SPECIAL ARTICLE

H. G. SCHMIDT, Ph.D., G. R. NORMAN, Ph.D., and H. P. A. BOSHUIZEN, Ph.D.

A Cognitive Perspective on Medical Expertise: Theory and Implications

Abstract—A new theory of the development of expertise in medicine is outlined. Contrary to existing views, this theory assumes that expertise is not so much a matter of superior reasoning skills or in-depth knowledge of pathophysiological states as it is based on cognitive structures that describe the features of prototypical or even actual patients. These cognitive structures, referred to as "illness scripts," contain relatively little knowledge about pathophysiological causes of symptoms and complaints but a wealth of clinically relevant information about disease, its consequences, and the context under which illness develops. By contrast, intermediate-level students without clinical experience typically use pathophysiological, causal models of disease when solving problems. The authors review evidence supporting the theory and discuss its implications for the understanding of five phenomena extensively documented in the clinical-reasoning literature: (1) content specificity in diagnostic performance; (2) typical differences in data-gathering techniques between medical students and physicians; (3) difficulties involved in setting standards; (4) a decline in performance on certain measures of clinical reasoning with increasing expertise; and (5) a paradoxical association between errors and longer response times in visual diagnosis. *Acad. Med.* 65 (1990): 611–621.

Nearly two decades ago, a major emphasis in medical education was the teaching and assessment of medical problem-solving *skills*. The notion was simple—physicians' daily activity was directed to the solution of patient problems, so it seemed reasonable that students should acquire the skills required to complete this task—data gathering by history taking and physical examination, data interpretation, diagnosis, clinical reasoning, management, and so on. Evaluation methods such as the patient management problem[1] evolved to assess these skills, and new curricula were developed to teach problem solving and problem-based learning.[2] These approaches to teaching and assessment had in common the assumption that clinical skills essentially consisted of a set of reasoning strategies or heuristics that were largely domain-independent and that would enable those who had acquired them to solve problems successfully, even problems that would be new to them. Basic and applied research were initiated to elucidate the components of the skills,[3,4] and related research was conducted in domains ranging from preclinical medical education to quality-of-care appraisal in practice.

Dr. Schmidt is a professor of educational psychology, Department of Educational Development and Research, University of Limburg, Maastricht, The Netherlands; Dr. Norman is a professor of medical education, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada; and Dr. Boshuizen is an associate professor in the Department of Educational Development and Research, University of Limburg, Maastricht, The Netherlands.

Correspondence and requests for reprints should be addressed to Dr. Schmidt, Department of Educational Development and Research, Faculty of Medicine, University of Limburg, Peter Debijeplein 1, 6229 HA Maastricht, The Netherlands.

Recurrent Problems

From this research, a number of recurrent problems emerged, casting doubt on some of the fundamental assumptions about the nature of clinical competence. Some of these are summarized below.

Content Specificity

A consistent finding across many studies was that correlation of performances across problems was low. This was apparently true regardless of the nature of the problems or the scoring method used.[5] Elstein and colleagues[3] labeled this phenomenon "content specificity," implying that (biomedical) content knowledge could not be separated from problem-solving heuristics as easily as had once been supposed. In fact, the low correlation across different problems seemed to indicate that problem-solving performance is highly dependent on the availability of knowledge relevant to a specific problem and that availability of knowledge for one problem does not automatically imply that adequate knowledge for another problem is also available. Interestingly, however, although this explanation has become an axiom of current research, there have been few demonstrations of a positive association between knowledge and problem solving at the level of the individual problem. Certainly, if sufficient problems are used and the sampling of knowledge is extensive, a high correlation between knowledge and problem solving can be demonstrated.[6] Nevertheless, at the level of the individual problem it has proven difficult to demonstrate a relation between the knowledge deemed necessary to solve that problem and its successful solution.[5,7]

Expertise and Data Gathering

A second basic assumption of early problem-solving research—and indeed an assumption that pervades many other domains of assessment related to competence, such as quality-of-care appraisal—is that the expert will gather more data, or, if not more data overall, will gather more of the significant, critical, or essential data, than the novice. This assumption has also found little support. Early studies of quality of care by Clute[8] and Peterson[9] showed that general practitioners did far less history taking and physical examination than was deemed necessary, and observational studies of medical students and physicians showed virtually no difference between the amounts of critical or significant data they gathered.[4] Resolution of the issue is particularly important for the use of patient management problems (PMPs), since most of the scoring options for PMPs are based on data gathering, and studies using PMPs have shown that experts gather less, rather than more, data, and as a result score lower than relative juniors.[10]

Criterion Setting

It seems self-evident that there are both an optimal approach and a minimum approach to the solution of a class of medical problems such as chest pain or otitis media. Indeed, the development of objective methods for standard setting and quality-of-care appraisal rests on this assumption. However, it appears that the development of such standards is not nearly as straightforward as might have been imagined. Friedman and colleagues[11] recently commented on the issue for standard setting in medical schools. Several years ago, Bligh[12] showed that score weights derived from random numbers (with the correct sign) did as well as expert weights in scoring PMPs, and Swanson and coworkers[13] have commented on the marginal gains derived from elaborate scoring schemes. Moreover, the problem is not only one of dealing with the various ways a clinical problem may present: Norman and colleagues[14] asked general practitioners to develop criteria for specific simulated patients and found that the same physicians did approximately half of what they recommended when the patients were introduced into their own practices.

Intermediate Effects

A common approach in the validation of assessment methods is to compare the performances of students at various levels. A consistent finding of such studies is that it is relatively easy to demonstrate differences between junior and senior students but much more difficult to show differences between final-year medical students or beginning residents and practicing clinicians.[15,16] In fact, quite often *declines* in performances on certain measures of clinical reasoning have been demonstrated.[17,18] Several hypotheses explain this intermediate effect. It is possible that indeed students possess the most formal knowledge at graduation, or that they are the most motivated then, or that they have acquired "testmanship" skills. Nevertheless, the notion that expertise should peak at graduation appears at variance with common sense. The literature does not provide a satisfactory explanation for this phenomenon; at the least it would seem likely that experts are acquiring some form of specialized knowledge that is not being assessed by the testing methods used.

