

The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates

EUGÈNE J. F. M. CUSTERS, HENNY P. A. BOSHUIZEN, and HENK G. SCHMIDT
University of Limburg, Maastricht, The Netherlands

The present study investigated the influence of medical expertise, case typicality, and illness script component (enabling conditions vs. consequences) on the speed of case information processing and subjective disease probabilities. It was hypothesized that expert subjects would process case information faster than nonexpert subjects, that typical information would be processed faster than atypical information, and that an interaction would be found between expertise level, typicality, and illness script: Experts were expected to be sensitive to typicality of both illness script components, while advanced students would be sensitive only to typicality of consequences. This sensitivity would also be reflected in assigned probability estimates. The data supported the predictions concerning the effects of expertise level and typicality; it was also found that expert physicians are particularly sensitive to a combination of prototypical enabling conditions and prototypical consequences. Implications of these results for the illness script theory are discussed.

Up until about 1980, the essence of medical expertise was generally conceived of as good medical problem solving (see, e.g., Elstein, Shulman, & Sprafka, 1978). Expert physicians' diagnostic problem-solving processes, expressed by parameters like thoroughness, quality of problem-solving strategies, efficiency, number of hypotheses considered, and use of critical cues, were assumed to account for their diagnostic superiority over students or less experienced physicians. This emphasis on the quality of the medical problem-solving process was to a large extent a consequence of the general interest at that time in the formal aspects of human problem solving: the important influence of the work of Newell and Simon (Newell, Shaw, & Simon, 1958; Newell & Simon, 1972) attests to this view. Furthermore, expectations were high as to the possibility of implementing medical diagnostic knowledge in expert systems (see, e.g., Clancey & Shortliffe, 1984; Lusted, 1968; Swanson, Feltoovich, & Johnson, 1977). In line with the general design of problem-solving systems, in these early medical expert systems, a distinction was made between disease knowledge and diagnostic control knowledge (see Clancey & Shortliffe, 1984). This distinction was paralleled in medical education by efforts to teach students to become good medical problem solvers, that is, to develop their diagnostic control

knowledge, rather than to stuff them with medical facts, which were thought to become obsolete very quickly.

However, in a number of studies specifically designed to investigate problem-solving differences between outstanding and average physicians, Elstein et al. (1978) failed to find most of the relevant predicted differences in problem-solving characteristics. No differences were found between excellent and average physicians on variables like number of information search units, moment of generating first hypothesis, number of currently activated hypotheses, total number of hypotheses generated, percentage and number of cues discovered, and percentage and number of critical findings. Furthermore, generalizability across similar problems turned out to be quite low, a finding that also contradicts the view that emphasizes procedural aspects of problem solving rather than knowledge-related aspects.

Elstein et al. (1978) also found that, in contrast with the expectations, early hypothesis generation turned out to be a feature of both medical experts and preclinical students. In problem-solving terms, the function of early hypothesis generation is to constrain the search space and to reduce cognitive load. Differences between experts and nonexperts were revealed mainly by the quality of the diagnostic hypotheses generated. Indeed, Barrows, Norman, Neufeld, and Feightner (1982) found that if the correct diagnostic hypothesis is included in the initial set of working hypotheses, it is almost always recognized as the definite solution, while if the correct hypothesis is not included in this set, it is missed in about 6 out of 7 cases. Since the ability to generate correct hypotheses in an early stage of the diagnostic process is probably dependent on the way diagnostic knowledge is organized in memory, it may be assumed that this knowledge organi-

Parts of this article were presented at the annual meeting of the American Educational Research Association, Atlanta, GA, April 12–16, 1993. The authors are grateful to A. Bosman and B. van Swol from the University of Amsterdam, The Netherlands, for their permission to use the program PET. Correspondence should be addressed to H. P. A. Boshuizen, Department of Educational Development and Research, University of Limburg, P.O. Box 616, 6200 MD Maastricht, The Netherlands (e-mail: boshuizen@educ.rulimburg.nl).

zation is a critical factor in medical diagnostic expertise. This is also the upshot of the work of, among others, Patel and Groen (1986, 1991); Feltovich, Johnson, Moller, and Swanson (1984); and Barrows and Feltovich (1987). Thus, experts' diagnostic superiority should be attributed to their knowledge organization, rather than to variables associated with the problem-solving process.

Structuring of Medical Diagnostic Knowledge:

The Illness Script Concept

In cognitive psychology, there are several approaches to describing the organization of complex knowledge structures involved in problem solving: production systems (Anderson, 1983, 1993), connectionistic models (Rumelhart & McClelland, 1986), prototypes (Rosch, 1978), frames (Minsky, 1975), schemas (Bartlett, 1932/1954; Rumelhart & Ortony, 1977), mental models (Gentner & Stevens, 1983; Johnson-Laird, 1983), scripts (Bower, Black, & Turner, 1979; Schank & Abelson, 1977), individual instances (Brooks, Norman, & Allen, 1991), and cases or prior problem-solving episodes (Kolodner, 1988; Kolodner & Simpson, 1986, 1989).

In accordance with Feltovich and Barrows (1984), we propose the *illness script* as the knowledge structure most pertinent in medical diagnostic situations. An illness script is a narrative structure consisting of three main components: (1) enabling conditions, that is, patient contextual factors that influence the probability that someone has a specific disease (e.g., age, sex, previous medical history, current medication, occupation, hereditary and environmental influences, risk behavior); (2) the fault, that is, the major real malfunctions in illness, characterized in biomedical terms, like the invasion of tissue by pathogenic organisms; and (3) the consequences, that is, the different complaints, signs, and symptoms a specific fault might lead to. According to Feltovich and Barrows, an illness script is *constructed* for each individual patient; however, in line with Schank and Abelson's (1977) original script idea, we prefer to conceptualize illness scripts as precompiled packages of diagnostic and clinical knowledge describing a general sequence of events that may be expected to occur when someone suffers from a particular disease. Like the Schank and Abelson scripts, illness scripts are activated as integrated wholes and instantiated by the data available in the current case: Script instantiation consists of the substitution of default slot values and characteristics by the actual features and values of the patient. In a practical diagnostic situation, the information available in the initial stages of the diagnostic encounter determines to a large extent which illness scripts are activated and how they are instantiated. The finding of Elstein et al. (1978) that the superior quality of early diagnostic hypotheses is an important feature of excellent physicians' diagnostic performance suggests that the ability to activate the appropriate illness scripts as a consequence of the information available in the early stages of the diagnostic process is an important feature of superior diagnostic achievement.

According to the illness script theory, the better part of the information available in such an early stage of the diagnostic encounter consists of enabling conditions: the patient background and contextual factors (Hobus, Boshuizen, & Schmidt, 1990; Hobus, Schmidt, Boshuizen, & Patel, 1987; Schmidt, Norman, & Boshuizen, 1990). Although enabling conditions consist partly of nonmedical, aspecific background information, as defined by Weber, Böckenholt, Hilton, and Wallace (1993), it should be emphasized that the enabling conditions of illness scripts can also include medical and specific information, as long as this information influences the probability that a specific disease is present. Information about hereditary conditions and current medication, for example, may form part of the enabling conditions for a disease. It is also important to note that the relationship between the enabling conditions and the fault is of a psychological and probabilistic, rather than of a medical and causal nature: For example, risk behavior (e.g., alcoholism) may in general increase the possibility that a particular disease (e.g., pancreatitis) is present, even though in a particular case, these events may be unrelated. The two most important features of enabling conditions are their potential availability in an early stage of the diagnostic consultation and their probabilistic relationship to the presence of certain diseases. Of course, if a physician is not familiar with a particular patient's context and background, and no medical file is present, it will be impossible for him/her to exploit enabling conditions other than those immediately visible. Indeed, as Hobus et al. (1987, 1990) have shown, physicians may be seriously handicapped if information about a patient's background and context is missing.

In addition to the enabling conditions, some consequences may also be available in an early stage of the diagnostic encounter. The most pressing complaint is usually voiced by the patient, and occasionally, a sign may be immediately visible (e.g., yellow sclera). However, much of the relevant information about consequences will be revealed only as a result of leading questions or physical examination, and hence will become available in a much later stage of the consultation.

