12

LCS: The Role and Development of Medical Knowledge in Diagnostic Expertise

Paul J. Feltovich, Paul E. Johnson, James H. Moller, and David B. Swanson

After a visit by Herbert Simon to the University of Minnesota in 1972, Paul Johnson (then a professor of educational psychology at the Center for Research in Human Learning) and associated graduate students began applying information-processing concepts to the study of expertise and problem solving. This investigation was consistent with their view that psychology is the study of contextually dependent phenomena. That is, the psychology of human behavior is most fully understood in domains of use and practice.

Johnson then met James Moller (a professor of pediatrics) who had similar interests in problem solving within medicine and medical education, and the collaboration started. David Swanson was Johnson's graduate student and wrote a simulation program called DIAGNOSER as part of his Ph.D. dissertation. Paul Feltovich also studied with Johnson, and this chapter reports on his dissertation research, a formal psychological study. The development of DIAGNOSER and the design of Feltovich's study took place in tandem, and each contributed to the other, although the simulation was completed first. The whole group at Minnesota, over this period of time, evolved a conception of expertise in terms of the organization and

This chapter is based on the doctoral dissertation of Paul J. Feltovich, which was submitted to the graduate school of the University of Minnesota under the advisorship of Paul E. Johnson. The first version of this paper was presented at the annual meeting of the American Educational Research Association, Boston, 1980. Used with permission.

manipulation of knowledge and the adaptation of inner environment (knowledge and reasoning) to task environments.

The roots of Feltovich's study are interesting and illustrate the changing nature of psychological investigations over the past decade. The major empirical studies of clinical expertise (Elstein et al., 1978; Barrows et al., 1978) had focused on the process of clinical reasoning and found no differences between experts and novices. At the same time, psychological studies of expertise [e.g., Chase and Simon (1973)] had also found no differences at process levels (e.g., number of moves considered, depth of search). They were pointing to elements of the quality of reasoning and knowledge as the main contributors to expertise. The work by Barrows et al. in medicine also cited quality of reasoning as the only discriminator they could find. This previous work, in conjunction with the Minnesota group's view of expertise as the adaption of a knowledge base to a task environment, led Feltovich, Johnson, and Swanson to study the organization and representation of knowledge in medicine, focusing on the determinants of quality. In this pursuit they were influenced by related AI work in knowledge representation, including early writings about frames [e.g., Minsky (1975)] and collections [e.g., Bobrow and Collins (1975)].

Thus, in sharp contrast with traditional psychological studies, Feltovich and his colleagues attempted to ferret out how the structure of an individual's knowledge affects his or her problem solving. This level of analysis asks how particular hypotheses *come to mind, not just how many hypotheses are considered at once or how soon the first one is vocalized. The experiments reported here are of considerable value as scientific support for the structuring schemes that have been derived more intuitively by AI researchers. These include schemes for articulating strategies and principles in program explanations (Chapters 11 and 16) and factoring a knowledge base into "specialists" (Chapter 13). Such an analysis also provides a basis for eliciting knowledge from an expert and for teaching students (Chapter 15).*

The reported study investigates the contribution of case-related medical knowledge to clinical diagnosis. Subjects, varying in their training and clinical experience in pediatric cardiology, diagnosed four cases of congenital heart disease while thinking aloud. Each case was designed to assess a different aspect of the subjects' medical knowledge. Consistent differences in performance among diagnosticians at different levels of experience were found, and inferences were made to sources of medical knowledge responsible for performance. Recurrent sources of error were identified for the less experienced diagnosticians.

Unlike the other chapters in this volume, this chapter does not report on a working computer program. In a narrow sense, this is not a report of medical AI research. However, the contribution to AIM research is evident in the kinds of questions asked and in the form of the model of reasoning. In this respect Feltovich's work is distinguished in the depth and controlled nature of his investigation. Moreover, research that followed (Johnson et

al., 1981) made good use of the DIAGNOSER simulation model for testing and experimenting with conjectures about knowledge structures and reasoning.

The approach taken by Feltovich and colleagues in this study continues to evolve. Besides seeking generality in diverse areas such as law and physics, they are investigating the implications of their findings for the assessment of clinical competence and expertise, as well as the implications for teaching basic science for clinical problem solving.

12.1 Overview: Studies on the Nature of Knowledge and Reasoning

Knowledge influences reasoning and other cognitive skills. In recent years distinctions between knowledge and reasoning have blurred. That knowledge influences the quality and nature of reasoning that can occur has been suggested. That reasoning uses knowledge as a substrate is evident, and even the idea that reasoning constitutes a form of knowledge has been entertained.

Recent laboratory research has indicated that knowledge contributes to even the most fundamental cognitive skills. The knowledge base possessed by an individual has been shown to influence fundamental intellectual skills such as induction and analogy (Glaser and Pellegrino, 1980), basic memory mechanisms such as grouping and rehearsal (Chi, 1978), and even the functional size of short-term memory (Chi, 1976). Voss and his colleagues (Chiese et al., 1979; Spilich et al., 1979) have extended work of this sort beyond basic laboratory environments into domains of complex subject-matter learning. Within a given subject matter, high-knowledge individuals have greater recognition and recall of new material than do low-knowledge individuals, can make useful inferences from smaller amounts of partial information, and are better able to integrate new material within a coherent and interconnected framework of knowledge (organized, for example, around a common goal structure).

Reasoning itself has been shown to be highly dependent on the individual's knowledge base for the task environment in which the reasoning occurs. Subjects show dramatic improvement in testing the implications of logical inference rules (e.g., if p then q) when these are couched in terms of a familiar setting, as opposed to when the expression is stated in a more purely symbolic form (Rumelhart, 1979; Wason and Johnson-Laird, 1972). This content-constrained conception of formal reasoning is in contrast to structural developmental theories (Piaget, 1972) that claim cross-situational, content-free, and maturationally determined general reasoning skills. Yet even within these theories, evidence is emerging for the import of accumulated knowledge as a contributor to these abilities (Carey, 1973).

Artificial intelligence research has also shown an evolution from systems in which knowledge (declarative) and reasoning (procedures) were clearly separated to systems in which these components strongly interact. Early systems such as Green's QA3 (Green, 1969) and Quillian's TLC (Quillian, 1969) relied on data bases of uniformly formatted declarative knowledge and a few general-purpose reasoning algorithms for operating on these knowledge bases. These systems have given way to ones in which the separation between knowledge and reasoning components is much less distinct and in which general reasoning algorithms have considerably less status in comparison to specific (local) reasoning strategies associated with specific domains of knowledge (Norman et al., 1975; Sacerdoti, 1977; vanLehn and Brown, 1979). Reasoning is seen not so much as a general but as a task-specific skill.

The role of knowledge and its organization have been emphasized in recent work on expertise and expert/novice differences in problem solving in complex domains. The findings of groupings in expert perception of a chess board is taken as evidence that guidance in the choice of chess moves is provided by knowledge representations for configurations in the board (Chase and Simon, 1973). Similarly, Larkin (1978) has proposed a construct of "chunked procedures" for expert physics problem solvers, whereby expert categorization of a problem leads to a relatively integrated problem plan and associated "bursts" of equations applied in solution. Feltovich and colleagues have shown that differences in problem-solving processes among expert and novice physics problem solvers result both from differences in the structure of knowledge representations for problem types and from differences in memory organization among these types (Chi et al., 1981). Simon and colleagues (Hinsley et al., 1978; Paige and Simon, 1966) have shown that schemata, which are knowledge structures representing problem types, strongly influence the nature of the problem-solving process in algebra.

In light of developments such as those outlined in this section, Greeno (1979) has proposed that knowledge and its effects on problem solving constitute a relatively neglected and important direction for research. Others have turned attention to the problem of how knowledge bases change and develop with experience so as to become better suited to problem-solving demands (Anderson et al., 1979; Lenat et al., 1979; Rumelhart, 1979; Rumelhart and Norman, 1977). Among implications from this work important to the present study are that knowledge bases change in the directions of: (1) accretion or, simply, augmentation of knowledge, (2) knowledge reorganization, and (3) changes and refinements in the condition tests by which knowledge is judged applicable to situations.

The present study investigates the effects of medical knowledge on the clinical reasoning process and the changes in such knowledge as individuals gain experience with the task of medical diagnosis and with the subject matter of a subspecialty of medicine.

12.2 Introduction: Clinical Diagnostic Reasoning and Expert/Novice Studies

Recent research in clinical diagnosis (Barrows et al., 1978; Elstein et al., 1978; McGuire and Bashook, 1978) contributed to a consensus about the *general* form of the process of clinical diagnostic reasoning. Cues in patient data suggest hypotheses, which are, in turn, tested against subsequent data of the case. The basic hypothetico-deductive process is shared by experienced and inexperienced diagnosticians alike, as are numerous parametric characteristics of the process, such as the percentage of data items to first hypotheses, the average number of hypotheses maintained in active consideration, etc.

These studies, however, have generally neglected the content of diagnostic reasoning, that is, the knowledge base of medical subject matter involved in the diagnostic process. Yet, despite prevalent findings of lack of differences in the form of diagnostic reasoning as a function of experience, the few differential findings from these research efforts implicate the importance of the knowledge base. The Michigan State group (Elstein et al., 1978) found that expert and less expert physicians differ in the "accuracy of interpretation" of patient data with respect to the hypotheses they consider, a finding that shows the importance of the physician's knowledge of patient data that present in particular diseases. Barrows's group (Barrows et al., 1978) found that experience can be discriminated by the *actual* hypotheses (as opposed to the number of hypotheses) that physicians use. This suggests that experienced and less experienced physicians differ in their knowledge store of diseases or the conditions by which they judge that particular diseases are likely to apply to a case. The same projects have also confirmed the case-specificity of skill in diagnostic reasoning. The same physician may show different profiles of competence depending on his or her particular experiential history with different types of cases, a further indication that clinical reasoning is not a general skill, but rather a process that is strongly dependent on the contents of knowledge to which it is applied.

Research at the University of Minnesota has concentrated on diagnostic problem solving in the medical subspecialty of pediatric cardiology and has resulted in a theory of diagnosis in this field that attempts to explicate the knowledge and knowledge organization necessary for expert diagnostic performance (Johnson et al., 1979b). Extensive experimentation and consultation with an expert pediatric cardiologist has resulted in a computer-runnable instantiation of the theory for this expert that represents knowledge explicitly and shows strong correspondence to the subject's performance over a broad range of cases (Swanson, 1978; Swanson et al., 1979).

Within the constructs of this theory, the present experimental study investigates the development of the knowledge base, as exemplified by individuals with different levels of experience with pediatric cardiology, and the implications of developmental differences for diagnostic performance. The particular theoretical construct of focus is prototype, or disease knowledge (Johnson et al., 1979b).[1] Disease knowledge refers to a memory store of disease models, each of which specifies, for a particular disease, the pathophysiology of the disease and the set of clinical manifestations that a patient with the disease should present [see also Rubin, (1975), disease "templates," and Pople (1977), "disease entities"]. In the theory of the expert, this set of disease models is extensive [see also deGroot (1965) and Simon and Chase (1973)] and organized hierarchically [see also Wortman (1972) and Pople (1977)]. At upper (more general) levels of the hierarchy are *disease categories*, sets of diseases that present similarly because of physiologic or clinical similarity. Particular diseases occupy middle ranks of the hierarchy, and these, in turn, are differentiated at the lowest hierarchical levels into numerous variants of each disease, each of which presents slightly differently in the clinic for reasons of subtle underlying difference in anatomy, physiology, severity, or age of presentation in a patient.