Errors in Clinical Diagnosis

It is commonly assumed that errors in clinical diagnosis by experts are a result of taking shortcuts, inattention to detail, or lack of knowledge of the explicit rules. However, some recent evidence in visual domains of medicine, specifically radiology and dermatology, casts serious doubt on this assertion.[19,20] One would expect to observe inaccurate diagnoses mainly in those cases where experts make quick decisions about the nature of the problem. Errors in these visual domains, however, are associated with *longer*, not shorter, viewing times. Also, novices appear to recognize abnormalities as frequently as experts; expertise is often associated with recognizing normal variations, and with reduced rates for false positives.[21]

All these issues represent continuing dilemmas for those involved in assessment in medicine at all levels from undergraduate to continuing education. Although there are technical solutions, the most obvious being to increase the sample sizes of cases used for assessment,[22] there is no evidence yet as to why these phenomena arise. Certainly, these issues are not restricted to medicine,[23] and considerable research has been mounted in other domains of expertise to elucidate the nature of expertise.[24] Recent research in medicine has examined the development of expertise in medicine, focusing on issues of memory and using theories and methods strongly influenced by approaches in cognitive psychology. Although these studies have demonstrated many interesting performance differences between novices, intermediates, and experts, no attempt has been made to interpret these differences within one unifying framework, nor have these efforts yet resulted in new insights into the issues raised above.

We feel that a synthesis of research findings that *would* shed light on the issues raised in this introduction is now possible and that the findings can be understood in terms of a single theoretical framework. To that end, we present a theory of expertise in medicine based in part on recent research in that area but also grounded in contemporary theories in cognitive psychology. We then interpret the issues and findings raised in the first section in light of the theory. Finally, we describe some implications of the theory for medical education.

The present discussion is restricted to those investigations in medicine using cognitive methods. As such, we are deliberately excluding from the analysis artificial intelligence approaches,[25] decision analysis,[26] and policy-capturing methods.[27] This exclusion is not meant to imply that such methods do not have potential application to theory or assessment, but rather we intend to circumscribe our attention to relevant findings from the methods of cognitive psychology.

A Stage Theory of Clinical Reasoning

The theory we elaborate here rests on three assumptions. First, in acquiring expertise in medicine, students progress through several transitory stages, characterized by distinctively different knowledge structures underlying their performances. Second, these representations do not decay or become inert in the course of developing expertise but rather remain available for future use when the situation requires their activation.

Third, experienced physicians, while diagnosing routine cases, are operating upon knowledge structures that we call "illness scripts." These illness scripts emerge from continuous exposure to patients and are in that sense entirely the result of extended practice. They contain relatively little knowledge about pathophysiological causes of symptoms and complaints but a wealth of clinically relevant information about disease, its consequences, and the context in which it develops. Illness scripts exist at various levels of generality, ranging from representations of categories of diseases to representations of individual patients seen before. Indeed, one of the salient features of our theory is the presupposition that physicians actually use the memories of previous patients while diagnosing a new case.[28]

For those readers who are not familiar with psychological theories of knowledge representation, we intersperse the discourse with brief comments concerning the structure and functioning of human memory: from our theoretical perspective, memory is the overriding source of differences in diagnostic performances between medical students and physicians having different amounts of experience. We begin by describing four developmental stages that are characterized by the emergence of four different knowledge structures.

Stage 1: The Development of Elaborated Causal Networks

Cognitive psychologists generally assume that the knowledge people acquire about the world is organized or packaged into cognitive models.[29] These are mental representations at some level of abstraction of objects or events encountered in, or learned about, the outer world. They are activated whenever we are engaged in understanding a relevant part of our world, be it in reading a newspaper article on U.S.-Soviet relations, studying a text on osmosis and diffusion, watching an ongoing soccer game, or diagnosing a patient complaining of pain in the upper abdomen. These models provide meaning and structure to an otherwise chaotic influx of information and give coherence to experiences that would otherwise seem incomplete or fragmentary. In the course of their medical training, students rapidly develop rich, elaborated causal networks explaining the causes and consequences of disease in terms of general underlying pathophysiological processes.

Of particular interest to the present discussion is a special type of structure called a "propositional network."[30] It represents how things—objects, events, or concepts—are related to each other. It can be thought of as a set of nodes (representing the "things") connected by links (representing the relations). These relationships may be causal, temporal, spatial, part-whole, or family-type. Since understanding medical phenomena often involves the use of causal propositional networks, we confine our discussion to this category. Propositional networks are commonly derived from subjects' minds by requiring them to recall everything they know about a certain topic, or by asking them to explain something relevant to the domain under study.

Here is an example. In a study by Schmidt and colleagues,[18] subjects of different levels of expertise were presented with the following case description:

> A 27-year-old unemployed male was admitted to the emergency room. He complained of shaking chills and fever of 4 days' duration. He took his own temperature, and it was recorded at 40°C on the morning of his admission. The fever and chills were accompanied by sweating and a feeling of prostration. He also complained of some shortness of breath when he tried to climb the two flights of stairs in his apartment.
>
> The patient volunteered that he had been bitten by a cat at a friend's house a week before admission. Functional inquiry revealed a transient loss of vision in his right eye which lasted approximately 45 seconds. This he described the day before admission to the emergency ward.
>
> Physical examination revealed a toxic-looking young man who was having a rigor. His temperature was 41°C. Pulse was 124 per minute. BP was 110/40. Mucous membranes were clear. There were puncture wounds in his left antecubital fossa but no other skin findings.
>
> Examination showed no jugular venous distention. The pulses were collapsing, regular, and equal. The apex beat was not displaced. Auscultation of the heart revealed a 2/6 early diastolic murmur in the aortic area. There was a flame-shaped hemorrhage in the left eye. There was no splenomegaly. Urinalysis showed numerous red cells. There were no red cell casts on microscopic urinalysis.

The subjects were asked to explain the signs and symptoms described in terms of the underlying pathophysiological processes. The following is an example of a verbatim protocol produced by a fourth-year medical student (corresponding to a second-year student in the United States) when he was asked to explain the case, which is one of acute bacterial endocarditis due to intravenous drug use.