Findings by Hobus et al. (1987, 1988, 1990) suggest that expert physicians can arrive at quite accurate diagnoses on the basis of the scarce initial information available in an early stage of the diagnostic process (see Elstein et al., 1978) because they are much better able to use enabling conditions to activate the appropriate illness scripts than are less experienced physicians. For example, Hobus et al. (1988) found an increase of about 37% in diagnostic accuracy for recently graduated MDs as a consequence of the availability of enabling conditions (compared to a situation in which only a complaint was provided), while this increase amounted to about 75% for experienced family physicians. These findings were essentially confirmed by the Hobus et al. (1990) study.

If the patient to be diagnosed shows all or most of the expected enabling conditions and consequences, as in the

Hobus et al. (1988, 1990) studies, activation and instantiation of the appropriate illness script usually do not form a problem, particularly for experienced physicians. In fact, this is equivalent to a routine diagnostic situation. However, patients, even if they suffer from the same disease, may show a large variability as to their background, context, complaints, signs, and symptoms. Illness scripts are assumed to be particularly tuned to this real-life variation. Thus, the illness script slots allow for a certain range of values of variables, and for the absence or presence of certain features. However, this does not preclude the possibility that illness script instantiation on the basis of default or typical values and features is generally easier than illness script instantiation on the basis of less common values and features, given a particular disease.

Consequences of the Script Properties of Illness Scripts

The distinction between typical and atypical information in illness scripts has certain processing implications. Scripts are activated as integrated wholes, that is, in an all-or-none fashion. Whether script activation is described in terms of tagging atypical script actions and concepts to a generic structure (see Graesser, Woll, Kowalski, & Smith, 1980; Schank & Abelson, 1977; Smith & Graesser, 1981) or in terms of the activation of an underlying network (see Walker & Yekovich, 1984; Yekovich & Walker, 1986, 1987), the resulting predictions are highly similar. Activation and default instantiation of a script results in activation of the most typical features and instantiation of the most likely slot values; consequently, new incoming information that is highly typical for the script in question will be processed with a minimum of effort, because it is already activated. Information of a more atypical nature, on the other hand, will not coincide immediately with typical concepts or actions, or with the default values of the slots; hence, a certain amount of judgment, reflection, or elaboration will be required to adjust the current, default, script instantiation. This process may be expected to take more time than will the "smooth" processing of typical features and slot values. Thus, it may be expected that processing time for atypical script information, as evidenced by reading time, will be longer than processing time for typical script information. Indeed, Bellezza and Bower (1981) found that atypical script actions, in which objects or entities appeared that could not immediately be fit into script slots, took longer to process than typical script actions. If illness scripts can be conceived of as a specific kind of the Schank and Abelson (1977) scripts, it may be expected that the results of Bellezza and Bower (1981) will also apply to illness scripts. An important goal of this article is to test this hypothesis.

Another issue is also relevant here. In contrast with Schank and Abelson (1977) scripts, the activation and instantiation of illness scripts in practical situations is generally a probabilistic affair. One cannot always be sure that a patient indeed suffers from a particular disease (i.e., that a specific illness script is applicable). In

most studies investigating Schank and Abelson (1977) scripts, the appropriateness of the script is merely a matter of fact (e.g., "John and Mary went to a restaurant." "Jack went to the dentist."). In short, script competition is common in medical diagnosis, but not in daily social situations. Thus, information that is at variance with the activated script will be handled differently in both cases: For illness scripts, it will result in a decrease of the probability that the currently activated script is applicable, whereas for Schank and Abelson (1977) scripts, it will be tagged or explicitly activated as an unexpected, probably interesting event (see Smith & Graesser, 1981). Thus, it may be expected that illness script instantiation with atypical case information will lead to a decrease of the subjective probability estimation concerning the script's correctness. More specifically, if physicians are asked to provide a likelihood estimation with respect to a particular activated script, the value of this estimation will show a monotonic relationship with the degree of typicality of the patient described in the case. Highly typical cases will be assigned a high likelihood value (e.g., "The chances that this patient has disease X are 90% or higher."), whereas more atypical cases will receive lower probability estimates. In the present study, we will try to confirm this expectation.

Expertise Level and Illness Scripts: Development of Scripted Knowledge

The results of several studies (e.g., Davidson & Hoe, 1993; Hudson, 1988; Nelson & Gruendel, 1986) suggest that even preschool children have relatively well developed scripted knowledge structures, but for illness scripts, the situation appears to be different. There is evidence that it takes several years of practical, clinical experience for full-fledged illness scripts to develop (Custers, Boshuizen, & Schmidt, 1992; Hobus et al., 1987). In addition, the aforementioned findings of Hobus et al. (1987, 1990; see also Hofstra, Hobus, Boshuizen, & Schmidt, 1988; Schmidt, Boshuizen, & Norman, 1992) that the diagnostic performance of expert family physicians, unlike that of inexperienced physicians, improves considerably when information about the enabling conditions of a patient is presented, provide evidence for a developmental divergence of the different components of an illness script. This difference seems to be due not to the inexperienced subjects actually lacking knowledge of the enabling conditions of diseases in general, but to their inability to use this information when it comes down to diagnosing a case (see Hobus et al., 1990; Custers et al., 1992). The integration of enabling conditions into full-fledged illness scripts is something that appears to occur in a relatively late stage of the development of medical expertise: Less experienced physicians rely mainly on consequences as a source of activation of diagnostic hypotheses, whereas experienced physicians are able to exploit a combination of both enabling conditions and consequences.

It may be hypothesized that these developmental differences in illness script structure are reflected in infor-

mation processing times. Since illness scripts of less experienced physicians are much less well established, less coherent, and less integrated (Custers et al., 1992), these physicians will have to rely, at least to a certain extent, on more elaborate processing of medical information in order to comprehend and assimilate case information, even if it is highly typical. In contrast, experienced physicians' comprehension is of a much more immediate nature (see Schmidt & Boshuizen, 1993a, 1993b). Thus, it may be expected that case processing in general, and processing of typical cases in particular, will take inexperienced physicians more time than expert physicians. Atypical cases will be processed more slowly by subjects at all levels of expertise; probably, for these cases, expert physicians' processing speed will even decrease relatively more than inexperienced subjects' processing speed. On the other hand, it cannot be excluded that the effects of typicality of case information and expertise level on information processing times are additive, with subjects at different levels of expertise being about equally impeded in processing speed by atypicality of information.

However, a more detailed prediction is also possible. As suggested by Custers et al. (1992), advanced students or inexperienced physicians will be sensitive mainly to typicality of the *consequences* of a disease because their illness scripts are relatively well developed with respect to this component. Typicality of *enabling conditions*, on the other hand, will not exert a large influence on reading times for these subjects. Unlike less experienced subjects, expert physicians, whose knowledge about enabling conditions is integrated into their illness scripts, will also be sensitive to the typicality of enabling conditions and hence will show shorter reading times for cases with typical enabling conditions than for cases with atypical enabling conditions.

Similarly, experienced physicians will tend to take typicality of both enabling conditions and consequences into account in determining a probability estimate for a specific case, while less experienced physicians will be sensitive mainly to the typicality of consequences and will not attach great weight to the typicality of enabling conditions: They will be relatively insensitive to typicality of patient background and contextual factors.

An experiment was designed to test these hypotheses. Short case descriptions were constructed on the basis of a number of different diseases. Each case description consisted of a number of statements in which information about enabling conditions or consequences of the patient in question was provided. The case presentations were computer based so that we could measure reading times and record probability estimations. As typicality of both enabling conditions and consequences was experimentally manipulated, the influence of this factor on both reading times and probability estimations could be experimentally investigated. Subjects at two levels of expertise—advanced students and experienced physicians—participated. The following hypotheses were tested.

