology from an endocrinologist practitioner. Textual cues: ☐ CAU: = causal relation. COND: = conditional relation. RSLT: = resultant.

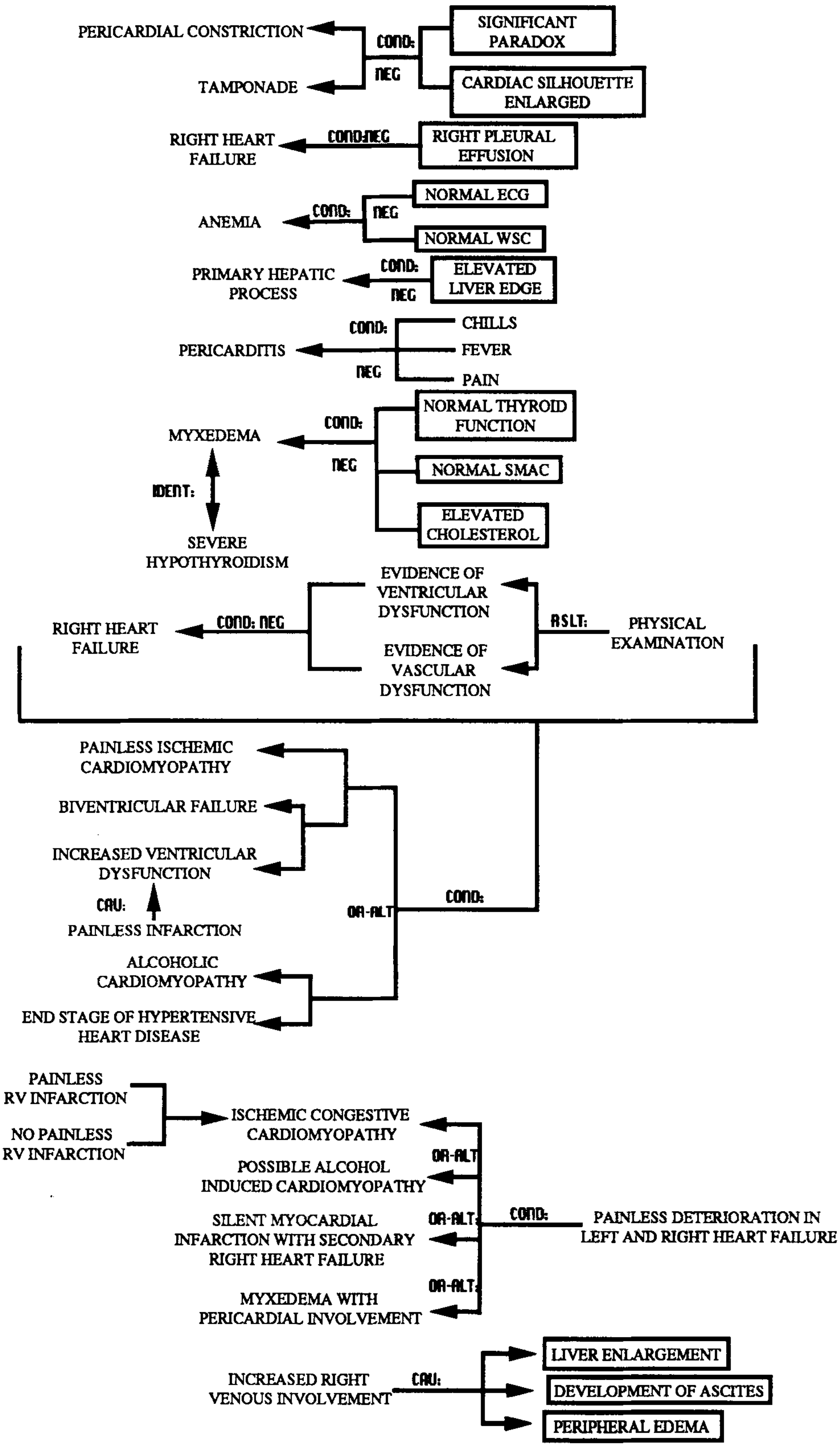

Figure 3. Representation of the relational structure of the patient problem in cardiology from a cardiologist practitioner. Textual cues: ☐ OR-ALT: or alternative. COND: NEG = negative conditional relation. CAU: = causal relation.

ditional reasoning strategies other than forward reasoning depended on the alternative hypothesis being considered.

