TRANSITORY STAGES IN THE DEVELOPMENT OF MEDICAL EXPERTISE: THE "INTERMEDIATE EFFECT" IN CLINICAL CASE REPRESENTATION STUDIES[1]

Henk G. Schmidt, Henny P. A. Boshuizen & Pie P. M. Hobus

University of Limburg, Maastricht, The Netherlands

One of the best-established empirical phenomena in research on the development of expertise in medicine, is the "intermediate effect " in clinical case representation studies. The experimental paradigm which produces this phenomenon is described as follows: Subjects differing in level of expertise are requested to study, for about two or three minutes, half a page of text describing a patient's history, presenting complaints and some additional findings--results of laboratory tests and/or physical examination. The text is removed, and the subjects are asked to recall everything they can remember from the text[2]. In addition, they are required to state a most likely diagnosis for the case. Subjects with intermediate levels of expertise consistently produce more extensive recalls than either experts --e.g. experienced physicians-- or novices (Patel & Groen, 1986a). This phenomenon has been demonstrated under various conditions, with different cases and in different populations. In a study using a cardiovascular respiration case, Muzzin, Norman, Feightner, & Tugwell (1983) found that internists recalled significantly fewer items from the case than did residents or pre-residency medical students. Claessen & Boshuizen (1985) presented typical and atypical cases of pancreatitis and prostatitis to family physicians and students at three levels of expertise. The residents participating in the experiment showed superior recall on all cases as compared with both the family physicians and the pre-clinical students. Patel & Medley-Mark (1985) demonstrated the same phenomenon in final year medical students as compared with both internists and novices, using an acute bacterial endocarditis case and a stomach cancer case. The intermediate effect has, in addition, been demonstrated in expertise-related tasks other than text processing (Grant & Marsden, 1988; Patel, Evans, & Kaufman, 1988). In all studies reviewed however, diagnostic accuracy of the most experienced group exceeded that of the other levels of expertise, suggesting that the intermediate effect cannot be explained away by assuming lack of ecological validity of the experimental tasks.

These results appear to be counterintuitive. Research into the way in which subjects with different levels of prior knowledge process text has shown that the richer the knowledge base used in the interpretation of new information, the better recall of that information. Spilich, Vesonder & Voss (1979), for instance, found that subjects with high knowledge of baseball remembered more, and more relevant information from a report of a baseball game than low-knowledge individuals. In the domain of chess, grandmasters were able to recall positions of pieces on the board more accurately than players with less experience (De Groot, 1946). It is generally assumed that prior knowledge provides scaffolding for the new information to be encoded and retrieved. In other words: Instead of the invertedly U-shaped curve commonly found in developmental studies in medicine, one would expect a monotonous increase of recall as a function of increasing expertise.

[1]Preparation of this paper was enabled by grant # 6266 of the Dutch Foundation for Educational Research (SVO) to H. G. Schmidt and H.P.A. Boshuizen. The authors like to acknowledge the contributions of Steve Forster and Lorence Coughlin.

[2]The resulting recall protocol is segmented into propositions. The propositions recalled, and their interrelations, are considered the subject's mental representation of the patient's problem, a representation on which his diagnosis of the underlying pathology is based. One simple way to deal with a recall protocol is to count the number of propositions correctly recalled.

