Medical Education 1987, **21**, 471–476

Contextual factors in the activation of first diagnostic hypotheses: expert–novice differences

P. P. M. HOBUS†, H. G. SCHMIDT‡, H. P. A. BOSHUIZEN‡ & V. L. PATEL§

†Department of Family Medicine and ‡Department of Educational Development and Research, University of Limburg, Maastricht and §Centre for Medical Education, McGill University, Montreal

Summary. According to Feltovich & Barrows (1984), the general frame used by medical experts to construct a mental representation of a particular patient problem contains a component part for those illness features that are associated with the acquisition of the illness. These 'enabling conditions' are related in several ways to the patient's disease. Examples are conditions like sex and age, or risk factors originating from work, behaviour and hereditary taint. Because of the sequential nature by which patient data become available during a clinical interview, contextual information is expected to play an important role in the generation of initial diagnostic hypotheses. To investigate the hypothesis that experienced doctors better utilize this kind of information, a group of 18 experts and 17 novices was confronted with 32 short case histories each presented on three slides: a portrait of the patient, the patient chart containing previous disease history, and a slide with the presenting complaint. It was hypothesized that differences in the number of correct diagnoses would emerge between the two groups, because the experts would use the contextual information, implicitly provided by picture and patient chart, in a more extensive way. If so, this would show in the amount of information that is explicitly recalled later. The data confirmed these predictions. The experts produced almost 50% more correct hypotheses as compared to the novices and were able to reproduce a larger amount of contextual information in particular information that was directly relevant to the patient's problem. Only the expert group showed a high correlation between accuracy of diagnoses and recall measures.

Correspondence: Dr P. P. M. Hobus, Department of Educational Development and Research, University of Limburg, PO Box 616, 6200 MD Maastricht, The Netherlands.

Key words: *Diagnosis; students, medical/*psychol; physicians, family/*psychol; *recall; clinical competence; information theory; problem-solving

Introduction

Investigations into expert–novice differences in the ability to solve medical diagnostic problems are useful because they may suggest ways to teach this skill. Research in this area indicates that medical experts perform better than novices on a large variety of expertise-related tasks. They are better at diagnosing dermatological diseases presented on slides (Norman *et al.* 1985), judging X-rays (Lesgold *et al.* 1981) or recalling clinical cases (Patel & Groen 1986). They are faster at processing patient information (Claessen & Boshuizen 1985) and show superior recall of texts derived from a medical textbook (Patel *et al.* 1984). However, it is largely unclear which factors mediate this superior performance on these tasks. One of the reasons for this shortcoming may be that an articulated theory of expertise in this area, despite several attempts (e.g. Elstein *et al.* 1978), is virtually absent.

The purpose of this research is not so much to demonstrate differential abilities between doctors of different levels of expertise in yet another domain, but to look for reasons why these differences arise. In particular we were interested in finding an explanation for the known differences in accuracy of diagnostic hypotheses between experts and novices that emerge during the first moments of a clinical encounter. This expertise-related feature has our special interest because the accuracy of the first hypotheses to a large extent determines the accuracy of the final differential diagnosis (Barrows *et al.* 1978, 1982; Neufeld *et al.*1981). Also, very little experimental research has been done as yet to explore factors involved in hypotheses generation during the initial problem-solving stages.

Characteristic of the early stage of an encounter is the fact that apart from the complaint with which patients confront their doctors little information is explicitly provided. In contrast, a large amount of information is implicitly available in the situation. One source of information is the patients' non-verbal behaviour and appearance, another is doctors' prior knowledge about their patients. They know the kind of diseases the patients have suffered from in the past, the drugs they use or have been using, and the nature of their work environment. This kind of information provides a context for the complaint which doctors can apply in their search for a likely diagnostic hypothesis explaining the patients' problem. Because of the sequential nature by which patient data become available during a clinical interview, this contextual information is expected to play an important role in the generation of initial hypotheses.

Feltovich & Barrows (1984) developed a theory that may apply to the present experiment. They suppose that experienced doctors use general knowledge frames (e.g. illness-scripts) to construct a cognitive representation of a particular patient's problem. These illness-scripts consist of three component parts that are interrelated with each other. Those illness features that are associated with the acquisition of illness are termed 'enabling conditions'. Instances of enabling conditions are predisposing factors like alcohol and nicotine abuse, or boundary conditions like sex and age. The contextual information discussed previously can be classified as enabling conditions. The second category of features, 'faults', contains the major malfunctions that lead to the third category of features, 'consequences'. An example is the inadequate oxygen supply (which can be considered the fault) of myocardial tissue, leading to anginal symptoms and eventually cardiac or systemic complications (which are the consequences). In the model of Feltovich & Barrows (1984) basic science knowledge (e.g. pathophysiology, anatomy) plays an integral role in the construction of the problem representation by guiding the way in which illness features are structured together.

