THE JOURNAL OF THE LEARNING SCIENCES, *4*(4), 355–384

Novice Diagnostic Reasoning in Medicine: Accounting for Evidence

José F. Arocha and Vimla L. Patel
Centre for Medical Education
McGill University

This study investigated the strategies used by beginning, intermediate, and advanced medical students when solving clinical cases containing inconsistent data. Written clinical cases were used in which the initial patient data suggested a very common and prototypical disease. Subsequent data in the cases were inconsistent with this first suggestion. The cases were presented on a computer so that the subjects had the opportunity to reason aloud about the evidence. Verbal protocols were collected and analyzed. Our results showed that the beginning students generated the suggested hypothesis and either ignored or reinterpreted the inconsistent evidence to fit their initial hypotheses. Advanced novices generated several hypotheses with a similar underlying problem; this allowed them to narrow their initial hypothesis set in the face of inconsistent evidence and make fewer data reinterpretations. Intermediate novices showed a similar pattern but failed to change their hypotheses with inconsistent data. They maintained several hypotheses of a diverse nature concurrently, without evaluating them efficiently. They showed a deterioration in performance—commonly known as the intermediate effect—which we interpreted as the result of a process of knowledge reorganization that interferes with problem solving. Results are discussed in terms of change in reasoning strategies as a function of the type of learning in medical school.

Learning in medical school is notoriously hard. To a large extent, the difficulty stems from the fact that medical students have to memorize a large number of clinical signs and symptoms and associate them with a substantial number of diseases. The problem is that the signs and symptoms of many diseases frequently overlap, overwhelming students who are in the midst of

Requests for reprints should be addressed to José F. Arocha, Cognitive Studies in Medicine, Centre for Medical Education, McGill University, 1110 Pine Avenue, Montreal, Quebec, Canada H3A 1A3. E-mail: francisco@medcor.mcgill.ca

their training. Nonetheless, with experience, physicians are capable of correctly diagnosing most clinical cases they encounter in their everyday practice. To reach the level of performance of an experienced physician, however, medical students must somehow learn to evaluate overlapping signs and symptoms and to approach inconsistent information correctly. This learning may take years of experience in hospital settings, where most patients present a heterogeneous set of signs and symptoms that do not always correspond with those learned during medical education.

The general purpose of this research is to understand novice diagnostic reasoning—specifically, how medical students evaluate hypotheses generated to account for patient problems. In a previous article (Arocha, Patel, & Patel, 1993), we examined the hypothesis generation strategies employed by medical students when solving clinical cases; in particular, we dealt with the directionality of reasoning and its relation to hypothesis generation. In this article, we focus on the evaluation process. More specifically, we investigate the process of evaluating inconsistent evidence by investigating how medical students accounted for two clinical cases: one in which the initial patient data suggested a disease, and another in which subsequent data contradicted the first suggestion.

MEDICAL KNOWLEDGE AND THE COORDINATION OF HYPOTHESES AND EVIDENCE

Coordinating hypotheses and evidence is a basic component of many reasoning tasks in scientific and technological domains like medicine (Chinn & Brewer, 1993; Kuhn, 1989). An important question to investigate is the following: How do students evaluate hypotheses that they believe to account for the data at a given point in time? To answer this question, it is necessary to investigate the hypothesis evaluation strategies that subjects use in the process of problem solving (Arocha et al., 1993). A number of such strategies have been thought to play an important role in both scientific and everyday reasoning (Bruner, Goodnow, & Austin, 1956; Dunbar & Schunn, 1990; Kuhn, 1989; Patel & Groen, 1986). The strategies that are used may range from local operations that connect a single piece of data to another to global operations that connect pieces of data to a general hypothesis, such as a medical diagnosis (Arocha & Patel, 1991; Braccio, 1988; Chinn & Brewer, 1993).

That the deployment of cognitive skills depends, to a large extent, on domain knowledge, has been supported in numerous studies (Klahr & Dunbar, 1988; Kuhn, Amsel, & O'Loughlin, 1988; Kuhn, Schauble, & Garcia-Mila, 1992). Previous research in medical problem solving has shown that the medical knowledge of the expert is organized hierarchically (Braccio, 1988; Patel, Evans, & Kaufman, 1989). At the lowest level of the knowledge hierarchy, there are *observations,* which are data with no particular medical interpretation. At a higher level there are *findings,* which are observations that have clinical significance (e.g., symptoms). The next level is that of *facets,* which are categories that are not explicit enough to form diagnoses but may suggest one or more diagnoses (e.g., ischemic disorder). In this sense, a facet may serve as an intermediate construct between findings and diagnoses, associating diagnoses with similar underlying pathophysiologies. At the subsequent level, there are *diagnoses,* which are classificatory concepts that subsume facets and findings and which correspond to differentiated clinical pathologies. It is useful to think of these levels as varying depending on the goals and the circumstances of particular medical problems. In this regard, a diagnosis may be a finding in another case.[1]

It has been found from previous research (Patel, Arocha, & Kaufman, 1994) that expert physicians, when solving cases with partial information, begin their problem-solving process by generating hypotheses at the facet level, whereas subexperts (i.e., physicians solving cases outside their specialty) generate specific diagnostic hypotheses from the beginning. The strategy used by the experts has the advantage that they can generate sets of closely related hypotheses before deciding on the final diagnosis of a case. These results contrast with findings from a study by Braccio (1988) showing that, when only partial information about a case is available, medical students perform cognitive operations at all levels of the hierarchy, ranging from observations to diagnoses. Whereas first- and second-year medical students perform more operations at the observation and finding levels, intermediate and final-year students perform operations at the finding, facet, and diagnosis levels. The difference between novices and experts in both of these studies indicates a change in diagnostic performance of medical students in terms of the level of the cognitive operations they apply to the clinical case as a function of their level of experience.

This change in diagnostic performance seems to correspond to a change in the nature of medical students' learning experiences. In a typical medical school, beginning students (mostly during the first 2 years of medical training) learn primarily from textbooks and lectures, but more advanced students learn primarily from direct contact with patients in the hospital settings (mostly in the last 2 years of their training). The shift from textbook-based to experience-based learning has been suggested to account for some experimental findings in diagnostic categorization. Brooks, Allen, and Norman (1991) argued that reliance on prototypes (i.e., idealized representations of diseases, which include typical finding–disease associations) is characteristic of novices, whereas reliance on whole exemplars (i.e., remembering particular patients seen in medical practice) is characteristic of experts. The

[1]This character of the medical knowledge hierarchy was brought to our attention by one anonymous reviewer.

important point for our purposes is that the different types of learning seem to be governed by different processes. The first one is governed by the use of rules linking specific clinical cues to diagnostic categories (often the most common ones) based on textbook descriptions; the second, by the use of more realistic exemplars, in which the links between findings and diagnosis are more "messy" and include both typical and atypical findings.

Given the high variability of associations between findings and diseases, large sampling through clinical experience may be necessary to "tune" the clinician's disease schema to the naturally occurring range (Feltovich, Johnson, Moller, & Swanson, 1984). A novice's disease schema may be too general or too specific to account for cases that deviate from classical textbook presentations. Obviously, this tuning is not a matter of all-or-none, but a matter of degree. It is developed gradually through increasing experience-based learning, possibly starting in medical school. Because beginning medical students are more exposed to typical clinical problems presented in textbooks, they may use different strategies for interpreting more complex clinical cases. A likely strategy is one based on rule mappings from textbook presentations to a current clinical case. In this situation, medical students may discard all information in the current case that is inconsistent with the textbook presentation. Another likely strategy is identifying all the case information and linking it to various disease categories, generating several explanations for the same case (as in differential diagnoses). Thus, when presented with inconsistent data, one would expect to find differences in the way more experienced and less experienced subjects evaluate the data.

The investigation of hypotheses evaluation may be done at various levels. Once a hypothesis has been generated to account for the case findings, one can identify several kinds of different operations that are used to evaluate the hypothesis with data. Work by Kuhn (1989), Kuhn et al. (1988), and Chinn and Brewer (1993) identified various types of responses to inconsistent data that are commonly used in scientific and everyday reasoning. Among them, we find the following: ignoring data, excluding data, reinterpreting data, reinterpreting hypothesis, modifying a hypothesis to fit data, and changing a hypothesis altogether. These are termed *coordinating operations.*

Research suggests that the process of evaluating hypotheses is not as straightforward as it may seem. Kuhn (1989), among others, found that children and even adults sometimes fail to separate their hypotheses from the evidence supporting those hypotheses. This separation is, in part, determined by the type of evidence the subjects have to coordinate. The weaker the evidence, the more difficult it is to achieve such a separation. In this sense, the difficulty of problem solving increases as the strength of the evidence decreases. This is because the process of problem solving with weak evidence seems to require an integration process in which various pieces of evidence are combined. This process then serves as a basis for providing coherence to the problem (Patel, Groen, & Arocha, 1990). Pathognomic findings in medicine—those that with no doubt indicate a given diagnosis—

are rare. Most medical patients present clinical signs and symptoms that are weak and may suggest a number of alternatives.