Pathophysiology Protocol of an Inexperienced Medical Student

> Probably this young man is an intravenous drug addict. And he is bitten by a cat. His resistance is not too good, so probably his immune system has not been able to eliminate the bacteria that have entered the man's body through the cat bite. As a consequence, his blood is invaded with bacteria. These bacteria produce toxins that are rendered harmless by antibodies. The complement system is activated and through this

mechanism vasoactive substances are released, such as histamine and serotonin, etc. This is what is called the "hot phase" of the septic shock. The temperature control center is disturbed as well and is reset at a higher point. The body loses much heat and slowly the "cold phase" of the anaphylactic shock is entered. This is characterized by shivers, pallor and a coldness of the extremities of the body, disseminated intravascular coagulation (because of a vicious circle in which several influences play a role, such as deteriorated oxygen supply, toxins, endothelium damage caused by a hypoxemia, etc.), I think, retinal detachment can be caused by this as well. At the moment of entrance his blood pressure was low (namely 110/40 mm Hg), then he was in the hot phase, in which reduced filling of the vascular bed caused this hypotension. Probably, the red cells in his urine can be explained by the disseminated intravascular coagulation that, accompanied with too much use of thrombocytes, causes a hemorrhagic diathesis; so this is caused by the consumption of platelets. In the end all organs are affected.

This protocol contains 42 propositions, in contrast to the protocol of an experienced student that is shown below.

This study[18] clearly demonstrated that, during the first four years of medical training (medical schools in The Netherlands offer six-year programs), these causal networks, explaining the causes and consequences of disease in terms of general underlying pathophysiological processes, become increasingly complex and elaborated as a result of learning. The learning that takes place in these first four years is largely based on books, and hence the resulting perspective on disease is rather prototypical, with only limited understanding of the variability with which disease manifests in reality.

Stage 2: The Compilation of Elaborated Networks into Abridged Ones

Through extensive and repeated application of acquired knowledge, and particularly through exposure to patient problems, these aforementioned declarative networks become compiled into high-level, simplified causal models explaining signs and symptoms and are subsumed under diagnostic labels.

This transformation begins the moment that a student is exposed to real patients. By applying his knowledge, he begins to take certain shortcuts in his reasoning. The first time he sees a patient, he will be forced to consciously relate symptoms he encounters to concepts in the relevant pathophysiological networks he possesses and to reason through the case, using whatever knowledge he has available. Diagnosing a first clinical case requires quite a lot of mental effort and involves extensive reasoning based on the elaborate causal networks available to the student, but when he sees his second or third similar case, shortcuts will emerge. He will no longer have to activate all possibly relevant knowledge in order to understand what is going on in his patient; only knowledge pertinent to understanding the case will be activated. With repeated application, this knowledge-in-use will reorganize itself so that accessibility and efficient use are assured. His original elaborate knowledge base becomes compiled into simplified causal models explaining signs and symptoms that contain only the higher-level concepts from the original pathophysiological networks.

The effects of this compilation of pathophysiological knowledge can be illustrated using material from the same study.[18] The following is the protocol of a sixth-year student (the equivalent of a fourth-year U.S. student).

Pathophysiology Protocol of an Advanced Student

Through port of entry either venous puncture or a cat bite a sepsis that secondarily produces damage to the aortic valve (endocarditis), kidneys (glomerulonephritis), and the retina (extravasation of the blood in the retina).

The effects of the compilation are evident when you compare this experienced student's protocol with the inexperienced student's protocol shown earlier. The inexperienced student used many words to explain the mechanisms that are active in shock due to sepsis (and this explanation is in some respects inadequate). In contrast, the experienced student does not mention shock—the whole case is explained in terms of sepsis and its secondary effects. An elaborate knowledge base has been compiled into a few high-level concepts and their interrelations.

Stage 3: Emergence of Illness Scripts

Simultaneously with compilation, a transition takes place from a causal type of knowledge organization to list-like structures that can be called "illness scripts." When a student sees his first patients, he will attempt to understand the signs and symptoms presented to him by consciously activating relevant pathophysiological representations that, in the act, compile. Then, after he has seen more patients with the same or comparable symptoms, further compilation takes place. The pathophysiological networks used extensively at first in the workup of patients are now gradually compiled into diagnostic labels or simplified mental models that sufficiently explain the phenomena observed (sufficiently in terms of their utility for diagnosis and treatment). By meeting many patients the student gets a feeling of how disease manifestations may vary. On the other hand, these different ways by which disease manifests merge with the abridged causal networks, and students begin to pay attention to the contextual factors under which disease emerges. Instead of causal processes, the different features that characterize the clinical appearance of a disease become the anchor points around which the physician's thinking evolves. Gradually, illness scripts for different diseases develop.

The concept of the illness script is adapted from Feltovich and Barrows.[31] According to them, physicians reorganize their knowledge of pathology, clinical manifestations of disease, variability in signs and symptoms, and the constraints under which certain diseases may occur into scripts that are as much tied together by temporal links as they are by causal relationships. Hence, the use of the word "script," which implies a scenario of events that occur in a certain order. Table 1 contains a generic illness script. "Enabling conditions" are factors that generally make the occurrence of a certain disease or family of diseases more likely. The "fault" is a description of the malfunction. It may consist of a diagnostic label or a simple (or simplified) description of a pathophysiological mechanism, for example, invasion of a pathogenic organism into body tissue. The "consequences" are the signs and symptoms that arise from the fault. An illness script in its most general format is a grammar that provides rules enabling one to construct mental models of a family of diseases (cardiovascular diseases, for instance), a specific disease (endocarditis) or even a concrete patient having the disease (the 27-year-old male).

Table 2 contains an endocarditis illness script extracted from the memory of a family physician (P. H.) with about five years' experience in general practice. The script was produced by asking the physician to retrieve whatever came to his mind with respect to the kind of enabling conditions, causal agents, symptoms, and complications one could expect in those cases.

It is assumed here that when a physician sees a patient such as the 27-year-old unemployed man in the emergency room, he searches his memory for an appropriate illness script, like the one in Table 2, and instantiates it by filling it in with the specific information provided by the present case. So, problem solving in routine cases is a process of script search, script selection, and script verification.