Based on an extensive review of the literature, Schmidt & Norman (1988) suggest that the intermediate effect may result from the fact that experts and intermediates apply different knowledge in the representation of a clinical case. In their view, the development of expertise in medicine progresses through several transitory stages, each of which is characterized by functionally different knowledge structures underlying performance. When applied to the understanding of clinical cases, these structures produce quite different effects. Novices, by their nature, have little more than a layperson's idea of illness. Their knowledge is limited and consists mainly of an understanding of basic biological processes and structures, without much reference to the consequences of disease as exemplified in a clinical case. Students however, who enter residency, have already developed rich and elaborated causal networks explaining the signs and symptoms associated with a disease in terms of underlying pathophysiological processes, principles or mechanisms. Since, by that time, their exposure to "real" patients still has been limited, they have to process information extracted from a new case consciously and elaborately, reasoning through the causal pathophysiological networks available to them, in order to arrive at an understanding of that case. This reasoning process may take (considerable) time.
For experienced physicians however, causal pathophysiological knowledge has become compiled into diagnostic labels or simplified causal models explaining signs and symptoms, as a result of extensive use. Compiled knowledge is, by its nature, automatically and effortlessly activated by relevant cues in a case, because repeated activation in response to these same cues has caused its compilation (Anderson, 1983). There is some evidence supporting the idea of the limited use of elaborate pathophysiological explanation by experienced clinicians. Boshuizen, Schmidt & Coughlin (1987), using a think-aloud methodology, showed that general practitioners only rarely refer to pathophysiology while reasoning about a sequentially presented case. Whereas students use pathophysiological concepts extensively when explaining the information presented to them. Patel, Evans & Groen (1988), reviewing research on pathophysiological explanation of clinical cases by subjects of different expertise, also conclude that experts appear to rely less on causal biomedical knowledge while diagnosing a case. These findings support the idea that the development of expertise involves extensive compilation of knowledge acquired during training. However, Schmidt & Norman postulate that, in addition to the process of compilation, another process takes place as a result of frequent exposure to patients. Experience adds something to the knowledge base of physicians that is only superficially taught in medical school but appears to be most relevant while diagnosing a patient: The constraints under which disease occurs in humans. There is some evidence that expert behavior is determined in particular by the extent to which rich and elaborated knowledge about these constraints is acquired (Schmidt, Hobus, Patel, & Boshuizen, 1987). To accommodate this point of view, Feltovich & Barrows (1984) have suggested that in the course of years of practice, physicians develop cognitive structures of various diseases which they call "illness scripts". These illness scripts contain the physician's idiosyncratic and compiled knowledge of the disease and its consequences, in addition to knowledge of the constraints under which a disease occurs. ("Enabling conditions" is the term they use to refer to these constraints.) Illness scripts are frame- or list-like structures, containing prototypical information about a disease, which, when activated, guide a clinician through a case and support him in looking for cues that are relevant.

This theory allows for a number of predictions about the nature of differences between subjects of various levels of expertise. Of these, the following are relevant to the experiment described here:
1. Assuming that conscious application of causal pathophysiological knowledge in explaining all the information embedded in a case can be considered more extensive than just looking for relevant cues in order to match an applicable illness script, it may be expected that the recall of that case by intermediates will be superior to the experts' recall. This would explain the source of the intermediate effects.
2. Conscious processing, however, takes more time than automatically activating a relevant script and filling in slots. Hence, it is predicted that the intermediate effect will disappear when processing time is restricted. The experts' recall performance however, will be less affected by a

decrease in the time available to process the case, simply because experts need less time to match the symptoms to a relevant script. The script in turn, will facilitate subsequent retrieval of case-related information.

3. If intermediates usually process a case activating elaborate pathophysiological knowledge, whereas the experts only apply compiled knowledge, it is predicted that post-hoc pathophysiological explanations provided by intermediates will be more extensive than those provided by experts.

4. Prediction 3 will only apply to those circumstances in which subjects have enough time to apply causal knowledge. Thus, a decrease in processing time available will also affect the elaborateness of the post-hoc pathophysiological explanations provided by intermediates. The amount of explanation provided by the experts however, will not be influenced by manipulation of the processing time, since their knowledge is compiled and easily available, even under highly restricted processing conditions.

5. Limitation of processing time will probably affect the accuracy of the diagnostic hypotheses offered by all groups. However, the expert group will display smaller losses in accuracy for the same reasons as stated before: Experts reason less and therefore, have more time available to scan briefly the case in search of a solution.

In order to test these hypotheses, subjects having different levels of expertise were requested to study a case history under varying time constraints, recall the information, provide a diagnosis and produce a pathophysiological explanation.

METHOD

Subjects

Subjects were 120 students and physicians of the University of Limburg: 24 first year allied health sciences students, 72 medical students (24 second-, 24 fourth- and 24 sixth-year undergraduate students) and 24 internists with at least two years of experience. Each group of 24 was randomly subdivided in three groups of eight who studied the clinical case under different time constraints. Subjects received a small compensation for their participation.