Evaluating hypotheses, then, seems to require the ability to integrate the various pieces of weak evidence with a view toward reaching a diagnosis. This task is not easy, even for experienced physicians (Joseph & Patel, 1990). Such an ability may be the most important distinguishing feature between novices and experts. Joseph and Patel (1990) compared low-knowledge with high-knowledge subjects in the process of evaluating a clinical case in endocrinology. The low-knowledge subjects were physicians who had general knowledge of the domain but lacked specialized knowledge; the high-knowledge subjects were specialists in endocrinology. The results showed that the most dramatic differences between the groups were in the way the subjects evaluated their hypotheses. High-knowledge subjects evaluated their hypotheses by starting with a broad class of disease, then further refining their hypotheses until they reached a final diagnosis. On the contrary, low-knowledge subjects generated alternative hypotheses concurrently even after producing the correct hypothesis, instead of attempting to evaluate those diagnoses that had already been generated.

Given the importance of data evaluation in the development of expertise, we designed this study to investigate further what medical students do with the evidence they face in a diagnostic task. The study addresses the following general questions: How are strategies and operations for evaluating hypotheses and evidence used by novice diagnosticians? Is there a change in the use of these strategies and operations when novices become more experienced in the clinical domain? Are more experienced students also different from inexperienced students in hypothesis evaluation rather than in hypothesis generation? To answer these questions, we used the "garden path" methodology, which is characterized by presenting findings suggestive of one diagnosis early in the case and findings suggestive of another diagnosis (and contrary to the first diagnosis) later in the case. In medicine, this methodology has been used by Feltovich et al. (1984) to investigate medical problem solving.

METHOD

Subjects

The subjects selected for the study were 12 second-year, third-year, and fourth-year medical students at McGill University (4 subjects at each level). At the time of testing, the second-year students (early novices) had just started an introduction to clinical medicine but had no experience in the hospitals. They had taken courses on clinical medicine in the form of lectures and small group discussions. Therefore, their knowledge of disease was principally from lectures and textbooks. Third-year students (intermediate novices) were at the beginning of clinical training in the hospitals. Their knowledge of clinical medicine received during the first 2 years of medical training had been complemented by in-hospital training under tutor supervision (the tutoring is shared by senior residents and physicians). Fourth-year students (advanced novices) had completed all of the requirements for obtaining their medical degrees (as general practitioners) and were preparing for their provincial and federal certification examinations. Thus, the learning of the three groups represented a continuum from lecture-only medical training to 2 completed years of hospital-based training, plus 2 years of lecture-based training.

Materials

Three clinical cases describing patients who presented a similar complaint (i.e., chest pain) were developed for the study. These are shown in the Appendix. Two cases, Case 1 (VP45) and Case 2 (AD45), were written such that the initial presentations suggested a typical case of cardiac pain. Subsequent information, however, presented data that were inconsistent with such a suggestion. The initially suggested diagnosis for these cases, but not the correct one, was myocardial infarction (MI). MI is the most common cardiac disease in North America and is caused by ischemia, or lack of blood to the heart, to the point of resulting in dead heart tissue. Also related to this is angina, or cardiac pain, which is also produced by lack of blood but does not necessarily result in dead heart tissue.

Case 3 (AD62) was written such that the initial complaint presented an atypical chest pain problem. Contrary to the first two cases, the initial data did not suggest any particular diagnosis and, therefore, did not contain any inconsistent data. The correct diagnosis for this case was the same as for Case 2 (AD45) and different from that of Case 1 (VP45). That is, Cases 2 and 3 presented the same underlying disease, but differed in the initial presentation of the problem. The correct diagnosis in Case 1 was viral pericarditis (VP), and in Cases 2 and 3, it was aortic dissecting aneurysm (AD). The cases were based on real clinical problems. The patient profiles were borrowed from the patient files of the internal medicine unit and cardiology units of a teaching hospital associated with the Faculty of Medicine. These cases were modified with the help of an expert collaborator who was not a subject in the study. The modifications consisted of shortening the cases, changing the beginning of the case descriptions such that they had the same type of chest pain presentation (suggesting an initial diagnosis of MI), and moving the findings suggestive of the alternative and correct diagnosis to the latter part of the cases. A schematic of the design of the initial presentation of the cases is presented in Figure 1.

The cases were presented in written form. Clinical training and practice often include the use of written case presentations that are discussed in

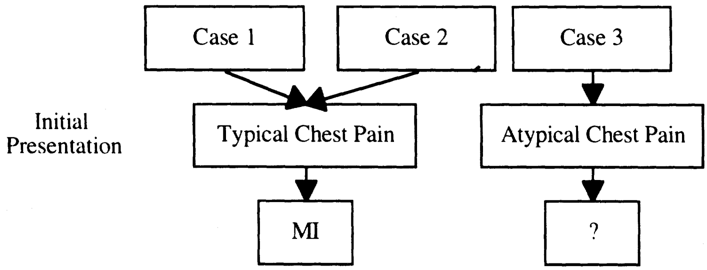

FIGURE 1 Design of the initial presentation of the cases. Cases 1 and 2 were designed such that the initial presentation (in the first segment) suggested a typical chest pain presentation most commonly associated with the diagnosis known as myocardial infarction (MI; i.e., heart attack). Case 3 was designed such that it suggested an atypical chest pain presentation and lacked inconsistent data, which can be associated with a number of different diagnostic hypotheses.

teaching sessions or medical rounds. Thus, written cases are commonly used in medical training and practice. There is also research that suggests that medical performance does not differ when written accounts or actual patients are used (Patel et al., 1989). Also, the cases were segmented into the standard format in which the clinical interview is conducted; the segments were (a) presenting complaint, (b) history of the problem, and (c) physical examination. No laboratory results were presented.

The clinical problems. The main complaint in each of the three cases was central chest pain. The quality of the pain (i.e., duration, location, severity) suggests a case of MI, but other findings presented later in the case suggest the correct diagnoses of dissection of the aorta and pericarditis, probably of a viral type. Alternative diagnoses in which chest pain is present are pulmonary embolism, pneumonia, abdominal aortic aneurysm, pancreatitis, and pleuritis. It is important to note that the cases share features that overlap different cardiac disorders such as ischemic disorders (those caused by an insufficient blood supply to an organ like the heart), pericardial diseases, or vascular diseases. It is also important to note that even though the cases suggest MI, this disorder often results in complications like cardiac arrhythmias or heart failure.

The first two cases (VP45 and AD45) describe 45-year-old men who present the complaint of severe retrosternal chest pain of 4 hr duration accompanied by sweating and faintness. The main complaint, which is presented at the beginning of the cases, is suggestive of MI because this affliction is the most common and is also very serious. The reason for giving the same presentation in Cases 1 and 2 was to make sure that the subjects started the evaluation of the cases with an hypothesis already in mind—MI. The rest of the cues, however, are suggestive of other diseases, namely, dissecting aorta in one case and acute pericarditis, most likely of a viral origin, in the other. Therefore, although the patients have the same presenting complaint and may seem to have the same disease, they actually have very different diseases. However, none of the cases is a clear-cut instance of their respective category. On the contrary, the cues are only weakly suggestive of alternative hypotheses.

To make sure that the cases were comparable in terms of the salience of the findings, we estimated the predictive value of the findings for their respective correct diagnoses by asking an expert cardiologist and the Quick Medical Reference (QMR) medical reference system (Miller, Masarie, & Myers, 1986) to assess them. QMR is a medical reference system implementation of the Internist I expert system, developed at the University of Pittsburgh, that runs on a personal computer. QMR uses a set of weights, which range from 0 to 5, to assess the strength of the clinical findings to specific diagnoses. A weight of 0 indicates that the presence of a given finding has no predictive value for the diagnosis, whereas a weight of 5 indicates that the finding is highly predictive of the diagnosis. All of the findings in the three cases had a strength value of 3 or less for the initially suggested diagnosis and for the correct diagnosis, which supports the equivalence of the cases.

The underlying problem in Case 1 is VP, which consists of an inflammation of the pericardium (i.e., the external layer of the heart), which is located within the thoracic cavity. The inflammation results in a rubbing of the pericardium against the abdominal wall, which causes pain that changes with shifts in the pericarditis position. The accumulation of fluids in the pericardial sac also produces coughing.