Unlike experts working within their domains of expertise, experts solving problems outside their area of specialization do not show patterns dominated by forward reasoning. These subexperts regularly appeal to rules that introduce new details, which expand rather than constrain the problem space.

The responses from physicians who are actively involved in biomedical research are quite different from those of practitioners. In general, a greater and more detailed use is made of biomedical scientific knowledge rather than clinical or patient-oriented knowledge. The direction of reasoning was found to be mostly forward with some backward reasoning at the end of the protocols. Once again, as with our practitioners in their own domain of specializations, the researcher also used backward reasoning to causally explain the propositions that were left unexplained, so as to provide a general coherence to their explanation. All four researchers showed the same pattern of reasoning. It should be noted that all the endocrine researchers made accurate diagnoses.

In contrast with the researchers in their own domain of specialization, researchers out of their domain do not make use of much of the textual cues. Cues from the text are selected that provide evidence of biochemical and physiological mechanisms rather than the clinical manifestations of the disease. There is a greater use of backward reasoning in all explanations. They were all associated with inaccurate or incomplete diagnoses.

To show precisely the kinds of information—biochemical/physiological versus clinical/patient—that were included in the underlying problem explanation, the semantic network generated for each protocol was broken down into the components according to the graph theoretic analysis described in Experiment 1. The total number of components that related to the biomedical aspects and ones that related to the clinical aspects were recorded. For example, the relational network generated from the explanation protocol of an endocrine practitioner working on the endocrine problem was decomposed into four components, as is shown in Figure 4. Each of the four components was identified as a clinically related component, in which all propositions were associated with clinical manifestations of the disease. Similarly, four components generated from the cardiologist's protocol working on an endocrine case are shown in Figure 5. In contrast, each of these components is associated by means of detailed physiological mechanisms rather than clinical ones.

The results for the endocrinologists are shown in Table 2. The endocrinologist practitioners in their own domain use equally the clinical and the biomedical components to explain the case. Some endocrinologist practitioners solving a problem in cardiology were shown to have emphasized biomedically oriented rather than clinically oriented components in their explanations. However, the researchers, both in and out of their domain, focused more on the biomedical than on the clinical aspect of the disorder. Furthermore, there was an increase in the total number of components with researchers, indicating a certain lack of global coherence. Global coherence is defined here as a semantic coherence that provides some form of overall relatedness or unity in the explanation of the case (van Dijk & Kintsch, 1983).

The cardiologist practitioners utilized biomedically oriented more than clinically oriented information in their explanations of the cardiology case. This is shown in Table 3. There was even a greater tendency among the cardiology researchers than among the endocrine researchers to use biomedically based information.

Overall, when endocrinologists were compared to the cardiologists in terms of component types, there were no statistical significant differences in either of the two case types. However, significant results were found in the use of clinical information in both case types when practitioners were compared with researchers, regardless of specialization. This result was significant for the use of clinical information in the endocrine case ($u < 7, p < .01$) and for the cardiology case ($u < 4, p < .05$). There were no significant differences for the use of physiological information in either of the two clinical cases.

The results suggest that when experts who are practitioners explain a problem in their own domain of expertise, they focus on patient-oriented components of the problem. When they are explaining a problem outside

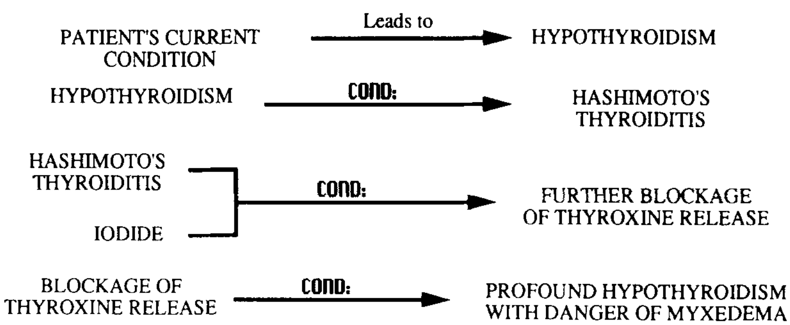

Figure 4. Semantic components of pathophysiological explanation of endocrinology case by Endocrinologist Practitioner 3.