First, in general, experienced subjects were expected to process case information more rapidly than less expe-

rienced subjects. Second, completely prototypical¹ cases (i.e., cases with prototypical enabling conditions as well as prototypical consequences) will generally be processed more rapidly than completely atypical cases (i.e., cases with atypical enabling conditions as well as atypical consequences); processing times of cases with either, but not both, atypical enabling conditions or atypical consequences, will fall somewhere in between. Third, an interaction will be found between expertise level and typicality of enabling conditions: Experienced physicians will be sensitive to typicality of enabling conditions, whereas advanced students will not, or they show such sensitivity to a lesser extent. No such interaction will be expected for typicality of consequences; thus, the results will show a three-way interaction among expertise level, typicality, and illness script component: For expert subjects, the effects of typicality and illness script component on processing times will be additive, whereas for less experienced subjects, an interaction between typicality and illness script component will be found, with typicality of consequences, unlike typicality of enabling conditions, exerting an influence on processing times.

For the probability estimates, similar predictions will hold. Completely prototypical cases will receive higher probability estimates than will completely atypical cases, with the probability estimates of partly prototypical, partly atypical cases falling somewhere in between. Experienced physicians will also be more sensitive to prototypicality of enabling conditions than less experienced subjects, who will mainly take consequences into consideration in determining the probability of an instance of a specific disease.

The role of script-inconsistent information (as opposed to mere atypicality) was also investigated. For this purpose, some cases were constructed in which a patient was described as having both enabling conditions and consequences that were completely contradictory to the activated illness script. It was predicted that such cases would be processed relatively quickly, because in an early stage subjects are able to decide that the case is not an instantiation of the announced disease (i.e., it is impossible to instantiate the activated illness script with the case data). Consequently, probability estimates for these cases will be close to zero.

METHOD

Subjects

Subjects were thirty 6th-year students from the Maastricht Faculty of Medicine and 30 experienced family physicians recruited from one of three small towns in the southeastern part of The Netherlands. The 6th-year students had either completed their residency or were about to complete it; they had on the average 2 years experience in a clinical setting.² The family physicians had on the average 11.75 years experience as general practitioners, ranging from 1 to 25 years (clerkship and postgraduate training years not included). Since it takes at least 2 years of additional education after medical graduation to become a family physician, the least experienced family physician had at least 5 years experience in a practical setting (clerkship years included).

Material

Case descriptions were constructed on the basis of 24 different diseases (see Appendix A). Each case description consisted of a number of statements in which information about the patient's background, the setting (e.g., consultation hour, emergency telephone call, requests for a visit), the introductory complaint, and some symptoms and findings from physical examination was provided. The first three cases, which were used for training purposes only, were identical for all experimental conditions.

For 16 of the remaining 21 diseases, 4 different case variants were constructed. Each case variant consisted of a set of enabling conditions statements followed by a set of consequences statements. Such sets could be either prototypical or atypical. Typicality of enabling conditions and consequences was independently manipulated: Prototypical enabling conditions could be followed by either prototypical or atypical consequences, and atypical enabling conditions could be followed by either prototypical (P) or atypical (A) consequences. Case variants of these types were called PP, PA, AP, and AA, respectively; thus, the first A or P of the type indication refers to the nature of the enabling conditions, the second to the nature of the consequences. Prototypicality and atypicality of case information were determined by analyzing the responses given by the subjects in a previous study (Custers et al., 1992) in which they were asked to describe a prototypical patient with each of the diseases used in the present study. It was assumed that the enabling conditions and consequences that were mentioned by a large proportion of the experienced subjects in this study might be considered prototypical, whereas features mentioned by only few subjects would be atypical. The case variants constructed on the basis of this material were checked by three experienced physicians who were instructed to judge whether the enabling conditions and consequences were indeed prototypical or atypical for the respective diseases.

An example of a disease with its four case variants can be found in Appendix B. The total number of statements per case description ranged from 7 to 18. The four different variants of each disease case description always contained the same number of statements, although the number of words could differ somewhat between these variants. In Appendix C, information about the length of the case descriptions is given.

Since only one variant of each case could be presented to every subject, 4 different sequences of cases had to be constructed, with each of the 4 variants of any case (PP, AP, PA, and AA) appearing in only 1 of the sequences. These sequences were constructed by assigning the 16 experimental diseases randomly to 4 groups of 4 diseases. For the first group, the 4 PP cases were assigned to 1 of these sequences, the AP cases to the next sequence, and so on. This procedure was repeated for the other 3 groups of 4 diseases, under the constraint that in the end each experimental sequence consisted of 4 cases of each type: PP, AP, PA, and AA. The order in which the individual cases appeared in the sequences was randomly determined, but it was identical for the 4 sequences and remained fixed during the experiment.

For the remaining 5 diseases, so-called no-cases (NO type cases) were constructed. These were cases in which a patient was described with a disease that was completely different from that suggested by the illness script header (i.e., the presumed diagnosis). Appendix A shows the diseases on which these cases were actually based, as well as the incompatible script headings. Since there was only one variant of the NO type cases, they were the same for all 4 sequences, and they were inserted at random positions in the sequences.

To control for possible order effects, an A-form and a B-form were constructed for each of the 4 case sequences. In the A-form of each sequence, Cases 4–13 were shown first and then Cases 14–24; in the B-form, this order was reversed. Practice Cases 1–3 were always presented first. Appendix D provides a complete overview of the sequence design.

Screen texts appeared in black on a white background (Palatino 18 point); no colors were used. The text lines were centered on the screen. Using a press-button device with three buttons (a button labeled *yes*, a button labeled *no*, and an unlabeled button for proceeding through the case description), together with an experimental program called PET, made it possible to measure reading times of individual statements with an accuracy of 1 msec. The main program was written in Authorware.

Procedure

According to the order of participation, subjects in both experiential groups were assigned to 1 of the 8 case sequences. They were tested individually, the 6th-year students at the university office, and the family physicians in their consulting rooms. Subjects were seated in front of a computer screen and given a short introduction. Next, the experimenter started the session by saying, "Read the first case and decide whether this is a patient with X" (X being the disease on which the first case presentation was based). Subsequently, the experimenter started the case presentation and the first statement of the first case appeared on the screen. Subjects had to push a button on the press-button device to proceed through the case. They were instructed to read each case description as thoroughly and as fast as possible and to judge whether the patient described in the case actually suffered from the disease announced by the experimenter at the beginning of the case presentation. Every time they pushed the button to proceed through the case description, the current case statement disappeared from the screen and the reading time for that particular statement was recorded. Subsequently, the next statement was presented.

At the end of every case description, subjects first had to answer a case evaluation question, phrased as a forced-choice question: "Is this a patient with X, yes or no?" ("X" being the name of the disease announced at the beginning of the case presentation). Subjects could make their choice by pressing either a *yes* or a *no* button on the press-button device they had at their disposal. At that point, no other options were available. Next, they had to estimate the probability, expressed as a percentage value, that the presented case description actually matched a patient with that specific disease.

All reading times were measured in milliseconds. Since these times were in the order of magnitude of at least 500–600 msec, but often even lasted several seconds, between-subject differences in the time needed to perform the physical response of pressing a button might be considered negligible. Reading times were also measured for the case evaluation question, but not for the probability estimations: Deciding on the "exact" quantitative size of a likelihood was not considered part of case information processing. After the subject had made the probability estimation, a pause screen appeared. Subsequently the experimenter announced the next disease in the sequence, and started the presentation of the next case.

After finishing the task, subjects were debriefed. The main purposes of the study were explained, and subjects were given an opportunity to ask questions or to make remarks. Finally, they received a reward for participating in the study.

Analysis

Average reading times for separate case statements, for the enabling conditions and consequences parts of cases, for complete cases, and for the case evaluation questions, were computed. To control for differences in length of corresponding case statements, reading times were expressed in milliseconds per word for individual statements as well as for complete cases. Besides reading times, the mean percentage estimates for each subject over the cases of the same type were calculated.