In both cases of AD, the patients had a long history of hypertension, in which the elastic fibers of the aortic wall became fragile and were eventually destroyed by the solidification of the wall. This solidification is caused by the accumulation of deposits of fat on the wall (a long history of hypertension in Cases 2 and 3). Eventually, the wall breaks and blood dissects into the aortic wall. The dissection causes the obstruction of major arteries connected to the aorta (such as the renal arteries) and accounts for the possible anuria (no urine voided in last 4 hr), the difference in blood pressure between the right and left arms in Case 2, and the very high blood pressure in Case 3 caused by the inability of the blood does to circulate freely. Finally, the blood can dissect back into the root of the aorta, which produces aortic regurgitation and accounts for the 1/6 aortic diastolic murmur over the aortic area in Case 2.

Procedure

The subjects were tested individually in a single 90-min session. Their task was to explain each case as completely as possible. The subjects were instructed to think aloud while they read the segments. They were asked to make a tentative diagnosis after each segment. Cases were randomly assigned to each subject and presented on three consecutive cards, each containing one segment of information. The subjects then went to the next card and repeated the procedure for the second and third cards.

Analysis

The verbal protocols were transcribed and analyzed for (a) diagnostic accuracy, (b) changes of diagnostic hypotheses across segments, and (c) the utilization and mapping of cues to generate the diagnostic hypotheses. The mapping of cues was used to represent the cases and protocols in terms of propositions (van Dijk & Kintsch, 1983). From the case propositions, a reference frame for each case was constructed as a semantic network, a method which has been used in cognitive research in medicine by Patel and Groen (1986). The reference frames contained the information that is needed to account for the case findings. The reference frames served as a comparison for the assessment of individual protocols constructed from the subjects' verbal reports. The procedure for the construction of the reference frames was the same as the one for analyzing the subjects' protocols (which is described later).

Diagnostic Accuracy

Subjects were asked to produce a diagnosis at the end of each segment and at the last segment of each case; the last one, made after all of the pertinent information had been interpreted, was the final diagnosis. The categories for the diagnosis were *accurate, partially accurate,* and *inaccurate.* An accurate diagnosis corresponds to the main diagnosis of VP in Case 1 and AD in Case 2. A diagnosis was considered partially accurate whenever the underlying process or some subcomponents of the diagnosis were identified, for instance, pleuritis is a component of the diagnosis of pericarditis in Case 1, and aneurysm is a component of the diagnosis of aortic dissection. Last, an inaccurate diagnosis is any diagnosis other than VP or AD or one in which no component of any of these diseases is identified.

Protocol and Analysis

The first part of the protocol analysis consisted of representing the subjects' protocols in terms of their underlying propositions (Frederiksen, 1975; van Dijk & Kintsch, 1983). In our analysis, we used semantic categories expressed in Frederiksen's propositional analysis system, such as conditional and causal relations. Most of the other relations were attributive, locative, and temporal. Each proposition consists of a proposition number, a predicate, and an argument to the predicate. From the propositional analysis, a list of rules was developed that linked the data cues to the hypotheses, and then a representation for each subject was constructed as a semantic network. These network representations were constructed by displaying the relations from cues to hypotheses in the subjects' protocols. Consider the following example of propositional analysis:

1. He lifted a box ...	1. ACT lift (PAT: He;[patient], OBJ: box)
2. and it is sudden onset,	2. ATT: (pain, sudden)
3. so it could be either something wrong musculoskeletal	3. CAT (musculoskeletal, [pain])
4. that persisted for 2 hours	4. *DUR* (2 hours, pain)
	5. COND: (1, 3)
	6. COND: (2, 3)
5. but I think ...	
6. it's a retrosternal chest pain	7. ATT: (pain, retrosternal)
7. that goes right to the back,	8. ATT: (pain, go to back)
8. which would be probably more a cardiac pain	9. CAT (cardiac, [pain])
	10. COND: (1, 9)
	11. COND: (2, 9)
	12. COND: (7, 9)
	13. COND: (8, 9)
9. brought on by the exertion.	14. IDENT: (1; exertion)
10. OR-EXCL	15. [3], [9]

In this example, the subject maps the first two propositions (patient lifted a box; sudden onset) to two categories— musculoskeletal pain and cardiac pain. Thus, the subject generates two diagnostic hypotheses to account for both observations. In our analysis, each relation from one observation to one diagnostic hypothesis is counted as a relation. Thus, for this set of observations, we have four relations that correspond to Propositions 5, 6, 10, and 11 and two more relations that correspond to Propositions 12 and 13:

5. lifted box → musculoskeletal pain
6. sudden onset → musculoskeletal pain
10. lifted box → cardiac pain
11. sudden onset → cardiac pain
12. retrosternal pain → cardiac pain
13. pain going through back → cardiac pain

Some of these rules contained data that were inconsistent with the disease suggested at the beginning of Cases 1 and 2. When referring to these inconsistent pieces of data, these rules were further categorized into four types of operations: Data operations, which consist of either *ignoring* data or *reinterpreting* data, and hypothesis operations, which consist of *modifying* hypothesis and *changing* hypothesis. Ignoring data was coded whenever the subjects either failed to mention the inconsistent evidence or dismissed it as unimportant (for instance when there is no mention in the subject's protocol of the inconsistent piece of data). Reinterpreting data was coded whenever the subject took the data as valid but gave a different pathophysiological or clinical interpretation from the one expected (for instance, when a piece of inconsistent data was interpreted in a manner consistent with the current hypothesis). Modifying hypothesis was coded whenever the subject made modifications to the hypothesis without changing the nature of the disease (for instance, when a given piece of data is interpreted correctly but is associated with a variant of the current diagnosis rather than with the correct diagnosis). Changing hypothesis was coded whenever the subject changed the hypothesis as a response to the inconsistent evidence (for instance, when the piece of data triggers the abandoning of the current hypothesis and replaces it with another hypothesis).

RESULTS AND DISCUSSION

We first present our results concerning diagnostic accuracy, follow with the strategies used in hypothesis generation, and then present some examples of the kinds of coordinating operations used.

Diagnostic Accuracy

The number of subjects who gave inaccurate, partially accurate, and inaccurate diagnoses is presented in Table 1. These diagnoses constitute the main final diagnosis of each case, that is, the one the subjects considered to be the most likely once all of the evidence was presented.

Two early novices gave the correct diagnoses for Case 2 (AD45). The rest of the diagnoses of this group were inaccurate. The early novices were more homogeneous in their responses, giving only two diagnoses, MI or angina, to all three cases. The exception was in Case 2, in which 2 subjects were able to change their diagnosis by correctly interpreting a single finding. In general, the final interpretation of the cases given by early novices did not seem to be influenced by the data presented later in the cases. This is supported by the fact that their final diagnosis was the same as the ones they had generated at the beginning even though the data that followed were different for each case. A possible explanation is that early novices had knowledge only of

TABLE 1
Number of Subjects Who Generated Accurate, Partially Accurate, and Inaccurate Diagnoses as a Function of Clinical Experience

	Expertise Level of Novices		
Diagnostic Accuracy	*Early*	*Intermediate*	*Advanced*
Case 1			
Completely accurate	0	0	0
Partially accurate	0	2	3
Inaccurate	4	2	1
Case 2			
Completely accurate	2	0	1
Partially accurate	0	0	2
Inaccurate	2	4	1
Case 3			
Completely accurate	0	0	1
Partially accurate	0	0	1
Inaccurate	4	4	2

prototypical diseases associated with chest pain, like MI. When the early novices reached the correct diagnoses in Case 2, it was triggered by one or two findings that were directly mapped to the correct diagnosis in a forward directed manner. This diagnosis seemed to have been retrieved from memory because there is no evidence that these subjects had reasoned through the pathophysiology of the case.

All intermediate novices produced mostly inaccurate diagnoses to both cases with the exception of Case 1, for which they generated two partially accurate diagnoses. The novices generated a large range of diagnoses which included ischemic disease (angina or MI), pulmonary embolism, pleuritis, cancer, and AIDS.

Like the intermediate novices, advanced novices generated different final diagnoses for the cases. Three advanced novices gave the correct diagnosis, all of them to Case 2 (AD45). All other subjects generated inaccurate diagnoses. Overall, the advanced novices introduced more kinds of diagnostic hypotheses than the two other groups, but the diagnoses were closer in nature than those generated by the intermediate novices. All of the diagnoses made by these subjects corresponded to known diagnostic pathologies, and not only were they able to recognize the pleuritic character of the chest pain in Case 1, they were also able to recognize the important role of hypertension in the two cases of AD (Cases 2 and 3).

