Hypothesis Generation and the Coordination of Theory and Evidence in Novice Diagnostic Reasoning

JOSÉ F. AROCHA, PhD, VIMLA L. PATEL, PhD, YOGESH C. PATEL, MD, PhD

This study investigates hypothesis generation and evaluation in clinical problem solving by medical trainees. The study focuses on 1) directionality of reasoning and 2) use of confirmation and disconfirmation strategies in generating and evaluating hypotheses. Two clinical problems were divided into segments of information containing presenting complaint, past history, and physical examination. The initial information indicated a typical myocardial infarct but subsequent information contradicted it. The results showed that the participating students predominantly used forward reasoning and confirmation strategies. When faced with contradictory evidence: 1) second-year students ignored cues in the problem or reinterpreted them to fit the hypothesis; 2) third-year students generated concurrent hypotheses to account for different sets of data; and 3) fourth-year students generated several initial hypotheses and subsequently narrowed the hypothesis space by generating a single coherent diagnostic explanation. The results are discussed in terms of coordination of clinical evidence and its relationship to scientific reasoning. *Key words:* hypothesis generation; diagnostic reasoning; novice strategies; medical problem solving. **(Med Decis Making 1993;13:198–211)**

The study of diagnostic reasoning has a long history in medical education research,[1] where it has taken a number of different forms. First, there is the normative study of diagnostic reasoning based either on the rules of deductive logic or on the experimental knowledge and intuition of physicians. Second is the psychometric study, mostly descriptive, of the way successful physicians diagnose clinical cases. Third is the cognitive tradition, which in the study of medical reasoning has emphasized the view that diagnosticians reason like scientists do. As the scientist's task can be seen as predicting results from a theory and testing those predictions, the diagnostician's task can be seen as the reverse: finding a theory that best accounts for the results.

Given the apparent similarities between the two kinds of tasks and between the processes underlying their cognitive performances, and the possible implications of such similarities for developing a unified account of cognition, it is important to investigate aspects that are common to both the scientific discovery tasks and the medical diagnostic tasks from a single perspective. In this study, we looked at how novice diagnosticians coordinated diagnostic hypotheses and clinical case evidence.

Hypothesis Generation and Directionality of Diagnostic Reasoning

At a detailed level of analysis, hypothesis generation can be seen as the utilization of two inference types, forward reasoning and backward reasoning, sometimes referred to as data-driven and hypothesis-driven, respectively. The term "forward reasoning" is used to refer to an inference pattern that acts on a data item or on sets of data items to generate a hypothesis. Forward reasoning is present when data serve to trigger a hypothesis ($d_d \ldots d_c \rightarrow H1$). If further data trigger a second hypothesis ($d_d \ldots d_n \rightarrow H_2$), there may be competition between the hypotheses H1 and H2 or there may not be. In the latter case, the two hypotheses are kept separate, accounting for two different data sets. The pattern of forward reasoning is not broken down. Simply, different cues are linked to different hypotheses.

However, if the two hypotheses H1 and H2 compete for the same data set, a conflict is created. The conflict may be resolved by backward reasoning; for instance, by predicting the consequences of one hypothesis or both hypotheses and examining already observed data against the hypotheses or collecting new, possibly decisive, data. Backward reasoning is, then, a pattern of

Received July 16, 1991, from Cognitive Studies in Medicine, Centre for Medical Education, McGill University, Montreal, Quebec, Canada. Revision accepted for publication October 9, 1992. Presented in part at the meeting of the Cognitive Science Society, Chicago, Illinois, August 7–10, 1991. Supported by a grant from the Medical Research Council of Canada (MH0004) to Vimla Patel and by a fellowship award from the Royal Victoria Hospital Research Institute (McGill University) to José Arocha.

Address correspondence and reprint requests to Dr. Arocha: Cognitive Studies in Medicine, Centre for Medical Education, McGill University, Lady Meredith House, 1110 Peel Avenue West, Second Floor, Montreal, Québec, Canada H3A 1A3.

inference in which a hypothesis serves to generate or evaluate a data item that may or may not be currently available to the diagnostician. The direction, then, is from hypothesis to data. In this case, the hypothesis H serves to constrain the kind of data searched ($H \rightarrow d_a, d_b, d_c \ldots d_n$). In short, this inference type serves to generate a hypothesis and to try to determine whether the data conform to the hypothesis.

Research shows that forward reasoning is characteristic of successful expert performance in some domains and that it is used when the subject possesses the relevant knowledge to solve the problem. It is highly efficient and fast. Backward reasoning is used mainly when there is lack of sufficient knowledge to reach the solution from the data alone. It is slow and makes heavy demands on working memory. It is most characteristic of the novice and is frequently tied to unsuccessful performance. However, as several researchers have noted, the use of forward or backward reasoning is a function of the knowledge available to the problem solver, the problem to be solved, and the nature of the task contraints.[2–4]

Notice Performance and Directionality of Reasoning

Although forward reasoning is frequently associated with expert performance, there is evidence that it is also used by subjects who are less than experts. Recent research has provided evidence that the directional pattern of reasoning may be linked to factors related to curriculum and instruction in novices.[5] Patel, Groen, and Norman compared the diagnostic reasoning processes of students who had been trained in two different curricula. One group (the problem-based–learning group) had been trained in a problem-solving environment, which emphasizes the utilization of a hypothetico-deductive approach to medical diagnosis, and in which the acquisition of basic science is concurrent with clinical practice. The other group had been trained in a more classical medical curriculum, where most of basic science is taught independently of, and prior to, clinical training. The results showed that the problem-based–learning group relied more on backward reasoning, whereas the more traditional group relied more on the use of forward reasoning, in the explanation of patient problems.

Braccio[6] investigated medical students' problem solving in an on-line diagnostic reasoning task. The task consisted of explaining and diagnosing a clinical case presented by segments on a microcomputer. Each segment corresponded to a sentence, which presented some observation about a patient, either a sign or a symptom. The results of Braccio's study showed a developmental continuum in terms of diagnostic accuracy; that is, in the correctness of the diagnoses proposed by the students. But more importantly, there was a difference in the types of inferential operations the subjects used to solve the problem. The naïve subjects attempted to solve the problem by linking single cues to hypotheses, which allowed them to construct only a fragmentary representation of the case. The sources of knowledge used by these students combined medical knowledge from textbooks with their personal knowledge of illness. They were not able to organize their knowledge into a hierarchy. Early novices, the group next more advanced in their training, demonstrated an ability to organize their knowledge at higher levels of abstraction. They identified disease components, but also explored various hypotheses to account for different findings. However, they did not construct a coherent explanation of the case. Advanced novices were able to cluster the information in a hierarchical way, from cues to diagnosis, thus more closely resembling expert performance. However, they were not able to diagnose the case in a pure forward-directed manner; rather, they frequently used backward reasoning in the form of evaluation and elaboration and also made use of their knowledge of physiology to reach the diagnosis.

Hypothesis Generation and Evaluation in Scientific and Diagnostic Reasoning

Several attempts have been made by researchers using various approaches to investigate the reasoning processes operative in scientific discovery tasks. Langley et al.[7] have developed simulation models characterizing how scientists solve scientific problems. The matter has also been investigated by psychologists such as Dunbar and colleagues[8–10] and Kuhn,[11] who have studied lay people working on simulated scientific problems. Finally, it has been investigated by Tweeney and Hoffner[12] in their reconstruction of scientific reasoning from scientists' notes.

Two kinds of the strategies that have been widely investigated are confirming and disconfirming strategies. The former refer to the search for, or the use of evidence to conclude, a hypothesis. The latter refer to the search for, or the use of the evidence to rule out or reduce the likelihood of, a hypothesis. Although, logically, disconfirmation is more conclusive and potentially efficient,[13] psychological research has found in a variety of situations in both lay problem solving and scientific problem solving that people have a preference for confirmation strategies over disconfirmation strategies.[14]

Skillful utilization of both of these strategies requires coordination between the data and the hypotheses that are being evaluated, which, in turn, is a function of the amount and quality of knowledge possessed by the subject. Research by Kuhn[11] has found that children, more than adults, fail to differentiate the evidence from the hypothesis. Hypothesis and evidence become integrated, undifferentiated, in a single "whole."