Speculations about characteristics of novices' disease knowledge can be garnered from analysis of the training experiences that novices encounter, the training materials they use, as well as psychological theory pertaining more generally to the development of knowledge bases. The first postulate for the novice's knowledge base of diseases is that it is *classically centered.* Initial training materials (Moller, 1978), as well as the probability distribution of diseases presenting in the hospital, accentuate the most common versions of diseases that constitute "anchorage points" for subsequent elaboration of the store of diseases [see also Rosch et al. (1976), "basic objects"]. A second postulate for novices is that the disease store is *sparse* in the sense that it lacks extensive cross referencing and connection among the diseases in memory (Elstein et al., 1971; Shavelson, 1972; Thro, 1978). It is with experience that the starting-point set of diseases is augmented and both generalized into categorical clusters, as similarities among diseases are discovered, and discriminated into finer distinct entities, as differentiation points among and within diseases are learned (Reed, 1978; Wortman and Greenberg, 1971). A third postulate about novice disease knowledge refers to the internal structure of the disease models themselves; this involves *imprecision* in the clinical expectations associated with diseases. Given that there is a range of natural variability associated with the clinical findings that can occur with any disease, large sampling,

[1]The term *disease knowledge* will be used in the present paper instead of the term *prototype knowledge.* It was decided to abandon the latter designation because of its suggestion of entities particularly typical of a class (Rosch, 1975). While some disease models are prototypic, not all of them are.

through clinical experience or other training devices, is probably necessary to "tune" (Rumelhart and Norman, 1977; Anderson et al., 1979) clinical expectations in disease models to the naturally occurring range. Novice expectations may be either overly general, allowing clinical findings that should not occur, or overly specific, not allowing the legitimate range.

In contrast to the novice, whose disease store is assumed sparse, imprecise, and classical, the expert's knowledge base of disease models, by converse arguments as well as by our prior research findings, is assumed *dense*, *precise*, and *penumbral*. The device for studying these claims in the present study is the careful selection of naturally occurring patient cases, each of which, through the structure of patient data it contains, provides a focused test of a different aspect of disease knowledge. In a laboratory setting, these cases were diagnosed by subjects at different levels of experience with pediatric cardiology.

12.3 Method

12.3.1 Materials

Stimulus materials for the study were sets of patient data, each representing a different patient case, extracted from medical records of clinical cases seen at the University of Minnesota Hospitals. Clinical and laboratory findings from the medical record for each case were assembled in a typed "patient file." The file arranged these data in the typical clinical order of history findings, followed by those from physical examination, x-ray, and electrocardiogram (EKG).

Four cases were used, each of which was chosen to assess a different characteristic of subjects' disease knowledge, for example, the differentiation of a disease into subtypes. In addition, the case design employed a "garden path" methodology; some chosen cases showed early strong cues for erroneous diseases but had later critical, disconfirmatory evidence for these same diseases. This device had two functions. First, it brought all subjects to a common starting point in their thinking about possible explanations for the case. Second, because the true diseases were physiologically and clinically similar to the initially induced diseases, it provided a test of the precision in a subject's model of the initial disease (if it was to be rejected), and it established an environment for assessing the diseases that subjects considered as plausible competitors to the original disease. Hence the "garden path" is a means for studying subjects' "conceptual competitor" sets (Elstein et al., 1971).

Case 1. The operative (true) disease in this case is subvalvular aortic stenosis, an uncommon variant of aortic stenosis, the "classic" or most com-

mon version of which is valvular aortic stenosis. The case was meant to assess subjects' differentiation of diseases into subtypes and the precision in their models of the classical variant.

Case 2. The operative disease in this case is total anomalous pulmonary venous connection (TAPVC). The case contains classic auscultatory findings for atrial septal defect (other findings are discrepant), a highly common congenital heart disease, findings that are also perfectly consistent with TAPVC, and, in fact, also consistent with any disease in the category of diseases with volume overload in the right side of the heart (including, in addition to the diseases mentioned, partial anomalous pulmonary venous connection and some forms of endocardial cushion defect). The case was designed to assess subjects' knowledge of and use of disease clusters corresponding to disease categories.

Case 3. This case is a straightforward presentation of the operative disease, patent ductus arteriosus, a highly common congenital heart disease. The case was intended to assess the relationships of this disease to other similar diseases within a subject's disease knowledge and the diagnostic use of these related diseases in a case where the correct diagnosis seems clear.

Case 4. The operative disease in this case is pulmonary atresia, one of a group of physiologically similar diseases (including, in addition, tricuspid atresia and Ebstein's malformation) that constitute a category of "cyanotic diseases with decreased pulmonary blood flow." Like Case 2, this case was designed to assess subjects' knowledge and use of disease clusters corresponding to categories.

12.3.2 Subjects

Subjects were 12 individuals from the University of Minnesota Medical School and were chosen to span a dimension of training and clinical experience in the diagnosis and management of congenital heart disease. Except for faculty experts, so few subjects existed at the prespecified experience levels that the subjects chosen comprised nearly all of them. There were four subjects from each of the following three groups:

- *Students.* These were fourth-year medical students who had just completed a six-week course in pediatric cardiology. As part of this training, each had held primary responsibility for diagnosis and management of 25–30 patients with congenital heart disease.

- *Trainees.* Subjects in this group were either in the third year of a general pediatrics residency or were beginning their first year of fellowship in pediatric cardiology. Subjects in this group estimated that they had held primary responsibility for about 150 patients with congenital heart disease. Residents and fellows did not differ in their estimates.
- *Experts.* This group was composed of two faculty members in the division of pediatric cardiology with upwards of 20 years of active practice as pediatric cardiologists and two fourth-year fellows in pediatric cardiology, one of whom was board-certified at the time of the study. The two fellows estimated that they had held primary responsibility for about 400 patients with congenital heart disease. The best estimates the faculty subjects could give were somewhere between 5,000 and 10,000. The experience discrepancy in this group enabled assessment of the effects of very long-term experience in the faculty members.

12.3.3 Procedure

Each subject diagnosed all four cases, and every subject diagnosed the cases in the same order. The subject was presented the patient file for each case and was instructed to read aloud each numbered data segment in the order in which data were given in the file.[2] The subject could review findings but could not skip ahead. The subject was instructed to report *aloud* any thoughts he or she had at any time toward formulating a diagnosis for the patient's condition. At four points in the case, after history, physical examination, x-ray, and EKG, the subject was asked for an explicit reporting of any "hunches" he or she might have about the patient's condition. After EKG, the subject was also asked for a primary diagnosis and as many as two alternatives.

12.3.4 Data and Analyses

Basic data from the study were typed transcriptions (protocols) of tape recordings made while subjects diagnosed the cases and reported aloud their thinking toward a diagnosis for each. Particular analyses of these data vary somewhat according to the objectives of each case. In general, analyses are organized according to a concept of *logical competitor sets* (LCS), which are sets of diseases targeted as important from the choice of cases for the study (see Section 12.3.1). Diseases in the competitor set for each case share

[2]Order and content of patient findings presented to subjects were fixed in order to compare inferences, interpretations, and evaluations by subjects in a uniform "stimulus" environment. While this deemphasized some components of the diagnostic process, primarily those associated with data collection (e.g., strategy) and first-order interpretations of patient data (e.g., reading x-rays), "fixing" of the input was important to the control needed to investigate the knowledge-based issues of interest in the study.

major underlying physiology with the operative or true disease in the case and hence have similar clinical presentation.

In concentrating analyses on the logical competitor set for each case, a commitment was made to focus analyses on diseases specified in advance to be plausible but easily confused alternatives for the case. Hence they constitute a set of *good* hypotheses to be considered in a case. There were two major motivations for concentrating on LCS diseases. One motivation comes from prior work on expert/novice differences, which suggests that unless a dimension of *quality* is built into the "dependent variables" measured, expert/novice differences are not likely to be revealed (Chase and Simon, 1973; Barrows et al., 1978). The second motivation was the case design itself (see Section 12.3.1).[3] It was assumed that the structure contained within the cases (e.g., garden-path notions) would greatly control and delimit subjects' performance so that the important dynamics of each case would center around the prespecified hypotheses (the LCS) and their management. (This turned out not always to be true for some subjects/cases—as will be noted.)

The LCS for each case was developed from two major sources. First, for the operative disease in each case, an expert in pediatric cardiology and collaborator on the project (the third author, a faculty member in pediatric cardiology at the University of Minnesota) was asked to specify the set of alternative diseases most similar to the true disease and likely to be confused with it. Because these are diseases that are highly similar in clinical presentation, he was also asked to specify the items of patient data that, if interpreted correctly, could be used to discriminate among diseases in the LCS. These judgments were then cross-checked against a major disease reference for pediatric cardiology (Moss et al., 1977). Specifically, for each disease described in this reference, the authors provide a "differential diagnosis" section that discusses diseases similar to and difficult to discriminate from the target disease, as well as differential data points. Based on the reference, no diseases were deleted from the consultant's list, although some were added.

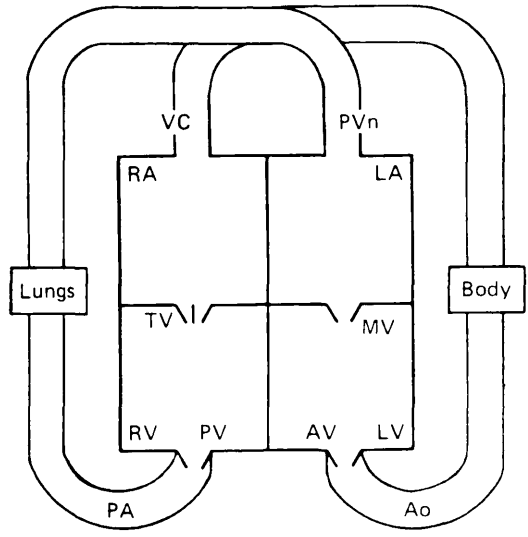
For each case, protocols were coded for two general kinds of uses of the logical competitor set. The first of these is the use of LCS members as hypotheses by subjects at each patient data point of the case. To the extent LCS members are used together, this is taken as evidence that these diseases are being used as competitors and are clustered in memory. The second is the evaluations of LCS members with respect to a set of selected data items. These evaluations yield evidence of the precision in subjects' individual disease models, and also can be used to discern characteristic kinds of errors among the subjects and the loci of these errors in disease knowledge.

---

[3]"Design" was through selection and not construction. Cases in the study are naturally occurring clinical cases and should not be considered oddities. According to the logic of the study, most cases, say, of TAPVC will have atrial septal defect as a naturally occurring garden-path foil.

12.4 Results

In this section, the results from the study will be presented in a case-by-case manner. The presentation of results from each case will follow the same general format. First, there is an introduction to each case that discusses the knowledge-based issue of interest and introduces the operative disease, its logical competitors, and key data points of the case. Since these discussions of congenital heart diseases refer to abnormal modifications to the normal heart and cardiovascular system, a depiction of the normal cardiovascular system is given for comparison as Figure 12-1. After the case discussion, two kinds of results are presented for each case. The first involves the use by subjects of LCS members as hypotheses during the course of the case. The second addresses diagnostic errors and their possible loci in disease knowledge.



LEGEND

Ao	= Aorta	PV	= Pulmonary Valve
AV	= Aortic Valve	PVn	= Pulmonary Veins
LA	= Left Atrium	RA	= Right Atrium
LV	= Left Ventricle	RV	= Right Ventricle
MV	= Mitral Valve	TV	= Tricuspid Valve
PA	= Pulmonary Artery	VC	= Vena Cavae

FIGURE 12-1 The normal heart and cardiovascular system.

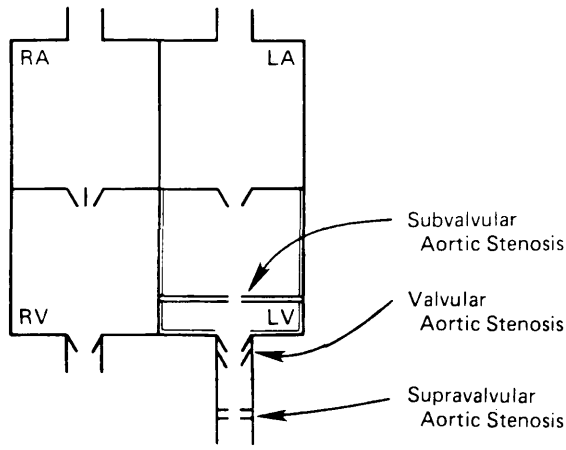


FIGURE 12-2 Logical competitor set for Case 1: three types of aortic stenosis.