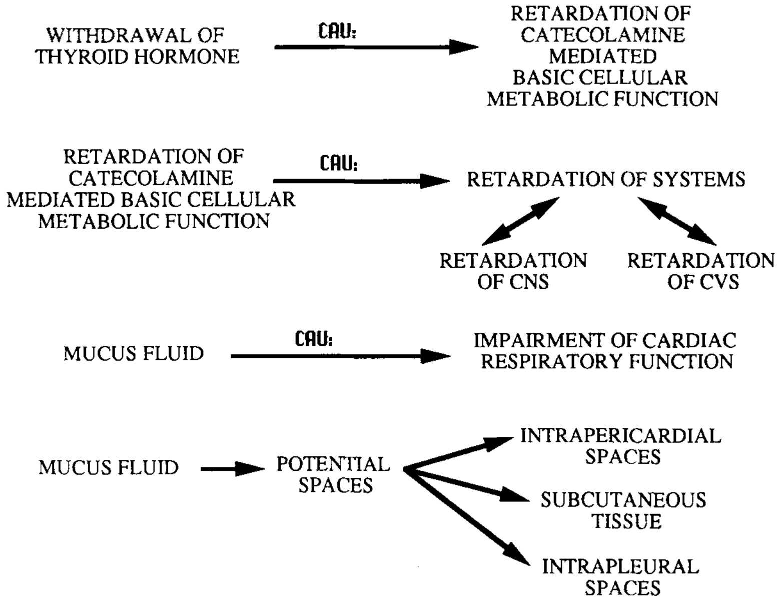

Figure 5. Semantic components of pathophysiological explanation of endocrinology case by Cardiologist Researcher 2.

their domain of expertise, they focus on both the biomedical and the patient-oriented component. This is true for both routine and difficult cases with unconnected information. In all cases, the researchers, regardless of their expertise, focus on the biomedical aspect rather than the clinical/patient aspect.

Discussion

These results support the hypothesis proposed in the introduction that properties of the structure of the clinical case can result in a breakdown of forward reasoning. However, only one type of property considered in this experiment seems to induce breakdown. Contrary to expectation, the ruling out of alternative diagnoses was generated by forward reasoning. The breakdown of forward reasoning appears to occur when irrelevant facts are tested for consistency against the main diagnosis. Apart from this, the results once again provide a close replication of the results of Patel and Groen (1986). The subjects with accurate diagnoses always used pure forward reasoning in their explanations of the principal component of the diagnosis. The subjects with inaccurate diagnoses all exhibited an absence of pure forward reasoning.

There is, however, one additional respect in which the results of this experiment go beyond those of Experiment 1. This is the indication of a highly generic type of explanatory bias, in the sense that researchers tend to make more use of biomedical concepts than of clinical concepts.

GENERAL DISCUSSION

These results indicate that two phenomena may have a considerable amount of generality. The first is the lack of relationship between accuracy of recall and accuracy of diagnosis. It is reasonably clear that recall encounters a ceiling effect, and that this ceiling occurs at relatively low levels of expertise. This is important, because many theoretical accounts of expertise emphasize the importance of processes of comprehension and short-term memory encoding that might be expected to result in recall differences. This may reflect the fact that most problem-solving tasks studied in the literature do not require the high degree of expertise demanded by the cases presented in this experiment. Because of this, Patel and Groen (in press) have suggested a distinction between generic expertise and specific expertise. All experts and subexperts

Table 2
Number and Type of Components in Pathophysiological Explanations by Endocrinologists and Case Type

	Number of Components		
Subject	Total	Physiology/ Pathophysiology	Clinical/ Patient
	Endocrinology		
Practitioner			
1	2	0	2
2	5	2	3
3	4	2	2
Researcher			
1	4	4	0
2	9	8	1
3	8	7	1
	Cardiology		
Practitioner			
1	6	4	2
2	8	6	2
3	4	3	1
Researcher			
1	8	6	2
2	10	9	1
3	4	3	1

Table 3
Number and Type of Components in Pathophysiological Explanations by Cardiologists and Case Type

Subject	Total	Physiology/ Pathophysiology	Clinical/ Patient
	Number of Components		
Cardiology			
Practitioner			
1	6	5	1
2	10	4	6
3	8	6	3
Researcher			
1	4	3	1
2	5	5	0
3	5	5	0
Endocrinology			
Practitioner			
1	8	6	2
2	6	2	4
3	4	2	2
Researcher			
1	4	4	0
2	3	3	0
3	5	5	0

appear to have the generic expertise required for the recall of relevant information in a clinical case. Only the domain experts normally have the specific expertise required for an accurate diagnosis. It is, however, possible for a nonexpert to possess this expertise, as was indicated by the surgeon in Experiment 1, whose explanation was both accurate and generated by pure forward reasoning. This distinction between generic and specific expertise is consistent with the distinction proposed by van Dijk and Kintsch (1983) between a text-base and a situation model. Groen and Patel (1988) have shown that the absence of recall differences between experts and subexperts can be explained through the use of a model in which the text base can be constructed solely on a basis of familiarity with clinical situations, where the domain-specific rules reside in the situation model.