This makes it impossible to evaluate competing hypotheses. This differentiation is, in part, determined by the knowledge available to the subjects, by a preference for positive evidence, and by the type of evidence the subjects have to coordinate. The weaker, less salient, the evidence, the more difficult it is to achieve such a separation.

It has been also found that the use of confirmation or disconfirmation strategies is linked to the number of hypotheses that are generated.[15] Consideration of a single hypothesis at a time has been called depth-first search; considering several hypotheses simultaneously has been called breadth-first search.[2,16] Freedman found that when subjects were instructed to consider a single hypothesis, they showed preference for a confirmation strategy, but when they were instructed to consider several hypotheses (i.e., breadth-first search), they tended to use a disconfirming strategy. This has been supported by research by Klahr and Dunbar[10] and Dunbar and Schunn.[9] Dunbar and Schunn investigated the problem-solving strategies of subjects who were asked to solve two problems in a scientific discovery task. The authors found that the generation of multiple hypotheses was associated with successful performance of this task. The process of hypothesis generation and evaluation they uncovered was characterized as follows: 1) subjects search for a hypothesis that accounts for known instances that are similar to the present problem; 2) the hypothesis is then tested; 3) if this test fails, then the subject searches for an alternative hypothesis in terms of an underlying mechanism, and 4) if a mechanism is primed, then the current problem is explained; and finally, 5) if no mechanismic hypothesis is found, then the subject adjusts the present hypothesis to conform to the data.

In a medical context, the process of hypothesis generation in the diagnostic process has been studied by Joseph and Patel.[17] They compared low-knowledge and high-knowledge subjects in the process of evaluating a clinical case in endocrinology. Low-knowledge subjects were physicians who had general knowledge of the domain but lacked specialized knowledge; high-knowledge subjects were specialists in endocrinology. The results showed that the most dramatic differences between high- and low-knowledge subjects were in the ways they evaluated their hypotheses. High-knowledge subjects evaluated their hypotheses by searching for confirmatory evidence that led them to refine their initial hypotheses. Low-knowledge subjects, on the contrary, generated alternative hypotheses after producing the correct hypothesis, even though the data could not discriminate among them. Joseph and Patel suggest that diagnostic reasoning can be described as a two stage process in which hypotheses that depend on general domain knowledge are generated and then the hypotheses are evaluated by using specialized domain knowledge. This suggests that although general knowledge is sufficient for generating the correct hypothesis, specialized knowledge is needed for its evaluation.

In the present study, we examined the process of problem solving by novice diagnosticians. A novice is defined here as someone who is still a medical student. The results reported here address three aspects of hypothesis generation: 1) the number of initial diagnostic hypotheses generated and their effect on problem resolution; 2) the use of confirmation and disconfirmation strategies; and 3) the utilization of forward and backward inferences. We also look at how these three aspect of hypothesis generation relate to one another.

We expected to find that novice diagnosticians, when presented with a clinical case where the initial data suggested a single hypothesis but further evidence suggested that the initial hypothesis was wrong, would use different strategies according to their levels of training. Although overall the most commonly used strategy is that of confirmation, a result consistent with other studies, there may be specific differences with various levels of training. Thus, one issue that this study attempted to explore was the extents to which the uses of confirmation and disconfirmation strategies varied with training. The second issue investigated was the extents to which forward and backward inferences changed with training and how they related to evaluation strategies.

Method

SUBJECTS

The subjects who volunteered for the study were at four levels of expertise. There were four students in each of the second, third, and fourth years of undergraduate training and one senior resident in cardiology, who served for comparison with the novices, totalling 13 subjects. The subjects were all associated with the medical school at McGill University. All the student subjects had undergraduate science degrees.

The second-year medical students had just begun, during the period of testing, a course entitled "Introduction to Clinical Medicine," which bridges basic science to clinical training and serves as a first course in disease classification. They did not have experience in the hospitals, and therefore their knowledge of disease came principally from lectures and textbooks. Their training in medical school consisted of a year and a half of basic science courses, covering physiology, biochemistry, anatomy, histology, and pharmacology. The clinical teaching of these students usually begins with the more common and typical diseases, such as diabetes, hyperthyroidism, angina, and myocardial infarction. They also have knowledge of infectious diseases.

Third-year medical students were one full year more advanced than the second-year students and had one year of clinical experience in the hospitals. They had completed all the basic science courses and their training consisted mostly of the core clinical rotations of the different aspects covered in the clinical curriculum: general medicine (cardiology, endocrinology, respirology), pediatrics, surgery, and psychiatry. Thus, their textbook knowledge had been complemented with one year of practical experience as junior clerks under the supervision of clinical tutors.

At the time of testing, the fourth-year medical students had completed their undergraduate medical degree requirements and were preparing for their provincial and federal certification examinations. Overall, they had one and a half years of biomedical sciences training and two and a half years of clinical clerkship training in the hospital setting (including junior and senior clerkships). The resident had two years of training in the cardiology residency training program.

MATERIALS

Two real-life clinical problems in cardiology were selected and modified in such a way that the relevant information that distinguished the two cases was presented later in the cases. The cases were similar in their initial presentations, but differed in subsequent information. The cases were divided into three segments: 1) presenting complaint; 2) past history; and 3) results of the physical examinations of the two patients.

Clinical Problems

The main complaint in each of the cases was central chest pain. The quality of the pain (i.e., duration, location, severity) served as a distinguishing factor between competing diagnoses. One characteristic of the cases presented is that they could be easily confused with other types of problems, given the overlap among cues. Figure 1 presents a space of possible hypotheses that might account for the chief complaint of severe central chest pain.

The text passages describe two patients who have come to the hospital suffering from severe central chest pain. Although such pain may be caused by a life-threatening disease such as myocardial infarction or dissecting aneurysm, it can also be caused by minor problems such as dyspepsia or musculoskeletal afflictions, which have less clinical significance.[18] It is customary, however, to take such complaints seriously, given the possibility that they may lead to fatalities. When facing a patient who has chest pain, a physician will generally consider the possibility of ischemic heart disease due to angina pectoris or myocardial infarction.

The nature and location of the pain, however, may

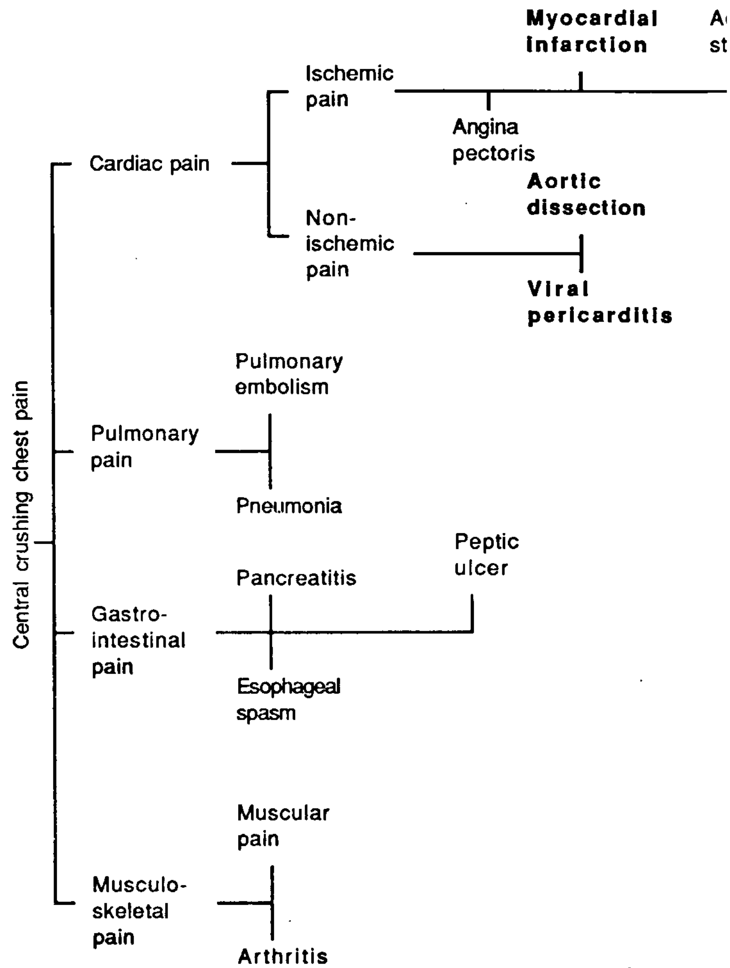

FIGURE 1. The presenting complaint of chest pain and the space of diagnostic hypotheses, including the suggested hypothesis (myocardial infarction) and the correct diagnoses (aortic dissection and acute pericarditis).