The data on one of the NO cases (i.e., the one in which a patient actually suffering from pernicious anemia was announced as someone with iron-deficiency anemia) were discarded because many subjects apparently confused both ailments. Because of occasional disturbances during an experimental session, a small number of

reading times of individual statements had to be removed from the analysis; if this had to be done, average reading times of enabling conditions and consequences were computed using the remaining, undisturbed reading times of the case statements. For 1 subject, 3 cases, and for another subject, 1 case had to be completely removed; hence, their reading times for the particular case types were computed over 3 instead of 4 cases. Furthermore, if a *no* answer on the case evaluation question was followed by a percentage estimate higher than 50, or if a *yes* answer on this question was followed by an estimate below 50, the respective probability estimates were also excluded from the analysis.³ In all, for every subject for all case types, reaction times were available, though in a few cases these were based on fewer than 4 cases for each type. As far as the probability estimates are concerned, the data of two 6th-year students could not be included in the overall within-subject analysis because the probability estimates for 2 of the 4 cases of one specific type could not be computed, and probability estimates based on only 2 cases of a type were considered too unreliable. The data of these subjects were included in those analyses in which the missing case type data were not involved, though.

The procedure described above finally yielded the following measures per subject and per case type:

1. Average reading times for individual statements, for clusters of statements, and for complete case descriptions, expressed in milliseconds per word. Reading times for the case evaluation question are included in the average reading times for the complete case descriptions, since they are considered to reflect additional information processing of scripted information. These times were also analyzed separately as case evaluation latencies.

2. Average probability estimates expressed as percentages. The reading times of complete cases, including the case evaluation latencies, were analyzed in a 2×4 analysis of variance (ANOVA) with expertise level as the between-subjects factor and case type as the within-subject factor. The NO cases were not included in this analysis because we had no preconceived notion about the case processing times for these cases and could not exclude the possibility that they might cloud a possible interaction effect between expertise level and typicality. To test the hypothesized three-way interaction among expertise level, typicality, and illness script component, the reading times of the separate enabling conditions and consequences parts of cases were analyzed in a $2 \times 2 \times 2$ ANOVA with expertise level as the between-subjects factor and typicality and illness script component (i.e., enabling conditions or consequences) as within-subject factors. In a second $2 \times 2 \times 2$ ANOVA, the influence of typicality of preceding enabling conditions on the reading times of subsequent prototypical and atypical consequences was investigated.

The case evaluation latencies and the percentage estimates were analyzed in a 2×5 ANOVA, with expertise level as the between-subjects factor and case type as the within-subject factor. In this analysis, the NO cases were included because they were expected to yield extreme values (i.e., short case evaluation latencies and low probability estimates).

Finally, reaction times and probability estimates for the NO cases were analyzed separately, as an outside control, and contrasted with PP and AA type cases in separate 2×2 ANOVAs.

RESULTS

Case Reading Times

Analysis of variance of the average reading times for complete cases, expressed in milliseconds per word, with expertise level as the between-subjects factor and case type (the NO cases not included) as the within-subject factor showed a significant main effect of expertise level [$F(1,58) = 4.244, MS_e = 13,648.488, p < .05$],

a significant main effect of case type [$F(3,174) = 18.556, MS_e = 920.194, p < .0001$], but no significant interaction between expertise level and case type. Table 1 shows the results.

From Table 1 it can be seen that consistently for all case types, family physicians were on the average faster than 6th-year students. It took 6th-year students about 14% more time to read a case description than it did family physicians. Furthermore, reading times increased from PP type cases via AP and PA type cases to AA type cases, while the NO type cases, which were not included in this analysis, appeared to fall somewhere in between. These results support at least one important prediction of the illness script theory: Experienced physicians have well-formed illness scripts that enable them to process case information more rapidly than can 6th-year students; generally, the better the fit between the incoming information and the activated illness script, the faster this new information is processed. Therefore, completely prototypical cases are processed most rapidly, and an increasing degree of atypicality results in longer processing times.

The absence of an interaction between expertise level and case typicality for the reading times of the 4 case types suggests that typicality has to a large extent the same effect for both 6th-year students and family physicians. This finding is not in line with the illness script theory. However, what we are particularly interested in is a possible three-way interaction between expertise level, typicality, and illness script component (i.e., enabling conditions or consequences). Hence, a $2 \times 2 \times 2$ ANOVA of reading times with expertise level as the between-subjects factor and typicality and illness script as the within-subject factor was performed. Case statements that were either a mixture of enabling conditions and consequences, or that consisted of enabling conditions occurring *after* the first consequence in a case,⁴ were not included in this analysis; neither were the case evaluation latencies. Thus, in contrast to the previous 2×4 ANOVA, the present analysis involved, for each case, only "pure" enabling conditions and "pure" consequences.

The results are displayed in Figure 1. Apart from the already mentioned significant main effect of expertise level significant main effects were found for illness script

Table 1
Average Reading Times (in Milliseconds/Word) for Whole Cases as a Function of Expertise Level and Case Type

Expertise Level	Case Type					M
	PP	AP	PA	AA	NO	
Sixth-year students*	230	240	253	268	246	247
Family physicians*	190	221	224	232	218	217
M	210	230	238	250	232	232

Note—PP, enabling conditions and consequences prototypical; AP, enabling conditions atypical, consequences prototypical; PA, enabling conditions prototypical, consequences atypical; AA, both enabling conditions and consequences atypical; NO, case description inconsistent with illness script header. *N = 30.

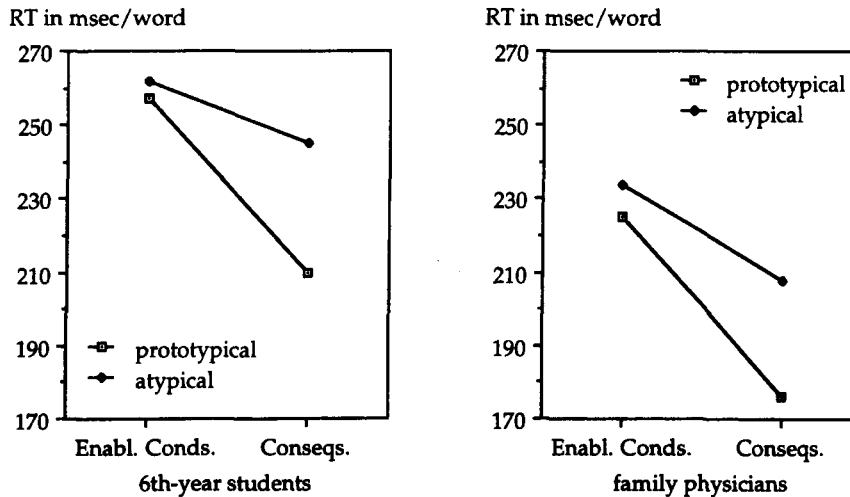


Figure 1. Reading times (RT, expressed in milliseconds/word) of individual enabling conditions and consequences case parts as a function of typicality of enabling conditions and consequences for 6th-year students (left-hand panel) and experienced family physicians (right-hand panel).

component [$F(1,58) = 92.548$, $MS_e = 797.538$, $p < .0001$] and typicality [$F(1,58) = 37.067$, $MS_e = 671.971$, $p < .0001$], and a significant interaction between typicality and illness script component was also found [$F(1,58) = 16.332$, $MS_e = 632.876$, $p < .0005$]. The two-way interactions between expertise level and illness script component and between expertise level and typicality were not significant; nor was the three-way interaction among expertise level, typicality, and illness script component. Figure 1 shows that in general, typical case parts were processed more rapidly than were atypical case parts, and consequences more rapidly than enabling conditions. From this figure, it can also be seen that the typicality \times illness script component interaction can be accounted for by prototypical consequences being processed more rapidly than were atypical consequences, while typicality of enabling conditions does not seem to have exerted a significant influence on processing speed. However, as the three-way interaction among expertise level, typicality, and illness script component was not significant, it cannot be concluded from the data that the typicality \times illness script component interaction should be attributed differentially to the two expertise levels: Apart from a general effect of expertise, reading speed of both expertise groups for enabling conditions information was scarcely influenced by typicality. Thus, the data do not suggest a processing advantage of expert physicians for prototypical enabling conditions.