12.4.1 Case 1: Subvalvular Aortic Stenosis

The purpose of this case is to investigate subjects' differentiation of a disease into subtypes. The vehicle for doing this is a diagnostic problem that encourages subjects to display, in a diagnostic setting, their working knowledge of a set of disease variants.

The logical competitor set for Case 1 includes three variants of aortic stenosis: valvular aortic stenosis (ValvAS), subvalvular aortic stenosis (SubAS), and supravalvular aortic stenosis (SupAS). Figure 12-2 depicts the anatomical abnormalities within the heart that define each of these disease variants. All involve obstruction to left ventricular outflow with different variants defined by slight differences in the locus of obstruction: ValvAS is obstruction at the aortic valve itself; SubAS is an obstruction slightly "upstream" from the valve; SupAS is obstruction slightly "downstream" from the valve. Because these disease variants are only subtly different anatomically and physiologically, they differ only slightly in clinical presentation. ValvAS is the most common of the three and receives the greatest amount of exposition in introductory training materials of pediatric cardiology (Moller, 1978). Hence it might be expected that subjects' knowledge for ValvAS would develop more rapidly than for the others and that ValvAS may function as a "foil" for some subjects. SubAS, however, is the operative disease in the case and the correct diagnosis.

In the patient file presented to subjects for Case 1, patient data items 17 and 19, a "thrill" and a "murmur" respectively, are strong cues for valvular aortic stenosis, although they are compatible with other variants. Hence it was suspected that all subjects would raise at least ValvAS as a

hypothesis by the time of these data points. Data item 18, a finding of "no systolic ejection click," is very strong evidence against ValvAS. Data items 10, "normal facies," and 22, "prominent aorta," are evidence against SupAS. All data of the case are compatible with the operative disease, SubAS.

Use of the Logical Competitor Set in Case 1

Table 12-1 shows the variants of aortic stenosis that were used as hypotheses by individual subjects at all patient data points where any variant was mentioned by any subject and at the four points of the case where "hunches" were actively solicited from the subjects.[4] Numbers representing data from the patient file are listed across the top in the left-to-right order in which they were presented to subjects. An X in this table simply indicates that the subject mentioned a particular aortic stenosis variant in the protocol at the data point where the X appears.

Table 12-1 shows an increase in the use of variants of aortic stenosis, other than ValvAS, from medical students to experts in pediatric cardiology. In particular, only one student (S2), ever raised both of the less classic variants of aortic stenosis at all, during the entire course of the case, and he mentioned SubAS and SupAS only once each. Two trainees (T1, T3) and three experts (E1, E3, E4) used all three variants at some time during the case. If one considers the number of subjects in each group who not only used all three variants but used each more than once, no students, one trainee (T1), and, again, three experts meet this criterion.

While simple mention (as reflected in Table 12-1) of the aortic stenosis variants as hypotheses is one indication of whether these were considered by subjects, a measure of how *actively* these hypotheses were considered is the prevalence with which they were evaluated with respect to data items. Table 12-2 shows all evaluations by subjects of the aortic stenosis variants with respect to the set of data items that are central to successful solution of the case. A mark (+, −, 0) under a disease variant and data item in this table indicates that the data item was judged to be positive, negative, or ambivalent evidence for the disease variant as a hypothesis.[5] For ex-

---

[4]Subjects E3 and E4 are the faculty subjects with upwards of 20 years of experience. They are noted with asterisks in this and all subsequent tables.

[5]There is no absolute correspondence between the use of a hypothesis at the point of a particular data item (Table 12-1) and the evaluation of the hypothesis with respect to that data item (Table 12-2). Subjects could evaluate a hypothesis with respect to a data item long past (e.g., evaluate with respect to data item 10 having reached, say, data point 17 of the case) and could also mention a hypothesis at a data point without necessarily evaluating the hypothesis with respect to that data item. Hence, for example, even though subject S2 mentioned all three variants at data point 10, he only ever evaluated one of these (SupAS) with respect to data item 10. The mention of the other variants at 10 was part of a puzzled attempt to recall the variants of aortic stenosis.

TABLE 12-1 Case 1: Subjects' Use of LCS Hypotheses in Response to Patient Data Items

		Patient data items																	
		History						*Physical exam*								*X-ray*		*EKG*	
Subjects/hypotheses		*1*	*3*	*4*	*7*	*8*	*HHx*	*10*	*13*	*14*	*17*	*18*	*19*	*20*	*HPEx*	*22*	*Hxray*	*23*	*HEKG*
S1	ValvAS			X	X		X		X			X	X		X	X	X	X	X
	SupAS							X											
S2	ValvAS				X		X				X	X	X	X	X	X	X	X	X
	SubAS							X											
	SupAS							X											
S3	ValvAS						X		X	X	X	X	X	X	X	X	X	X	X
S4	ValvAS												X		X	X	X	X	X
T1	ValvAS						X				X	X	X		X	X			X
	SubAS												X	X	X	X	X	X	X
	SupAS												X						X
T2	ValvAS												X		X	X	X	X	X
T3	ValvAS				X	X	X		X		X	X	X	X		X		X	
	SubAS				X		X					X		X	X	X	X	X	X
	SupAS											X							
T4	ValvAS										X	X	X		X	X	X	X	X
E1	ValvAS	X	X		X		X				X	X	X	X	X	X	X	X	X
	SubAS										X		X		X	X	X		
	SupAS										X		X		X	X			
E2	ValvAS								X			X	X	X	X	X	X	X	X
	SupAS												X		X		X	X	
E3*	ValvAS		X				X				X	X	X	X	X	X	X		X
	SubAS										X	X	X	X	X	X	X		X
	SupAS							X			X		X		X	X			
E4*	ValvAS										X	X	X			X	X	X	X
	SubAS											X	X	X	X	X	X		X
	SupAS											X	X						

Note: X indicates a subject's use of a hypothesis at the time of a patient data item. HHx, HPEx, etc. refer to points in the case where subjects are asked for hunches.

* The two experts with more than 20 years of experience.

TABLE 12-2 Case 1: Evaluations of Target Data Items in Relation to LCS Hypotheses

	Target patient data items																	
	10 Normal facies			*17 Thrill*			*18 No click*			*19 Murmur*			*20 Aortic insuff.*			*22 Prominent aorta*		
Hypotheses	*ValvAS*	*SubAS*	*SupAS*	*ValvAS*	*SubAS*	*SupAS*	*ValvAS*	*SubAS*	*SupAS*	*ValvAS*	*SubAS*	*SupAS*	*ValvAS*	*SubAS*	*SupAS*	*ValvAS*	*SubAS*	*SupAS*
Subjects																		
S1			−				−			+						+		
S2			−	+			+			+			+			+		
S3				+			+			+			+			+		
S4										+						+		
T1				+			−			+	+	+		+		+	+	
T2										+						+		
T3			0	+	+	+	−	+		+			+				+	
T4				+			−			+						+		
E1			−	+	0	+	−	+	+	+	0	+	+	−	−	+	−	−
E2			−				−		+	+			+					
E3*			−	+	+	+	−	+		+	+	+	+	+		+	+	
E4*				+			−	+	+	+	+	+		+		+	+	

Note: + indicates subject judged data item as confirmatory for a hypothesis. − indicates subject judged data item as disconfirmatory for a hypothesis. 0 indicates subject judged data item as ambivalent in relation to a hypothesis.

* The two experts with more than 20 years of experience.

ample, a negative evaluation of "no click" with respect to ValvAS would be "The lack of a systolic ejection click is against valvular aortic stenosis."

Table 12-2 shows an increase, from students to experts, in the active evaluation of data items as evidence for or against the variants of aortic stenosis. In particular, no student evaluated all three of the variants with respect to a data item (of course, only one student, S2, ever mentioned all three variants at all). The two trainees (T1, T3) and three experts (E1, E3, E4) who used all three variants in the case also evaluated all three variants with respect to at least one data item. While this suggests activeness in the evaluation of variants by more experienced subjects, it does not necessarily reflect comparative evaluation. However, when a subject evaluates all variants with respect to the *same* data item, this is an indication that the subject is actively attempting to weigh the variants against each other to determine which is the best explanation for the data item and case. In this regard, no students, the two trainees (T1, T3), and again, the three experts (E1, E3, E4) evaluated all three variants with respect to a common (the same) data item. These same experts, but *not* the trainees, evaluated all variants in relation to *more* than one data item in common (E1, 5 items; E3, 2 items; E4, 2 items).

The analysis thus far suggests that with increasing diagnostic experience subjects know and actively utilize nonclassical variants of a disease as hypotheses in a diagnostic setting. Examination of the protocols of the two most experienced subjects, E3 and E4, yields some clue as to the knowledge structure that supports this performance. Figure 12-3 shows the protocols of these subjects at two data points: 17, which is the first strong evidence for valvular aortic stenosis and other variants; and 18, which is the strongest evidence against ValvAS. E3 raises all three variants together at the time of the first *strong* evidence. These hypotheses are then available to be evaluated comparatively against subsequent data, in particular, data item 18. This same form characterizes expert E1 (see Table 12-1). Expert E4, however, aggressively focuses on the "classic" member of the competitor set at 17, but immediately expands to the *full set* upon receiving strong negative evidence at 18. This form is shared by subject T3 and, less clearly, by subject T1 (see Table 12-1) and suggests that for these subjects LCS hypotheses other than the classic disease may have undergone at least partial activation earlier.

One explanation for these patterns is that in the expert a disease and its set of subtle variations come to constitute an interconnected memory unit, a kind of category; when one of the members is strongly activated in memory, the category and other members are also activated. The expert can then choose to consider category members in two modes. In the first mode, he or she tests all members simultaneously. This first mode might be termed *precautionary* since if any hypothesis encounters disconfirmatory evidence, alternative explanations for which the same evidence might be compatible are already under consideration. In the second mode, the expert tests only the most likely (in his or her current judgment) member.

(17) There is systolic thrill felt below the right clavicle, along the mid-left sternal border and in the suprasternal notch.

E3: This thrill is most consistent with a diagnosis of bicuspid aortic valve or aortic valvular stenosis. It would also be consistent with supravalvular stenosis and discrete subaortic stenosis.

E4: Until proved otherwise, now, he must have valvular aortic stenosis.

(18) The first heart sound is normal, and there is no systolic ejection click.

E3: The absence of a systolic ejection click in the presence of what I would consider to be an aortic outflow thrill makes aortic valvular stenosis and bicuspid aortic valve less likely. Aortic valvular stenosis of a very severe degree might be associated without a click. On the other hand, uh, it makes us think more seriously of discrete membranous subaortic stenosis.

E4: Absence of the click is against valvular aortic stenosis. Then perhaps instead he has subvalvular or supravalvular aortic stenosis.

FIGURE 12-3 Protocols from experts E3 and E4 at data points 17 and 18 in Case 1.

This mode might be termed one of *extraction* because its general success depends heavily on rejection of the target disease when appropriate, which, in turn, depends heavily on the precision in the diagnostician's model for the disease. In instances where the target disease is rejected, other category members provide a ready back-up set of alternative hypotheses. Further evidence for these speculations will be addressed as results from other cases are presented.