Although knowledge of the basic sciences puts constraints on how the various parts of a script can be related, the absence of pathophysiology in the script (apart from a simple model explaining the cause of the disease) implies that the clinician, while building a representation of the patient's problem, normally does not activate or use many (patho)physiological concepts in understanding the case. This may be the reason that experienced clinicians, while thinking aloud, solving a case presented to them, produce fewer pathophysiological remarks than do those with intermediate levels of expertise.[32,33]

An important feature of an illness script is its serial structure: items that make up the script appear in a specific order. This order closely matches the way in which physicians inform other physicians about their patients' conditions; as such, one could say that illness scripts obey certain conventions regarding an optimal story structure in medicine. But what is the psychological reality of these ordered lists? Claessen and Boshuizen[34] presented students of different levels of expertise and physicians with patient information. The cases were presented in a sequential way. After processing the items and stating a differential diagnosis, the participants were requested to recall all the items they could remember. The serial order in which they reproduced the items provided information about the way the information was stored in memory and about the memory structure that regulates storage and retrieval. The researchers were able to show that, as the participants' expertise increased, the clustering of items in the recall protocols became more in accordance with the ordering they had in the illness scripts for the cases presented. Groen and Patel[35] reported similar and consistent script-like order effects in the free recall of case histories.

A methodological weakness of these studies, however, is that the cases were presented to the participants in an order that could be considered similar to the underlying scripts, and hence the findings could also be attributed to simple input-output similarity. A study by Coughlin and Patel,[36] however, provides direct evidence for the existence of illness scripts as organizers of information in memory. They presented medical students and physicians with both organized and disorganized (random) texts describing disease histories. In recalling both the organized and the disorganized versions of the same cases, the physicians reported the information in the same script-like order. Because the texts for the disorganized cases were scrambled, there was no similarity between input order and output order for these cases. In contrast, the students were affected to a larger extent by the random presentation order. Apparently, they did not yet have strong illness scripts available for reorganizing the information presented. A similar finding was reported in a study by Norman and colleagues[37] in which

Table 1
Generic Illness Script

Illness script	
Enabling conditions ⟶	Predisposing factors, boundary conditions, hereditary factors, etc.
Predisposing factors ⟶	Compromised host factors, travel, drugs, etc.
Boundary conditions ⟶	Age, sex, etc.
Fault ⟶	Invasion of tissue by pathogenic organism, inadequate nutrient supply, inability of tissue to survive, etc.
Consequences ⟶	Complaints, signs, symptoms
Complaints ⟶	Etc.
Signs ⟶	Etc.
Symptoms ⟶	Etc.

Table 2

Endocarditis Script Derived from P. H., a Family Physician with Five Years' Experience

Enabling Conditions	Fault	Consequences
Personal features: age usually above 35; drug or alcohol addiction	Sepsis Infection of endocardium	Onset: insidious, acute health condition; malaise; weight loss
Previous health condition: decreased resistance, malnutrition, tuberculosis, long period of infection	Causative agent: staphylococcus; streptococcus; gram-negative bacillus	Fever: remittent low-high, chills, night sweats Cyanosis, anemia
Previous diseases: rheumatic fever, congenital heart disease, heart valve prosthesis, dental caries		Heart condition: failing heart (left), not responding to medication, collapsing pulse, exertional dyspnea alternating with resting dyspnea, tachycardia, overactive heart and pulse
Medication: corticosteroids		
Provoking factors: dental, urological, or gynecologic surgery; ulcerating skin lesions; osteomyelitis; monarthritis; contaminated invasive instruments or needles		Heart sounds: change in preexisting murmur; murmur of aortic insufficiency, mitral insufficiency or stenosis
		Lungs: inspiratory rales and expiratory wheezes
		Optical field defect Hematuria (red cells but no casts) Small hemorrhages in conjunctivae, nails, fingertips, and toes Spleen enlarged, painful
		ESR: normal to high WBC count: normal to high; increased proportion of young granulocytes
		Renal failure: increased urea, creatinine

students and expert physicians were asked to read and recall 20 laboratory values for a series of hypothetical patients. The laboratory data were based on coherent clusters — blood chemistry, urine chemistry, blood gases — but were presented in random order. Again, the experts' recall showed greater grouping into meaningful clusters than did the students'.

These studies suggest that physicians actually use illness scripts for storing and organizing in memory information about the conditions of patients. There is also evidence that enabling conditions do not emerge as the result of the compilation process described above but develop quite slowly in response to the constraints of daily practice. Schmidt and colleagues[38] presented experienced and novice family doctors with short case histories on three slides. The first slide showed a picture of the patient, enabling the subjects to deduce age and sex (a woman who is about 75 years old, for instance). The second slide displayed a patient chart containing information about profession, previous diseases, medication, marital status, and so forth. These slides provided the physicians with information on enabling conditions such as impaired kidney function (derived from the patient's age); risk of drug intoxication (derived from the daily application of a thiazide diuretic in combination with digoxin); and poor patient compliance (derived from a note from the patient's cardiologist). The third slide contained the complaint of the patient, for example, "Doctor, I have been unwell several times; then my heart was pounding. And I have thrown up more than once." Given this limited information, the physicians were required to state a diagnosis, digitalis intoxication in this case. The expert physicians were able to accurately diagnose 38% of the cases, whereas the novice physicians diagnosed about 27% accurately. The authors attributed this difference in diagnostic performances to a differential use of the enabling-conditions information, because the experts also remembered 40% more of this information when asked to report whatever they could remember about the patients presented. In an attempt to test this hypothesis more directly, Hobus and colleagues[39] presented expert and novice physicians with the same cases but without the enabling-conditions information. The novices performed the diagnostic task at the same level of accuracy as the novices in the original study, but the experts also performed at the novice level, suggesting that indeed experts make a more extensive use of constraining information than do physicians without extensive practical experience. In addition, in another study, Hobus and colleagues[40] showed that when asked to describe typical patients with certain diseases, experienced physicians gave richer descriptions, in particular with respect to the contextual factors that facilitate the emergence of these diseases.

Illness scripts and the way in which physicians develop them allow for the possibility that, because of different experiences, different physicians may develop quite different scripts for the same disease. In other words, the scripts are highly idiosyncratic and bear only a superficial relationship to the descriptions of "prototypical" cases as they occur in clinical textbooks. The distinction made here is the extent to which the typical signs and symptoms manifest themselves in a concrete case. In addition to the prototypical textbook cases, many disorders manifest in varied forms, making an accurate diagnosis a matter of expertise. In short, we suggest here that in the course of professional practice, and based on his or her unique experiences with a certain disease, each physician develops rich, idiosyncratic scripts for that disease, which may or may not resemble the scripts of other physicians or the textbook. This may explain why some doctors have difficulty diagnosing some diseases where others immediately recognize the essential patterns. It may also explain why extensive exposure to many different cases may be the crucial factor in developing expertise.