However, since the experimental setup did not allow for a complete separation of enabling conditions and consequences because the enabling conditions always occurred first in the case descriptions, it is in fact not possible to independently assess the reading times of consequences statements: They will probably be influenced by the typicality of the preceding enabling conditions. For example, the processing of atypical consequences may be accelerated by the preceding prototypical

enabling conditions. Of course, any influence the other way around is excluded. Thus, in order to investigate the effect of typicality of preceding enabling conditions on the processing of subsequent consequences, an additional $2 \times 2 \times 2$ ANOVA was performed, with expertise level as the between-subjects factor, and typicality of enabling conditions and typicality of consequences as within-subject factors; the dependent variable in this analysis was the average reading time of the consequences part of the cases.

Figure 2 shows the results. The analysis revealed the expected main effects of expertise level [$F(1,58) = 4.975$, $MS_e = 15,307.430$, $p < .05$], of typicality of enabling conditions [$F(1,58) = 4.742$, $MS_e = 11,097.602$, $p < .05$], and of typicality of consequences [$F(1,58) = 42.410$, $MS_e = 1,587.712$, $p < .0001$]. In addition, a significant two-way interaction between typicality of enabling conditions and consequences [$F(1,58) = 5.559$, $MS_e = 1,477.799$, $p < .05$], and a borderline significant three-way interaction among expertise level, typicality of enabling conditions, and typicality of consequences [$F(1,58) = 3.833$, $MS_e = 1477.799$, $p < .06$] was found. Inspection of Figure 2 shows that reading times of atypical consequences were relatively independent of the typicality of the preceding enabling conditions, while reading times of prototypical consequences were influenced by typicality of the preceding enabling conditions: If the latter were prototypical, subsequent prototypical consequences were processed particularly rapidly; the two-way interaction between typicality of enabling conditions and consequences attests to this conclusion. However, a comparison of the two panels of Figure 2 also shows that this enabling conditions \times consequences interaction probably can be attributed completely to the data of the experienced physicians: For the 6th-year students, the effects of typicality of enabling conditions and consequences seem to have been additive, whereas the experi-

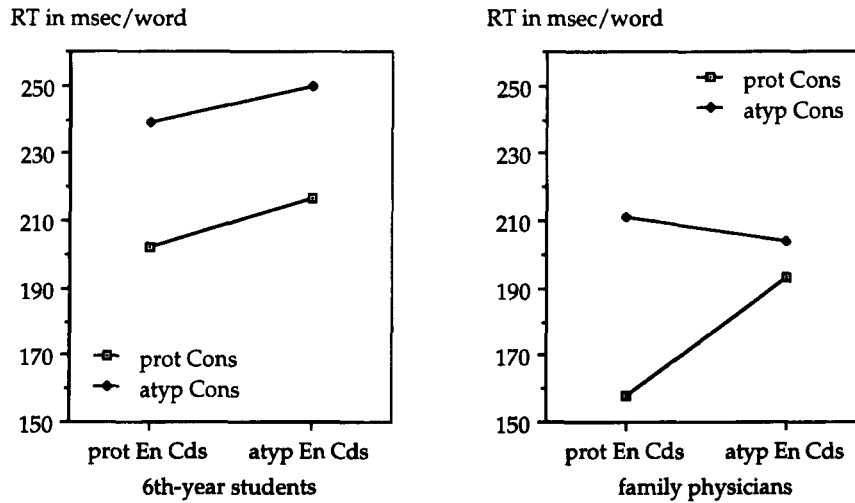


Figure 2. Reading times (RT) of prototypical and atypical consequences as a function of expertise level and typicality of the preceding enabling conditions.

enced physicians showed particularly rapid processing times for prototypical consequences preceded by prototypical enabling conditions (see the right-hand panel of Figure 2). Therefore, though we did not find evidence for the hypothesis that the influence of typicality of enabling conditions and consequences on reading times was different for subjects at different expertise levels (Figure 1), the data do suggest that prototypical enabling conditions had an accelerating effect on expert physicians' reading speed of subsequent prototypical consequences, an effect that was absent for 6th-year students. The speed by which 6th-year students processed (prototypical or atypical) consequences seems to have been comparatively independent from the nature of the preceding enabling conditions.

Finally, the influence of inconsistent information was investigated in separate analyses comparing the NO type cases with either completely prototypical or completely atypical cases. An ANOVA of reading times of complete cases, with expertise level as the between-subjects factor and case type (PP versus NO) as the within-subject factor showed a significant effect of expertise level [$F(1,58) = 4.969, MS_e = 6878.022, p < .05$] and a significant effect of case type [$F(1,58) = 15.944, MS_e = 911.881, p < .0005$], but no significant interaction. Thus, NO type cases were processed significantly more slowly than completely prototypical cases by subjects at both expertise levels. Although a similar ANOVA of AA versus NO type cases showed only a borderline significant effect of expertise level [$F(1,58) = 3.48, MS_e = 9093.184, p < .07$], again a significant effect of case type was found [$F(1,58) = 12.304, MS_e = 793.138, p < .001$], and no significant interaction. Apparently, NO type cases were processed significantly more rapidly than were AA type cases. These results suggest that, on the one hand, deciding to reject a script takes more processing time than just assimilating prototypical, script-fitting information into an already

active script, but rejecting takes less time than does processing of cases for which it remains inconclusive whether they are an instantiation of an activated script. Indeed, this conclusion was supported by analyses of the separate enabling conditions and consequences parts of the NO cases. These analyses revealed what might be called a "processing speed gradient": The enabling conditions of these cases were processed at a speed of 278 msec/word, while the consequences took only 203 msec/word (computed over both expertise groups). Thus, at the beginning of the NO cases, the evidently contradictory information, probably as the consequence of a surprise effect, takes relatively more time to be processed, whereas by the time the first consequences appear on the screen, subjects will have decided that the patient described in the case is not suffering from the announced disease. This applies to 6th-year students as well as experts. Apart from the family physicians having a general speed advantage over their less experienced colleagues, performances of both groups on the NO cases were highly similar.

Case Evaluation Latencies

Once faced with the case evaluation question, subjects had to decide as quickly as possible whether the presented case description actually would match someone with the specified disease. Since subjects knew that the case evaluation question would eventually be presented, it seems reasonable to expect that they had already tried to decide during the case presentation whether the patient description actually matched the announced illness. It might be expected that for relatively clear-cut cases (i.e., PP cases and NO cases), they would have been able to give a rapid answer to the case evaluation question. For other case types, they might have wanted to ponder some time before opting for a definite answer.

The results are displayed in Table 2. Since differences in the length of the case evaluation question were negli-

gible (the only difference being the length of the name of the illness), raw decision times are presented instead of the decision times in milliseconds/word. The data in Table 2 indicate that this decision was generally not taken too rashly. A 2×5 ANOVA of the raw reading times, with level of expertise as the between-subjects factor and case type as the within-subject factor revealed no significant main effect of level of expertise. It revealed a significant main effect of case type [$F(4,232) = 25.007$, $MS_e = 4,597,629.911$, $p < .0001$], but no significant interaction. From Table 2 it can be seen that, in ascending order, the case evaluation question led to increasing response latencies from NO cases, via PP, AP, and PA cases to AA cases, for both expertise levels. A separate comparison between NO type cases and PP type cases revealed that the case evaluation latencies for the former type of cases were indeed significantly shorter than for the latter type [$F(1,58) = 20.29$, $MS_e = 2,314,294.047$, $p < .0001$]. This finding also supports the conclusion that by the time the case evaluation question was presented for a NO case, subjects were fairly confident that the case was not an instantiation of the activated illness script. It should be remarked that although PP type cases were generally processed more rapidly than were NO type cases (Table 1), the case evaluation latencies for PP type cases were longer than those for NO type cases (Table 2). Apparently, subjects at both expertise levels were more strongly inclined to reject NO cases than to accept PP cases.