Diagnostic Errors in Case 1

A final analysis of the results of this case involves an attempt to discern the causes for subjects' errors in final diagnosis. Table 12-3 gives the final primary diagnosis for each subject. Among unsuccessful subjects, six subjects (S1, S3, S4, T2, T4, E2) never considered subvalvular aortic stenosis at all (see Table 12-1), although all generated and concluded valvular aortic stenosis. At least three explanations could apply to this lack of activation. First, and most basically, it could be that subjects do not know about SubAS at all. However, postexperimental interviews with all these subjects confirmed that they had some knowledge of this disease and could describe it. A second possible explanation is that these subjects have built up no

TABLE 12-3 Case 1: Subvalvular Aortic Stenosis —Final Diagnoses

Subjects		*Final diagnosis*
Students	S1	Valvular aortic stenosis
	S2	Valvular aortic stenosis
	S3	Valvular aortic stenosis
	S4	Valvular aortic stenosis
Trainees	T1	Subvalvular aortic stenosis
	T2	Valvular aortic stenosis
	T3	Subvalvular aortic stenosis
	T4	Valvular aortic stenosis
Experts	E1	Valvular aortic stenosis
	E2	Valvular aortic stenosis
	E3*	Subvalvular aortic stenosis
	E4*	Subvalvular aortic stenosis

*E3 and E4 are the two experts with more than 20 years of experience.

strong "bottom-up" association in memory between any data item of the case and the subvalvular disease. Even lacking such a "trigger" or recognition rule for SubAS itself, it would have been possible for subjects to generate SubAS as a side effect of their activation of ValvAS, if these two diseases were related in a memory unit, through a process of "spreading activation" (Anderson, 1976) or "top-down" activation (Rumelhart and Ortony, 1977; Bobrow and Norman, 1975). This suggests the third explanation—that for these subjects knowledge representations for the variants of aortic stenosis exist more in isolation than they do in the more experienced subjects. This is the issue of sparseness in disease knowledge.

For those subjects who generated ValvAS as a hypothesis but failed to abandon it in the face of strong negative evidence, examination of their handling of this disconfirmatory evidence yields insight into the nature and precision of their disease models for ValvAS. Discussion will focus on data item 18, the strongest evidence against ValvAS. Two students (S2, S3) evaluated 18, "no click," as confirmatory for ValvAS (Table 12-2). This appears to reflect, simply, an error in important factual knowledge about this disease. Two subjects (S4, T2) did not evaluate 18 at all with respect to ValvAS (Table 12-2). Significantly, they also did not generate *any* variant of aortic stenosis until after data item 18 (Table 12-1). This suggests that for these subjects the memory store of bottom-up associations between data items and aortic stenosis variants is not as extensive as for other subjects and, in particular, that data item 17 is not recognized as a strong cue for aortic stenosis–type diseases. A further implication is that the physical examination finding of a "systolic ejection click" and its import in ValvAS are not represented in the ValvAS disease models of these subjects, since, if

(18) The first heart sound is normal, and there is no systolic ejection click.

S1: Ah, well this, the fact that there is no systolic ejection click present, tells us that there is probably not a poststenotic dilation of the aorta, which one would expect with the presence of aortic stenosis and some aortic insufficiency. However, this does not necessarily rule it out.

T4: Love it. Um, well, okay. I wonder if there is..., no click, that's funny. I would expect it if he has AS. I wish they had said whether the murmur went up into his neck, okay.

(22) The chest x-ray shows normal cardiac size and contour and normal vascularity, but prominence of the ascending aorta.

S1: Ah, well this is what one would expect with ah, aortic stenosis with secondary aortic insufficiency. One would expect that the aorta, ascending aorta distal to the ah, to the stenosis, would be dilated due to the changes in the wall tension across the gradient. Therefore, ah, the fact that ah, a click was not heard on physical exam, may have been a subjective finding of the person examining. But, the x-ray does indeed suggest that there is some poststenotic dilation.

T4: Ha ha! AS-AI.

FIGURE 12-4 Protocols from subjects S1 and T4 at data points 18 and 22 in Case 1.

they were, the model itself should have led the subjects to reexamine this finding.

Finally, there were four subjects (S1, T4, E1, E2) who, although evaluating 18 as negative for ValvAS, still maintained ValvAS as a final diagnosis. The protocols of subjects S1 and T4 yield some insight into an explanation for these subjects. Figure 12-4 shows the protocols for these two subjects at data points 18 and 22, the latter consisting primarily of the finding of a "prominent aorta" on x-ray. Both subjects question ValvAS at 18, but are much more satisfied with this diagnosis at 22 and thereafter. Why might this be?

Figure 12-5 shows the causal relationship between a "tight" or stenotic aortic valve and an enlarged or prominent aorta. To open the tight valve, the left ventricle (LV) of the heart must generate abnormally high pressure. Blood expelled under this high pressure forces against the aortic wall and expands it. For the two subjects under discussion, it appears that their causal knowledge attributes the "systolic ejection click" in ValvAS to the enlarged aorta itself; that is, the click is caused by the large chamber into which the valve is opening, perhaps some kind of resonance phenomenon. For these subjects the causal chain from the valve to the click is as follows:

tight valve → big aorta → click

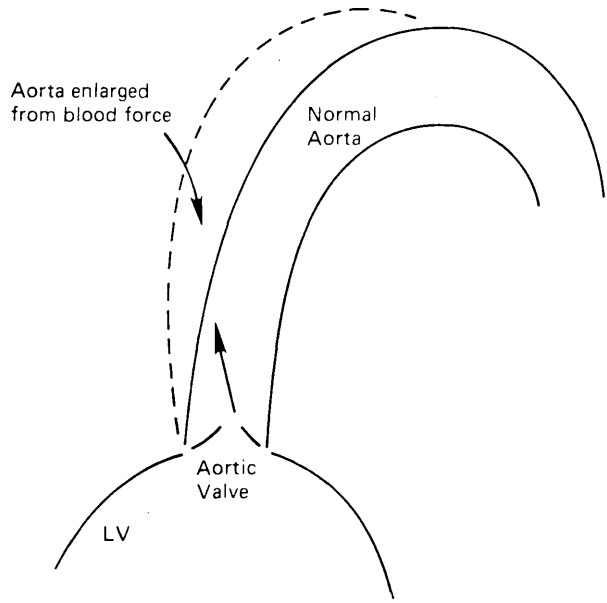


FIGURE 12-5 Aorta enlarged from the force of blood ejecting from a stenotic aortic valve.

Hence, for these subjects, the big aorta itself is predominant over the click as evidence for ValvAS, with the click just additional evidence for a big aorta. Once they receive their best evidence for a big aorta, data item 22, they are no longer worried about the lack of a click.

The true state of affairs appears to be that a tight valve causes both the click and the enlarged aorta at the same level of cause (Friedman and Kirkpatrick, 1977, p. 180). The systolic ejection click is associated with the opening of the tight valve itself as shown below:

tight valve → click
↳ big aorta

Hence both of these effects must be proved. Why might a number of subjects have misconstrued this relationship? One need look no farther than the introductory textbook these subjects use (Moller, 1978, p. 96) where the erroneous causal relationship is stated or at least strongly implied.

The subjects just discussed raise two important issues. First, they demonstrate how "small" knowledge errors can have major repercussions for the handling of a case, and they shed some insight into the case-specificity of a clinician's diagnostic performance found elsewhere (Elstein et al.,

1978). Second, they suggest a sensitivity in less experienced clinicians to specific training experiences, for example, training materials, particular patient cases, etc. As experience increases, so does the sample of "inputs" and the effects of particular experience might be expected to lessen.

12.4.2 Case 2: Total Anomalous Pulmonary Venous Connection

The purpose of this case is to investigate the aggregation by subjects of a set of physiologically similar diseases into a memory grouping or category. The case is different from Case 1 in that while Case 1 dealt with a set of variants of one disease, Case 2 is concerned with a set of diseases.

The logical competitor set for Case 2 includes four diseases: total anomalous pulmonary venous connection (TAPVC), partial anomalous pulmonary venous connection (PAPVC), atrial septal defect (ASD), and endocardial cushion defect (ECD). Figure 12-6 shows the anatomical and physiologic abnormalities within the heart that define each of these diseases.

In TAPVC, all four pulmonary veins (PVn in Figure 12-6) connect to the right atrium (RA) of the heart rather than to the left atrium (LA), their normal site of connection. All oxygenated blood coming back to the heart from the lungs mixes with deoxygenated blood coming back to the heart from the body. Hence, all blood subsequently pumped back to the body is a mixture of oxygenated and deoxygenated blood, which causes the patient to appear cyanotic, that is, to take on a mildly "blue" skin coloration.

In PAPVC, only a subset of the pulmonary veins connect abnormally to the right atrium, with the remainder connecting, as they should, to the left atrium. A result is that some already oxygenated blood is recirculated through the lungs. Blood pumped to the body, however, is oxygenated, and the patient retains a normal "pink" coloration.

Both ASD and ECD consist of a defect (a hole) in the atrial septum of the heart. They differ in the particular site of defect; ASD is a defect in the upper portion of the septum (the ostium secundum) while ECD is a defect in the lower portion of the septum (the ostium primum). In both diseases, the presence of the hole in the septum allows blood to shunt from the left atrium to the right atrium. While some oxygenated blood shunts to the right side to be recirculated to the lungs, blood expelled to the body is oxygenated, and the patient is pink.

A feature common to all four diseases in the LCS is an increased volume of blood in the right-sided chambers of the heart. This common element is a candidate feature on which diagnosticians might base a disease category, for example, "diseases with right-sided volume overload." A clinical manifestation related to volume overload that all these diseases produce in common is a set of three auscultation findings. One is a murmur

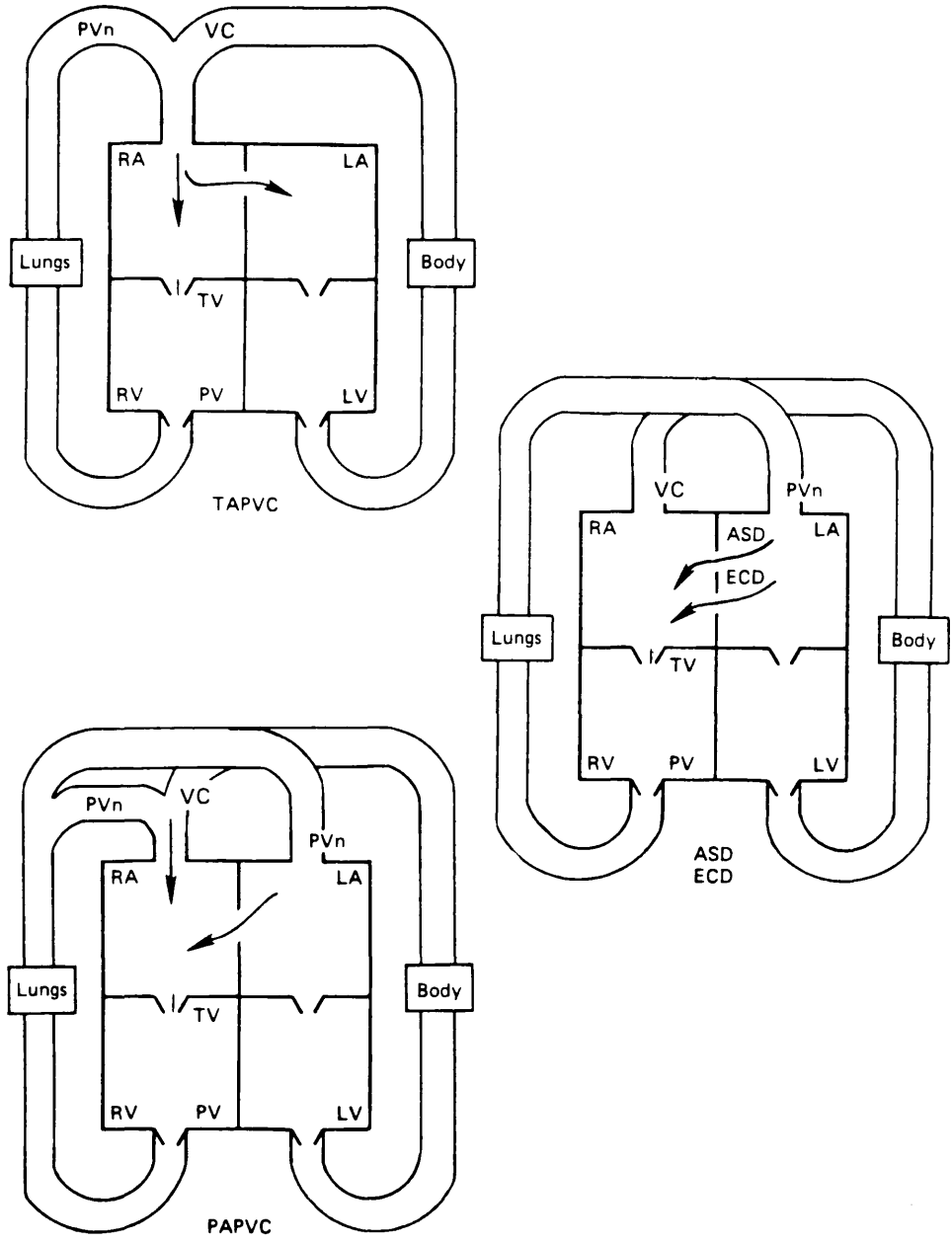


FIGURE 12-6 Logical competitor set for Case 2: total anomalous pulmonary venous connection, partial anomalous pulmonary venous connection, atrial septal defect, and endocardial cushion defect.

associated with increased blood flow across the tricuspid valve (TV). The second is a murmur associated with increased flow across the pulmonary valve (PV). The third is wide, fixed splitting of the second heart sound. The third finding is nearly pathognomonic for conditions of this type.