Stage 4: Storing Patient Encounters as Instance Scripts

The previous section described how experts make extensive use of background, contextual information in their approach to patients. The fourth component of our theory extends this argument. We suggest that memories of previous patients are retained in memory as individual entities and are not merged into some prototypical form.[41] Moreover, we hypothesize that expert clinical reasoning is, to a large extent, based on the similarity between the presenting situation and some previous patient available from memory.

As an example, one of the authors was discussing the idea of memory for individual patients with a respirologist. The respirologist proceeded to recall the story, which occurs frequently, of being awakened in the night by a resident:

> "Hello, this is Dr. Smith. We have just admitted Mr. Jones in respiratory failure."
> "Who?"
> "Mr. Jones."
> "Who is he? What does he look like?"
> "Brown hair. Scar on left ear."
> "Oh yes. Age 45. Sodium 132, potassium 5.8, Po_2 85, Pco_2 76."

The laboratory values were incorporated in the picture of the patient as much as his hair color and could be recalled from memory over periods of months or years. We suggest that this availability of a store of possible hundreds or thousands of previous patients is not simply an interesting curiosity but is instead a central feature of expertise in medicine. Recently, Hassebrock and Pretula[42] demonstrated that physicians retain vivid "autobiographical memory" for cases seen as long as 20 years earlier. Van Rossum and Bender[43] showed the impact of a single vivid case on diagnosis of similar examples two years later. Experienced physicians rarely draw on their knowledge of pathophysiology or general script-like descriptions of clinical syndromes, and only when ordinary similarity judgments fail. Rather, they usually find themselves in the position of rapidly recognizing a new presentation simply on the basis of its similarity to one previously encountered. In short, pattern recognition is not a shortcut process—it is an essential skill.

The notion that a clinician must search through many examples stored in memory to find a close match with the new presentation would seem to place intolerable demands both on the ability to store and retrieve information and on the capacity to rapidly match the presenting information with the stored data. However, this theory has been well substantiated in psychology using both artificial materials[44,45] and common, familiar objects.[46] Further, the primary role of instances in expert medical reasoning has been demonstrated.[28]

There is little doubt that in highly visual domains, such as dermatology or radiology, an important component of expertise is the accumulation of countless prior examples and their rapid and effortless retrieval when confronted with a similar situation. Although there is not yet solid evidence that these phenomena occur in the more usual verbal domain of clinical medicine, Van Rossum and colleagues[47] have showed contextual effects on diagnostic probabilities by using written protocols and simply interchanging such information as age and occupation.

The different representations we have described coexist in the mind of the physician. In other words, the ways in which a disease expresses itself in human beings are represented both as "generalized experience" in the form of illness scripts for the disease, pathophysiologic descriptions, and so forth, and as an elaborate set of lively recollections of specific patients who suffered from that disease. These representational formats have a synergistic effect. The recollections of prior patients, indexed by the relevant illness scripts, are stored in episodic memory, which makes them very easily accessible.

The four components we have presented are described as occurring in a developmental sequence, reflecting the stages of education. Students begin with an understanding of physiological processes, then acquire illness scripts as their knowledge bases begin to be compiled in the context of clinical problems, and then finally supplement these scripts with elaborated instances, which proceed to carry the normal work load with relatively little effort. Although we contend this is a reasonably accurate description of the developmental sequence, we also suggest that previously acquired knowledge structures remain available and that expert clinicians may move from one to another as the complexity of the problem demands. Thus the representations should be seen not as competing theories but as partners, available to be brought to bear on a problem. In contrast to previous investigations of expertise (such as that of Larkin[48]), we are not implying that experts work at some "deeper" level of processing but rather that expertise is associated with the availability of knowledge representations in various forms, derived from both experience and formal education. Whether the levels represent an interaction between memory characteristics and the process of instruction or some more fundamental level of organization, and what conditions result in the activation of a particular representation, remain to be determined.

Interpreting Research

The theory we have described provides an organizational framework for considering many of the conclusions reached by individual investigators. A synthesis of the main findings is given below.

Content Specificity and Problem Solving within Content Domains

As we have indicated, the change in direction that resulted in the substantial research in cognitive psychology was stimulated largely by the finding that correlations of performance across problems were consistently low. However, we suggest that the label "content specificity" is misleading. It remains to be demonstrated that, first, the successful solution of a clinical problem requires a body of biomedical knowledge or, second, that different bodies of knowledge are largely unrelated to each other in memory, both of which are prerequisites for a "content specificity" explanation. As we have indicated, some evidence points the other way.[5,7]

Our proposed stage model appears to provide an explanation. If expert (successful) problem solving depends to a large degree on script- or instance-based knowledge, it would be reasonable to expect a relatively low correlation with the relevant biomedical knowledge. Such knowledge is largely irrelevant to an expert's solution of a problem because the expert will tend to use idiosyncratic experience-based knowledge to tackle the problem. This issue is related to that of the relative absence of basic science information in clinicians' protocols.[32,33] In addition, studies in dermatology[20] highlight the importance of contextual features in the effect of prior similar instances. Thus, for instance-based knowledge, similarity in one case may be based on idiosyncratic manifestations of the disease or even such unrelated factors as occupation or age.[47] Under these circumstances, there would be no reason to expect a high correlation across different problems, however similar they may be in terms of causal pathophysiological processes underlying the symptoms presented.

Data Gathering and Expertise

Why do experts gather less data overall, and less important data, than intermediates? The phenomenon is easily understood in the context of our model. Under most circumstances, experts rapidly and routinely recognize the nature of a problem because it is similar to previous cases, and then gather additional data for confirmation. In any case, expert knowledge is available in a compiled form, focusing on critical elements of the problem. By contrast, intermediates, dealing with elaborated knowledge, have to confirm each issue suggested by the pathophysiological networks activated in response to the signs and symptoms presented to them. They are forced to consciously "think their way" through these networks and in doing so stumble over many points that need clarification, hence requesting more additional information than experts.