Material

The materials consisted of a booklet containing a 270-word description of a clinical case and three blank response sheets. The case was a Dutch translation of the acute bacterial endocarditis case used by Patel & Groen (1986b) and consisted of 71 propositions.

Procedure

First, subjects were requested to study the case carefully. Dependent upon the experimental condition, they were given the opportunity to study the case for 3 minutes and 30 seconds (3' 30"), which was the amount of time allotted to subjects in the original Patel studies, 1 minute 15 seconds, or 30 seconds[3]. After reading the text for the duration of time allowed, the experimenter asked them to turn to the next page. This page contained the following instruction: "Would you be so kind as to write down everything you recall from the case? Write complete sentences and avoid abbreviations". Sufficient space was given to produce a free recall protocol. On the next page, subjects were requested to provide a diagnosis for the patient. The last page contained this instruction: "Describe the pathophysiology which, in your opinion, underlies the case. Write complete sentences and avoid abbreviations. Only use schematic representations if strictly necessary". Subjects were free to use as much time as they needed for the assignments.

[3]Preceding the processing of the case, subjects in the 30" condition were given the opportunity to read an unrelated text of exactly the same length to provide them with some experience in scanning a text in a very short time. This was done in order to minimize variability in the way subjects would undertake the experimental task.

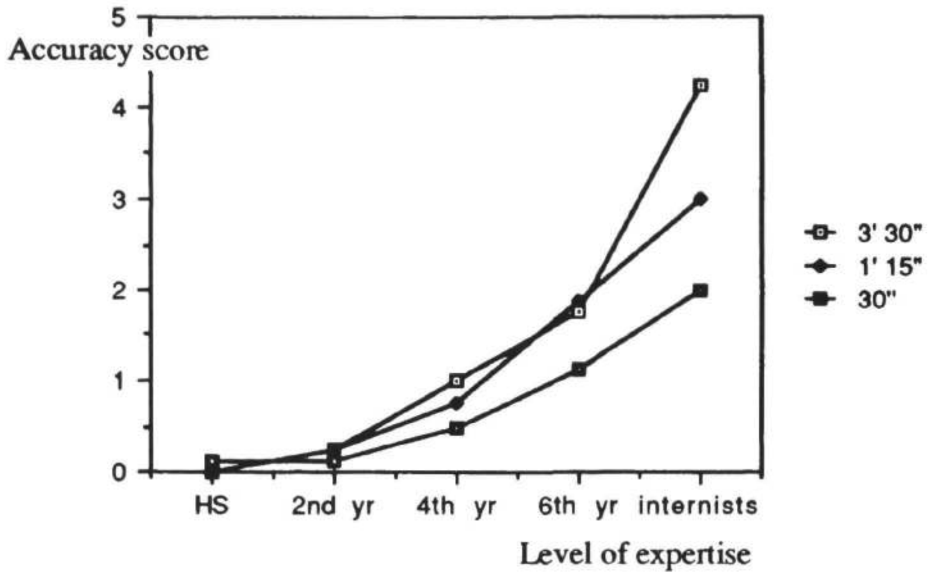

Figure 1. Average accuracy of diagnoses as a function of expertise and processing time

Subsequently, the free recall and pathophysiology protocols were segmented into propositions, applying a technique adapted from Frederiksen (1975). The number of propositions correctly recalled was recorded. In addition, the total number of propositions in the pathophysiology protocols was established. The accuracy of the diagnosis was determined attaching weights to each of the elements. If the diagnosis contained the term "endocarditis" 2 points were given. The presence of "acute", "bacterial" and "embolisms" each contributed 1 point. So, a maximum score of 5 could be obtained.

In addition, extensive qualitative analyses were carried out which are beyond the scope of this paper and whose results will not be discussed here.