Probability Estimates

The results of the analysis of the probability estimates are depicted in Table 3 and Figure 3. A 2×5 ANOVA of the data in Table 3 (NO type cases included) revealed no significant main effect of expertise level, but it did not re-

veal a significant effect of case type [$F(4,224) = 235.961$, $MS_e = 156.263$, $p < .0001$]. The interaction between expertise level and case type was not significant. The most right-hand column of Table 3 shows that the expertise groups did not differ in terms of their general inclination to assign probability estimates. Thus, all differences between expert and nonexpert probability estimates within the respective columns of Table 3 may be considered a consequence of different likelihood of illness script instantiation with the same case information by subjects at different expertise levels. Comparison of the column means in the table shows that manipulating typicality of enabling conditions and consequences had a profound effect on probability estimates: For both expertise levels, the probability estimates declined from PP type cases via AP, PA, and AA type cases to the NO type cases. Thus, subjects generally assigned a high disease probability to the patients described in the PP case variants (about 76.5% average over both levels of expertise), while for the AA case variants, these estimates were much lower (about 41% over both levels of expertise). The AP and PA case types fell somewhere in between, with mean probability estimates of 61.19% and 55.30%, respectively. Disease probability estimates of the NO cases, averaged over both expertise groups, did not exceed 10%. A separate comparison of AA and NO type cases revealed a significant difference between these two types [$F(1,57) = 150.869$, $MS_e = 184.075$, $p < .0001$]. Therefore, it can be concluded that even completely atypical cases had a significantly higher likelihood of instantiating an activated illness script than contradictory cases; the difference in probability estimates amounted to 30% (Table 3).

It was also hypothesized that for 6th-year students, typicality of enabling conditions, unlike typicality of con-

Table 2
Average Raw Reading Times (in Milliseconds) for the Case Evaluation Question as a Function of Expertise Level and Case Type

Expertise Level	Case Type					M
	PP	AP	PA	AA	NO	
Sixth-year students*	4,136	4,401	5,797	6,252	2,488	4,615
Family physicians*	3,426	4,509	5,069	5,852	2,572	4,286
M	3,781	4,455	5,432	6,052	2,530	4,450

Note—PP, enabling conditions and consequences prototypical; AP, enabling conditions atypical, consequences prototypical; PA, enabling conditions prototypical, consequences atypical; AA, both enabling conditions and consequences atypical; NO, case description inconsistent with illness script header. *N = 30.

Table 3
Average Probability Estimates (Expressed as Percentages) as a Function of Expertise Level and Case Type

Expertise Level	Case Type					M
	PP	AP	PA	AA	NO	
Sixth-year students*	73.11	62.33	53.11	40.86	12.13	48.31
Family physicians†	79.67	60.12	57.34	40.77	7.74	49.13
M	76.50	61.19	55.30	40.81	9.86	48.73

Note—PP, enabling conditions and consequences prototypical; AP, enabling conditions atypical, consequences prototypical; PA, enabling conditions prototypical, consequences atypical; AA, both enabling conditions and consequences atypical; NO, case description inconsistent with illness script header. *N = 28. †N = 30.

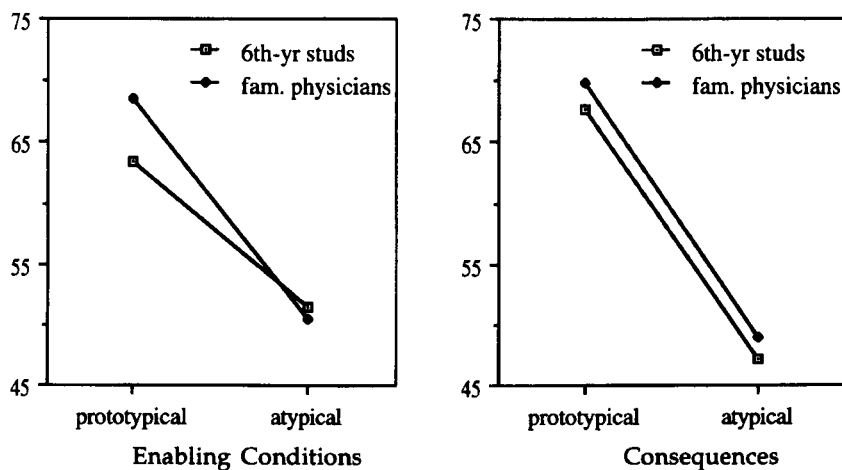


Figure 3. Probability estimates as a function of typicality and expertise level for enabling conditions (left-hand panel) and consequences (right-hand panel).

sequences, would not contribute much to the probability estimates assigned to a case. For experienced physicians, on the other hand, both typicality of enabling conditions and typicality of consequences would influence the probability estimates. Thus, an interaction was expected between typicality of enabling conditions and expertise level, with the more experienced group showing sensitivity to typicality of enabling conditions, and the less experienced group not, or, if anything, at least not to the same extent. However, as we have no probability estimates for individual case parts, but only for complete cases (i.e., for 4 specific typicality combinations of enabling conditions and consequences), it was not possible to analyze the data in a $2 \times 2 \times 2$ ANOVA; we had to revert to two separate 2×2 ANOVAs instead. In the first of these analyses, the average probability estimates of PP cases and PA cases were contrasted with those of AP and AA cases in order to assess the effect of typicality of enabling conditions; in the second analysis, the average probability estimates of PP cases and AP cases were contrasted with those of PA cases and AA cases in order to assess the effect of typicality of consequences.

The first 2×2 ANOVA, with expertise level as the between-subjects factor and typicality of enabling conditions as the within-subject factor, revealed, apart from the expected main effect of typicality [$F(1,57) = 97.993$, $MS_e = 67.129$, $p < .0001$], indeed a significant interaction between expertise level and typicality [$F(1,57) = 4.301$, $MS_e = 67.129$, $p < .05$]. In Figure 3 (left-hand panel), the results are displayed graphically. It can be seen from this figure that experienced family physicians generally assigned higher probability estimates to cases with prototypical enabling conditions and lower probability estimates to cases with atypical enabling conditions, whereas 6th-year students showed less sensitivity to this typicality manipulation, even though it cannot be concluded that the 6th-year students were altogether insensitive to typicality of enabling conditions.

In Figure 3 (right-hand panel), the results of the second 2×2 ANOVA, with expertise level as the between-subjects factor and typicality of consequences as the within-subject factor, are presented. Apart from the general typicality effect [$F(1,57) = 96.911$, $MS_e = 130.601$, $p < .0001$], no other significant results were obtained for the consequences; indeed, visual inspection of Figure 3 (right-hand panel) easily confirms that the effects of expertise level and typicality of consequences are additive. A comparison of the two panels in Figure 3 also suggests that typicality of consequences in general played a slightly more important role in determining probability estimates than did typicality of enabling conditions, especially for the 6th-year students.

Thus, unlike the case reading time data, the results of the analysis of the probability estimates support the hypothesis that experienced family physicians were more inclined than were 6th-year students to take typicality of enabling conditions into account when determining the probability of a case being an instantiation of a specific illness script.

DISCUSSION

In the present study, the influence of expertise level and typicality of case information on case processing times and case probability estimates was investigated. In this section, we will elaborate on the implications of the present results for the illness script theory. First, it can be concluded that manipulating typicality of case information turned out to be a powerful experimental tool to investigate processing speed and judgment of medical (clinical) information: Consistently large differences were found in processing times and probability estimates as a function of case typicality. Thus, the present study underscores the findings of Bellezza and Bower (1981) and Yekovich and Walker (1986) that experimental manipulation of typicality of story information can be used to investigate script processing characteristics.