vary from patient to patient and suggest alternative diagnoses. The pain in myocardial infarction is typically a pressing or burning pain, frequently associated with difficulty in breathing, not precipitated by respiration or by coughing, and not relieved by rest. The pain may radiate to the jaws and arms, most frequently the left arm. It is typically accompanied by sweating, nausea, vomiting, or anxiety.

Case 1. Viral Pericarditis (VP45). Table 1 presents the text of case 1 (VP45). The first segment gives the presenting complaint of central crushing chest pain accompanied by sweating and faintness. The second segment introduces the history of a viral infection the patient had the previous week. This information alone is not very important, but in the presence of the chest pain, may suggest the diagnosis of pericarditis (it is common for pericarditis to be preceded by an infection about ten days before its presentation), although there are a number of diagnoses that may have the same presentation. The third segment gives the cue of fever and tachypnea, which together point more

Table 1 • Text of Case 1: Viral Pericarditis (VP45)

A 45-year-old male business executive was admitted with a four-hour 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 had symptoms of an upper respiratory tract infection with a sore throat, running nose, and a mild, generalized myalgia. Past history included an appendectomy at age 17.

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

strongly to pericarditis, leading away from myocardial infarction. This suggestion is further supported by the lack of abnormalities on examination, particularly the absence of pulmonary edema, which is a common factor of myocardial infarction.

Case 2. Aortic Dissecting Aneurysm (AD45). Table 2 presents the text of case 2 (AD45). The first segment describes central crushing chest pain suggestive of myocardial infarction. The second segment introduces a prior history of untreated hypertension that supports the original hypothesis, as this is a risk factor for myocardial infarction. However, the presence of asymmetric blood pressures in the two arms and of an aortic diastolic murmur, in the third segment, suggests that the cause is aortic dissection or aneurysm.

Procedure

The first step in the procedure involved the construction of the clinical cases in conjunction with a physician collaborator. Besides constructing the cases, the physician collaborator provided an explanation for the cues that accounts for all the important observations in each case. Each explanation was translated to a conceptual network that formalized the semantic relations underlying the disease problem.

The cases were presented on three consecutive cards, each containing one segment of information. The subjects were instructed to think out loud while they read the segments. They were asked to imagine helping a fellow student in interpreting clinical cases. Their task was to explain to this imaginary student each case as completely as possible. They were asked to make a tentative diagnosis after each segment. Then, they moved on to the next card and repeated the procedure for the second and third cards. The subjects were tested in a single session that took about 90 minutes. All were tested at the same time.

Table 2 • Text of Case 2: Aortic Dissection (AD45)

A 45-year-old male business executive was admitted with a four-hour 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 four hours before admission.

Past medical history included an appendectomy at age 17 and a history of hypertension for the last five years, but he had not returned to his physician for follow-up treatment, nor had he been treated for hypertension during this period of time.

Physical examination showed that the patient was tense and sweaty. Pulse was 140/min, respirations 24/min. The temperature was 37°C. Blood pressure was 200/120 mm Hg in the right arm and 110/70 mmHg in the left arm. There was a 1/6 early diastolic murmur over the aortic area.

Data Analysis

The data were analyzed for 1) hypotheses generated at each segment in relation to the final diagnosis, and 2) the presence in the subjects' think-aloud protocols of various problem-solving strategies including confirmation and disconfirmation and forward and backward inferences.

The protocol analysis consisted of representing the clinical cases and the subjects' protocols in terms of propositions. From these propositions, a reference frame for each case was constructed as a relational structure of propositions in the form of a semantic network, as previously used in our cognitive studies.[3,17] The reference frames so developed served as models for the assessment of individual networks constructed from the subjects' protocols. The network representations contained mostly conditional (if–then implications) and causal (cause–effect) relationships. Other relationships were mostly attributive, locative, and temporal.

These network representations were constructed by displaying the relations from cues to hypotheses in the subjects' protocols. As an illustration, consider the protocol and its propositional representation presented in tables 3a and 3b.

The relations can be forward or backward. A forward relation goes from data (i.e., a cue mentioned in the case) to a hypothesis (cue → hypothesis). A backward relation goes from a hypothesis to the data in the case (hypothesis → cue). Here, the arrows denote directionality. These relations could also be either confirming or disconfirming. A confirming relation joins the presence of a cue to the presence of a diagnostic hypothesis (cue → diagnostic hypothesis), whereas a disconfirming relation links the presence or absence of a cue to the absence of a diagnostic hypothesis (cue → no diagnostic hypothesis or no cue → no diagnostic hypothesis). In this instance, a case cue either rules out a diagnosis or makes it less likely. Notice that the arrows in this situation do not denote directionality.

Table 3a • Sample Propositional Analysis Showing How the Subjects' Protocols Were Analyzed*

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

*The basic propositions describe an action or a property, whereas complex propositions relate two or more basic propositions.

Table 3b • Example*

Proposition Number	Relation
5	Lift box → musculoskeletal pain
6	Pain sudden → musculoskeletal pain
10	Lift box → cardiac pain
11	Pain sudden → cardiac pain
12	Pain retrosternal → cardiac pain
13	Pain goes to back → cardiac pain

*In this example, the subject maps the first two propositions from Table 3a (*lift box; pain sudden*) to two categories: musculoskeletal pain (proposition 3) and cardiac pain (proposition 9). Thus, the subject generates two diagnostic hypotheses in response to both observations. In our analysis, each relation from one cue to one diagnostic hypothesis is counted as a relation. Thus, for the sample protocol, we have six relations, which correspond to propositions 5, 6, 10, 11, 12, and 13.

Results and Discussion

First, a summary of the results for all groups of subjects is presented in terms of the accuracy of the diagnosis, the use of hypothesis generation and the evaluation strategies, and the directionality of reasoning. Next, detailed analyses of three novices' protocols for the same problem are presented, together with the resident's protocol. This elucidates the processes by which these subjects dealt with the evidence for or against the diagnostic hypotheses in a case and the processes they used to solve the problems.

DIAGNOSTIC ACCURACY

Table 4 presents the final diagnoses generated by all the novices. Most of the subjects ultimately arrived at the initially suggested hypothesis, namely myocardial infarction as their main diagnoses. However, five subjects—two second-year students and three fourth-year students—reached the correct diagnostic hypothesis in case 2. No third-year student correctly diagnosed this case. No subject generated the correct solution to case 1. The resident successfully diagnosed both cases.

Table 4 • Main Final Diagnostic Hypotheses* Generated by All Novices

	Clinical Case 1 (VP45)	Clinical Case 2 (AD45)
Early novices		
2.1	MI	AD
2.2	MI	MI
2.3	MI	MI
2.4	MI	AD
Intermediate novices		
3.1	MI	MI
3.2	MI	MI
3.3	MI	Hypertension
3.4	MI	MI
Advanced novices		
4.1	Angina	Hypertension
4.2	MI	AD
4.3	MI	AD
4.4	MI	AD

*M = myocardial infarction; AD = aortic dissection. The correct or operative diagnosis for case 1 (VP45) is viral pericarditis and that for case 2 (AD45) is aortic dissection.