Criterion Setting

Although few studies document difficulty in setting performance criteria, consensus often remains difficult to achieve. Parallels also exist in the attempts to establish artificial intelligence machines by using explicit rules derived from expert judgment.[49] Such a paradox is difficult to understand if one presumes that the expert is distinguished by the possession of complex analytical rules. However, in light of the data of Allen and colleagues[28] (cited earlier) and the model we propose, the explanation is straightforward. The final stage of expertise is non-analytical, to use a formulation by Brooks,[50] and based strongly on prior instances. In this perspective, the role of individual features in expert diagnosis is secondary, and strong redundancy exists among features. As a result, since experience is idiosyncratic, it is not surprising that it is difficult to achieve consensus among experts. More fundamentally, it may be that the attempt to develop explicit rules of performance is peripheral to the essential components of expertise.

Intermediate Effects

As we have indicated, it is a common phenomenon that a test of some aspect of competence will demonstrate little or no improvement beyond the completion of medical school. This intermediate effect has parallels in the basic research in cognition. Schmidt and collaborators[18] have conducted studies of the phenomenon. They suggest that, since expert knowledge is compiled and intermediates are actively elaborating mechanisms, intermediates recall more but require more time to process the text. With unrestricted time (a typical feature of these experiments and in direct contradistinction to the chess studies), students' recall exceeds experts'. Under time restrictions, the trend reverses and experts recall more than novices. These investigators have also demonstrated that activating relevant pathophysiological knowledge enhances recall of a clinical case at the lower levels, but it suppresses or interferes with recall at more advanced levels, suggesting that the senior students they studied were no longer processing information at the level of causal mechanisms.

It should be noted that not all studies have been unsuccessful in demonstrating expert-novice differences with unrestricted time. Norman and colleagues,[37] using cases based exclusively on numerical laboratory data, showed that instructing subjects to solve the problem and then incidentally recall the individual data resulted in large differences in favor of experts. This finding may reflect the complexity of the particular task, resulting in the experts' using knowledge at an elaborated level, while the novices, faced with the large task demands, were unable to integrate beyond the dispersed level.

These findings and interpretations in the basic research, consistent with the theory we have advanced, are paralleled by the initial paradox. We would presume that, except under special conditions where the task demands are particularly heavy (such as the manipulation of electrolyte data), expert knowledge is in a compiled form and based on illness scripts. Ordinary testing methods, which explore mastery of causal mechanisms, may well put experts at a disadvantage and do not exploit their special and largely experience-based knowledge.

Expertise in Visual Domains

As shown in the studies in dermatology and the Van Rossum study,[47] instance-based reasoning is remarkably sensitive to contextual clues. Similarity can be based on both features salient to the problem and features, such as hair color or location of lesion, that are clearly unrelated. As such, it would seem that becoming an expert is, at least in part, related to an implicit understanding of the potential bias that may result. There is some evidence that experts are less vulnerable to contextual effects. Allen and colleagues[28] showed, in dermatological diagnosis, that when students were provided with a correct similar instance before the test, their performance relied on similarity, and they did no worse than general practitioners. However, when there was no similar instance, giving them a rule increased accuracy significantly. An indirect way this may be manifested is that experts in visual domains may acquire skills in dealing with normal variation. There is evidence that this is indeed the case in radiology. Myers-Worsley and coworkers[21] showed that expertise in radiology was primarily evidenced in a reduction of false-positive calls, and Lesgold[51] commented on the ability of expert radiologists to compensate for normal variations, for example, a rotated chest or poor inspiration.

Application to Teaching and Assessment

The model of cognitive development we have presented has important implications for teaching and assessment.

Teaching

The qualitative distinction between the two types of knowledge—experiential and conceptual—has potentially revolutionary implications for teaching, particularly in the clinical domain. If one views the interaction with the problem as simply an opportunity to practice one's "problem-solving skills," then any problem is equivalent to any other. If one accepts the traditional doctrine that becoming an expert is a matter of applying scientific biomedical knowledge to the solution of a patient problem, then the nature of the problem is a little more important, for it is then necessary to ensure that the problems permit practice of all the important concepts.

But the view emerging from this line of inquiry is that the problems "have a life of their own." A large part of expertise appears to consist of matching a problem with similar ones seen before. If this is so, then the number of problems, their sequence in the curriculum, and the information extracted from each become of paramount importance. Presently, students in the cardiovascular component of a problem-based curriculum know that all the chest pains they see are cardiovascular problems, not esophagitis or pleurisy, so they may not be challenged to consider how these might be differentiated. The answers are not in yet; we do not know how to assess the number or order of problems necessary to achieve competence in a domain. But it is clear that the number is far larger than one of each type; perhaps our intuition as patients to avoid practitioners who are too old *or* too young is well founded.

Student Assessment

In discussing the potential of the theory for student assessment, some of the measures used for research purposes may be useful. There are few published data on the reliability and validity of the methods, with the exception of a study by Norman[52] exploring the reliability and discrimination of a number of these measures and showing very encouraging results in comparison with conventional testing methods. In particular, some of the measures used in these research studies were able to distinguish among levels of training after graduation as well as or better than among undergraduate levels and were also able to reliably distinguish among individuals at a particular level of training.

The notion that expertise is associated with a qualitative transition from a conceptually rich and rational knowledge base to one comprised of largely experiential and non-analytical instances is a radical departure from conventional views of clinical competence. The synthesis of these two representations presents a challenge for assessment. Conventional approaches to assess problem solving as a skill have focused on the former and been frustrated in that (1) measurement is complex and time consuming and (2) most of the time most experts never do problem solving, so experts always appear to do less, get there quicker, and so on.

If we accept the assumptions that (1) there are at least two separable levels or stages—a rapid, non-analytical dimension, which is used in the majority of problems, and a slower, analytic approach, applied to a minority of problems that present difficulties; (2) neither is to be preferred, since both may lead to a solution; (3) it is not now possible to predict which kinds of problems will cause difficulty for an individual, since difficulties arise from individual experience; and (4) both types of solution strategy should be assessed; then there is an obvious strategy—staged testing. Students may be challenged with a large variety of presenting situations, each with minimal information, under time constraints, and asked to arrive at a solution as quickly as possible. The existing literature suggests that it will be possible to achieve reliable test results on this component with reasonable testing time; then, on those scenarios for which students are unable to reach a correct solution, they are allowed to pursue the problem in more detail, asking for additional information. In this component the focus is strongly on the use of underlying conceptual knowledge to reason about the problem. By restricting the second test form to those areas that are difficult for the individual, we hope to avoid the confounding of two thinking strategies.