Of the four diseases, ASD is more common than the others. Hence it might be expected that subjects' knowledge for this disease would develop more rapidly than for the others. More importantly, ASD is the disease that is used instructionally to introduce the concepts of atrial level left-to-right shunting of blood in the heart and right-sided volume overload. Therefore, it might be expected that the three auscultation findings (especially the splitting) reflecting overload would be more strongly associated with ASD than with the other diseases. TAPVC, however, is the operative disease in the case.

There are six particularly important data items in the patient file presented to subjects for Case 2. Data items 17, 18, and 19 contain the set of three findings discussed above that are salient results of increased right-sided heart flow. Item 17 contains the "wide, fixed, split second heart sound." Hence, it was expected that all subjects would raise at least ASD, the classic instance of this type of disease, by the time of these data points. Data item 7 (also 11), which reports that the patient is mildly cyanotic, represents disconfirmatory evidence for all members of the LCS except TAPVC. Data item 21, which contains an x-ray description of "an unusual vascular shadow on the right side," is evidence against ASD and simultaneously constitutes a classic cue for PAPVC. In fact, one variant of PAPVC, scimitar syndrome, derives its name from its presentation of such a finding on x-ray (Lucas and Schmidt, 1977, p. 442). The EKG, item 22, contains a finding of "right-axis deviation" on the EKG and constitutes strong disconfirmatory evidence for ECD. All data of the case are compatible with the operative disease, TAPVC.

Use of the Logical Competitor Set in Case 2

Table 12-4 shows all uses by all subjects of the four diseases in the logical competitor set for Case 2 at all patient data points where any of the four was mentioned by any subject.

For reasons discussed above, it was assumed that most subjects would consider ASD in relation to the three data items, 17, 18, and 19. The use of other LCS members at these points is taken as evidence that the other diseases are associated in memory with ASD and this set of cues. Table 12-4 shows a decrease from students to experts in the number of subjects who considered only ASD at these points. All of the students considered only ASD, the disease we presume to be the classic exemplar of right-sided volume overload, at data items 17–19. Three of four trainees (T1, T2, T3) and the two least experienced experts also considered only ASD. Of the

TABLE 12-4 Case 2: Subjects' Use of LCS Hypotheses in Response to Patient Data Items

		Patient data items													
		History					*Physical exam*					*X-ray*		*EKG*	
Subjects/ hypotheses		*1*	*3*	*5*	*7*	*HHx*	*17*	*18*	*19*	*20*	*HPEx*	*21*	*Hxray*	*22*	*HEKG*
S1	ASD					X	X				X				
	PAPVC											X	X	X	X
S2	ASD						X		X		X				
	ECD											X	X	X	
	PAPVC											X			
S3	ASD						X				X		X		X
	ECD													X	X
	PAPVC											X			X
S4	ASD						X	X	X			X			X
	ECD													X	
	TAPVC											X			
T1	ASD						X		X						
	PAPVC											X	X	X	X
T2	ASD										X				
	ECD								X		X				
	TAPVC												X	X	X
T3	ASD			X	X		X	X	X		X	X		X	
	PAPVC													X	
	TAPVC													X	X
T4	ASD						X			X				X	X
	ECD									X	X	X	X	X	
	PAPVC													X	
	TAPVC						X					X			X
E1	ASD	X	X	X		X	X		X	X		X			
	PAPVC											X	X		X
E2	ASD								X		X			X	
	PAPVC										X				X
	TAPVC														X
E3*	ASD	X					X	X	X						
	ECD					X					X		X	X	
	PAPVC						X	X							
	TAPVC					X	X	X	X		X	X	X	X	X
E4*	ASD						X		X		X				
	ECD								X			X			
	PAPVC											X			
	TAPVC											X			X

Note: X indicates a subject's use of a hypothesis at the time of a patient data item. HHx, HPEx, etc. refer to points in the case where subjects are asked for hunches.

*The two experts with more than 20 years of experience.

two highly experienced experts, E3 utilized three LCS members (ASD, PAPVC, TAPVC) and E4 used two (ASD, ECD) at these points.

From the point of view of the entire case, no students, one trainee (T4), and two experts (E3, E4) generated *all four* members of the LCS during the course of the case. While this shows no obvious general trend toward increased use of the LCS with experience, it is perhaps significant that the full competitor set was used by the two high-level experts, E3 and E4.

In utilizing the full logical competitor set, the two most experienced subjects, E3 and E4, demonstrated the same patterns of *precaution* and *extraction* respectively as they did in Case 1. E3 considered three of the four LCS members (ASD, PAPVC, TAPVC) at item 17, the *first* strong cue for right-sided volume overload. E4 raised only ASD at this point and maintained this hypothesis until data item 21, which contains strong evidence against ASD. At this point, he expanded to the remainder of the LCS.

Diagnostic Errors in Case 2

Table 12-5 gives the final primary diagnoses for all subjects on Case 2. Only four subjects (trainees T2 and T3 and the two most experienced experts, E3 and E4) diagnosed the case correctly. Subjects who diagnosed the case incorrectly demonstrate informative types of errors.

Student S3 diagnosed the case as endocardial cushion defect (ECD). The strongest evidence against this disease is the finding of right-axis deviation on the EKG (data item 22). ECD uniformly presents with *left*-axis deviation and, in fact, is one of a very few congenital heart diseases that does; hence left-axis deviation is a nearly pathognomonic finding for ECD. S3 not only evaluated the *right axis* as positive evidence for ECD, but, in addition "triggered" or proposed ECD for the first time at this point (see Table 12-4). This is, simply, imprecision in the subject's disease model for ECD. It is as though the subject remembered that the EKG axis is important in ECD but could not remember the details.

The final diagnosis of subject T4 was ASD, even though she had considered TAPVC during the case. She correctly evaluated cyanosis (blueness—items 7 and 11) as negative for ASD, but maintained ASD nonetheless. Her primary difficulty was that she did not believe that TAPVC could present in a child as old as the one in the case (5 years old), although it certainly can—as the case itself, a real case, attests. This suggests that the allowable age range specified in the subject's disease model for TAPVC is overly restrictive, probably reflecting a limited sample of experiences with this disease.

Four subjects (S1, T1, E1, E2) diagnosed the case as PAPVC. Three of these subjects (S1, T1, E1) show a pattern in which only ASD (among the LCS members) is considered prior to data item 21, a classic x-ray cue for PAPVC, and only PAPVC is considered at that point and thereafter

TABLE 12-5 Case 2: Total Anomalous Pulmonary Venous Connection—Final Diagnoses

	Subjects	*Diagnosis*
Students	S1	Partial anomalous pulmonary venous connection
	S2	Transposition of the great vessels + pulmonary stenosis + atrial septal defect + partial anomalous pulmonary venous connection
	S3	Endocardial cushion defect
	S4	Pulmonary stenosis + atrial septal defect + ventricular septal defect
Trainees	T1	Partial anomalous pulmonary venous connection
	T2	Total anomalous pulmonary venous connection
	T3	Total anomalous pulmonary venous connection
	T4	Atrial septal defect
Experts	E1	Partial anomalous pulmonary venous connection
	E2	Partial anomalous pulmonary venous connection
	E3*	Total anomalous pulmonary venous connection
	E4*	Total anomalous pulmonary venous connection

*The two experts with more than 20 years of experience.

(see Table 12-4). This indicates a strong data-driven dependence in the diagnosis by these subjects; that is, the subjects are pushed from hypothesis to hypothesis depending on the most recent strong disease cue in the data, and when new hypotheses are generated, these are not strongly enough associated in memory with other LCS members to activate these other diseases. Some support for this claim can be seen in subject T1's protocol, taken from the point in the case where he offers his final diagnosis:

> *T1:* I am sort of drawing a blank on how to fit all this information together. And ah, I am just sort of guessing right now. I would say just scimitar syndrome [PAPVC] primarily based on the chest x-ray, and ah, I'm not really sure whether the whole thing fits together well. That is all I can say.

Of the four subjects, student S1 never evaluated PAPVC with respect to cyanosis; hence this finding had no opportunity to detract from his PAPVC hypothesis. Subject T1 evaluated cyanosis as *confirmatory* evidence for PAPVC, and this erroneous evaluation reinforced this disease interpretation. Expert subjects E1 and E2 evaluated cyanosis appropriately as negative evidence for PAPVC, but this evaluation was probably overridden by the strength of the cue for PAPVC on the x-ray.

Finally, two students (S2, S4) proposed configurations of multiple diseases as explanation for the case. Both of these composite diagnoses in-

TABLE 12-6 Case 2: Interpretations of Data Item 18

		Interpretation	
Subject		*Pulmonary stenosis*	*Increased flow pulmonary valve*
Students	S1	+	
	S2	+	
	S3	+	+
	S4	+	
Trainees	T1		
	T2	+	
	T3	+	+
	T4		+
Experts	E1	+	+
	E2		
	E3*		+
	E4*	+	+

Note: + indicates that a subject interpreted the murmur of data item 18 as pulmonary stenosis or increased flow over the pulmonary valve.
*The two experts with more than 20 years of experience.

cluded the disease pulmonary stenosis (PS), and it is this component of the final diagnosis that is the key to understanding the performance of these two subjects. Table 12-6 shows the interpretations by all subjects of data item 18, a systolic murmur in auscultation of the heart. Such a murmur results whenever there is too much flow over the pulmonary valve, relative to its orifice size. This situation prevails in either of two conditions:

1. When there is normal amount of flow but an abnormally small orifice. This is the disease pulmonary stenosis, which refers to an abnormally tight valve.
2. When there is a normal-sized orifice but abnormally high flow, the situation that prevails in the diseases of the LCS. A + under one of these two interpretations in Table 12-6 indicates that a subject attributed this interpretation to the murmur of data item 18.

Table 12-6 shows that most of the students (three of four) interpreted the murmur only as pulmonary stenosis, while most of the expert group (three of four) interpreted the murmur as increased flow *or* a tight valve. While student S1 (and subject T2) was eventually able to extract himself from his interpretation, students S2 and S4 were not. Once these students introduced PS into their diagnoses, they were forced to propose rather unusual combinations of multiple diseases to account for some of the findings of the case. For example, subject S2, in order to reconcile PS with other data of the case indicating increased blood flow in the lungs, simply transposed the great vessels of the heart; that is, he detached the pulmo-

nary artery from its normal mooring at the pulmonary valve and reattached it at the aortic valve and did the opposite with the other great vessel, the aorta. While this rather creative causal explanation represents a congenital heart disease, transposition of the great vessels, it is highly unlikely that a child with the combination of abnormalities proposed by the subject could have lived for five years untreated.

The interpretations of the systolic murmur by the students in Case 2 is another example of error, or at least limitation, in causal knowledge. It represents a situation where there are multiple causes for a finding and the novice considers only a subset. This is not unlike what has been shown at the disease and disease variant levels; that is, when multiple diseases in the logical competitor set can produce a finding, the novice seems limited to the most salient members. This suggests the import of grouped or clustered memory organization not only for diseases but also for "low-level," pathophysiologic interpretations for data.