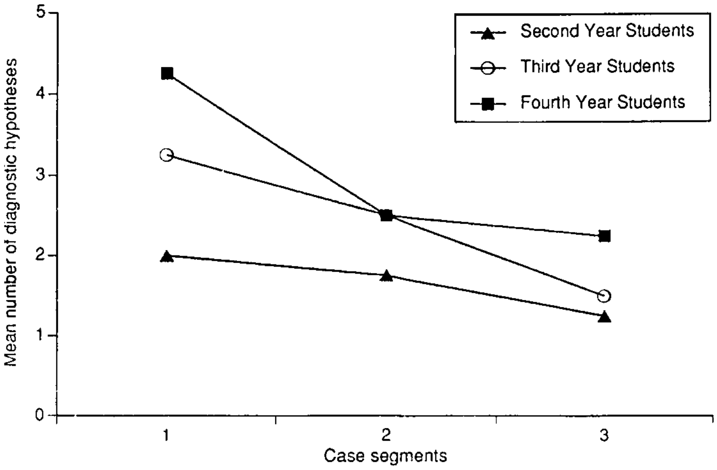

FIGURE 2. Mean numbers of diagnostic hypotheses generated at each segment of clinical case 1 (VP45) by the novice groups.

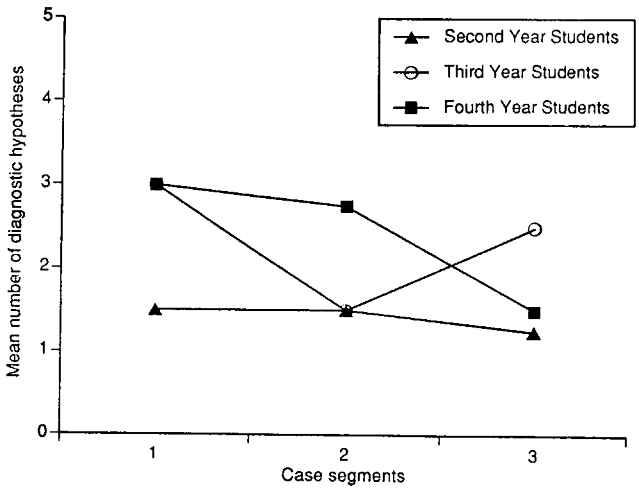

FIGURE 3. Mean numbers of diagnostic hypotheses generated at each segment of clinical case 2 (AD45) by the novice groups.

HYPOTHESIS GENERATION

The mean numbers of initial diagnostic hypotheses generated averaged over both cases varied across groups, with the second-year students generating ($\bar{X}$ = 1.5) than either the third- or the fourth-year students ($\bar{X}$ = 3.14 and $\bar{X}$ = 3.75, respectively). For comparison, the resident generated the same number of diagnostic hypotheses for both cases ($\bar{X}$ = 5).

Figures 2 and 3 present the mean numbers of diagnostic hypotheses generated or evaluated at each segment for cases 1 (VP45) and 2 (AD45), respectively. The fourth-year students generated the largest numbers of diagnostic hypotheses at the beginning of the cases. This suggests that even though most subjects concluded the same diagnosis, different groups did not generate the same numbers of hypotheses during problem solving. This difference also suggests that the novices were more susceptible to the initial presentation of the cases than were the more senior subjects. As information was further presented, all subjects exhibited a tendency to narrow the numbers of hypotheses they considered. One exception was the third-year students in case 2, where they showed an increase of diagnostic hypotheses in the last segment. This group's generation of more hypotheses at the end of case 2 may mean that they generated independent hypotheses to account for different sets of cues. The generation of hypotheses, then, seems to be related to training, as there was an increase of initial hypotheses from more novice to more advanced subjects. Fourth-year students tended to use a form of breadth-first search as they generated more competing diagnostic hypotheses to account for the same data cues than did either second- or third-year students.

EVALUATION STRATEGIES

Table 5 gives the evaluation strategies used. The most widely used evaluation strategy was confirmation. A confirmation strategy was one in which a positive cue was linked to a hypothesis as support for the

Table 5 • Mean Percentages of Confirming and Disconfirming Strategies Used by All Novices in Solving Cases 1 (VP45) and 2 (AD45)

	Confirming Strategies	Disconfirming Strategies
Case VP45		
Early novices	70	30
Intermediate novices	93	7
Advanced novices	87	13
Case AD45		
Early novices	96	4
Intermediate novices	98	2
Advanced novices	98	2

*The proportion was calculated out of the total numbers of confirming and disconfirming strategies produced by each subject.

Table 6 • Mean Percentages of Forward and Backward Inferences Used by All Novices in Solving Cases 1 (VP45) and 2 (AD45)*

	Forward Inferences	Backward Inferences
Case VP45		
Early novices	95	5
Intermediate novices	85	15
Advanced novices	98	2
Case AD45		
Early novices	89	11
Intermediate novices	70	30
Advanced novices	83	17

*The proportion was calculated out of the total numbers of forward and backward strategies produced by each subject.