RESULTS and DISCUSSION

Diagnostic accuracy

Figure 1 shows the average accuracy of the diagnoses proposed by the subjects. In line with other studies, the number of accurate diagnoses significantly covaries with expertise, resulting in the classical monotonically increasing performance curve. These data suggest that the case representation task is ecologically valid, because expertise-related differences to be expected are actually found.

Processing time, however, does not have a significant effect on diagnostic performance. This conclusion applies to all levels of expertise, including expert performance: $\underline{F}\ (2, 21)= 2.62$, $\underline{p}< .10$. These results are somewhat at variance with the expectation that, in particular, performance of subjects with less experience would suffer from the constraining of time available for processing. However, since the average scores of these groups are extremely low under all conditions, failure to find significant differences may be due to the overall difficulty of the case.

Recall data

Figure 2 shows the results of the analyses of the free recall protocols. The number of propositions recalled is displayed as a function of expertise and processing time. Overall differences between levels of expertise, and within different processing time conditions, are statistically significant ($\underline{p}< .001$). An exception is the performance of the different groups under the intermediate --1' 15"-- processing time constraint; $\underline{F}\ (4, 35)= 2.33$, $\underline{p}< .07$. Pairwise comparisons within conditions show the following pattern: Under the 3' 30" processing time condition, the internists and the health sciences students produce significantly less propositions than the three intermediate levels of expertise ($\underline{p}< .001$), nicely illustrating the existence of an intermediate effect in these data. Among

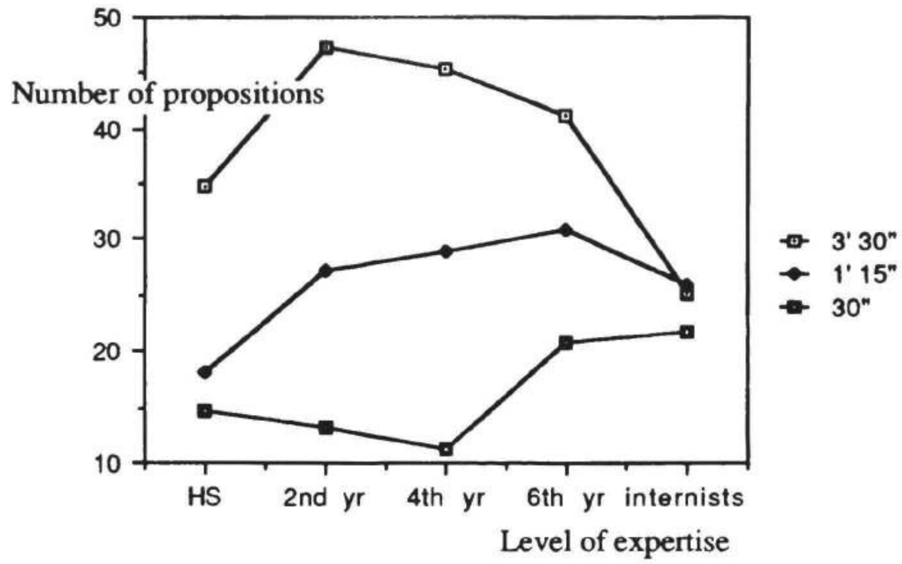

Figure 2. Number of propositions recalled from the acute bacterial endocarditis case as a function of expertise and processing time.

the groups that were required to process the case within 1 minute and 15 seconds, only the fourth- and sixth-year students produce significantly more recall, indicating that the intermediate effect, although still traceable, is vanishing. Under the shortest processing condition, the two highest levels of expertise differ significantly from the other groups, but not from each other.

In summary, as predicted by Schmidt & Norman (1988), the emergence of the intermediate effect appears to be dependent on the amount of time available for processing clinical information. Given a sufficiently short period of time, the phenomenon disappears, and is replaced by an increase in performance as a function of expertise. It is interesting to note that differences within levels of expertise as a result of differences in processing time are highly significant (all show p-values smaller than .0001), with the exception of the internists; $F(2, 21) = 1.04$, $p < .37$. These data suggest that in contrast to students of all levels of expertise, the performance of experienced physicians is relatively insensitive to manipulations of time, at least within the limits of the present experiment.