12.4.3 Case 3: Patent Ductus Arteriosus

The purpose of this case is to test the robustness of expert grouping of hypotheses in a straightforward case in which there are no data discrepant with an initially induced disease interpretation. Interest is in whether subjects, even in a case with a very common disease, strong cues for this disease, and no data discrepant with this interpretation, still investigate a related set of physiologically similar alternatives.

The operative disease in the case is patent ductus arteriosus (PDA), a schematic for which is shown in Figure 12-7. This disease is an extracardiac shunt, that is, an abnormal communication between vessels, the aorta (Ao) and the pulmonary artery (PA), outside the heart. There are four other "disease" conditions in the logical competitor set. The congenital heart diseases arterio-venous fistula (AVF) and aorto-pulmonary window (APW) are other extracardiac shunts. Venous hum (VH) is a benign condition that presents a murmur similar to PDA, and ruptured sinus of valsalva (RSV) is a heart condition that has a clinical presentation similar to that of PDA. In the patient file presented to subjects for Case 3, the most important patient data item is number 19, a classic murmur of patent ductus arteriosus. It was assumed that all subjects would generate PDA as a hypothesis no later than this point. No data of the case are incompatible with PDA.

Use of the Logical Competitor Set in Case 3

Table 12-7 shows all uses of members of the logical competitor set by all subjects during the course of the case. It is clear that only one subject, E3, one of the two high-level experts, considered the full competitor set, al-

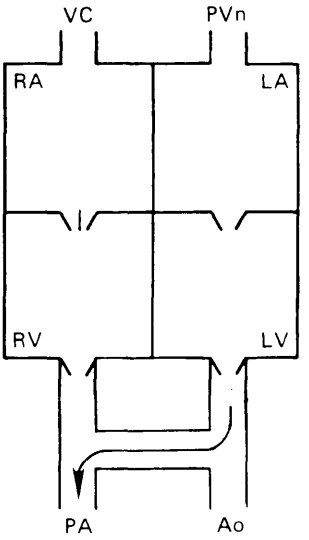


FIGURE 12-7 Patent ductus arteriosus.

though expert E2 considered three of the five—more than any of the remaining subjects. Since it was assumed that all subjects would consider PDA, a criterion far less stringent than "full use" for the LCS is the number of subjects in each group who considered even one additional LCS member and used it more than once. This condition holds for only one student (S3), one trainee (T1), but three of the experts (E1, E2, E3).

Expert E3 considered the full LCS in a precautionary pattern consistent with his performance on other cases (see Table 12-7). He used three of the five LCS members as hypotheses at data item 19, a strong cue for PDA. The remainder of the LCS was filled out two items later, after an intervening, uninformative data item, at the point where the subject was asked for "hunches." The other high-level expert, E4, looks in all respects like a novice in this case, in that he considered only PDA. However, if our earlier interpretations of an extraction method are correct for this subject, we would not expect him to expand to other members of the competitor set unless he encountered data discrepant with his target hypothesis; of course, there are none in this case.

The diseases in this case constitute a category of extracardiac communications and related conditions. An interpretation of the results from this case is that with high-level experience, it is this category, and not isolated individual members, that is generated and tested when a strong cue for a category member is encountered. No subject diagnosed this case incorrectly; hence analysis of subject errors is uninformative.

TABLE 12-7 Case 3: Subjects' Use of LCS Hypotheses in Response to Patient Data Items

		Patient data items												
		History					*Physical exam*				*X-ray*		*EKG*	
Subjects/ hypotheses		*3*	*4*	*5*	*7*	*HHx*	*14*	*19*	*20*	*HPEx*	*21*	*Hxray*	*22*	*HEKG*
S1	PDA					X		X		X	X	X	X	X
	APW												X	
S2	PDA					X		X		X	X	X	X	X
S3	PDA		X	X	X	X		X		X	X	X	X	X
	AVF									X				X
S4	PDA					X	X	X		X	X	X	X	X
	AVF						X							
T1	PDA						X	X	X	X	X	X	X	X
	APW						X	X		X		X		X
T2	PDA							X		X	X	X	X	X
T3	PDA							X	X	X	X	X	X	X
T4	PDA	X					X	X	X	X	X	X	X	X
E1	PDA						X	X		X	X	X		X
	AVF							X				X		X
E2	PDA						X	X		X	X	X	X	X
	AVF						X	X						
	VH							X		X				
E3*	PDA							X		X	X		X	
	AVF									X				
	VH							X						
	APW							X		X	X			
	RSV									X				
E4*	PDA							X		X		X	X	X

Note: X indicates a subject's use of a hypothesis at the time of a patient data item. HHx, HPEx, etc. refer to points in the case where subjects are asked for hunches.

*The two experts with more than 20 years of experience.

12.4.4 Case 4: Pulmonary Atresia

The objective of this case is similar to that of Case 2, that is, to assess subjects' aggregation of physiologically similar diseases into categories. Case 4 is different from Case 2 in that no single cue serves to distinguish the members of the logical competitor set from diseases outside it (as did "wide, fixed, split second heart sound" in Case 2). In Case 4 the diagnostician must arrive at the LCS by partitioning the space of diseases, using multiple data items from widely separated parts of the case.

The logical competitor set for Case 4 includes three diseases: pulmonary atresia (PAT), tricuspid atresia (TAT), and Ebstein's malformation of

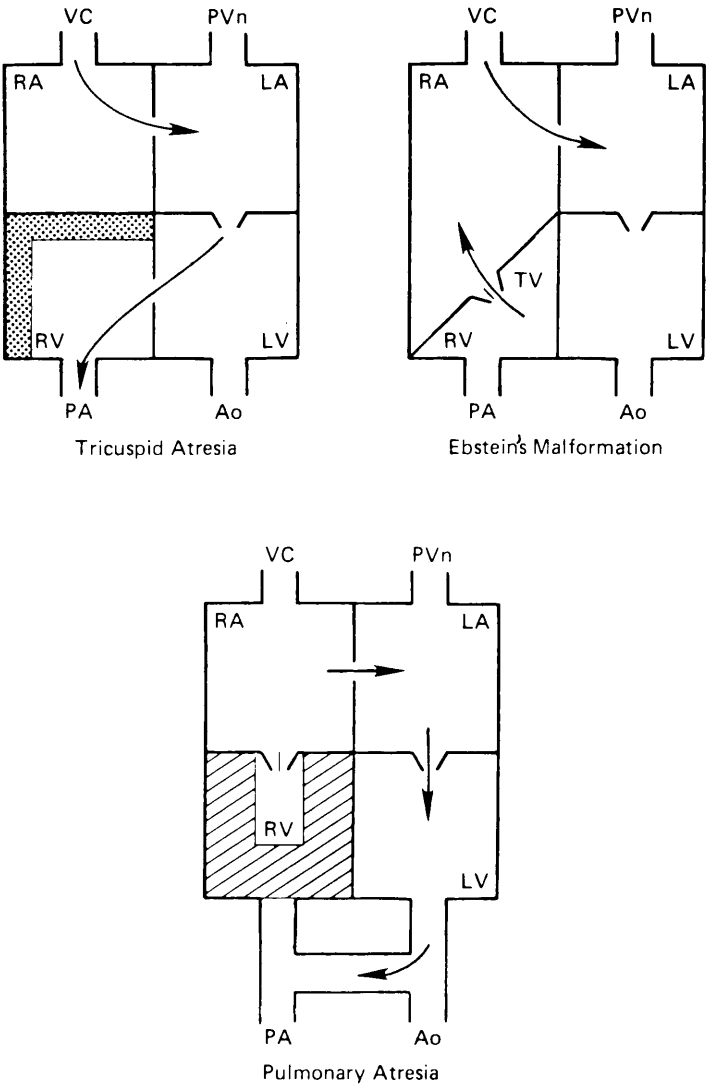


FIGURE 12-8 Logical competitor set for Case 4: pulmonary atresia, tricuspid atresia, and Ebstein's malformation

the tricuspid valve (EBST). Figure 12-8 depicts the anatomical abnormalities within the heart that define each of these diseases. In pulmonary atresia and tricuspid atresia, the pulmonary and tricuspid valves respectively are "shut" (only tissue exists where the valves should be). In Ebstein's malformation, a diminutive and noncompliant right ventricle (RV) restricts inflow of blood to that ventricle. The net physiology of all these diseases

is one of obstruction to blood flow on the right side of the heart, resulting in reduced blood flow to the lungs and right-to-left shunting of blood at the atrial level within the heart. The right-to-left shunting and diminished blood flow to the lungs cause the patient to be cyanotic (blue skin coloration). In short, these diseases constitute a physiologic category of "cyanotic diseases with decreased pulmonary blood flow."

Pulmonary atresia is the operative (or true) disease in the case. The three members of the LCS are best discriminated on the EKG. Tricuspid atresia produces a finding of left-axis deviation on the EKG, while pulmonary atresia produces a normal EKG axis. Ebstein's, unlike the other two, produces an EKG finding of right bundle branch blocking. All other clinical manifestations of the three diseases are quite similar.

There are several key data items in the patient file presented to subjects for Case 4. The subject receives evidence of cyanosis during history and early physical examination (items 1, 3, and 8). The x-ray, item 17, contains evidence of diminished blood flow to the lungs and, with the cyanosis evidence, could enable the subject to narrow diagnosis to the three members of the LCS. The EKG, item 18, contains information to discriminate among these.

Use of the Logical Competitor Set in Case 4

Table 12-8 shows all uses of members of the logical competitor set as hypotheses by all subjects during the course of the case. Table 12-8 shows a clear increase in the use of the full LCS from students to trainees, but no clear difference in this regard between trainees and experts. In particular, no student considered the full LCS, and two students (S1, S3) considered only one member. All four trainees and three experts (E1, E2, E3) used all of the diseases in the LCS. Two experts (E2, E3) used all three diseases more than once, while no trainee did—suggesting somewhat more active consideration of the LCS by these experts.

While both trainees and experts considered the full LCS, their patterns of use of these diseases were different. Three of the four experts used all members of the LCS at data point 17 (the x-ray) or at the immediately succeeding point where subjects reported hunches. Since item 17 is the data item that allows specification of the category "cyanotic heart diseases" into the category "cyanotic diseases with decreased pulmonary blood flow," this pattern suggests that the expert subjects were using this category. In contrast, *no* trainees used all three LCS members at either of these points, suggesting that these three diseases do not, at least to the same extent, constitute a functional diagnostic category for these subjects.

Regarding the expert diagnostic modes of precaution and extraction, expert E3 again considered all three LCS members together before the onset of data useful for discriminating among them. Expert E4 considered explicitly only pulmonary atresia, the correct disease, at data item 17. How-

TABLE 12-8 Case 4: Subjects' Use of LCS Hypotheses in Response to Patient Data Items

		Patient data items								
		History	*Physical exam*				*X-ray*		*EKG*	
Subjects/hypotheses		*HHx*	*14*	*15*	*16*	*HPEx*	*17*	*Hxray*	*18*	*HEKG*
S1	TAT						X	X	X	
S2	TAT			X				X	X	X
	PAT			X	X	X		X	X	
S3	EBST								X	X
S4	TAT						X	X	X	X
	PAT			X		X	X	X		X
T1	EBST						X			
	TAT			X		X	X	X	X	X
	PAT					X				
T2	EBST								X	
	TAT									X
	PAT			X						
T3	EBST						X			
	TAT			X						X
	PAT			X	X	X	X	X	X	X
T4	EBST									X
	TAT						X		X	
	PAT						X		X	X
E1	EBST						X			
	TAT						X	X	X	X
	PAT	X		X		X	X	X	X	X
E2	EBST						X	X		
	TAT						X	X	X	X
	PAT						X	X		X
E3*	EBST						X	X	X	X
	TAT		X				X	X	X	X
	PAT							X	X	X
E4*	PAT						X	X	X	X

Note: X indicates a subject's use of a hypothesis at the time of a patient data item. HHx, HPEx, etc. refer to points in the case where subjects are asked for hunches.