In order to investigate the role of context in the activation of early hypotheses, family doctors and graduate or near-graduate medical students were presented with 32 short case histories, each containing a picture of a patient, the previous disease history and the present complaint. The subject read each case, and was asked to formulate a diagnosis. Reading and response time were fixed. It was expected that differences in the number of correct diagnoses would emerge between the two groups because the contextual information may be more meaningful to experienced doctors and accordingly processed in a more elaborate way. If the more experienced group uses the contextual information provided by the picture and previous disease history in a more elaborative way, leaving memory traces of greater distinctiveness, this would show in the information recalled, because more elaborative processing of information produces better recall (Anderson & Reder 1979). Therefore subjects were asked to recall all the information they considered to be important to the case.

Method

Subjects

Subjects were 35 doctors and graduate medical students. The group of experts consisted of 18 family doctors who had on average $11 \cdot 1 \pm 7 \cdot 7$ years of experience in health care. Average age was $38 \cdot 2 \pm 9 \cdot 1$ years. The novices were 12 final-year medical students who were about to graduate and five doctors who graduated less than 1 month prior to the experiment. Their average age was $25 \cdot 5 \pm 3 \cdot 3$ years.

Materials

As was already outlined, enabling conditions are not explicitly available in the early moments of a doctor–patient encounter. An attempt was made to simulate this condition by presenting the information with a portrait and medical card of the patient. In doing so, information could be given about sex, age, profession, previous disease history, etc. in an implicit way. The portrait, patient chart and the presenting complaint were presented on three separate slides. The stimulus materials consisted of 32 case histories selected from a larger set of case descriptions about patients available at the University of Limburg Medical School. The cases were selected so that contextual information could play an important role in the interpretation of the complaint. The pictures were selected from a set of 110 portraits especially produced for the experiment. They were only to convey information concerning age and sex of the patient. Therefore, only pictures were utilized on which the person showed a neutral facial expression and no signs of any disease (like exophthalmus or a drooping corner of the mouth). Previous disease history was typed on a patient chart. In addition, this chart contained information about the patient's profession, marital status, medication, family history of diseases, and risk factors. Data included both relevant and irrelevant information as related to the presenting complaint. The presenting complaint consisted of one or two sentences as expressed by the patient, for instance: 'I have had a cold fever for the past two days, doctor. I sometimes lie down shaking in my bed'. The complaints covered all organ systems.

Procedure

The 32 cases were presented in a standardized fashion. First, the subject was exposed to the portrait for about 4 seconds, then the patient chart appeared on the screen, followed by the presenting complaint. Since recall, as an important dependent variable, also depends on the duration of information input, exposure times were fixed. Some patient charts and complaints contained more information than others, therefore exposure times varied between 3 and 42 seconds for the chart and between 3 and 9 seconds for the complaint. They were established in several pilots with experts and novices. Mean exposure time per case was 32 seconds. Finally, a blank slide was projected for about 15 seconds. During that interval the subject was asked to state the *most likely* diagnosis, given the information presented. In the present experiment a one-diagnosis response was preferred rather than a differential diagnosis. By doing so an error-introducing procedure for weighing a list of possible diagnoses produced by a subject can be avoided. (A most likely diagnosis was asked for two extra reasons: Firstly, because it structures the subject's thoughts, when several possible hypotheses are considered. By adding an hierarchical element utmost utilization of given information can be expected. This was necessary to elicit the mechanisms we were in search of. And secondly, the concept of the most likely diagnosis is more than being only *one* diagnosis. It bears traces of all the alternatives in being the most likely of them).

After an instruction session, the 32 cases were presented in two series of 16 each. After both series each of the 16 presenting complaints was read back to the subject, together with the tentative diagnosis the subject had generated. Subsequently the experimenter asked the subject to recall which information embedded in the case gave rise to the particular hypothesis. All responses were audiotaped. Presentation-order of the two series was varied systematically over subjects.

Scoring

A verbatim transcription was produced from the audiotaped responses. For each subject and for each case it was determined whether the hypothesis generated was correct and how much information was retained.

Diagnostic accuracy. Because there was no 'objective' criterion for deciding whether a hypothesis is 'correct', given the restricted information provided to the subjects, the actual diagnosis of the particular patient was used as a criterion. Two judges compared the statement of the subject with the actual diagnosis. Interrater agreement for this task was 95·4%.

Accuracy and completeness of recalled information. In order to judge how much contextual information about each patient was recalled, the information on the patient chart was segmented into information units. Recall of complaint information was not included. An information unit was defined as a statement containing one singular fact or idea. The portrait of the patient and the birth-date on his chart were considered to contain two units of information: sex and age.

For each information unit it was determined whether it was relevant to the correct diagnosis or not.