The fact that advanced novices introduced more diagnoses than intermediate novices and intermediate novices introduced more than early novices shows that there was a lesser sensitivity to the initial case information and a greater sensitivity to subsequent information as experience was acquired. Not only did the quantity of different diagnoses vary across groups, the quality of diagnoses varied as well: The hypothesis sets generated for Case 1

(VP45) included more pleuropulmonary hypotheses, whereas those for Case 2 (AD45) included more cardiovascular ones. This is particularly true for intermediate and advanced novices. It may be that the subjects interpreted the cases similarly at the beginning, but when they were presented with more information, the subjects were differentially affected by the subsequent information. On the other hand, it may be that they interpreted the cases differently from the start. One way to inquire further into this issue was to find out whether the subjects were able to generate one or more initial hypotheses.

Initial Hypothesis Generation

The mean number of different types of hypotheses generated at the initial data set for all three cases is presented in Table 2. A higher mean number indicates the generation of more diagnoses simultaneously. All subjects generated the suggested hypothesis of MI, which reflects the fact that this is a very common disorder of chest pain among adult males. Although most subjects considered more than one diagnosis during the initial case interpretation, the early novices considered fewer than intermediate or advanced novices for Cases 1 and 2. They generated the greater number for Case 3. Recall that the initial data in Cases 1 and 2 suggested the diagnosis of MI, whereas Case 3 did not.

Early novices generated fewer initial diagnostic hypotheses to both typical presentations (Cases 1 and 2), but they generated the highest number to Case 3. This is to be expected given that Case 3 depicted a case of an uncommon pattern of chest pain. Sometimes the novices produced the single hypothesis of MI. In the cases for which they suggested more than one diagnosis, early novices generated the hypothesis of angina pectoris, a less severe form of ischemia than MI. The intermediate novices generated about the same number of initial diagnostic hypotheses for all three cases. Some of these hypotheses were imprecise (Feltovich et al., 1984) in the sense that they did not always correspond to known pathological processes but to disease components (i.e., syndromes). Nonetheless, the subjects were able to recognize that the types of problems underlying the cases were different

TABLE 2
Mean Number of Diagnostic Hypotheses Generated at the First Segment by All Groups of Subjects

	Clinical Problems		
Group	*Case 1 (VP45)*	*Case 2 (AD45)*	*Case 3 (AD62)*
Early novices	1.75	1.50	3.25
Intermediate novices	3.25	3.00	3.00
Advanced novices	4.25	2.75	2.25

(pleuropulmonary in Case 1 and cardiovascular in Case 2). The advanced novices generated several initial diagnostic hypotheses to all three cases. Like the intermediate novices, they were able to identify the kinds of problems involved in the cases by generating general hypotheses encompassing several related diagnoses (e.g., pleuropulmonary problems).

All subjects seemed to have used variations of two basic strategies: (a) breadth-first search, a strategy similar to the ones identified by Bruner et al., (1956) as simultaneous scanning; and (b) depth-first search, a form of successive scanning. In the more common chest pain presentations, early novices used the latter strategy; more advanced novices used the former.

Initial Case Explanation and Use of Backward Inferences

The percentage of backward inferences used in Cases 1, 2, and 3 are presented in Figures 2, 3, and 4, respectively. The use of forward and backward inferences is taken as an indication of the degree of confidence the subject has in his or her proposed hypothesis. Backward inferences are used mostly when there is uncertainty about the solution of a case (i.e., when the solution is not readily "seen"). The expectation is that backward inferences should be used in cases in which the subjects are uncertain about how the diagnoses that are generated account for the case data; they were expected to be used mostly in the second and third segments of Cases 1 and 2, in which inconsistent data were introduced. The reason for this is that, if the subjects initially interpreted the case in terms of MI, it was because the early case data strongly suggested this diagnosis. Thus, there should not be much use of backward inferences during the initial case presentation. However, in Case 3

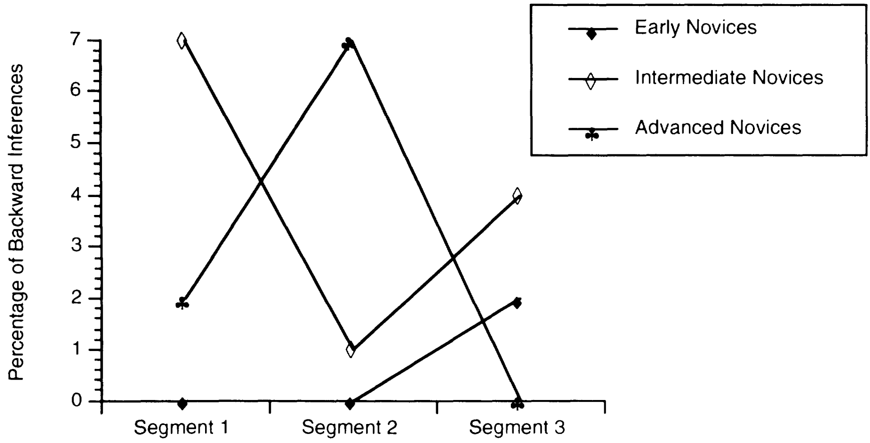

FIGURE 2 Percentage of backward inferences used by all groups of subjects at the different segments of the problem solution process for Case 1 (VP45).

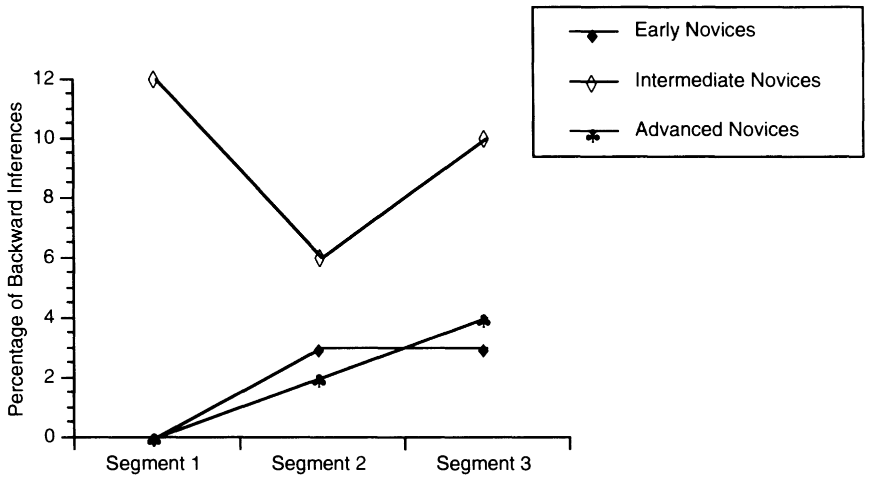

FIGURE 3 Percentage of backward inferences used by all groups of subjects at the different segments of the problem solution process for Case 2 (AD45).

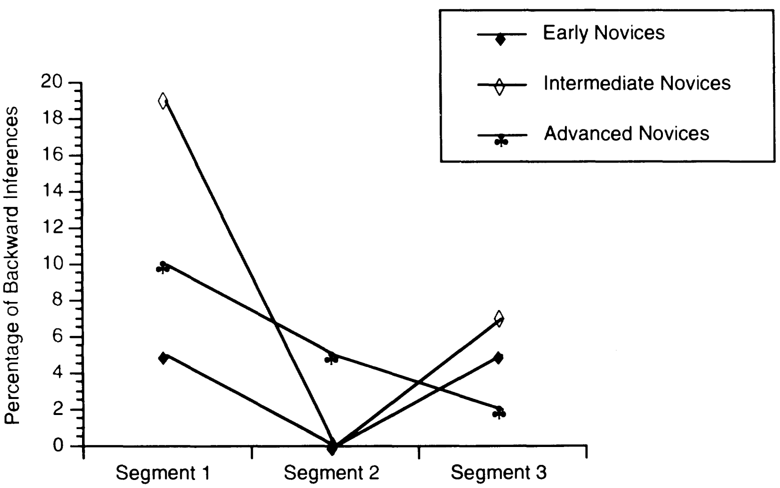

FIGURE 4 Percentage of backward inferences used by all groups of subjects at the different segments of the problem solution process for Case 3 (AD62).

(AD62), backward inferences were expected to be used more frequently in the first segment because it described a much less common chest pain presentation and, alternatively, they were expected to be used in all segments if no strong hypothesis had been generated.