*The two experts with more than 20 years of experience.

ever, his protocol from the immediately succeeding data point, Hxray (hunches after x-ray), shows explicit consideration of the category of "cyanotic disease with decreased pulmonary blood flow" with targeting for active consideration of the particular LCS member he judged most likely:

> *E4:* At this point the picture would be more likely that of cyanotic heart disease involving decreased pulmonary blood flow. The specific defect would seem to be pulmonary atresia with intact septum.

TABLE 12-9 Case 4: Pulmonary Atresia—Final Diagnoses

	Subjects	*Diagnosis*
Students	S1	Truncus arteriosus
	S2	Hypoplastic right ventricle
	S3	Truncus arteriosus
	S4	Pulmonary atresia
Trainees	T1	Tricuspid atresia
	T2	Tricuspid atresia
	T3	Tricuspid atresia
	T4	Pulmonary atresia
Experts	E1	Pulmonary atresia
	E2	Pulmonary atresia
	E3*	Ebstein's malformation
	E4*	Pulmonary atresia

*The two experts with more than 20 years of experience.

Since no succeeding data are discrepant with this target hypothesis, his performance is consistent with the extraction mode as we have proposed it. In addition, E4's overt consideration of the LCS category here lends credence to a speculation we have made about the extraction mode in Case 1 and Case 2, that is, that the subject covertly considered the LCS category in those cases before he overtly articulated the members.

Diagnostic Errors in Case 4

Table 12-9 gives the final primary diagnoses for all subjects. The final diagnoses of the students on this case are outside the logical competitor set, and the full explanation for their performance is not transparent. However, a partial explanation can be given.

Two students (S1, S3) gave a final diagnosis of truncus arteriosus. Truncus is a congenital heart disease in which the aorta and pulmonary artery, the two great vessels that normally lead out of the heart, are merged into one large outlet vessel with one outlet valve. The single valve results in a patient finding of "single second heart sound" on auscultation as presented in Case 4. While truncus produces a single heart sound, so do a number of other diseases, including all members of the logical competitor set. It is not even necessary that only one valve exist for only a "single sound" to be produced; the same finding is produced when there are two outlet valves but the blood flow across one of them is substantially diminished—the situation in Ebstein's malformation and tricuspid atresia.

(15) The second heart sound is single and perhaps slightly increased in intensity. There is no gallop or diastolic murmur.

S1: Well, this is a significant finding because ah, the fact that the second heart sound is not split ah, suggests that ah, we could be dealing with a truncus.

S3: It could be ah, ah. There is a single outflow tract, ah. It could be truncus arteriosus. Ah, that would fit with the single S2 [second heart sound] . . . So, I'll go with number one on my list as ah, truncus arteriosus, and I'm not sure what type. I'd have to do an angio, I guess, or I mean arteriography.

FIGURE 12-9 Protocols from subjects S1 and S3 showing interpretation of "single second heart sound"—Case 4.

One explanation for the performance of students S1 and S3 is that they judged the "single sound" to be more discriminating for truncus than it really is; in particular, they did not consider the multidimensional nature of this finding—number of valves *and* flow. Some evidence for this explanation can be seen in protocols from these two subjects showing interpretations of the patient finding of a "single sound" (Figure 12-9). It is clear that this finding had a substantial influence on the final diagnoses of these subjects. If our interpretation for these subjects is correct, it would be another example of how the beginning practitioner is restricted in the number of alternative explanations he or she can bring to bear on a finding, at the level of either alternative pathophysiological causes or alternative disease explanations. In addition, the restricted explanations of novices are the highly salient or "classic" ones, since the "common trunk" that *defines* truncus greatly highlights the single sound as an expected finding in that disease.

S2, the other student who misdiagnosed Case 4, gave as a final diagnosis (hypoplastic right ventricle) one of the patient findings presented in the case (the EKG); that is, the subject used one of the patient data items as a final diagnosis. This subject suggests a kind of constraint relaxation that interacts with interpretive restrictiveness in the novice. The usual or preferred constraint on a good diagnostic explanation is that it account for much of the case data. When the novice encounters severe difficulty in meeting this constraint, he or she relaxes to accounting for a few key data items (S1, S3 above) or, in the extreme, to a data item itself, which embodies a level of physiological/disease interpretation.

The trainees and experts were nicely split on this case with most trainees (three of four) judging tricuspid atresia and most experts (three of four) judging pulmonary atresia, the correct disease. Recall that TAT and PAT are distinguishable on the axis of the EKG where TAT presents left-axis deviation and PAT presents a normal, undeviated axis. It is in the subjects' evaluations of this particular data item that we might expect to find an explanation for the performance of these two groups.

TABLE 12-10 Case 4: Evaluations of EKG Axis in Relation to Tricuspid Atresia and Pulmonary Atresia

		Hypotheses	
Subjects		*Tricuspid atresia*	*Pulmonary atresia*
Students	S1		
	S2		
	S3		
	S4	+	
Trainees	T1	+	
	T2		
	T3		
	T4	+	
Experts	E1	0	+
	E2	−	
	E3*	−	
	E4*		

Note: +, −, or 0 indicate that the subject evaluated the EKG axis as confirmatory, disconfirmatory, or ambivalent evidence, respectively, in relation to the hypothesis.

*The two experts with more than 20 years of experience.

Table 12-10 shows all explicit evaluations by subjects of the EKG axis as confirmatory (+), disconfirmatory (−), or ambivalent (0) evidence with respect to pulmonary atresia and tricuspid atresia. All subjects below the expert level who explicitly evaluated the axis with respect to either of these two diseases evaluated the axis as confirmatory evidence for tricuspid atresia. All expert subjects who explicitly evaluated the axis evaluated it as either disconfirmatory for tricuspid atresia or confirmatory for pulmonary atresia.

The EKG axis as presented in the case is +50 degrees, which technically represents left-axis deviation [for a four-day-old child, as presented in the case (Moller, 1978, p. 24)] as one would expect in tricuspid atresia. So that if one were using the textbook rule for discriminating PAT and TAT (Moller, 1978, p. 137), tricuspid atresia *would* be the diagnosis of choice in the case. However, the expert evaluations of this finding, as well as postexperimental discussions with these subjects, confirmed that the experts judged +50 degrees to be "just not far enough leftward" for tricuspid atresia and that these subjects would require the axis to be "down around zero or negative" before they would choose TAT over PAT. We see here a nice example of overly general, textbooklike rules of evaluation and clinical expectations in less experienced subjects (imprecise disease models) and pinpoint refinement of these in more experienced diagnosticians, probably just reflecting their greater clinical experience with the two diseases and the contextually dependent manifestations.

12.5. Summary

For the cases of the study, an expert *form* and an expert *substance* for diagnosis were identified. The expert form involves the full, active use of a set of physiologically similar diseases (the logical competitor set) for each case, diseases that have similar physiological structure and clinical presentation. The use of this set by the experts, generally in close proximity to the strongest cues for any member of the set, is interpreted here as evidence that these diseases constitute a unit or category in memory. Since diseases in the LCS are likely to be confused with each other, it would seem that as a "long-run" strategy of diagnosis it would be adaptive for a diagnostician to consider (give a "hearing" to) other members of the set whenever there is reason to believe any one of them is a good candidate in a case. It appears that this is what the experts do. Expert substance refers to correct data evaluations, within the logical competitor set of diseases, necessary to isolate the correct member. This is taken as evidence for precision in these subjects' models for diseases.

For the two high-level experts in the study, two distinct methods of utilizing the LCS were also identified:

1. *Precaution.* This involves the generation and use *together* as hypotheses of the full set of logical competitors, enabling them to be weighed against each other and the data.
2. *Extraction.* This method involves more aggressive focus on a member of the set, with full expansion to the remainder of the set as disconfirmatory evidence for the target member is found.

Medical students, after six weeks of training and clinical practice in the field represented by the cases, generally showed neither expert form nor expert substance. Students hardly ever considered the full LCS and focused on the "classic" members in cases that encouraged this. This suggests that LCS members, when they exist at all, are represented in a more isolated form in memory. Errors of evaluation (shared at times with intermediate-level subjects) included several types:

1. *Mundane factual errors.* These are just factual errors about which findings "go with" which diseases.
2. *Causal errors.* These are errors concerning how observable data are related to underlying physiology.
3. *Imprecise tests.* These are either overly general or overly restrictive tolerances on the range of variability allowed in an expected clinical finding for a disease.

4. *Interpretive restrictiveness.* This refers to restriction in the number of interpretations that are made of a finding. In some instances, these errors can be interpreted simply as reflecting imprecision in subjects' models for diseases, but other errors suggest a deficiency in integrating disease models or data with their underlying causal or physiological mechanisms.

The trainees in the study showed performances that at times looked very much expertlike and at other times could not be distinguished from the students. The number of trainees in each case who used the full LCS generally fell between the number of students and the number of experts. Moreover, depending on subject and case, trainees at times exhibited the types of errors discussed above for the students. The ultimate diagnoses of the trainees, unlike those of the students, were generally at least within the LCS, if not correct. This suggests that for these subjects the main problems were lack of connectedness in memory among LCS members or imprecision in knowledge necessary for discriminating LCS members correctly.

12.6 Discussion

The study demonstrates that diagnosticians' disease knowledge, a memory store of disease models and the memory organization among them, is crucial to successful diagnosis and does discriminate expert from less expert performance. The major differences that have been demonstrated among subjects concern their handling of a set of "good moves," that is, the logical competitor sets. More experienced subjects tend to consider more of the good hypotheses in a case, consider them in groups, and evaluate them correctly.

The study did not set out to show that highly experienced practitioners are better diagnosticians than novices; this should go without saying. The intent was to learn something about the medical knowledge that diagnosticians use, the way this knowledge influences performance, and the ways this knowledge changes as people acquire experience in a field. Medical students, after only six weeks of training in the field of interest, were included because these individuals represent the "starting point" in a long learning process.

12.6.1 The Nature of Knowledge Change

What has been learned about the nature of knowledge change? It seems clear that the whole learning process starts with a small set of "classic" training concepts where these include particular diseases, descriptions of

expected patient findings under these diseases, and rules for disambiguating diseases in this starting set. The learning of these training concepts is encouraged by the selection of content for inclusion in introductory training materials, that is, introductory textbooks and classroom instruction. The diseases are the common ones, the patient data descriptions are prototypic, or average, and the rules of evaluation are overly simplified. We have seen several instances where the locus of novice errors could be traced fairly directly to such statements in the introductory textbooks to which the subjects had been exposed. Although students' initial exposure is limited, it provides the cognitive "anchorage points" to enable them to benefit from the experience to follow.

With experience, the practitioner is exposed to and adds to memory additional diseases beyond the starting point set. Within psychology, the expert's "large vocabulary" of discriminable instances is now assumed (Chase and Simon, 1973). Concurrently with the addition of disease models to memory, there is an embellishment of the compositional features of a disease that are encoded in each disease model. These are features representing the disease's internal physiology and clinical presentation. The expert simply knows more defining characteristics of a disease (Rosch and Mervis, 1975). In some of our own work, we have found that expert physics problem solvers actively use "transformed" or "abstracted" features of a physics problem statement that novices do not even seem to recognize (Chi et al., 1981).

In Case 1 of the present study, there were some inexperienced subjects who did not "pick up" any aortic stenosis hypothesis until after the presentation of the critical finding of "no click." The fact that they did not return to this finding after the aortic stenosis model was engaged suggests they may have had *no* expectation regarding a click. Recall that in Case 2 of the present study some inexperienced subjects seemed to view the pulmonary stenosis issue (Table 12-6) as involving only one dimension, that is, orifice size, when in fact the problem involves the two interacting feature dimensions of size and flow. This is highly reminiscent of the "dimensional restrictiveness" or paucity of encoded problem features reported by Siegler (1976; 1978) for inexperienced problem solvers.

As an individual encodes more features of a disease, this provides opportunity for discriminating the disease into subtypes, that is, variants that differ on a particular feature (Anderson et al., 1979). As an illustration of what we mean, if a person encodes only the features of height and weight for people, he or she is quite limited in the discriminations he or she can make among people. It is clear that the disease knowledge of the highly experienced diagnostician is highly differentiated within a disease type. In the present study the case explicitly designed to assess this was Case 1, where the increasing differentiation was demonstrated. It can be noted that for Case 2, TAPVC, expert E3 raised and considered no fewer than ten different subvarieties of TAPVC, where each of these was distinguished by slight anatomical difference.