The segmented case information was then compared with the verbatim transcripts of the subjects' responses. The accuracy and completeness of recalled information was determined as follows; of every unit recalled it was decided whether it could be considered literal or paraphrased recall, partial or inferred recall, or a summary of the contextual information presented. These responses were weighed. Each literal or paraphrased information unit recalled was scored as 3 points. Partial or inferred information units were scored 2 points, whereas a summary was scored as 1 point. For instance, the patient who presented with the complaint of shaking chills was a male born on 15 October 1931. He had had a haematuria, after which carcinoma of the bladder was diagnosed. This diagnosis led to a total cystectomy with diversion of the urine by ureteroileostomy. Examples of literal or paraphrased recall by subjects are: '55 years old ...' and 'they removed the bladder ...'. Examples of partial or inferred statements are: 'older patient ...' and 'bladder operation ...'. An example of a summary is 'there was something in the chart about urinary tract problems ...'.

In this way complete answers of subjects could be scored as follows. With regard to the above described patient case, expert 8 motivated his hypothesis with these statements: 'this man has had a total cystectomy, then you get ... there will be an isolated ileal loop with implanted ureters. Well, in such a case you often see ascending infections'. This subject obtained 9 points, that is to say, 3 for 'man', 3 for 'total cystectomy' and 3 for 'isolated ileal loop with implanted ureters'. Expert 1 was much shorter with his motivation: 'I believe there was something on the chart about urinary tract problems, that's why an infection came high on my list. Beside that, fever can be very high then'. This subject scored 1 point for his summary as already exemplified above.

Interrater agreement of two judges for this task was 93·4% on one fifth of the material. The remaining protocols were scored by one judge.

Group differences in diagnostic accuracy and in recall of contextual information were analysed by means of a one-way analysis-of-variance.

Results

Table 1 shows the average number of correct hypotheses produced by experts and novices. The observed difference is statistically significant: $F(1,33)=16{\cdot}75$, $P=0{\cdot}0003$. This result indicates that experts produce more accurate first hypotheses than novices do, given a very restricted amount of information about a patient.

Recall of information summarized over two series of trials is shown in Table 2. Recall of relevant and recall of irrelevant information units were distinguished. Total amount of information units recalled by the experienced doctors exceeds that of the novices: $F(1,33)=9{\cdot}41$, $P=0{\cdot}004$. This effect is particularly apparent for the information that was relevant for a correct interpretation of the complaint; $F(1,33)=11{\cdot}25$, $P=0{\cdot}002$. The difference in recall of irrelevant information units between the two groups was not significant; $F(1,33)=3{\cdot}35$, $P=0{\cdot}076$.

Table 3 contains correlations between these recall measures and the number of correct hypotheses in each group. This table shows high correlations between total number of accurate hypotheses and recall of contextual information for experts whereas this relationship is virtually absent among novices.

Table 1. Average number of correct hypotheses, produced by experts and novices (with standard deviations)

	n	Mean	SD
Experts	18	12·11	2·52
Novices	17	8·88	2·12

$F(1,33)=16{\cdot}75$, $P=0{\cdot}0003$

Table 2. Average recall scores by experts and novices (standard deviations between brackets)

	Experts	Novices
Total recalled	216·05 (52·84)	163·35 (48·56)[1]
Relevant	180·33 (40·99)	136·24 (36·50)[2]
Irrelevant	35·74 (14·68)	27·11 (13·02)[3]

[1] $F(1,33)=9{\cdot}41$ $P=0{\cdot}004$.
[2] $F(1,33)=11{\cdot}25$ $P=0{\cdot}002$.
[3] $F(1,33)=3{\cdot}35$ $P=0{\cdot}076$.

Table 3. Product-moment correlations between number of correct hypotheses and recall measures for each group

	Correct hypotheses	
	Experts	Novices
Total recall	0·54*	0·12
Relevant recall	0·63**	0·15
Irrelevant recall	0·17	0·03

$*p<0.05$.
$**p<0.01$

Discussion

The central thesis of this research is that experts are better able to utilize information implicitly available in an information-restricted environment than novices are, even if this information is not overtly related to the complaint at hand. The data seem to confirm this hypothesis. Firstly, the experts generally produced more accurate diagnoses than the novices. It is interesting to observe how well experts are able to solve a diagnostic problem, in the absence of additional data like results of physical examination or laboratory tests. Even before history-taking they are already on the right track in almost 40% of the cases. Secondly, Table 2 shows that this superior diagnostic performance is accompanied by better recall, in particular of relevant contextual information.