The percentage of backward inferences was low compared with the percentage of forward inferences. For the first segment, the percentages ranged from 0% to 7% in Case 1, 0% to 12% in Case 2, and from 5% to 20% in

Case 3. Early novices generated no backward inferences in the first segment of Cases 1 and 2. The pattern for advanced novices is very similar to that of the early novices. The percentage of backward inferences for the intermediate novices was the highest in all cases. If, as it has already been hypothesized (Patel et al., 1990), backward inferences serve as an indication of uncertainty, the results suggest that intermediate subjects are more uncertain of their hypotheses and that this uncertainty is independent of whether the case shows a typical presentation. This greater uncertainty of intermediate subjects may be related to the finding in prior research of an "intermediate effect" in clinical reasoning and comprehension (Schmidt & Boshuizen, 1993). When performance measures of novices, intermediates, and experts are plotted, a U or inverted U pattern is shown. This is typically interpreted as a decline in intermediates' performance as compared with novices' performance. We discuss this in a later section.

Coordination Operations

The number of different coordination operations used in the cases that contained inconsistent information, that is, Cases 1 (VP45) and 2 (AD45), is presented in Table 3. This inconsistent information includes a few pieces of evidence that point to alternative diagnoses. In this section, we also illustrate the use of coordination operations with selected excerpts from the protocols. The data suggest that, when faced with contradictory evidence against an initial hypothesis (a) early novices ignored or reinterpreted the data in the problem to fit the initial hypothesis or performed hypothesis modification to include variations of the same hypothesis that seemed more consistent with the inconsistent data; (b) intermediate novices mostly used reinterpretation of data and hypothesis change operations to concurrently evaluate different hypotheses that accounted for different sets of data; and (c) advanced novices, because they initially generated multiple hypotheses, gave several reinterpretations to a given piece of data to make it fit one of their preferred hypotheses and mostly made use of hypothesis change. Recall that hypothesis modification and hypothesis change were defined differently. The former refers to minor modifications of a hypothesis while keeping it essentially the

TABLE 3
Use of Hypothesis and Data Coordination Operations for Dealing With Inconsistent Data in Cases 1 and 2

	Data Operations		*Hypothesis Operations*	
Group	*Ignore*	*Reinterpret*	*Modify*	*Change*
Early novices	18	2	1	3
Intermediate novices	14	8	2	0
Advanced novices	5	3	4	12

same (e.g., variations of a disorder); the latter refers to shifting from one hypothesis to a different one (e.g., involving a different body system, organ, or diagnosis).

The pattern in the data suggests that there was a tendency for the early and the intermediate subjects to ignore the data that were inconsistent with the initially suggested diagnosis and that there was a tendency for the advanced novices to change their hypotheses. Early novices ignored more pieces of inconsistent information than did the other two groups of subjects, followed by the intermediate subjects and then by the advanced subjects. One can argue that more novice subjects do not have the knowledge to interpret the evidence, and therefore, they ignore it. As their knowledge of clinical medicine increases, medical students become more capable of interpreting more of the evidence. However, if this is the result of a monotonic increase in knowledge, then we should be able to also see an increase in hypothesis change by the intermediate subjects as opposed to the early novices. However, this is not the case: No inconsistent finding led the intermediate subjects to change their diagnosis. Furthermore, the intermediate subjects maintained more than one hypothesis concurrently and failed to see the inconsistencies between the hypotheses and the data. One can argue with Feltovich et al. (1984), that the tuning of the diagnostic process, which is provided by extensive direct experience with patients, starts to affect performance at some point beyond the time the students have begun their experience-based learning. However, it may take extensive exposure to real-life problems to make this tuning effective. It may also be possible that experience-based learning has some detrimental effect on diagnostic problem solving as evidenced by the poor performance of the intermediate novices.

As an illustration of the particular responses that the subjects gave to the inconsistent data, we shall present a few excerpts from selected protocols. In Case 1, the finding of cough suggests the presence of some inflammatory process that is not entirely consistent with an ischemic event and that may point to pleuritis, pericarditis, or pneumonia. This piece of data was ignored by all subjects except one advanced novice who suggested its association with pulmonary problems, more specifically, pulmonary embolism. A second piece of inconsistent evidence was presented in Segment 2. This was the presence of a viral infection, which suggests the possibility of pleuritis or pericarditis. One early novice changed her diagnosis to pneumonia, two intermediate novices also changed their diagnoses but reevaluated the data in terms of the crushing chest pain and concluded that the viral infection was possibly a cold. All advanced novices changed their diagnoses to pneumonia, miascitis, or pulmonary embolism. Although at this point there was little indication in the data for the correct diagnosis, a proper interpretation of the case should alert the subjects to the possibility of some pleural involvement.

The following excerpts illustrate the use of some of these operations during the solution to Case 2 (VP45). In these excerpts, the subjects interpret a piece of data that suggests the presence of some respiratory problem in a patient who shows signs of MI. First, we present an excerpt from an early novice that illustrates the use of hypothesis modification. This early novice has generated the diagnosis of MI and is considering the possible complication of left heart failure:

> The cough is incidental in the absence of sputum or hemoptysis. I would expect sputum if there is pulmonary edema secondary to left heart failure and I'd also expect shortness of breath. ... [Therefore, it is] right heart failure.

Left heart failure is the most common type of heart failure associated with MI (found in about 50% of MI cases); right heart failure is a much less common type. Findings for left heart involvement would provide evidence to strengthen the hypothesis of MI. Because the absence of sputum and hemoptysis is evidence against left heart involvement, the student rejects the significance of cough and generates the hypothesis of right heart failure. He reasons that, because the cough is not accompanied by sputum and hemoptysis and the patient does not have shortness of breath, the heart failure is not of the left side but of the right side, instead of generating another hypothesis that may account for the pattern of cues. This way, the subject maintains his interpretation of the data coherent with his initial hypothesis of MI. We now look at an excerpt from an advanced novice, who, when faced with the same data, concludes a completely different hypothesis from the one that she has initially generated:

> He had no shortness of breath. ... There was a mild cough, but no sputum or hemoptysis. ... OK, ... so that could be pulmonary embolism or something like that.

In this case, the subject changes her initial hypothesis to accommodate the evidence (or the lack of it) for MI and generates the alternative hypothesis of pulmonary embolism. Thus, in the context of this case, the advanced subject was able to change her initial hypothesis for another that seemed to fit better the information in the case, even though the inconsistent piece of data is weakly suggestive of other diseases.

There were three pieces of inconsistent information in Case 2. The first piece of evidence was the finding of headache, which suggests the presence of vascular abnormality (in the context of the patient). The two advanced novices who interpreted the finding of headache generated the hypothesis of hypertension. It is important to note that hypertension is a risk factor for MI and that the presence of hypertension (indicated by the finding of headache) is therefore consistent with it. This was noticed only by two intermediate and two advanced novices. All early novices ignored this piece of information (there were no statements about it in their protocols). The intermediate novices generated possible hypotheses to account for it. The following excerpt from an intermediate novice shows a hypothesis change to account for the headache. This pattern is very similar in hypothesis of the other intermediate novice:

> And the headache, ... it's nonspecific. ... There are so many things that can cause headaches, tension headaches ... a lot of things; ... when you think of headaches, you think of migraines, a mass, a tumor, things like that and then you think of cerebrovascular accidents, inflammatory reasons ... hmm ... hypertension.

This excerpt is typical of the intermediate novices who invariably interpreted the same finding as being consistent with several alternative hypotheses. This subject was unable to narrow the space of likely diagnoses, generating instead parallel hypotheses without making any attempt to evaluate them against the data for the best fit. It seems that, in the context of this case, this subject cannot establish which hypotheses are relevant and which ones are not.

Another piece of inconsistent evidence in Case 2 is an asymmetry in the blood pressures taken on both arms. This finding, which is suggestive of aortic dissection, was ignored by two early novices. Two other early novices changed their hypotheses and generated the correct diagnosis. Three intermediate novices recognized the importance of the inconsistent finding but were unable to account for it, as is apparent in the following excerpt:

> His blood pressure is certainly very high, but I don't understand why his left arm is so low, 110/70. ... Maybe it's anatomic. ... A cause for the problem may be stenosis or something of the sort. I would make sure that the blood pressure of the left side is not wrong. Given the history, I would tend to go with the 120/200. Certainly, I want to repeat both of them.

All advanced novices changed their hypotheses when they evaluated this finding. Two advanced novices changed their diagnoses to the correct one, whereas another interpreted the finding in terms of coarctation of the aorta, a related diagnosis. The other advanced novice changed his diagnosis to aortic dissection but hypothesized that the site of the dissection was in the abdomen.