Quality-of-Care Assessment

We have already demonstrated the difficulties that arise in establishing criteria for quality-of-care assessment. More fundamentally, this theory's perspective on medical diagnosis suggests that developing explicit rules for diagnosis, even when possible, is potentially irrelevant, since much diagnosis is based on holistic judgments about similarity. Certainly, it does seem reasonable to parallel the usual approaches to quality-of-care assessment with specific tests of this skill. Case and colleagues[53] have shown the practicality of such tests at the level of licensure; perhaps such tests could be extended into specialty certification and recertification.

This does not, of course, presume that all attempts to establish formal review criteria are irrelevant or futile. Certainly, in many areas of laboratory investigation or management, there are likely to be more explicit rules of conduct. Nevertheless, it may be more reasonable to restrict applications to these areas and to recognize the resulting deficit.

The authors acknowledge the contributions of Pie Hobus, who appears in this article as P. H., the family physician, and who gave them his personal illness script for bacterial endocarditis.

References

1. McGuire, C. H., and Babbott, D. Simulation Technique in the Measurement of Problem-Solving Skills. *J. Educ. Measur.* **4**(1967):1–10.
2. Neufeld, V. R., and Barrows, H. S. The "McMaster Philosophy:" An Approach to Medical Education. *J. Med. Educ.* **49**(1974):1040–1050.
3. Elstein, A. S., Shulman, L. S., and Sprafka, S. A. *Medical Problem-Solving: An Analysis of Clinical Reasoning.* Cambridge, Massachusetts: Harvard University Press, 1978.
4. Neufeld, V. R., Norman, G. R., Barrows, H. S., and Feightner, J. W. Clinical Problem-Solving by Medical Students: A Longitudinal and Cross-Sectional Analysis. *Med. Educ.* **15**(1981):315–322.
5. Norman, G. R., Tugwell, P., Feightner, J. W., and Muzzin, L. J. Knowledge and Clinical Problem-Solving Ability. *Med. Educ.* **19**(1985):344–356.
6. Maatsch, J. L., and Huang, R. L. An Evaluation of Four Alternative Theories of Clinical Competence. In *Research in Medical Education: 1986. Proceedings of the Twenty-Fifth Annual Conference,* pp. 69–74. Washington, D.C.: Association of American Medical Colleges, 1986.
7. Ronteltap, C. F. M., and Imbos, T. The Use of Knowledge in Diagnostic Problem-Solving. Paper presented at the TICTAC Conference, Groningen, The Netherlands, 1989.
8. Clute, K. F. *The General Practitioner.* Toronto, Ontario, Canada: University of Toronto Press, 1963.
9. Peterson, O. L., Andrews, L. P., Spain, R. S., and Greenberg, B. G. An Analytical Study of North Carolina General Practice: 1953–1954. *J. Med. Educ.* **31**, Part 2 (December 1956).
10. Marshall, J. Assessment of Problem-Solving Ability. *Med. Educ.* **11**(1977):329–334.
11. Friedman, M., Prywes, M., and Benbassat, J. J. Variability in Doctors' Problem-Solving as Measured by Open-Ended Written Patient Simulations. *Med. Educ.* **23**(1989):270–275.
12. Bligh, T. Written Simulation Scoring: A Comparison of Nine Systems. Paper presented at the American Educational Research Association meeting, New York, 1980.
13. Swanson, D. B., Norcini, J., and Grosso, L. Assessment of Clinical Competence: Written and Computer-Based Simulations. *Assess. Eval. Higher Educ.* **12**(1987): 220–246.
14. Norman, G. R., Neufeld, V. R., Woodward, C. A., McConvey, G. A., and Walsh, A. Measuring Physicians' Performances by Standardized Patients. *J. Med. Educ.* **60**(1985):925–934.
15. Friedman, R. B., et al. *Experience with the* Simulated Patient Physician Encounter. *J. Med. Educ.* **53**(1978):825–830.
16. McLeskey, C. H., and Ward, R. J. Validity of Written Examinations. *Anesthesiology* **49**(1978):224.
17. Grant, J., and Marsden, P. Primary Knowledge, Medical Education, and Consultant Expertise. *Med. Educ.* **22**(1988): 746–753.
18. Schmidt, H. G., Boshuizen, H. P. A., and Hobus, P. P. M. Transitory Stages in the Development of Medical Expertise: The "Intermediate Effect" in Clinical Case Representation Studies. In *Proc. 10th Conf. Cognitive Science Soc.,* pp. 139–145. Hillsdale, New Jersey: Erlbaum, 1988.
19. Kundel, H. L., Nodine, C. F., and Carmody, D. Visual Scanning, Pattern Recognition and Decision Making in Pulmonary Nodule Detection. *Invest. Radiol.* **13**(1978):175–181.
20. Norman, G. R., Brooks, L. R., Allen, S. W., and Rosenthal, D. The Development of Expertise in Dermatology. *Arch. Dermatol.* **125**(1989):1063–1068.
21. Myers-Worsley, M., Johnston, W. A., and Simons, M. A. The Influence of Expertise on X-ray Image Processing. *J. Exp. Psychol. Learn. Mem. Cogn.* **14**(1988):553–557.
22. Van der Vleuten, C. P. M., and Swanson, D. B. Assessment of Clinical Skills with Standardized Patients: The State of the Art. *Teach. Learn. Med.* 1990 (in press).
23. Perkins, D. N., and Salomon, G. Are Thinking Skills Context-Bound? *Educ. Researcher* **18**(1989):16–25.
24. Chi, M. T. H., Glaser, R., and Farr, M. J. *The Nature of Expertise.* Hillsdale, New Jersey: Erlbaum, 1989.
25. Clancey, W. I. Methodology for Building Intelligent Tutoring Systems. In *Methods and Tactics in Cognitive Science,* W. Kintsch, et al., eds. Hillsdale, New Jersey: Erlbaum, 1984.
26. Weinstein, M. D., et al. *Clinical Decision Analysis.* Philadelphia, Pennsylvania: W. B. Saunders, 1980.
27. Slovic, P., Rorer, L. G., and Hoffman, P. J. Analyzing Use of Diagnostic Signs. *Invest. Radio.* **6**(1971):18–26.
28. Allen, S. W., Brooks, L. R., and Norman, G. R. Effect of Prior Examples on Rule-based Diagnostic Performance. In *Research in Medical Education: 1988. Proceedings of the Twenty-Fifth Annual Conference.* Washington, D.C.: Association of American Medical Colleges, 1988.
29. Lakoff, G. Cognitive Models and Prototype Theory. In *Concepts and Conceptual Development,* U. Neisser, ed., pp. 63–102. Cambridge, England: Cambridge University Press, 1987.
30. Anderson, J. R. *Cognitive Psychology and Its Implications.* New York: Freeman, 1985.
31. Feltovich, P. J., and Barrows, H. S. Issues of Generality in Medical Problem Solving.