The second phenomenon is the strong relationship between forward reasoning and diagnostic accuracy on the one hand, and between breakdown of forward reasoning and inaccuracy on the other. Although the results of the experiments presented in this paper serve to extend the generality of the phenomenon in medicine beyond a single domain, they also indicate that diagnostic accuracy is not invariably accompanied by forward reasoning, even though inaccurate diagnoses appear to be invariably accompanied by a breakdown of forward reasoning. Our results indicate, however, that it is possible to be precise regarding how and why this pattern of results can break down in accurate diagnoses. The pattern is disrupted when a problem contains loose ends consisting of facts unrelated to the principal solution. In certain cases, subjects are evidently forced to resort to a form of backward reasoning to account for these loose ends.

It is tempting to speculate that this may reflect a general phenomenon that may also explain the breakdown of forward reasoning by subjects with inaccurate diagnoses. To be precise, we note that there are resemblances between the cardiology case of Patel and Groen and a traditional concept-identification task. Generating an accurate diagnosis involves recognizing a set of attributes that characterize the disorder category and making appropriate inferences from it. In this connection, it is important to note that clinical diagnoses are based on elaborate taxonomy that one can asssume to be overlearned by specialists in a domain. Thus, these attributes have many of the characteristics present in the artificial dimensional tasks studied in the 1960s and 1970s, as opposed to the much more unstable characteristics that are a focus of contemporary research on concept learning (e.g., Neisser, 1987). As a result, it is legitimate to make a distinction analogous to that of Bruner, Goodnow, and Austin (1956) between conjunctive and disjunctive concepts.

The medical taxonomy is essentially organized so that all these categories are conjunctive. Thus the formation of a disjunctive concept is essentially a sign that a patient has a disorder that does not easily fit the standard taxonomy. This is often phrased in terms of a differential diagnosis or one in which a patient has a main complaint and subsidiary complaints.

The main point is that an inaccurate or incomplete diagnosis will be represented as a disjunctive concept. It is reasonable to put forward a hypothesis that it is this disjunctiveness that induces the backward reasoning, since Bruner et al.'s (1956) classical comparison of strategies used to obtain conjunctive and disjunctive concepts implies that forward reasoning will be extremely efficient for conjunctive concepts, but inadequate for disjunctive concepts. It is possible that this kind of approach could be recast in a more contemporary form by combining it with the SOAR model of Laird, Rosenbloom, and Newell (1985). In a sense, SOAR represents a prototype for expert systems of the future, since it goes far beyond current "flat" systems by positing a nontrivial chunking mechanism, a learning mechanism, and methods for making transitions between forward and backward reasoning. In particular, there is a mechanism in SOAR that switches between different problem spaces in response to "impasses" in reasoning. Essentially, the presence and absence of forward reasoning can be viewed as indicating the use of different spaces. The issue is whether it could be shown that the presence of disjunctive concepts or loose ends is a trigger for a change in problem space.

REFERENCES

ANDERSON, J. R. (1988). The place of cognitive architectures in a rational analysis. In V. Patel & G. Groen (Eds.), *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 1-10). Hillsdale, NJ: Erlbaum.

BRUNER, J. S., GOODNOW, J. L., & AUSTIN, G. A. (1956). *A study of thinking*. New York: Wiley.

CHARNESS, N. (1985). Aging and problem solving performance. In N. Charness (Ed.), *Aging and human performance* (pp. 226-260). New York: Wiley.

CLANCEY, W. (1987). Methodology for building an intelligent tutoring system. In G. Kearsley (Ed.), *Artificial intelligence and instruction:*

Applications and methods (pp. 193-227). Reading, MA: Addison-Wesley.

COUGHLIN, L. D., & PATEL, V. L. (1987). Processing of critical information by physicians and medical students. *Journal of Medical Education*, **62**, 818-828.

FREDERIKSEN, C. H. (1975). Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology*, **7**, 371-458.

GREENO, J. G., & SIMON, H. A. (1988). Problem solving and reasoning. In R. C. Atkinson, R. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' Handbook of experimental psychology* (2nd ed.). New York: Wiley.