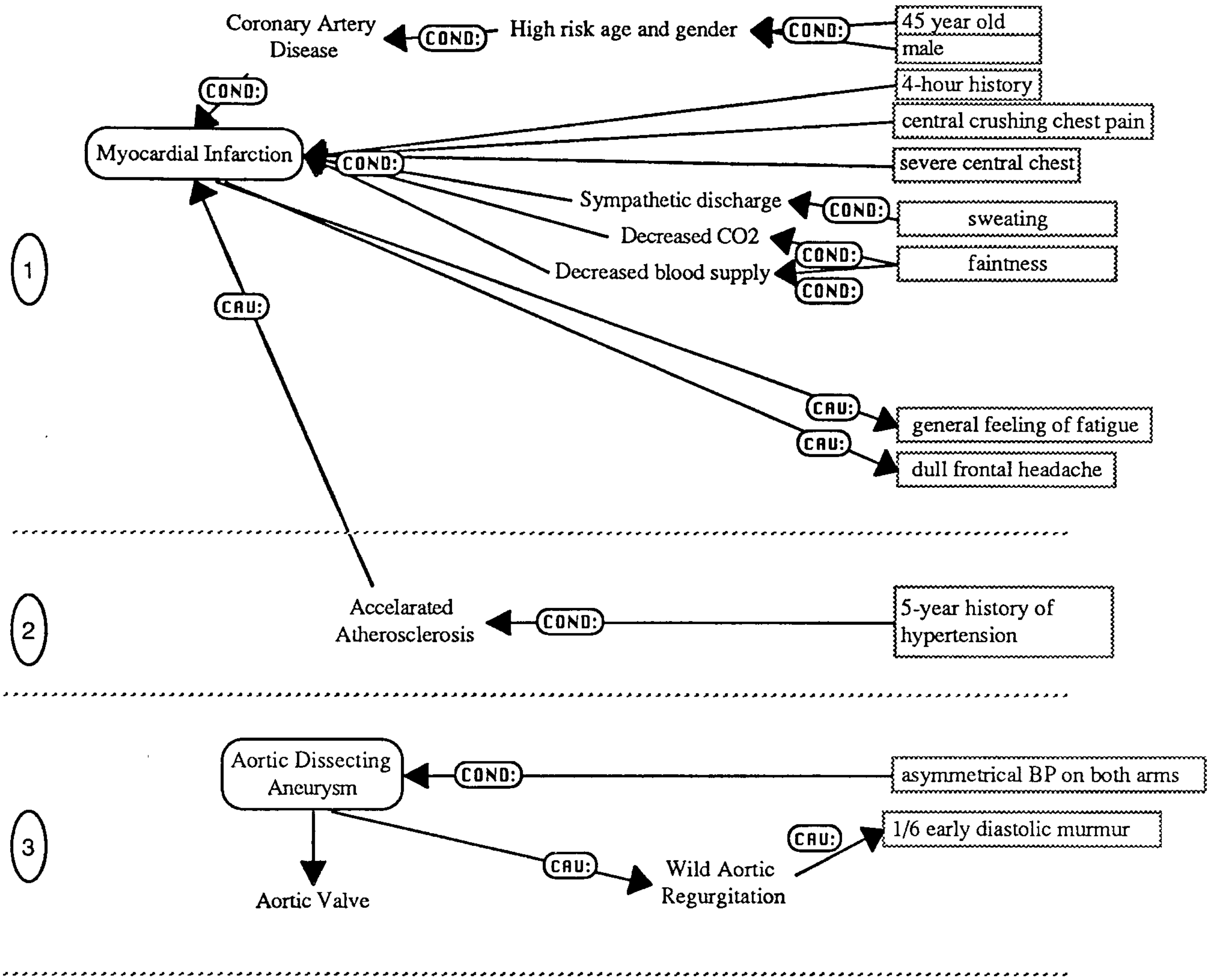

FIGURE 4. Schematic representation of the problem-solving process used by second-year student 2.1 to diagnose case 2 (AD45). COND: represents a conditional relation; CAU: represents a causal relationship; LOC: represents a locative relationship. Numbers on the left indicate segments. Boxed terms on the right represent the clinical cues in the case.

hypothesis. There were, however, differences between the two cases, with case 1 (VP45) triggering a higher percentage of disconfirmation strategies than case 2 (AD45). Recall that the first case was not solved correctly by any of the novices. The differences between the groups in terms of the use of confirming and disconfirming strategies in case 2 does not seem of major significance, as the vast majority were confirming strategies. Given the relatively higher percentage of disconfirming strategies in case 2, the second-year students seem to have generated alternative hypotheses that they later rejected. The third-year students generated larger numbers of diagnostic hypotheses, but seem to have failed to disconfirm many of them. The fourth-year students, although also generating more disconfirming strategies than the third-year students, generated a higher percentage than the second-year students. The greater use of disconfirming strategies may reflect the greater difficulty of case 1 in relation to case 2.

DIRECTIONALITY OF REASONING

Table 6 presents the proportions of forward and backward inferences generated by all novices. Mixtures of both forward and backward inferences were used. However, most reasoning inferences were forward inferences, a result consistent with prior research.[5] There was also a difference between the two cases: a higher use of forward inferences in case 1 (VP45) than in case 2 (AD45). Early and fourth-year students used similar percentages of forward and backward inferences, whereas third-year students used the highest percentage of backward inferences in both cases. The different pattern for this group may be accounted for by their greater amount of search in the hypotheses space.

DETAILED ANALYSIS OF PROTOCOLS

As the levels of diagnostic accuracy showed, two second-year students and three fourth-year students

Table 7 • Main Rules Used by Early Novice 2.1 in Solving Case 2 (AD45)

1.	IF	Crushing central chest pain & sweating & faintness
	THEN	Myocardial infarction
2.	IF	Asymmetric blood pressure between right and left arms & 1/6 aortic diastolic murmur
	THEN	Aortic dissecting aneurysm

concluded the correct hypotheses in case 2. The third-year students, although more knowledgeable than the second-year students, did not produce the correct hypothesis in this case. This section of the results compares the solution processes of several subjects in response to case 2 (AD45). We present the solution processes of one second-year student, one fourth-year student, one third-year student, and the resident.

The relational network structure of the problem-solving protocol for the second-year student, S2.1, is presented in figure 4. The problem-solving process of this subject can be described as follows: First, as soon as the subject sees the cues of severe central chest pain with faintness and sweating, he generates the single hypothesis of myocardial infarction; using a form of depth-first search. Second, he explains the mechanisms for sweating and faintness and states that the anuria and the frontal headache are consistent with this hypothesis. The attention paid by this subject to the headache and the anuria is consistent with that of other subjects. Other information inconsistent with the "classical myocardial infarction" is ignored. At the second segment, the subject comments on the relevance of a long history of hypertension as a risk factor for myocardial infarction, and suggests the possible presence of atherosclerosis in this patient, which again is consistent with myocardial infarction.

The third and last segment is where the subject changes the diagnosis to aortic dissecting aneurysm, which is triggered by the cue of asymmetric blood pressures in the right and left arms. Once the subject generates the hypotheses of aortic dissection, he then confirms it by noting the presence of 1/6 diastolic murmur over the aortic area. The subject does not account for the prior cues that suggest aortic dissection in the first and second segments, and that are somewhat inconsistent with myocardial infarction. If a specific strategy is being used by this subject, then it seems to be the retrieval from memory of rules that are linked to the two diagnoses he gives in segments 1 and 3. The rules used by the subject are presented in table 7.

In summary, this second-year student starts by generating the initially suggested hypothesis, namely, myocardial infarction, in a typical depth-first search. Then he uses further cues to confirm this hypothesis, without considering alternative hypotheses. Finally, a single cue, asymmetry of the blood pressures in the arms, triggers the correct diagnosis. Having generated the hypothesis of aortic dissection, he then accounts for the 1/6 diastolic murmur in a backward-directed manner. The fact that the subject was able to generate the correct diagnosis is evidence that he knew about other alternative hypotheses to myocardial infarction. However, he could only have generated it when confronted with strong cues of such disease.

The semantic network structure of the protocol from the third-year student 3.2 is given in figure 5. Unlike the other subjects whose detailed discussions are presented in this section, this third-year student did not reach the correct solution. As a matter of fact, no third-year student did so. This novice subject starts, as the second year student did, by generating the suggested hypothesis of myocardial infarction and angina. He then interprets other cues, not shown, in the case (i.e., sweating, faintness, no shortness of breath at rest, no orthopnea), without postulating any other hypothesis. It appears that this novice considers these cues to be consistent with myocardial infarction. But he introduces the possibility of viral illness given the cues no cough and hemoptysis. This is so because, assuming an infection, the presence of these cues would rule out a bacterial infection (a secondary, instead of a primary respiratory infection). A second hypothesis he generates is migraine from the cue dull frontal headache. This is then linked to the diagnosis of brain tumor, which, in turn, is used to account for the cue of no voided urine in four hours. Finally, the subject introduces renal artery stenosis to also account for the same cue of no voided urine in four hours.

In the second segment, the subject introduces hypertensive crisis in order to account for the cue five-year history of hypertension and hypertension not treated nor followed up. The third segment introduces two other diagnoses unrelated to the previous ones. These are subclavian steal syndrome and mitral stenosis, which are introduced to account for the cues asymmetry of the blood pressures in the arms and 1/6 diastolic murmur over the aortic area, respectively.

In summary, the third-year student, who gives the initially suggested hypothesis of myocardial infarction in response to the chief complaint of chest pain, also introduces several other hypotheses to account for different cues, maintaining each hypothesis separately from the others. Instead of generating a single hypothesis, as the second-year student did, the subject then generates various independent hypotheses, which leads to an extensive search of the problem space. The subject then seems to interpret different cues as being explained by different diagnoses. This has the effect of increasing confirmation strategies, as he does not use the generation of multiple hypotheses to rule out possibilities or to eliminate conflicting hypotheses.