Pathophysiological protocols

The hypothesis was that intermediates and experts use functionally different knowledge while representing a clinical case. According to Schmidt & Norman (1988), intermediates consciously process the information applying a rich base of causal pathophysiological knowledge, whereas experts almost automatically activate compiled knowledge directly relevant to the case. Thus, post-hoc pathophysiological explanations provided by the intermediates will be more extensive than those provided by the experts. However, if intermediates are hindered in unfolding their knowledge, by restricting time needed to activate and reason from relevant pathophysiological networks, information from the case cannot properly be processed. Under the latter condition, the elaborateness of post-hoc explanation by intermediates is expected to decrease, whereas pathophysiological explanation provided by the experts --although compiled and therefore less extensive-- will remain constant.

Figure 3 contains quantitative information concerning this issue. Again, overall differences are highly reliable, both within levels of the experimental treatment and within each of the levels of expertise. Exceptions are the 6th year students; $F(2, 21) = 2.28$, $p < .13$; and the internists; $F(2, 21) = .40$, $p < .67$. Among students, a decrease in processing time generally causes a decrease in the number of propositions produced in the pathophysiological protocols, suggesting that less extensive processing of pathophysiological knowledge has taken place while trying to understand the case.

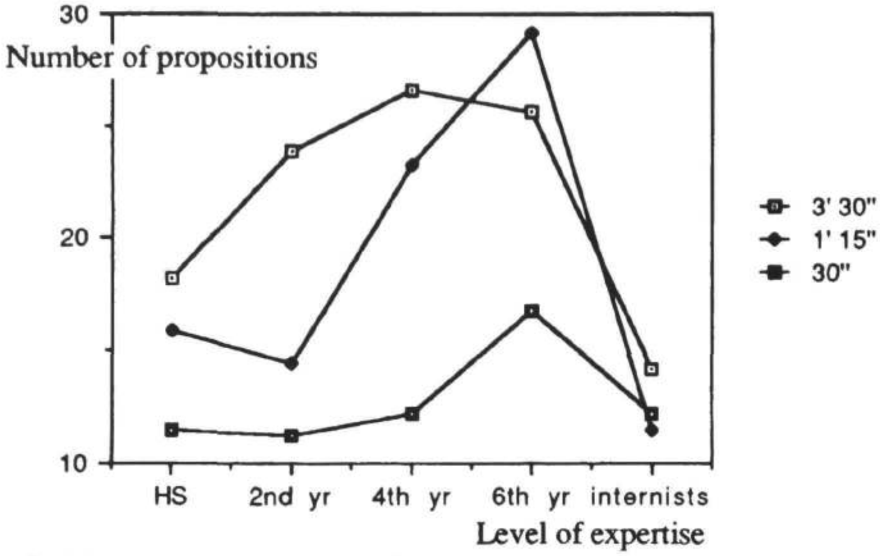

Figure 3. Number of propositions found in the pathophysiological protocols as a function of expertise and processing time

In many ways, these data resemble those of Figure 2. An intermediate effect is present, and processing time seems to have the same effect on pathophysiological reasoning as it has on recall. These similarities indeed suggest that the intermediate effect in free recall is the result of extensive and conscious processing of pathophysiological knowledge in the course of understanding the case. (Although the average within-group product-moment correlation between numbers of propositions produced in free recall and pathophysiology protocols is not impressive: $r = .34$, $p < .001$; largely due to the presence of zero correlations in two lowest expertise levels) Again, the performance of the physicians appears to be a separate issue. Their output is small and quite stable, irrespective of the constraints. Qualitative analysis shows that their protocols contain highly relevant high level explanatory concepts and compiled causal reasoning about the case. In addition, they incorporate in their explanations to a larger extent conditions that enabled the acute bacterial endocarditis to emerge (Boshuizen, 1988).

GENERAL DISCUSSION

The theory outlined rather simply explains a number of hitherto unexplained empirical phenomena, notably the intermediate effect repeatedly demonstrated in recall studies (Patel & Groen, 1986a; Muzzin, et al., 1983) and the relative absence of pathophysiological reasoning by expert-clinicians in think-aloud experiments (Boshuizen, et al, 1987; Patel, et al.,1988).