Of course, the data presented do not automatically imply that experts produce better hypotheses *because* they use contextual information in a more elaborative way. Whether these outcomes reflect a causal relationship between extensive elaboration of contextual information and better diagnoses remains to be questioned. An alternative explanation for the superior performance as shown in Table 1 might be that the experts have better lists of possible diagnoses stored in memory, which are activated by the presenting complaint alone. This would imply that their guesses are better, not so much because contextual factors are critical to their performance, but because they have developed more appropriate lists of diagnoses in relation to certain sets of complaints. However, this alternative explanation cannot account for the observed expert correlations. (Because recall of complaint information is not included in the scores, in our view these figures can only express a real relationship betweeen extensive recall of contextual information and the production of good hypotheses.) Since correlations between recall and problem-solving measures are absent in the novice group, this alternative hypothesis could eventually provide an accurate model for the well-performing novice.

One might argue that a relation between accuracy of hypotheses and amount of relevant information recalled is an artefact of the experimental procedure, in which subjects were asked to recall all information they considered relevant to the diagnosis. If such an effect had mediated the experimental outcome a strong correlation between diagnostic accuracy and relevant recall would have to be expected. Such a relation happens to occur in the expert group but correlation is absent in the novice group. So, although the elaboration hypothesis at the moment appears to be the best candidate for explaining the results, further research is necessary. For instance, in order to test the 'better list' hypothesis an experiment will be designed in which contextual information is deleted for some of the experimental groups. The 'better list' hypothesis predicts experts to perform better than novices regardless whether contextual information is provided or not.

Conclusion

The knowledge activated by the contextual information in the 32 cases covers a broad range of subject matter, from disease distributions to side-effects of drugs. Generally, in the training of medical students to become diagnosticians much attention is paid to complaint exploration and physical examination. This is mainly due to the fact that most of the time training occurs in clinical settings like hospital wards. Moreover, in medical education emphasis lies on anatomic and pathophysiological knowledge in order to let the students understand the patient's symptoms and signs. However, the results of this experiment indicate that in a critical phase of the diagnostic process another kind of knowledge is used, namely that needed to understand the context of a patient.

In conclusion, the data provide some support for the notion that experienced doctors make an extensive use of contextual information while attempting to solve diagnostic problems, at least when no additional information regarding the present condition of the patient is available. They confirm the assertion of Feltovich & Barrows (1984) that enabling conditions form an essential part in the construction of a mental representation of the patient's problem and help the doctor to reduce the number of alternatives to be taken into account. The results also indicate that this skill is acquired only through years of clinical practice.

Acknowledgements

The research reported in this paper was made possible by a grant from the University of Limburg Medical Faculty to Pie Hobus, and supported in part by grants from the SVO (Dutch Foundation for Educational Research) to Henk Schmidt (6626) and the Canadian Government (Medical Research Council Grant 299-61) to Vimla Patel.

References

Anderson J.R. & Reder L.M. (1979) An elaborative processing explanation of depth of processing. In: *Level of Processing in Human Memory* (ed. by L. S. Cermak & F. I. M. Craik). Lawrence Erlbaum, Hillsdale, New Jersey.

Barrows H.S., Neufeld V.R., Feightner J.W. & Norman G.R. (1978) *An Analysis of the Clinical Methods of Medical Students and Physicians*. McMaster University, Hamilton, Ontario, Canada.

Barrows H.S., Norman G.R., Neufeld V.R. & Feightner J.W. (1982) The clinical reasoning of randomly selected physicians in general medical practice. *Clinical and Investigative Medicine* **5**, 49–55.

Claessen H.F. & Boshuizen H.P.A. (1985) Recall of medical information by students and doctors. *Medical Education* **19**, 61–7.

Elstein A.S., Shulman L.S. & Sprafka S.A. (1978) *Medical Problem Solving: An Analysis of Clinical Reasoning*. Harvard University Press, Cambridge, Mass.

Feltovich P.J. & Barrows H.S. (1984) Issues of generality in medical problem-solving. In: *Tutorials in Problem-Based Learning* (ed. by H. G. Schmidt & M. L. de Volder). Van Gorcum, Assen, the Netherlands.

Lesgold A.M., Feltovich P.J., Glaser R. & Wang Y. (1981) *The Acquisition of Perceptual Diagnostic Skill in Radiology*. Technical Report No. PDS-1. Learning Research and Development Centre, University of Pittsburgh, Pittsburgh, Pennsylvania.

Neufeld V.R., Norman G.R., Feightner J.W. & Barrows H.S. (1981) Clinical problem-solving by medical students: a cross-sectional and longitudinal analysis. *Medical Education* **15**, 315–22.

Norman G.R., Muzzin L.J. & Rosenthal D. (1985) Expert–novice differences in perception and categorization in dermatology. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Illinois.

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

Patel V.L., Ho Ping Kong H. & Mark V. (1984) The role of prior knowledge in the interpretation of medical texts by medical students and physicians. In: *Proceedings of Research in Medical Education*. Association of American Medical Colleges, Washington, DC.

Received 5 November 1986; editorial comments to authors 22 January 1987; accepted for publication 22 May 1987