Figure 6 is the schematic representation of a fourth-year student's problem-solving process. The subject starts by generating the hypotheses of myocardial in-

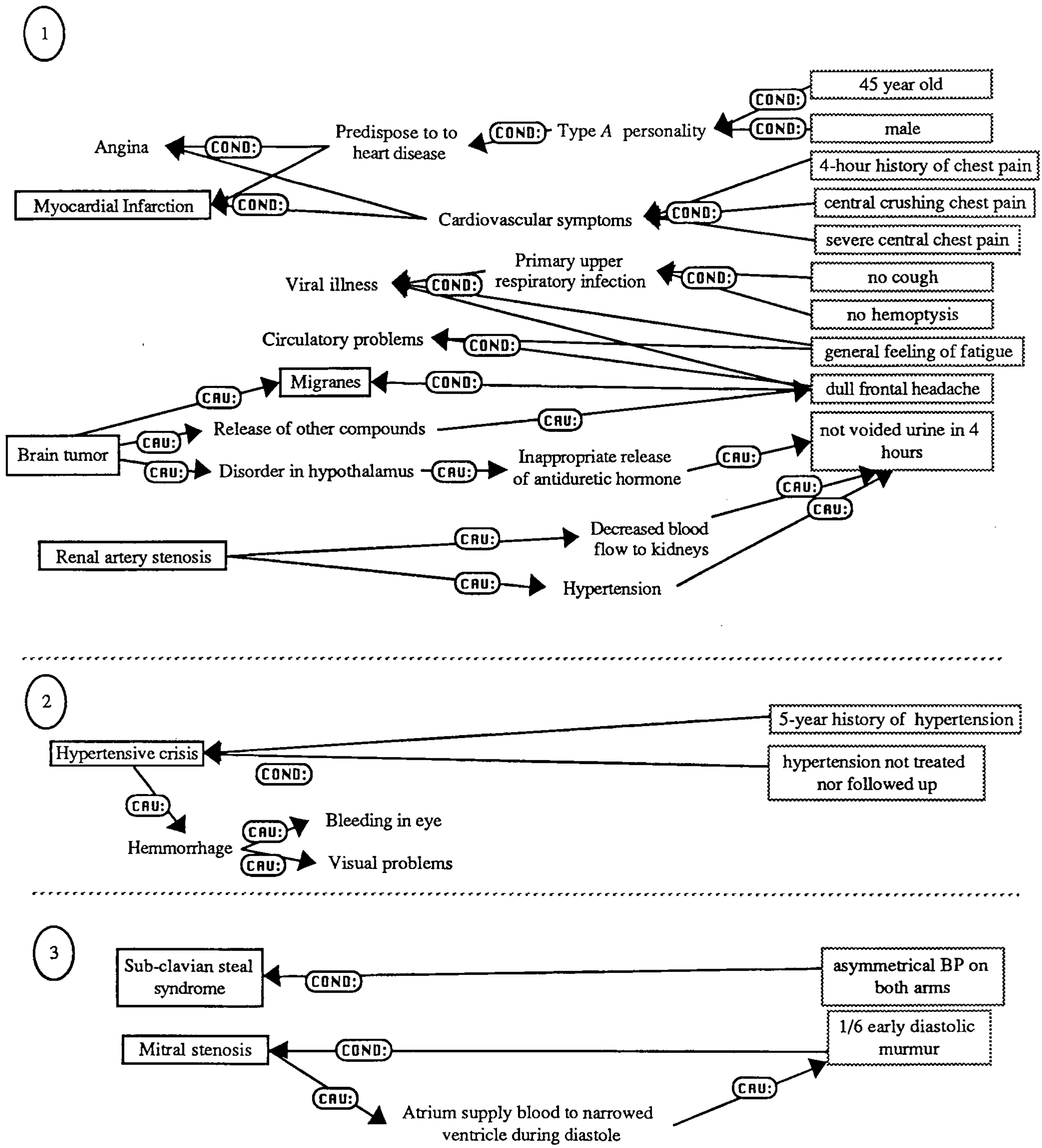

FIGURE 5. Schematic representation of the problem-solving process used by third-year student 3.2 to diagnose case 2 (AD45). COND: represents a conditional relation; CAU: represents a causal relationship; NEG: represents a negative relationship. Numbers on the left indicate segments. Boxed terms on the right represent the clinical cues in the case.

farction, aortic dissection, and hypertensive crisis as possible explanations for the central crushing chest pain accompanied by sweating and faintness, the anuria, and the frontal headache. In the second segment, given the evidence of uncontrolled hypertension, this fourth-year student introduces first the diagnosis of aortic dissection, then hypertensive crisis, and last, myocardial infarction, narrowing the space of hypotheses to cardiovascular diseases. The third segment introduces the cues of asymmetry of the blood pressures in the arms, which is evidence of a blocking of the major arteries, and 1/6 diastolic murmur over

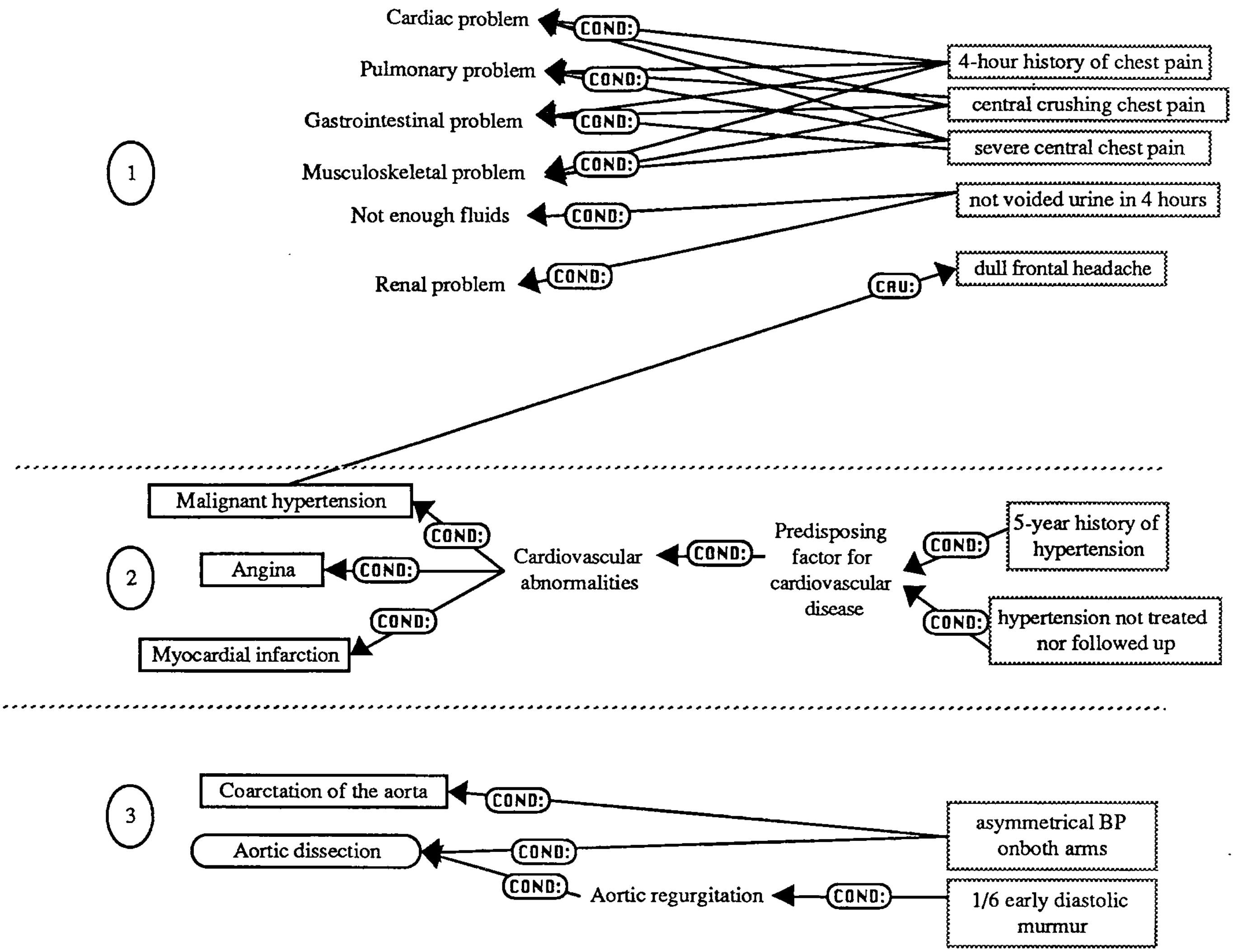

FIGURE 6. Schematic representation of the problem-solving process used by fourth-year student 4.3 to diagnose case 2 (AD45). COND: represents a conditional relation; CAU: represents a causal relationship; LOC: represents a locative relationship; NEG: represents a negative relationship. Numbers on the left indicate segments. Boxed terms on the right represent the clinical cues in the case.

the aortic area, which is an indication of aortic insufficiency, correctly explained by the subject as the result of the blood flowing back to the heart from the aneurysm. Thus, the subject concludes the diagnosis of aortic dissection.