As far as processing times are concerned, general effects of expertise level, typicality, and illness script component were found: Experts were faster than nonexperts, typical information was processed more rapidly than was atypical information, and consequences were processed more rapidly than were enabling conditions. This latter effect, however, may have been due to the fact that consequences were always preceded by enabling conditions, and illness script instantiation may have been already partly accomplished by the time the subjects were processing the consequences information. Although the three-way interaction of expertise level, typicality, and illness script component did not reach significance, and hence it cannot be definitely concluded from the data that experienced subjects processed enabling conditions information in a more script-like way than less experienced subjects, the results at least suggest that completely prototypical cases have a special status for expert physicians: Reading times for these cases for this group of subjects were much shorter than those for any other combination of case type and expertise level (Table 1). This conclusion is bolstered by the data on the case evaluation latencies. In a more fine-grained analysis, we took into consideration the fact that the processing times of case parts were not independent (i.e., processing of information appearing at a later point in the case description may have been influenced by the typicality of the preceding information), and found that processing of prototypical enabling conditions facilitated the processing of subsequent prototypical consequences for experienced physicians, but not for 6th-year students (Figure 2). Conversely, for experienced physicians, atypical enabling conditions seemed to decelerate the processing speed of subsequent prototypical consequences, whereas for the 6th-year students, this effect was small. Thus, typicality of enabling consequences indeed appears to have had a differential effect at the two expertise levels included in the present study, a finding that may reflect the hypothesized differential integration of enabling conditions into illness script structures and thus support the proposed developmental course of illness scripts. On the other hand, the data also support the assumption that advanced students already do possess script-like structures: Generally, they are influenced to about the same extent by typicality of case information as expert family physicians; the absence of an interaction between expertise level and typicality attests to this conclusion.

The results also showed that if case descriptions deviate too much from what is feasible given a specific illness script, reading times show a significant drop. Thus, after some initial hesitation or surprise reaction, NO cases were processed relatively quickly. Apparently, for this type of case, less processing is required, probably because already in an early stage of the case presentation, subjects have decided that the activated illness script cannot be instantiated by the case information; hence, they have rejected it and do not feel inclined to further try to fit incoming information into it. In fact, in the present study, an inverted-U relationship between "script-fitting qual-

ity" of case information and processing speed was found: Both information that can be easily integrated into an activated script *and* information that evidently contradicts allowed script features or slot values are processed fast (i.e., without much elaboration or consideration), whereas information of intermediate fitting quality (i.e., atypical information) requires additional processing time. This conclusion is supported by the long case evaluation latencies for partly and completely atypical cases (Table 2).

As far as probability estimates are concerned, a general effect of case typicality was found. Completely prototypical cases were assigned probability percentages in the 70s, and completely atypical cases were assigned probability percentages in the lower 40s. Cases with either prototypical enabling conditions or prototypical consequences (but not both) received probability percentages in the range of 55%–62%. Thus, both typicality of enabling conditions and typicality of consequences contributed to the overall subjective likelihood estimates. Separate comparisons of the influence of typicality of enabling conditions and typicality of consequences revealed that experienced physicians were somewhat more inclined to take typicality of enabling conditions into consideration than were 6th-year students, although probability estimates of subjects at both expertise levels were clearly influenced by this factor. As to the typicality of consequences, no differential effect of expertise level was found. This finding (i.e., the somewhat higher sensitivity of more experienced subjects for contextual and patient background information) is in line with the proposed development of illness scripts.

On the basis of the present results, no definite conclusions can be drawn concerning the role of illness scripts in diagnostic situations. It should be recalled that Hobus et al. (1987, 1990) found that experienced physicians who had access to enabling conditions were able to use this information in activating diagnostic hypotheses, whereas inexperienced physicians were much less able to do so. A previous study by Custers et al. (1992) showed that this phenomenon could not be accounted for by a mere lack of knowledge of enabling conditions on the part of the less experienced subjects. The results of the present study support the idea that integration of information about enabling conditions into illness script structures is a critical factor in the later stages of development of medical diagnostic knowledge, even though the size of the effects in an experimental task was relatively small and requires confirmation by future studies. Expert physicians seem to possess a particular sensitivity for completely prototypical cases, that is, patient descriptions in which both contextual factors and complaints, signs, and symptoms are in line with the default values for the applicable illness script. This particular sensitivity is not shared by advanced students, even though the latter subjects were evidently influenced by typicality of enabling conditions as well as by consequences. The present results also suggest that expert physicians may be more able to profit from prototypical consequences than

nonexperts, a possibility that was not included in the Hobus et al. (1987, 1990) studies because these authors did not investigate the influence of knowledge of consequences on diagnostic hypothesis formation.

We are inclined to think that our present results may have a broader range of application than the medical domain. Diagnosis is an important aspect of many real-life problem-solving situations. It remains a challenge for future studies to develop a proper conceptualization of the relevant knowledge structures with an emphasis on knowledge accessibility, rather than on the mere presence or absence of knowledge, and to discover why subjects at intermediate levels of expertise do possess the relevant knowledge, but are unable to use it in many diagnostic situations.

REFERENCES

- ANDERSON, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- ANDERSON, J. R. (1993). *Rules of the mind*. Cambridge, MA: Harvard University Press.
- BARROWS, H. S., & FELTOVICH, P. J. (1987). The clinical reasoning process. *Medical Education*, **21**, 86-91.
- 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 & Investigative Medicine*, **5**, 49-55.
- BARTLETT, F. C. (1954). *Remembering. A study in experimental and social psychology*. Cambridge: Cambridge University Press. (Original work published 1932)
- BELLEZZA, F. S., & BOWER, G. H. (1981). The representational and processing characteristics of scripts. *Bulletin of the Psychonomic Society*, **18**, 1-4.
- BOWER, G. H., BLACK, J. B., & TURNER, T. J. (1979). Scripts in memory for text. *Cognitive Psychology*, **11**, 177-220.
- BROOKS, L. R., NORMAN, G. R., & ALLEN, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General*, **120**, 278-287.
- CLANCEY, W. J., & SHORTLIFFE, E. H. (1984). Introduction: Medical artificial intelligence programs. In W. J. Clancey & E. H. Shortliffe (Eds.), *Readings in medical artificial intelligence* (pp. 1-17). Reading, MA: Addison-Wesley.
- CUSTERS, E. J. F. M., BOSCHUIZEN, H. P. A., & SCHMIDT, H. G. (1992, April). *The relationship between medical expertise and the development of illness scripts*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- DAVIDSON, D., & HOE, S. (1993). Children's recall and recognition memory for typical and atypical actions in script-based stories. *Journal of Experimental Child Psychology*, **55**, 104-126.
- ELSTEIN, A. S., SHULMAN, L. S., & SPRAFKA, S. A. (1978). *Medical problem solving. An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.
- 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. New directions in training for the health professions* (pp. 128-142). Assen/Maastricht: Van Gorcum.
- FELTOVICH, P. J., JOHNSON, P. E., MOLLER, J. H., & SWANSON, D. B. (1984). LCS: The role and development of medical knowledge in diagnostic expertise. In W. J. Clancey & E. H. Shortliffe (Eds.), *Readings in medical artificial intelligence: The first decade* (pp. 275-319). Reading, MA: Addison-Wesley.
- GENTNER, D., & STEVENS, A. L. (Eds.) (1983). *Mental models*. Hillsdale, NJ: Erlbaum.
- GRAESSER, A. C., WOLL, S. B., KOWALSKI, D. J., & SMITH, D. A. (1980). Memory for typical and atypical actions in scripted activities. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 503-515.
- HOBUS, P. P. M., BOSCHUIZEN, H. P. A., & SCHMIDT, H. G. (1990). *Expert-novice differences in the role of contextual factors in early medical diagnosis*. Paper presented at the annual meeting of the American Educational Research Association, Boston.
- HOBUS, P. P. M., HOFSTRA, M. L., BOSCHUIZEN, H. P. A., & SCHMIDT, H. G. (1988). De context van de klacht als diagnosticum [The context of the complaint as a diagnostic tool]. *Huisarts en Wetenschap*, **31**, 261-267.
- HOBUS, P. P. M., SCHMIDT, H. G., BOSCHUIZEN, H. P. A., & PATEL, V. L. (1987). Contextual factors in the activation of first diagnostic hypotheses: Expert-novice differences. *Medical Education*, **21**, 471-476.
- HOFSTRA, M. L., HOBUS, P. P. M., BOSCHUIZEN, H. P. A., & SCHMIDT, H. G. (1988). *Diagnostiek bij huisartsen: het belang van gegevens uit de context van de patiënt* [Family physicians' diagnosis: The importance of patient contextual information]. Maastricht: University of Limburg.
- HUBSON, J. A. (1988). Children's memory for atypical actions in script-based stories: Evidence for a disruption effect. *Journal of Experimental Child Psychology*, **46**, 159-173.
- JOHNSON-LAIRD, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.
- KOLODNER, J. L. (1988). Extending problem-solver capabilities through case-based inference. In J. L. Kolodner (Ed.), *Case-based reasoning: Proceedings of a workshop on case-based reasoning*. San Mateo, CA: Morgan Kaufmann.
- KOLODNER, J. L., & SIMPSON, R. L. (1986). Problem solving and dynamic memory. In J. L. Kolodner & C. K. Riesbeck (Eds.), *Experience, memory, and reasoning*. Hillsdale, NJ: Erlbaum.
- KOLODNER, J. L., & SIMPSON, R. L. (1989). The MEDIATOR: Analysis of an early case-based problem solver. *Cognitive Science*, **13**, 507-549.
- LUSTED, L. B. (1968). *Introduction to medical decision making*. Springfield, IL: Thomas.
- MINSKY, M. A. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211-280). New York: McGraw-Hill.
- NELSON, K., & GRUENDEL, J. (1986). Children's scripts. In K. Nelson (Ed.), *Event knowledge: Structure and function in development*. Hillsdale, NJ: Erlbaum.
- NEWELL, A., SHAW, J. C., & SIMON, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review*, **65**, 151-166.
- NEWELL, A., & SIMON, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- PATEL, V. L., & GROEN, G. J. (1986). Knowledge-based solution strategies in medical reasoning. *Cognitive Science*, **10**, 91-116.
- PATEL, V. L., & GROEN, G. J. (1991). The general and specific nature of medical expertise: A critical look. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 93-125). Cambridge: Cambridge University Press.
- ROSCHE, E. H. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.
- RUMELHART, D. E., & MCCLELLAND, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.
- RUMELHART, D. E., & ORTONY, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 99-135). Hillsdale, NJ: Erlbaum.
- SCHANK, R. C., & ABELSON, R. P. (1977). *Scripts, plans, goals and understanding. An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.
- SCHMIDT, H. G., & BOSCHUIZEN, H. P. A. (1993a). On acquiring expertise in medicine. *Educational Psychology Review*, **5**, 205-221.
- SCHMIDT, H. G., & BOSCHUIZEN, H. P. A. (1993b). On the origin of intermediate effects in clinical case recall. *Memory & Cognition*, **21**, 338-351.
- SCHMIDT, H. G., & BOSCHUIZEN, H. P. A., & NORMAN, G. R. (1992). Reflections on the nature of expertise in medicine. In E. Keravnou (Ed.), *Deep models for medical knowledge engineering* (pp. 231-248). Amsterdam: Elsevier, North-Holland.