In *Tutorials in Problem-Based Learning: A New Direction in Teaching the Health Professions*, H. G. Schmidt and M. L. De Volder, eds., pp. 128–142. Assen, The Netherlands: Van Gorcum, 1984.
32. Boshuizen, H. P. A., Schmidt, H. G., and Coughlin, L. D. On the Application of Medical Basic Science Knowledge in Clinical Reasoning: Implications for Structural Knowledge Differences Between Experts and Novices. *Proc. 10th Conf. Cogn. Sci. Soc.*, pp. 517–523. Hillsdale, New Jersey: Erlbaum, 1988.
33. Patel, V. L., Evans, D. A., and Groen, G. J. Biomedical Knowledge and Clinical Reasoning. In *Cognitive Science in Medicine*, D. A. Evans and V. L. Patel, eds., pp. 53–112. Cambridge, Massachusetts, MIT Press, 1988.
34. Claessen, H. F. A., and Boshuizen, H. P. A. Recall of Medical Information by Students and Doctors. *Med. Educ.* **19**(1985):61–67.
35. Groen, G. J., and Patel, V. L. The Relationship Between Comprehension and Reasoning in Medical Expertise. In *The Nature of Expertise*, M. T. H. Chi, et al., eds., pp. 287–310. Hillsdale, New Jersey: Erlbaum, 1988.
36. Coughlin, L. D., and Patel, V. L. Text Comprehension and Expertise in the Domain of Medicine. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California, 1986.
37. Norman, G. R., Brooks, L. R., and Allen, S. W. Recall by Expert Medical Practitioners and Novices as a Record of Processing Attention. *J. Exp. Psychol.: Learn. Mem. Cogn.* **13**(1989):1166–1174.
38. Schmidt, H. G., Hobus, P. P. M., Patel, V. L., and Boshuizen, H. P. A. Contextual Factors in the Activation of First Hypotheses: Expert-Novice Differences. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., 1987.
39. Hobus, P. P. M., Hofstra, M. L., Boshuizen, H. P. A., and Schmidt, H. G. De context van de klatch als diagnosticum [The context of a complaint as a diagnostic tool]. *Huisarts en Wetenschap.* **31**(1988): 261–267.
40. Hobus, P. P. M., Hofstra, M. L., Boshuizen, H. P. A., and Schmidt, H. G. Mental Representation of Prototypical Patients: Expert-Novice Differences. Paper presented at the First European Congress on Psychology, Amsterdam, 1989.
41. Bordage, G., and Zacks, R. The Structure of Medical Knowledge in the Memories of Medical Students and General Practitioners: Categories and Prototypes. *Med. Educ.* **18**(1984):406–416.
42. Hassebrock, F., and Pretula, M. Autobiographical Memory in Medical Problem Solving. Paper presented at the American Educational Research Association Meeting, Boston, Massachusetts, 1990.
43. Van Rossum, H. J. M., and Bender, W. W. What Can Be Learned from a Boy with Acute Appendicitis? Persistent Effects of a Case Presentation on the Diagnostic Judgement of Family Doctors. Paper presented at the Fourth Ottawa Conference, Ottawa, Ontario, Canada, 1990.
44. Allen, S. W., and Brooks, L. R. Specializing the Operation of an Explicit Rule. *J. Exper. Psychol.: General*, 1990 (in press).
45. Whittlesea, B. W. A. Preservation of Specific Experiences in the Representation of General Knowledge. *J. Exp. Psychol: Learn. Mem. Cogn.* **13**(1987):3–17.
46. Jacoby, L. L., Baker, J. G., and Brooks, L. R. The Priority of the Specific: Episodic Effects in Picture Identification. *J. Exp. Psychol: Learn. Mem. Cogn.* **15**(1989): 275–281.
47. Van Rossum, H. J. M., Briet, E., Bender, W., and Meenders, A. E. The Transfer Effect of a Single Patient Demonstration on Diagnostic Judgement of Graduate Medical Students: Both Better and Worse. In *Proceedings of the TICTAC Conference.* Groningen, The Netherlands: TICTAC, 1989.
48. Larkin, J. H. The Role of Problem Representation in Physics. In *Mental Models*, D. Gentner and A. L. Stevens, eds., 75–98. Hillsdale, New Jersey: Erlbaum, 1983.
49. Waldrop, M. M. The Necessity of Knowledge. *Science* **223**(1984):1279–1282.
50. Brooks, L. R. Decentralized Control of Categorization: The Role of Prior Processing Episodes. In *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, U. Neisser, ed., pp. 141–174. Cambridge, England: Cambridge University Press, 1987.
51. Lesgold, A. M. Context Specific Requirements for Models for Expertise. In *Cognitive Science in Medicine*. D. A. Evans and V. L. Patel, eds., pp. 373–400. Cambridge, Massachusetts: MIT Press, 1989.
52. Norman, G. R. Reliability and Construct Validity of Some Cognitive Measures of Clinical Reasoning. *Teach. Learn. Med.* **1**(1989):194–199.
53. Case, S. M., Swanson, D. B., and Stillman, P. S. Evaluating Diagnostic Pattern Recognition: The Psychometric Characteristics of a New Item Format. In *Research in Medical Education: 1988. Proceedings of the Twenty-Seventh Annual Conference*, p. 3–8. Chicago, Illinois: Association of American Medical Colleges, 1986.