In summary, the fourth-year student's problem solving is similar to that of the third-year student in that both generate several diagnostic hypotheses at the beginning of the case (i.e., again, as demonstrated by the previous subject, a form of breadth-first search). However, the fourth-year student generates all the hypotheses to the chief complaint and maintains them as competing hypotheses. As further information is presented, the subject reduces the problem space by focusing on cardiovascular diseases and no longer considers other types of hypotheses (e.g., pulmonary, gastrointestinal, musculoskeletal), until the final diagnosis is reached. The fourth-year subject's problem-solving process differs in this respect from that of the second-year student, in that the correct diagnosis is introduced early in the case and confirmed only at the last segment, whereas the second-year student generates the correct diagnosis only at the last segment.

Figure 7 gives the relational structure of the resident's protocol. He starts by stressing the importance of the chest pain, pointing out the long list of possibilities that can cause this affliction—aortic dissecting aneurysm, myocardial infarction, and pulmonary embolism—and giving special attention to those diseases that are life-threatening. Like the fourth-year student, the resident considers a large number of hypotheses, but the resident attempts to account for information that is ignored by the novice, i.e., lack of shortness of breath and cough, which are inconsistent with pulmonary edema, a characteristic of myocardial infarction. The resident also links the cue of dull frontal headache to the possibility of hypertensive crisis, but the presence of anuria remains unexplained. Thus, the resident attempts to make a coherent explanation of

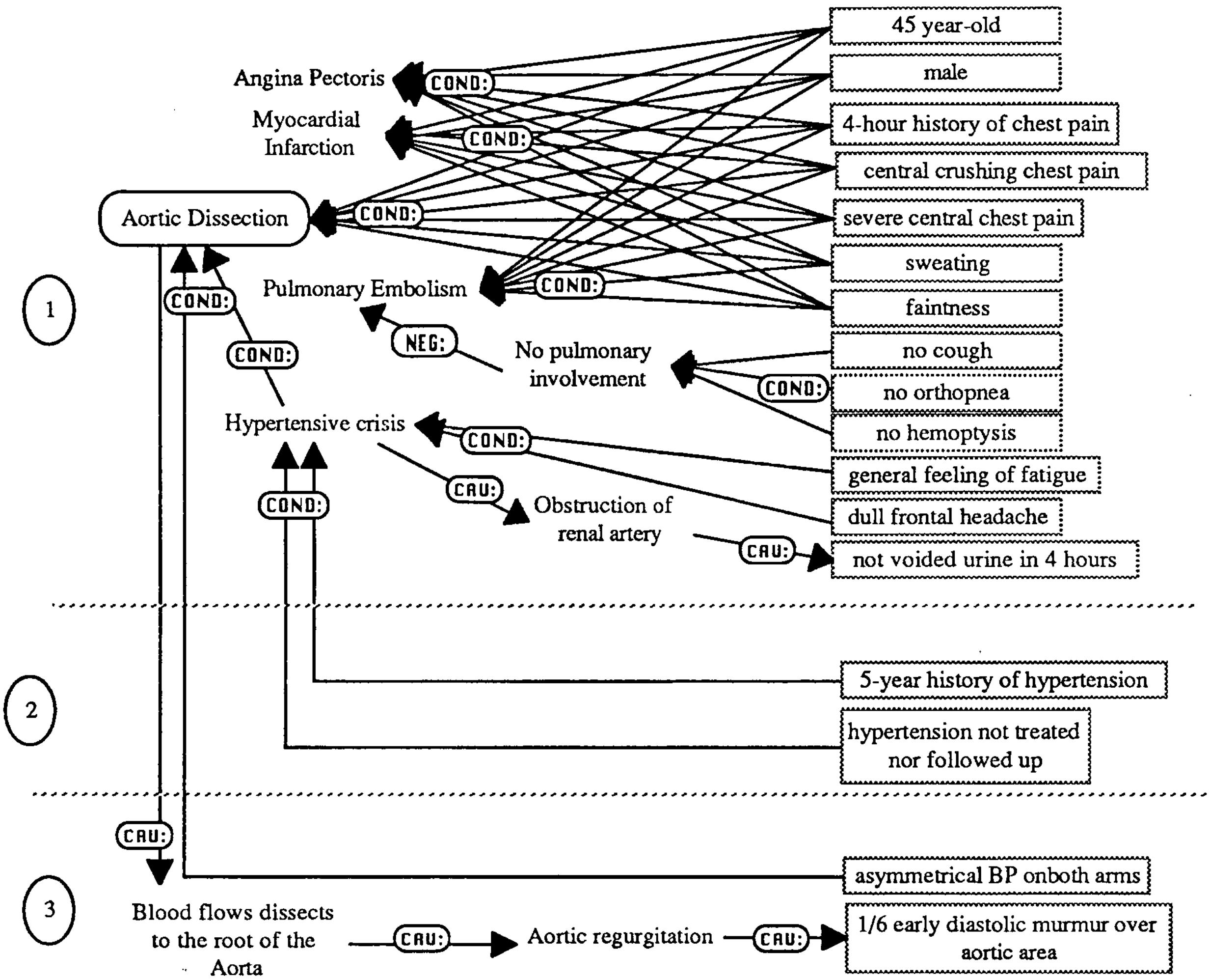

FIGURE 7. Schematic representation of the problem-solving process used by the resident to diagnose case 2 (AD45). COND: represents a conditional relation; CAU: represents a causal relationship; LOC: represents a locative relationship. Numbers on the left indicate segments. Boxed terms on the right represent the clinical cues in the case.

the case by generating hypotheses that may account for all the important symptoms in the patient. However, at this point, the resident cannot generate a hypothesis that accounts for chest pain, headache, and anuria taken together. The second segment is interpreted by the resident as supporting his hypothesis of hypertensive crisis, given the uncontrolled hypertension. The third segment introduces the cues of asymmetry of the blood pressures in the arms and 1/6 diastolic murmur over the aortic area, which suggest aortic dissecting aneurysm. The resident at this point goes back and explains the anuria as consistent with the diagnosis of aortic dissecting aneurysm, attempting to account for the "loose ends" in his explanation. This portion of the resident's protocol can be summarized in terms of rules, as presented in table 8.

In summary, the resident, like the fourth-year student, starts by generating a large number of diagnostic hypotheses, all compatible with the chief complaint,

Table 8 • Main Rules Used by the Resident in Solving Case 2 (AD45)

1.	IF	Severe central crushing chest pain & sweating and faintness & no pulmonary symptoms
	THEN	Myocardial infarction or angina or aortic dissection
2.	IF	Rule 1 & dull frontal headache & general fatigue
	THEN	Hypertensive crisis
3.	IF	Hypertensive crisis
	THEN	Anuria?
4.	IF	Asymmetric blood pressure & aortic regurgitation murmur
	THEN	Aortic dissecting aneurysm
5.	IF	Aortic dissecting aneurysm
	THEN	Anuria

using a form of breadth-first search. A difference can be seen between the problem-solving processes of the second- and the fourth-year students and that of the resident. The second-year student reaches the diagnosis at the third segment, whereas the fourth-year student reaches it at the second segment, and the resident generates the correct diagnosis at the first segment. Both the fourth-year student and the resident generate the diagnosis using the intermediate hypothesis of hypertensive problem and link it to the cues of dull headache and anuria. The resident, however, generates this hypothesis at the first segment, whereas the fourth-year subject generates it at the second segment and the second-year student at the last segment.