In summary, the most commonly used coordinating operation that early and intermediate subjects used was ignoring data. This may be the result of their lack of knowledge of disease classification because they know more about common diseases than they know about uncommon ones. With uncommon case presentations, early and intermediate novices may not be able to associate the findings to specific diagnoses, which produces either the pattern found in early novices of focusing on a single hypotheses while ignoring inconsistent findings or the pattern found among intermediate novices of generating different diagnoses to account for different findings. Advanced subjects may have been able to recognize nonsalient or untypical findings, but they were not be able to make a coherent representation of the cases that include findings that cannot be easily accounted for.

Up to this point in our exposition, we have presented aspects of the theory or data coordination process. In the next section, we look at some protocols in which the whole problem-solving process is presented. We present this to illustrate the general approach to the cases used by the selected students at the three levels of training. We do this in a schematic fashion due to the complexity and the length of the process.

Representation of Individual Protocols

The process by which the subjects mapped the case cues to the diagnostic hypotheses they generated is presented in this section. We first present a comparison of the solution processes given to Case 1 (VP45) by 2 subjects at different levels of training. The schematic representation of the problem-solving process generated from the protocol of an early novice is presented in Figure 5. As expected, the subject starts by generating the hypothesis of MI after being given the cues of severe central, crushing chest pain and chest pain for 4 hr. The subject adds support for this hypothesis by noting the age of the patient as being in agreement with a related affliction—coronary artery disease. A second hypothesis, which the student rejects, is rheumatic fever (streptococcal infection); this hypothesis is based on the finding of a history of 4 hr of chest pain. The negative findings in the case are used to reject pulmonary edema secondary to left ventricular failure, which may result from MI. Last, the subject concludes that the cough is irrelevant to the patient's main condition. The subject bases her conclusion on the negative findings of no hemoptysis and no sputum. Thus, up to this point, this early novice has generated a single diagnosis—MI—accompanied by rejecting left-ventricular failure. The rest of the hypotheses she generates are negative ones, that is, they are ruled out by the evidence.

The second segment introduces the finding of a history of upper respiratory infection with sore throat, runny nose, and generalized myalgia, which the subject correctly interprets as a viral infection. However, she uses this finding to conclude that it is not a streptococcal infection, which is related to the hypothesis of rheumatic fever; this finding was introduced and rejected in the first segment (streptococcal infection precedes the appearance of rheumatic fever). The subject then concluded that the infection was unrelated to the chest pain. Thus, the subject at this segment rules out only the relevance of the infection while maintaining the diagnosis of MI. The third segment introduces seven new findings of which the subject selects three for comment. These are a temperature of 38° C, chest clear to percussion and auscultation, and no splenomegaly. The same strategy she used in the previ-

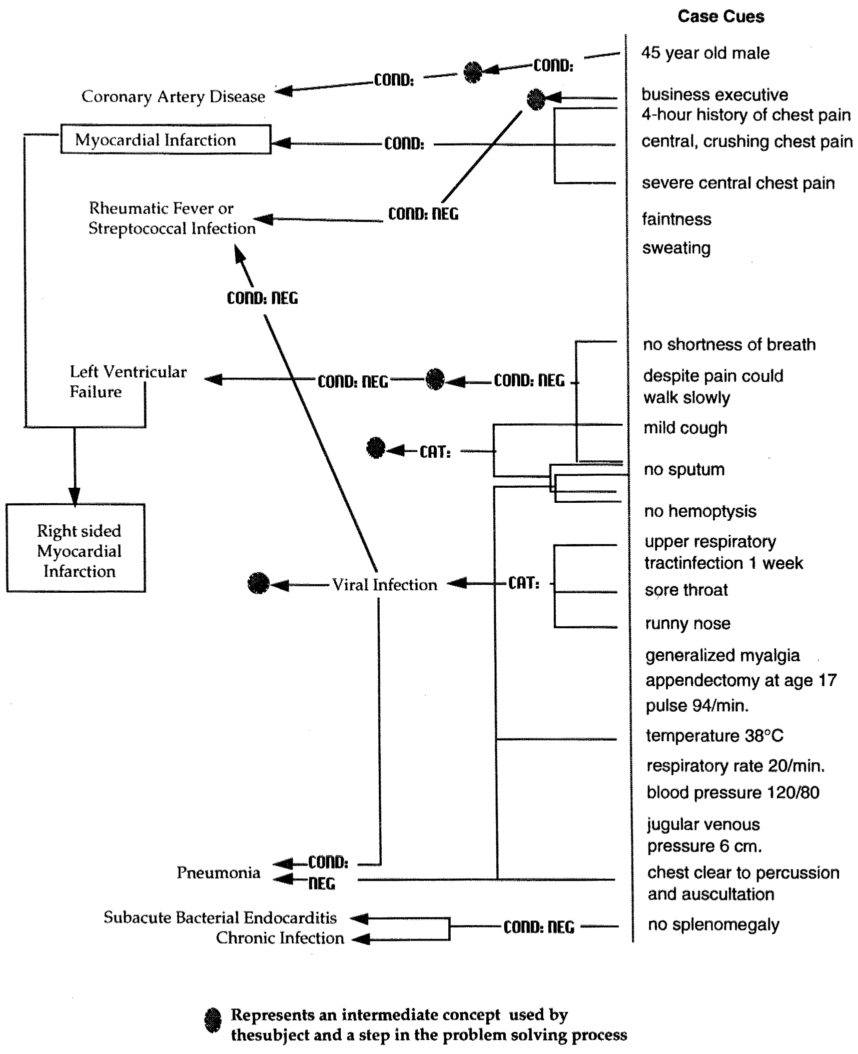

FIGURE 5 Schematic representation of the problem-solving process of Case 1 (VP45) by an early novice. COND = conditional relation; NEG = negation; CAT = category.

ous segment is used in this one: She uses the findings to rule out other diagnoses. This time pneumonia and subacute bacterial endocarditis are ruled out: Pneumonia implies a congested chest, and subacute bacterial endocarditis implies an enlarged spleen. In summary, the subject generated the initial diagnosis of MI and maintained it even in the presence of evidence that is not completely consistent. Further information that suggested other diagnoses served to modify the initial hypothesis without causing the subject to reject or question it. Other information is either acknowledged but interpreted as unrelated to the episode of chest pain or ignored.

Now we consider the explanation given by an intermediate novice, which is presented in Figure 6. The subject starts with the hypothesis of MI based on the presentation of severe central, crushing chest pain, which he recognizes as a typical presentation of such a disease. Having made this hypothesis, the subject searches for explanations for the chest pain and for the patient's current condition. In contrast to the early novice, this subject attempts to explain the findings by describing the process that is likely to cause them. Then he provides an explanation of a typical patient with MI that fits the picture of this patient (45-year-old male business executive). Note that the subject adds a number of default findings (findings not present in the case but that form part of what one would expect of a patient with MI) that he assumes are true of the patient but for which he has no evidence, such as obesity, smoking history, atherosclerosis, or increased cholesterol. These findings, one may interpret, are part of the subject's schema for MI.

The finding of cough, which was ignored by the early novice, was noticed by the intermediate novice. This he interprets as not being consistent with the diagnosis of MI. This finding causes the subject to generate a second diagnostic hypothesis of pneumonia, which is maintained parallel to the first one. This hypothesis is discarded because of the absence of shortness of breath, sputum, and hemoptysis. Then he rules out the possibility of muscular pain because this type of pain is not consistent with sweating and faintness. Another diagnosis he mentions is hyperthyroidism, which goes along with sweating and chest pain. Thus, this subject introduces four different diagnostic hypotheses, one of which he rules out (muscular pain) and one (hyperthyroidism) for which the patient does not present any evidence.

In the second segment, the subject introduces the diagnosis of bacterial pneumonia based on the finding of upper respiratory infection. This hypothesis is generated despite the absence of respiratory symptoms in this case. The other diagnosis the subject maintains is MI, in which case, the infection should be unrelated to the chest pain. In the third segment, the intermediate novice focuses on the finding of jugular venous distension, which he correctly interprets as being compatible with MI of the right side. However, the findings that are contrary to such a diagnosis—chest clear to percussion and auscultation as well as respiratory rate of 20/min—are ignored, and he concludes the diagnosis of MI.