The differentiation of disease knowledge aids the development of precision in the clinical expectations associated with any particular disease model. If possible distinctions among versions of a disease are not made, that is, if they are in a sense all seen as the same thing, then the associated variability in clinical manifestations among patients will be great. However, when an expert represents in memory, say, ten different versions of TAPVC, with each of these perhaps differentiated into more specific versions by severity and age of presentation in a child, then the clinical expectations associated with each of these "micro-models" can be highly specific.

Precise clinical expectations, in turn, contribute to precise rules of evaluation for patient data. This is the difference between the "left-axis deviation" rule used by less experienced subjects in Case 4 and the experts' "down around zero or slightly negative" rule used in evaluating the EKG axis in that case with respect to tricuspid atresia (see Table 12-10 and the discussion about it). Again, in Case 1, one can see a nice example of how differentiation of a disease contributes to correct evaluation. In the protocol given in Figure 12-4, expert E3 raises the one micro-version of valvular aortic stenosis in which a click is not expected. This is the version with a pressure gradient between the left ventricle and aorta (over the valve) of greater than 100 mm, that is, "aortic stenosis of a very severe degree." Under this version, other data of the case would have been different from those presented. The expert was able to bring the appropriate (i.e., moderate severity) version of valvular aortic stenosis to bear on the evaluation and to reject it.

The embellishment of the feature set in disease models aids generalization as well as discrimination. Every additional feature represented for a disease is a potential feature of similarity with another disease; hence the potential of a generalization to "diseases that share feature *x*" exists (Anderson et al., 1979). The LCS analyses throughout this paper are taken as evidence that such groupings are pervasive in the more experienced knowledge base.

Students and novices learn some disease groupings directly (Moller, 1978, p. 46). These, like other teaching concepts, might be thought of as a set of "starting-point" disease categories. With experience and embellishment of feature sets, a diagnostician augments this initial set, often creating useful categories that "cross over" the original classic set. Case 2 from the present study is a good example. One might wonder how it is that a number of subjects on this case could generate and consider extensively the hypothesis of partial anomalous pulmonary venous connection, and never once even think of the correct disease, *total* anomalous pulmonary venous connection, a disease that even in its name is so similar. In the classic categorization of diseases, PAPVC, ASD, and ECD, three members of the LCS for this case, all go together in a category of "acyanotic heart diseases" (see Figure 12-10), while the final LCS member, TAPVC, is in a different category, "cyanotic heart diseases." One explanation for these subjects is

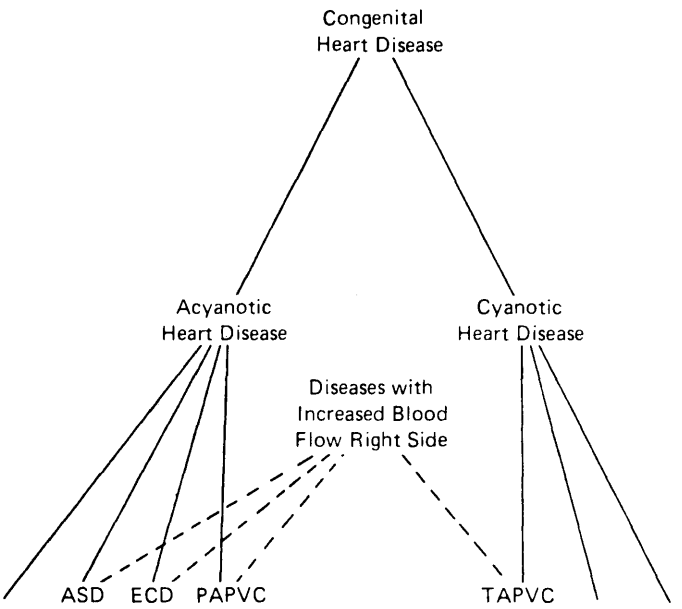


FIGURE 12-10 The classic categorization (solid lines) of the members of the logical competitor set for Case 2 and the expert regrouping (dashed lines) of these diseases.

that they became "stuck in a chunk"; that is, they were in the wrong branch of their classic hierarchy and were not able to benefit from associative triggering or hypothesis activation. The two high-level experts, on the other hand, had created a category for the LCS members that crosses the classic categorization scheme (see Figure 12-10). Creation of this category required them to represent a new disease feature, the feature of "increased blood flow on the right side."

The speculation is that many kinds of logical groupings exist for the expert, tailored to different problem contexts and even different phases of data collection, for example, "the not too sick two-day-old child" in the very early phases of diagnosis. The totality of these groupings for the expert need not be strictly hierarchical; that is, the groupings "cross over" each other in many different ways, forming more a lattice structure than a formal hierarchy (Pople, 1977).

The pervasiveness of groupings in the expert is a logical extension of the general "perceptual chunking hypothesis" of Simon and Chase (1973) and all of its ramifications (Chase and Chi, 1980). The cognitive "chunks" for an environment that people create with experience are those that serve their goals for functioning in that environment [see Egan and Schwartz, (1979) for "electronics trouble shooters"].

12.6.2 Knowledge and Problem Solving

One of the issues we set out to address with this study was the relationship between knowledge and general problem-solving processes. One way to address this issue is from a framework for problem-solving processes set out by Newell (1969). Newell proposed a power-generality dimension for problem-solving procedures. General procedures (weak methods) are those that apply widely, but offer little guarantee of success. Examples are means-ends analyses and "hill climbing." Powerful procedures (strong methods) are those that have well-specified conditions that must be met for their applicability, and hence are tailored to particular closed environments. An example is the formula for solving quadratic equations. Our work and that of others (Elstein et al., 1978) has shown that the *general* problem-solving procedure for diagnosis is one of hypothetico-deduction and that all subjects, regardless of experience, share this general approach. However, the present study has shown that this alone will not get one very far. The general process must be backed up by a rich body of accurate, well-organized medical content.

As problem-solving research has moved from semantically "lean" domains, for example, various toy problems such as the "Tower of Hanoi" and "cryptarithmetic" (Newell and Simon, 1972), to semantically rich domains, such as physics or "engineering thermodynamics" (Bhaskar and Simon, 1977), the role of domain knowledge has become increasingly important as a supplement to general procedures. We speculate that with development of disease knowledge as outlined above, corresponding sets of more powerful procedures, in Newell's sense, are concurrently created. Hence we would propose that as the diagnostician establishes various partitionings of the disease space, for example, the logical competitor sets of various kinds, he or she also establishes associated strong "local" procedures for working within abstracted regions of the space. This would mean, for instance, that the experienced diagnostician would have relatively intact or readily assembled "plans" (Sacerdoti, 1977; vanLehn and Brown, 1979) or "scripts" (Schank and Abelson, 1977) for discriminating hypotheses within conceptual groupings of various kinds and levels of generality.

While related domain knowledge is clearly critical to high-level skill in problem solving in any complex domain and, in particular, in medical diagnosis, this is still not the whole story. Knowledge must be utilized appropriately in the particular contexts where it is needed. What is happening when less experienced subjects fail to consider hypotheses (especially good ones) or evaluate data items poorly? One explanation is that knowledge is stored in memory incorrectly or not stored at all (knowledge "voids"). Another explanation concerns problems of access; subjects simply do not retrieve knowledge they need or retrieve it in some faulty manner.

Postexperimental discussions with the subjects from this study indicated that most subjects, when they failed to generate particular hypotheses or interpreted items poorly (e.g., the click in Case 1), "knew better" in some

sense. Under conditions outside the diagnostic task they could discuss subvalvular aortic stenosis or the import of the click in valvular aortic stenosis, etc. One subject called the experimenter on the day after his session, in which he had erroneously diagnosed Case 2, to tell him that the correct diagnosis had "dawned on him in the shower."

Psychology has long known that the ability to access and use knowledge that one "has" is situationally dependent (Melton, 1963; Tulving and Pearlstone, 1966). For example, knowledge that medical subjects might display on a paper and pencil test is not necessarily what they could display "on-line" in the diagnostic setting. (It was for this reason that the current study, despite its interest in knowledge, was conducted in a diagnostic context rather than in some other manner.) Yet it is this task-accessible knowledge that is crucial to successful performance.

To the extent less experienced diagnosticians have knowledge access problems, several implications for training would seem to follow: First, a disease, other diseases likely to be confused with it in a diagnostic setting, and cues for the grouping should be emphasized together in instruction and, to the extent possible, in the clinical experiences of the diagnostician in training. This encourages the memory unitization of these diseases in categories or other kinds of connected knowledge organizations. Unitization is a hedge against oversight since information in a unit has two modes of "on-line" access, associations from external events and activations directed by the unit itself (Anderson, 1980; Cohen, 1966). Because real clinical experiences are somewhat constrained by the distribution of patients in the training setting, simulated diagnostic encounters (McGuire and Solomon, 1971) could provide a vehicle for augmenting natural experience and for packaging prespecified sets of experiences. Second, tutorial instruction in the diagnostic process itself must attempt to interact with the "on-line" thought processes of the learner as he or she engages in diagnosticlike tasks. This is to help ensure that what is to be taught will be connected both to the situational cues and to the state of active memory likely to exist at some later time when the new material will be needed during a real diagnostic encounter. Expert-based instructional devices (computer-assisted instruction or decision-support systems) that contain expert knowledge and are capable of performing diagnosis in an expertlike manner could provide diagnostic practice exercises in which the device diagnoses a case in parallel with a "student," prompting alternative hypotheses when they are overlooked, correcting erroneous interpretations, and offering instruction when this seems necessary (Brown et al., 1975; Clancey, 1979c; Swanson et al., 1977; Johnson et al., 1979a) (see also Chapter 11). Finally, it would be advantageous if much of the learning of medical content for those in training could be tied as closely as possible to its conditions of ultimate use. "Problem-based learning" approaches to medical education (Barrows and Tamblyn, 1980) seem the prototype of such an endeavor. Under this type of program, much of the basic medical sub-

ject matter (e.g., physiology) that a student learns is organized within representative professional problems, including diagnosis. The problem directs what is to be learned.

12.6.3 Future Directions

Several directions for future research are suggested by the current work. The first of these is the problem of knowledge access and knowledge use. Not much is currently known about the structure of the knowledge base in memory that facilitates its situational use. Yet this is clearly a critical issue in problem solving within semantically rich domains. A second important focus is to investigate the "local" procedures or "scripts" that competent diagnosticians associate with the various partitions of the disease space that they recognize, for example, various disease and problem categories at different levels of generality such as "admixture lesions" or even "the healthy-appearing five-year-old." This appears to be the most promising avenue for studying the *procedures* and strategy of diagnosis that have hitherto been studied only at their *most* general level, that is, at the level of hypothetico-deduction. This will require a better mapping of the types of diagnostic partitions good diagnosticians use—where the current study is only a start. Finally, the current study can be viewed as one step in a cyclical research paradigm that involves experimentation and more formal cognitive simulation. The Minnesota Diagnostic Simulation Model (Swanson, 1978; Swanson et al., 1979) is a model of the expert, and its initial version was built based on studies similar to the current one. As a result of the present study, adjustments and additions to the initial expert simulation model have been made. In addition, the framework now exists for the creation of a more novice simulation. This may enable the study of learning mechanisms (Anderson et al., 1979) responsible for the transition from "noviceness" to expertise. The simulations will also direct a new cycle of more focused experimentation.

It is hoped that the present study provides some guidance for the study of problem solving in semantically rich domains. Such work requires both task-environment and knowledge-base analysis and the creation of problem-solving environments that make the interaction between the problem's information structure and the solver's knowledge structure comprehensible to the observer.

ACKNOWLEDGMENTS

The work on which this paper was based was finished while Feltovich was at the Learning Research and Development Center, University of Pittsburgh, and was supported in part by the Personnel and Training Research