GROEN, G. J., & PATEL, V. L. (1988). The relationship between comprehension and reasoning in medical expertise. In M. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 287-310). Hillsdale, NJ: Erlbaum.

HAYES-ROTH, R., WATERMAN, D. A., & LENAT, D. B. (Eds.) (1983). *Building expert systems*. Don Mills, ON: Addison-Wesley.

KINTSCH, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.

LAIRD, J. E., ROSENBLOOM, P. S., & NEWELL, A. (1985). *Chunking in SOAR: The anatomy of a general learning mechanism* (Tech. Rep. CME-CS-85-154). Pittsburgh, PA: Carnegie Mellon University, Department of Computer Science.

LARKIN, J. H., MCDERMOTT, J., SIMON, D. P., & SIMON, H. A. (1980). Expert and novice performances in solving physics problems. *Science*, **208**, 1335-1342.

NEISER, U. (Ed.) (1987). *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge, England: Cambridge University Press.

PATEL, V. L., AROCHA, J. F., & GROEN, G. J. (1986). Strategy selection and degree of expertise in medical reasoning. In C. Clifton (Ed.), *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* (pp. 780-791). Hillsdale, NJ: Erlbaum.

PATEL, V. L., AROCHA, J. F., & GROEN, G. J. (1987). Domain specificity and knowledge utilization in diagnostic explanation. In E. Hunt (Ed.), *Proceedings of the Ninth Annual Conference of the Cognitive Science Society* (pp. 195-202). Hillsdale, NJ: Erlbaum.

PATEL, V. L., & GROEN, G. J. (1986). Knowledge-based solution strategies in medical reasoning. *Cognitive Science*, **10**, 91-116.

PATEL, V. L., & GROEN, G. J. (in press). The generality of medical expertise: A critical look. In A. Ericsson & J. Smith (Eds.), *Study of expertise: Prospects and limits*. Hillsdale, NJ: Erlbaum.

PATEL, V. L., GROEN, G. J., & FREDERIKSEN, C. H. (1986). Differences between students and physicians in memory for clinical cases. *Medical Education*, **20**, 3-9.

SIMON, D. P., & SIMON, H. A. (1978). Individual differences in solving physics problems. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 325-345). Hillsdale, NJ: Erlbaum.

SOWA, J. F. (1983). *Conceptual structures: Information processing in man and machine*. Reading, MA: Addison-Wesley.

VAN DIJK, T. A., & KINTSCH, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

WINSTON, P. H. (1984). *Artificial intelligence* (2nd ed.). Reading, MA: Addison-Wesley.

APPENDIX
Propositional Representation of Pathophysiology Protocol by Surgeon 1

1. Puncture wounds on left arm leads to suspect intravenous drug abuse.

1.1.	COND: (leads to)	[1.2],[1.4] MOD: QUAL: suspect;P
1.2.	wound	ATT:puncture,LOC:arm;
1.3.	arm	ATT:left;
1.4.	suspect	ACT:1.5;
1.5.	abuse	ATT:drug;
1.6.	drug	CAT:intravenous;

2. Fever of 4-day duration with malaise suggest an infection process.

2.1.	COND: (suggest)	[2.2],[2.3],[2.4];
2.2.	fever	*DUR*day,NUM:4;
2.3.	PROX:	[2.2],[malaise];
2.4.	process	ATT:infectious;

3. Shortness of breath on exertion and aortic diastolic murmur suggest cardiac insufficiency.

3.1.	COND: (suggest)	[3.2],[3.3],[3.4];
3.2.	breath	ATT:short,TEM:(on) exertion;
3.3.	murmur	ATT:aortic,CAT:diastolic;
3.4.	insufficiency	ATT:cardiac

4. Transient visual loss on eye with flame shaped haemorrhage suggest embolic phenomena.

4.1.	COND: (suggest)	[4.2],[4.3],[4.5];
4.2.	loss	CAT:visual,*DUR*transient,LOC:eye;
4.3.	PROX:	[4.2],[4.4];
4.4.	shape	ACT:haemorrhage,ATT:flamed;
4.5.	phenomena	ATT:embolic;

Note—COND: = conditional relation. PROX: = proximity relation. TEM: = temporal relation. ATT: = attribute. LOC: = location. CAT: = category. *DUR* = duration. MOD: QUAL: = modality: qualify. ACT: = action.

(Manuscript received March 13, 1989;
revision accepted for publication October 25, 1989.)