This subject's case representation has similarities to and differences from the case the early novice examined previously. One identifiable difference is that the intermediate subject used a much more backward reasoning. However, similar to the early novice, the intermediate novice used information to rule out the alternative diagnoses while ignoring information that was inconsistent with the initial hypothesis. A second distinction is that the intermediate novice introduced a larger number of explanatory hypotheses. Whereas the early novice ruled out diagnoses and did not maintain any hypothesis

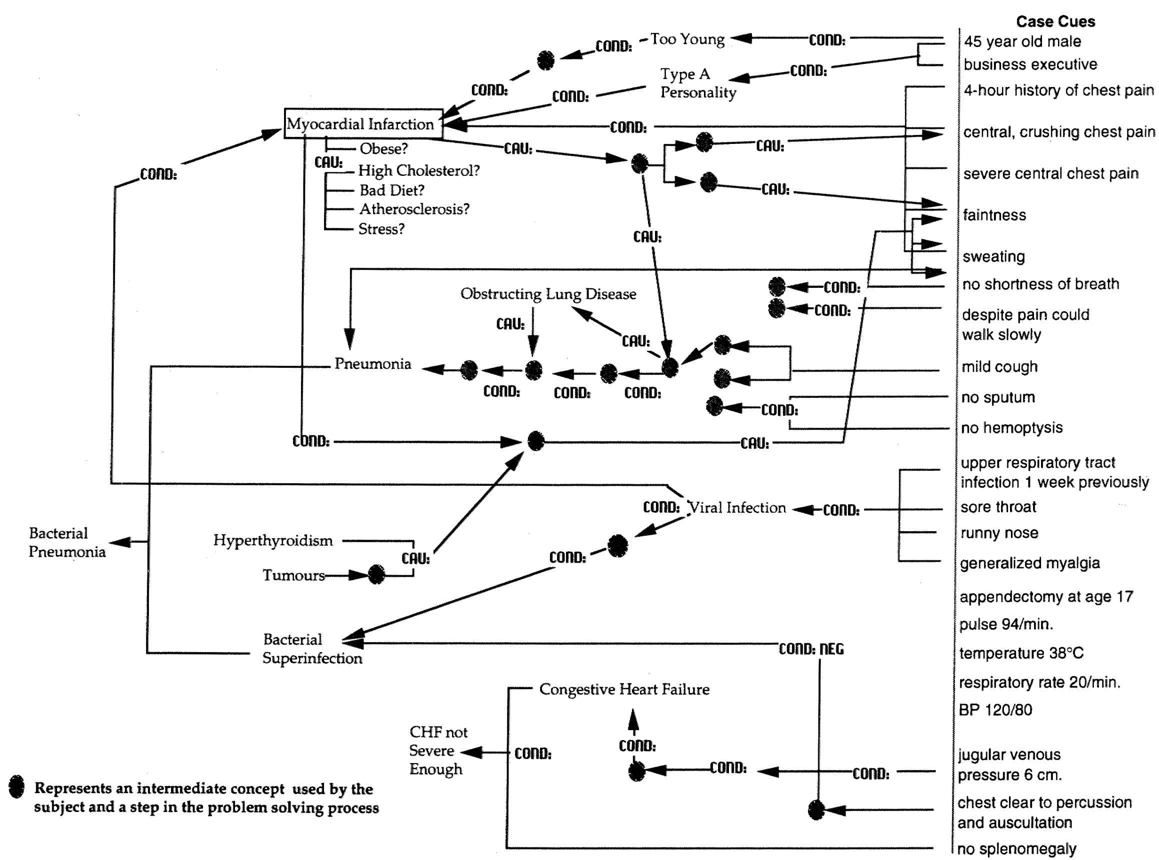

FIGURE 6 Schematic representation of the problem-solving process of Case 1 (VP45) by an intermediate novice. COND = conditional relation; CAU = causal relation; NEG = negation.

other than MI, this subject generated alternative hypotheses for different findings. A third distinction is that the intermediate novice made use of most of the findings in the case, whereas the early novice did not. In terms of the similarities, both subjects ignored some findings that were perceived as inconsistent with their initial hypothesis and also generated various hypotheses to explain specific unrelated findings. A second similarity is the fragmented character of their explanations.

GENERAL DISCUSSION

In this section, we summarize the main findings of the study and discuss their implications for learning and education in medical school. We investigated the strategies used by medical students for evaluating clinical hypotheses. We used a methodology that consists of presenting to subjects a clinical case in segments. The first segment of Cases 1 and 2 described a patient with a typical presentation of MI. Subsequent information showed clinical data that was not consistent with the information presented in the first segment and that suggested a different type of problem from MI. A third case presented a patient description that did not contain inconsistent information. We found that, in Cases 1 and 2, most early novices generated the suggested hypothesis, whereas intermediate and advanced novices generated a larger pool of initial hypotheses (often including MI). The results for Case 3 showed a high number of hypotheses during the initial segment for all groups.

The level of the hypotheses in the knowledge hierarchy generated at this initial segment also differed among the groups. Whereas all early and most intermediate novices generated hypotheses at the diagnostic level, most advanced novices generated them at the facet level. Facet-level hypotheses typically describe general categories of diseases with similar underlying processes. Because advanced students generated a large pool of hypotheses during the initial data processing, when confronted with inconsistent information in subsequent segments, they could reduce their hypotheses set from the beginning of the case to a smaller hypothesis set at the end. They achieved this in a more efficient way by evaluating the facet-level hypotheses and later specifying hypotheses at the diagnostic level. Although intermediate novices also generated a large pool of hypotheses during the initial case processing, their hypotheses were found to be at both the facet and the diagnostic levels, which included cardiac, pulmonary, and cancer problems. In comparison with advanced novices, intermediate novices had more difficulty reducing the initial set of hypotheses to a few final diagnoses.

The type of coordinating operations also showed differences among the groups. Early and intermediate novices performed more data operations and they more frequently ignored or reinterpreted inconsistent data, whereas advanced novices more often changed their hypotheses to account for the data. Furthermore, advanced novices' change of hypotheses seemed to have decreased the inconsistency with the data. This conclusion is based on the result that the advanced subjects interpreted inconsistent evidence in terms of a pleuropulmonary or a cardiovascular process for Cases 1 or 2, respectively.

The fact that advanced novices generated facet-level hypotheses suggests that their memory for clinical information is organized in a somewhat similar way to the expert's memory. Their seemingly hierarchical knowledge organization allowed advanced students to approach the case evidence in a more systematic way: starting with global hypotheses that were narrowed down as new information was processed. Ericsson and Staszewski (1989) argued that, when experts solve problems, they use intermediary constructs in memory that allow them to excel in tasks in which novices most often fail. These intermediary constructs, known as *retrieval structures,* are similar to facets in that they are intermediary representations that encode clinical case data into higher level constructs that can be used to generate more specific diagnoses (Ericsson & Kintsch, 1994).

The fact that typical case presentations in the first segment of Cases 1 and 2 did not trigger the generation of the most common diagnosis may suggest that the students' categories may not be textbook-like and may instead be fuzzier and less clearly delineated than those of novices. A similar finding was reported by Feltovich et al. (1984). Although the sample in their study consisted of final-year medical students, residents, and expert physicians, the domain that was investigated, pediatric cardiology, was new to many of the subjects. The researchers found that diagnostic categories generated by the novice subjects represented the typical and ideal categories, whereas the experts' categories were fuzzier, that is, they contained more attributes that did not belong to the ideal prototype. Our results suggest that this process of increasing "fuzzification" of disease categories begins during medical training and may not be a characteristic unique to the expert.

One can hypothesize that a change in the kind of learning experiences medical students go through induces a change in the way they approach the cases and in the way they respond to clinical data. As medical students gain more experience in hospital settings, they learn that most clinical cases present a variety of signs and symptoms that may not correspond to textbook descriptions. To deal with this increase in case complexity, students generate various hypotheses at the beginning of their problem-solving process. Having several hypotheses concurrently in memory allows the subjects to predict the diagnosis even when the critical information is not yet present. However, to take advantage of generating a variety of hypotheses, the students must learn to evaluate hypotheses effectively. One way to do this is by generating and evaluating sets of related hypotheses with similar underlying processes (e.g., hypotheses at the facet level of the knowledge hierarchy). This is a strategy frequently used by expert physicians (Kushniruk, Patel, & Marley, 1994). Experts seem to organize diagnostic knowledge in terms of small subsets of logically related diseases, which allows them to focus on findings that distinguish among the hypotheses in the set. This strategy can also be seen to some extent in the advanced novices, who with some success produced hypotheses with similar underlying processes. In contrast, Kushniruk et al. (1994) also found that nonexperts tended to generate many different hypotheses that could not be discriminated on the basis of the evidence available. The performance of intermediate subjects in our study is similar to that of the nonexperts in the sense that they generated various unrelated hypotheses concurrently but failed to discriminate among them.