General Discussion

The results of this study suggest that there may be three general strategies linked to diagnostic problem-solving tasks. As their experience with clinical medicine increased, our novice subjects appeared to become more aware of the role of providing coherence to the problem solution by coordinating hypothesis and evidence. The three general strategies that were used in the problem-solving tasks are: 1) generating a major hypothesis and either reinterpreting and/or ignoring contradictory data; 2) generating concurrent hypotheses to account for different sets of data; and 3) generating various initial hypotheses and then narrowing the hypothesis space and interpreting the data in terms of a single hypothesis or a few hypotheses by imposing coherence on the data.

To account for these three major strategies, we propose the following hypothetical explanation, to be tested in later research:

1. The second-year students searched for previous descriptions (probably textbook descriptions) similar to the current cases. The most common diagnosis for cases of chest pain is myocardial infarction. Second-year students have more knowledge of myocardial infarction than of any other cause of chest pain, which is why most of these subjects chose it as their main hypothesis. To account for pieces of data that are not compatible with any typical cues, a subject needs more knowledge of disease classification than what was available to these second-year students. As a consequence, they either ignored the contradictory cues or reinterpreted them as consistent with the most common diagnosis. This may be a reflection of their inability to differentiate hypothesis and evidence, as suggested by Kuhn.[11] In some instances, however, the second-year students could generate the correct hypothesis but were not able to evaluate it. To produce an answer to the problem, then, the second-year students adhered to their initially generated hypothesis. This result is consistent with those of Kuhn,[11] Klahr and Dunbar,[10] and Joseph and Patel.[17] Klahr and Dunbar concluded that when subjects cannot think of alternative hypotheses, they keep the original hypothesis, even if they know that their hypothesis may be incorrect. Once the subjects generate viable alternative hypotheses, they may be willing to change their original hypothesis.

2. The third-year students had accumulated more knowledge of clinical classification through their experience in the hospitals. They knew that "typical" presentations do not often match real-life cases, which are "fuzzier," and, in some cases, include multiple disorders. This knowledge, however, is not yet well articulated. This may produce unnecessary search through the hypothesis space. Because their hypothesis space was larger than that of the second-year students but it was not well organized, they "got lost" and ended up producing irrelevant hypotheses, some of which may account for sets of cues. Generating several hypotheses to account for the distinct sets of data had a detrimental effect on problem solving, especially when the subjects did not possess the specific knowledge to evaluate them[17] or when they generated multiple hypotheses and used mostly a confirmation strategy.[15]

3. The fourth-year students searched for previous instances similar to the current case. This search, however, was constrained by their knowledge of disease classification. This knowledge contained information about overlapping characteristics of different diseases, but also knowledge of underlying pathophysiologic principles and distinguishing clinical patterns. This allowed them to evaluate further cues in terms of their initial hypotheses. Thus their search became more focused, which resulted in developing a more circumscribed problem space during problem solving. The fourth-year students might have learned to generate multiple hypotheses, but their apparent lack of specific knowledge about real-life patients may have impaired their problem-solving process, resulting in keeping their initial hypothesis when the alternative hypotheses failed.

4. The resident had a more extensive knowledge of disease classification and of underlying pathophysiologic principles, which allowed him to better coordinate the hypotheses generated against the case data. This evaluation process seems to be decisive in the physician's problem-solving process[17]: it resulted in a more coherent account of the case, as the subject attempted to tie "loose ends" to a major hypothesis, through the use of backward reasoning. This is also consistent with studies conducted with experts.[19] The use of simultaneous multiple hypotheses from the beginning of the case may have facilitated the resident's

problem solving, because this strategy allowed him to eliminate competing hypotheses by use of disconfirming strategies.

A final note on the limitations of the study: It is always difficult to generalize the results of studies such as this, because of the lack of randomization and the low representativity of subjects and clinical cases. However, this should not be the goal of this kind of research. The strength of it comes from the detailed analysis of performance, which may be a source of hypotheses to test in further studies and may also give us clues as to where individual differences lie. Therefore, the hypothetical explanation proposed to account for the results should be taken not as an indication of their generalizability to other samples of subjects and clinical materials, but as a lead to possible hypotheses for later research to be carried out in more tightly controlled settings. Whether these results are general or not and whether our explanations hold, then, is a matter for future research to establish.

We have attempted to provide a reasonable account of our novices' performances and to propose hypotheses for further exploration. There appear to exist many factors affecting problem-solving skills in a complex domain such as medicine. This study has attempted to investigate some of these factors, but much more research is needed to provide a more valid and general account of novice performance. Future research will have to look more closely at these issues in more controlled settings. Experimental research, together with fruitful theories of the processes underlying performance, may guide us towards making our knowledge a useful tool for improving educational practice.

The authors thank the medical students and the resident for their participation in the study. They thank David Kaufman and the two anonymous reviewers for their helpful comments on the paper.

References

1. Feinstein AR. Clinical Judgement. Baltimore: Wilkins and Williams, 1967.
2. Larkin JH, McDermott J, Simon DP, Simon, HA. Expert and novice performance in solving physics problems. Science. 1980;208:1335–42.
3. Patel VL, Groen GJ. Knowledge-based solution strategies in medical reasoning. Cogn Sci. 1986;10:91–116.
4. Voss JF. Problem solving and the educational process. *In* Glaser R, Lesgold A, eds. Handbook of Psychology and Education. Hillsdale, NJ: Lawrence Erlbaum, 1989.
5. Patel VL, Groen GJ, Norman GR. Reasoning and instruction in medical curricula. Cogn Instruct. 1993;10.
6. Braccio A. On-line analysis of novice problem solving on medicine. Unpublished MA thesis. McGill University, Montreal, Quebec, Canada, 1988.
7. Langley P, Simon H, Bradshaw GL, Zytkow JM. Scientific discovery: computational explorations of the creative process. Cambridge, MA: MIT Press, 1987.
8. Dunbar K. Scientific reasoning strategies in a simulated molecular genetics environment. Proc 11th Annual Conf Cognitive Sciences Society. Hillsdale, NJ: Lawrence Erlbaum, 1989.
9. Dunbar K, Schunn CD. The temporal nature of scientific discovery: The roles of priming and analogy. Proc 12th Annual Conf Cognitive Sciences Society. Hillsdale, NJ: Lawrence Erlbaum, 1990.
10. Klahr D, Dunbar K. Dual space search during scientific reasoning. Cogn Sci. 1988;12:1–55.
11. Kuhn D. Children and adults as intuitive scientists. Psychol Rev. 1989;96:674–89.
12. Tweeney RD, Hoffner CE. Understanding the microstructure of science: an example. Proc 9th Annual Conf Cognitive Sciences Society. Hillsdale, NJ: Lawrence Erlbaum, 1987.
13. Popper KR. Conjectures and Refutations. New York: Basic Books, 1963.
14. Wason P, Johnson-Laird PN. Psychology of Reasoning. Cambridge, MA: Harvard University Press, 1972.
15. Freedman EG. Scientific induction: effects of test strategy and multiple hypothesis. Paper presented at the 32nd Annual Meeting of the Psychonomic Society, November 1991, San Francisco, CA.
16. Nilsson NJ. Principles of Artificial Intelligence. Palo Alto, CA: Tioga, 1980.
17. Joseph G-M, Patel VL. Domain knowledge and hypothesis generation in diagnostic reasoning. Med Dec Making. 1990;10:31–46.
18. Harrison TR. Principles of Internal Medicine. 12th ed. New York: McGraw-Hill, 1991.
19. Patel VL, Groen GJ, Arocha JF. Medical expertise as a function of task difficulty. Mem Cogn. 1990;18:394–406.