The data presented generally appear to support the notion proposed by Schmidt & Norman (1988) that students and experienced physicians represent clinical cases in different ways, because in the process of understanding the text, both groups use functionally different knowledge. Provided they have sufficient time, medical students consciously process causal pathophysiological knowledge activated by cues embedded in the text. Physicians, however, only pick up information that is directly relevant to the solution of the problem, because their knowledge of disease, its enabling conditions, causes and consequences, has been etched by continuous exposure to thousands of patients. The resultant structures are highly compiled in order to meet the demands of their profession.

Of course, the data presented only suggest a cause-effect relationship between the kind of knowledge used and recall performance. Further research is needed to establish the causal chain between the two. A particularly attractive option in this respect seems to be the experimental manipulation of prior knowledge of groups having different levels of expertise, in order to investigate effects of knowledge activation on the representation of clinical information in memory.

REFERENCES

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Boshuizen, H. P. A. (1988). De rol van kennis in de ontwikkeling van expertise in de geneeskunde (The role of knowledge in the development of expertise in medicine). Doctoral dissertation. Maastricht, The Netherlands: University of Limburg Press.

Boshuizen, H. P.A., Schmidt, H.G., & Coughlin, L.D. (1987). *On-line representation of a clinical case and the development of expertise.* Paper presented at the Annual Meeting of the American Educational Research Association. Washington, DC.

Claessen, H.F.A. & Boshuizen, H.P.A. (1985). Recall of medical information by students and doctors. *Medical Education; 19,* 61-67.

De Groot, A. D. (1946). Het denken van den schaker (Thinking processes in chess players). Den Haag: North Holland.

Feltovich, P.J. & Barrows, H.S. (1984). Issues of generality in medical problem solving. In H.G. Schmidt & M.L. de Volder. (Eds.). *Tutorials in problem-based learning.* Assen: Van Gorcum.

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

Grant, J. & Marsden, P. (1988). Primary knowledge, medical education and consultant expertise. *Medical Education, 22,* 746-753.

Muzzin, L. J., Norman, G. R., Jacoby, L. L., Feightner, J. W., Tugwell, P., & Guyatt, G. H. (1982). Manifestations of expertise in recall of clinical protocols. *Proceedings of the 21th Annual Conference on Research in Medical Education* . Washington, DC.

Muzzin, L. J. , Norman, G. R. , Feightner, J. W., & Tugwell, P. (1983). Expertise in recall of clinical protocols in two specialty areas. *Proceedings of the 22nd Conference on Research in Medical Education.* Washington, DC.

Patel, V. L., Evans, D. A. , & Groen, G. J. (1988). Biomedical knowledge and clinical reasoning. In D. A. Evans & V. L. Patel. (Eds.), *Cognitive science in Medicine.* Cambridge, MA: MIT Press.

Patel, V. L., Evans, D. A., & Kaufman, D. (1988). A cognitive framework for doctor-patient interaction. In D. A. Evans & V. L. Patel. (Eds.), *Cognitive science in Medicine.* Cambridge, MA: MIT Press.

Patel, V. L. & Medley-Mark, V. (1985). Knowledge, integration from clinical texts, use of factual, inferential and integrative questions. *Proceedings of the 24th Annual Conference of Research in Medical Education.* Washington, DC.

Patel, V. L. & Groen, G. J. (1986a). Non-monotonicity in novice-intermediate-expert comparisons. *Proceedings of the 27th Annual Meeting of the Psychonomic Society.* New Orleans, LA.

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

Schmidt, H. G., Hobus, P. P. M., Patel, V. L., & Boshuizen, H. P. A. (1987). *Contextual factors in the activation of first hypotheses: Expert-novice differences.* Paper presented at the Annual Meeting of the American Educational Research Association. Washington, DC.

Schmidt, H. G. & Norman, G. R. (1988). *On the development of expertise in medicine: Evidence from case-representation studies.* Technical Report #123. University of Limburg, Maastricht, The Netherlands.

Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18,* 352-373.