Implications for Learning and Education

The traditional medical curriculum was designed under the assumption that to learn to be effective, students must begin by studying and solving simple clinical problems and continue on to more complex material once these simpler cases are mastered. In keeping with this assumption, medical students first learn typical, idealized presentations given in textbooks and lectures, then they learn more complex cases. However, learning how to deal with complex cases is done in an unsystematic way by sheer exposure to patients in the hospitals. Although this exposure has been hypothesized to be crucial for the development of expert clinical reasoning (Brooks et al., 1991; Feltovich et al., 1984), the assumption of the need for increasing learning of complexity has been questioned in recent work by Feltovich, Spiro, and Coulson (1989) and work by Feltovich, Coulson, Spiro, and Dawson-Saunders (1992). They have provided evidence that simplifications during complex learning can have a detrimental effect on comprehension of complex concepts and mechanisms.

Because medical students are not typically taught strategies for diagnostic problem solving in an explicit manner, the strategies they use, therefore, are the result of their spontaneous interactions with clinical cases. This may point not only to the need to teach medical students the content of medical knowledge, but also to the need to teach the strategies to exploit that knowledge in more effective ways. For instance, explicitly generating sets of related hypotheses seems to be useful once the students have acquired some minimal knowledge of disease classification. This method would prevent them from carrying out excessive search through the hypothesis space, help them to be more systematic in their approach to diagnosis, and would increase their chances of reaching the correct solution to a problem.

The conventional curriculum in medicine hides another assumption—that learning is a matter of acquiring increasingly larger amounts of knowledge. The intermediate effect suggests that there are qualitative changes during the knowledge acquisition process; thus, intermediate subjects are faced with having to relearn in the "real world" what they have been taught in the classroom. If, as Brooks et al. (1991) argue, there are two phases of learning—rule-based learning (i.e., through textbooks and lectures) and experience-based learning (i.e., through exposure to real patients)—then there must be some form of knowledge reorganization that allows the transition to take place. For instance, how do students recognize that a typical presentation may not be so typical after all? How are they to discover that presentations that suggest one disease may in fact hide another disease? This does not seem to be a case in which students increase the quantity of information about diseases but one involving some form of qualitative conceptual change.

If a qualitative change is involved, how does the transition from rule-based learning to experience-based learning occur? Although we do not have an answer to this question, what seems to be the case is that it is not a simple process of incremental learning of new knowledge, rather it is a process that involves some kind of reorganization, relinking of connections, and the creation of new links in the knowledge network. The investigation of this transition process may lead to answers to why there is a decrease in performance when more knowledge in accrued. But, to give answers to these issues requires the combined action of a relevant theory of conceptual and strategic change and more experimental research. We have started to take steps in that direction (Arocha & Patel, 1995).

ACKNOWLEDGMENTS

Portions of this article were presented at the annual meeting of the Psychonomic Society, San Francisco, CA, 1991 and the meeting of the American Educational Research Association, Atlanta, GA, 1993. The research reported in this article was supported in part by Medical Research Council of Canada Grant MH0004 to Vimla L. Patel and by a fellowship award from the Royal Victoria Hospital Research Institute (McGill University) to José F. Arocha.

We thank Sheldon Magder, the medical consultant, and the medical students for their participation in the study. We also thank Lee Brooks, Paul Feltovich, Michael Leccisi, and two anonymous reviewers for their insightful comments and valuable suggestions on previous drafts of the article.

REFERENCES

Arocha, J. F., & Patel, V. L. (1991). Hypothesis generation and the coordination of theory and evidence in medical diagnostic reasoning. In *Proceedings of the 13th annual conference of the Cognitive Science Society* (pp. 623–628). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Arocha, J. F., & Patel, V. L. (1995). Construction-integration theory and clinical reasoning. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse processes: Essays in honor of Walter Kintsch* (pp. 359–381). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Arocha, J. F., Patel, V. L., & Patel, Y. C. (1993). Hypothesis generation and the coordination of theory and evidence in novice diagnostic reasoning. *Medical Decision Making, 13,* 198–213.

Braccio, A. (1988). *On-line analysis of novice problem solving on medicine*. Unpublished master's thesis, McGill University, Montreal, Canada.

Brooks, L. R., Allen, S. W., & Norman, G. R. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120,* 278–287.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63,* 1–49.

Dunbar, K., & Schunn, C. D. (1990). The temporal nature of scientific discovery: The roles of priming and analogy. In *Proceedings of the 12th annual conference of the Cognitive Science Society* (pp. 93–100). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ericsson, K. A., & Kintsch, W. (1994). *Long-term working memory* (ICS Tech. Rep. No. 94–01). Boulder: University of Colorado, Institute of Cognitive Science.

Ericsson, K. A., & Staszewski, J. J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235–267). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Feltovich, P. J., Coulson, R. L., Spiro, R. J., & Dawson-Saunders, B. K. (1992). Knowledge application and transfer for complex tasks in ill-structured domains: Implications for instruction and testing in biomedicine. In D. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice* (pp. 213–254). New York: Springer-Verlag.

Feltovich, P. J., Johnson, P. E., Moller, J. H., & Swanson, D. B. (1984). LCS: The role and development of medical knowledge in diagnostic expertise. In W. J. Clancey & E. H. Shortliffe (Eds.), *Readings in medical artificial intelligence* (pp. 275–319). Reading, MA: Addison-Wesley.

Feltovich, P. J., Spiro, R., & Coulson, R. L. (1989). The nature of conceptual understanding in biomedicine: The deep structure of complex ideas and the development of misconceptions. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 113–172). Cambridge, MA: MIT Press.

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

Joseph, G. M., & Patel, V. L. (1990). Domain knowledge and hypothesis generation in diagnostic reasoning. *Medical Decision Making, 10,* 31–46.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12,* 1–55.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96,* 674–689.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific skills*. San Diego, CA: Academic.

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction, 9,* 285–327.

Kushniruk, A. W., Patel, V. L., & Marley, A. A. J. (1994). *Small worlds and medical expertise: Implications for medical cognition and knowledge engineering*. Manuscript submitted for publication.

Miller, R. A., Masarie, F. A., & Myers, J. D. (1986). "Quick Medical Reference" for diagnostic assistance. *MD Computing, 3,* 34–48.

Patel, V. L., Arocha, J. F., & Kaufman, D. R. (1994). Diagnostic reasoning and medical expertise. *Psychology of Learning and Motivation, 31,* 187–252.

Patel, V. L., Evans, D. A., & Kaufman, D. R. (1989). Cognitive framework for doctor–patient interaction. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 253–308). Cambridge, MA: MIT Press.

Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory & Cognition, 18,* 394–406.

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On the origin of the intermediate effect in clinical case recall. *Memory & Cognition, 21*, 338–351.
van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic.

APPENDIX

Text of Case 1: Viral Pericarditis (VP45)

A 45-year-old male business executive was admitted with a 4-hr history of sweating, faintness, and severe central, crushing chest pain. He had no shortness of breath, and despite the pain, he could walk slowly to the washroom. There was a mild cough but no sputum or hemoptysis.

The week before coming to the emergency room, he had symptoms of an upper respiratory tract infection with a sore throat, running nose, and a mild generalized myalgia. His medical history included an appendectomy at age 17.

On physical examination, pulse was found to be 94/min, temperature was 38° C, and respiration was 20/min. Blood pressure was 120/80, and jugular venous pressure was 6 cm. Chest was clear to percussion and auscultation. There was no splenomegaly on abdominal examination.

Text of Case 2: Aortic Dissecting Aneurysm (AD45)

A 45-year-old male business executive was admitted with a 4-hr history of sweating, faintness, and severe central, crushing chest pain. He had no shortness of breath at rest. There was no orthopnea, cough, or hemoptysis. He also complained of a dull frontal headache and a general feeling of fatigue. He had not voided any urine during the 4 hr before admission.

His medical history included an appendectomy at age 17 and a history of hypertension for the last 5 years, but he had not returned to his physician for follow-up treatment, and he was not treated for hypertension during this period.

Physical examination showed that the patient was tense and sweaty. His pulse was found to be 140/min, and his respiration was 24/min. His temperature was 37° C. His blood pressure was 200/120 on his right arm and 110/70 on his left arm. There was a 1/6 early diastolic murmur over the aortic area. Fundoscopy was performed.

Text of Case 3: Aortic Dissecting Aneurysm (AD62)

A 62-year-old male lifted a box and felt a sudden onset of severe retrosternal chest pain going through to the back. The chest pain persisted for 2 hr. He felt a mild shortness of breath. The pain persisted at rest and with movement.

He had a long history of hypertension and also had a MI 5 years ago. He had no history of recurrent pain and no shortness of breath on exertion. He had had no previous surgery.

On physical examination, his blood pressure was 170/90; his temperature was 37° C. His pulse was 110/min, and his respiratory rate was 30/min. There were no abnormalities on physical examination.