- SCHMIDT, H. G., NORMAN, G. R., & BOSHUIZEN, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine*, *65*, 611-621.
- SMITH, D. A., & GRAESSER, A. C. (1981). Memory for actions in scripted activities as a function of typicality, retention interval, and retrieval task. *Memory & Cognition*, *9*, 550-559.
- SWANSON, D. B., FELTOVICH, P. J., & JOHNSON, P. E. (1977). Psychological analysis of physician expertise: Implications for design of decision support systems. In D. B. Shires & H. Wolff (Eds.), *MED-INFO 77. Proceedings of the Second World Conference on Medical Informatics* (pp. 161-164). Amsterdam: Elsevier, North-Holland.
- WALKER, C. H., & YEKOVICH, F. R. (1984). Script-based inferences: Effects of text and knowledge variables on recognition memory. *Journal of Verbal Learning & Verbal Behavior*, *23*, 357-370.
- WEBER, E. U., BÖCKENHOLT, U., HILTON, D. J., & WALLACE, B. (1993). Determinants of diagnostic hypothesis generation: Effects of information, base rates, and experience. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*, 1151-1164.
- YEKOVICH, F. R., & WALKER, C. H. (1986). Retrieval of scripted concepts. *Journal of Memory & Language*, *25*, 627-644.
- YEKOVICH, F. R., & WALKER, C. H. (1987). The activation and use of scripted knowledge in reading about routine activities. In B. K. Britton & S. M. Glynn (Eds.), *Executive control processes in reading* (pp. 145-176). Hillsdale, NJ: Erlbaum.

NOTES

1. In line with previous research (Custers et al., 1992), *prototypical* and *atypical* are defined as opposite points on the underlying dimension of typicality.
2. The Dutch medical curriculum consists of a 4-year training program in the biomedical and clinical sciences, followed by a 2-year clerkship program.
3. Some subjects evidently misinterpreted the probability estimation instruction in that they thought *their own answer* on the preceding case evaluation question, rather than the probability that the patient suffered from the announced disease, should be evaluated (thus, they responded, e.g., "90%," if they were fairly sure that their *no* answer had been correct).
4. To make case descriptions as natural as possible, we had to construct some statements with a mixture of enabling conditions and consequences (e.g., a current complaint that referred to a previous complaint), and sometimes we had the patient reveal an enabling condition at the end of the case (e.g., "I've been on holiday to Spain lately.>").

APPENDIX A

Names of the Diseases Used in the Study

Diseases on which the three practice cases were based:

- Perforated otitis media
- Monilia of the mouth
- Dyspepsia of a nervous origin (also called nervous gastritis)

Diseases on which the 16 experimental cases were based:

- Metastatic sigmoid cancer (metastases in the lungs)
- Aneurism of the aortic artery (threatening rupture)
- Urosepsis
- Dermatitis peri-oralis
- Vaginal candidiosis
- Kidney stones (colic)
- Carcinoma of the head of the pancreas
- Stomatitis aftosa (multiple small ulcers in the mouth)
- Digitalis intoxication
- Epidural hematoma
- Nervous abdominal pain
- Pediculosis pubis
- Herpes zoster

- Meningitis or encephalitis as a complication of mumps
- Hepatitis A
- Pre-infarct syndrome

Diseases on which the five No cases were based:

- Pfeiffer's disease (announced as, "Is this a patient with Hodgkin's disease?")
- Pernicious anemia (announced as, "Is this a patient with iron-deficiency anemia or hypochromic anemia?")
- Perforated gastric ulcer (announced as, "Is this a patient with reflux-oesophagitis?")
- Seborrheic dermatitis (announced as, "Is this a patient with impetigo?")
- Thyrotoxicosis (announced as, "Is this a patient with hypothyroid?")

APPENDIX B

Example of a Case Description With Four Variants

Case 6: Kidney stones (colic)

The number preceding each text phrase refers to the number of the screen image in which that particular phrase was presented.

PP variant:

1. Man, age 47, married, 3 children
2. Occupation: storekeeper
3. Medical history: bronchitis at age 30
4. Had his leg broken 6 years ago, as a consequence of a car accident
5. Four years ago: treated with medicaments for kidney stones
6. Some of his relatives are known to have coronary disease and diabetes mellitus
7. His wife rings up, asks the physician for an immediate visit:
8. Just like a few years ago, her husband is rolling across the room because of the pain. He is also vomiting almost continuously
9. When the physician arrives, the pain has just subsided. The patient is sitting on the sofa and recovering a bit
10. At the moment, the patient doesn't look very ill
11. He complains about having had a convulsive abdominal pain abreast of the navel, at the left side
12. The pain is radiating to his groin
13. The pain emerges very suddenly, and then gradually subsides. During an attack he almost can't stand it
14. Earlier that day he had already seen some blood in his urine, but had had no pain at the time
15. Is this a patient with a kidney stone colic? (*yes/no*)
16. Please estimate the probability that this person has a kidney stone colic in a percentage (range, 0%-100%)

Comment:

The prototypical patient with kidney stone colic is a male, aged 35 to 55 years (Statement 1), who has had this kind of complaint before (Statement 5). As the clinical picture is often rather impressive, it is not unusual that relatives are in panic (Statement 7). Typically, the pain is of a convulsive na-

ture, about half-way down the abdomen, and one-sided (left or right, Statement 11). Radiation to the groin is also very prototypical (Statement 12). The same holds for the pain-free intervals, during which the patient appears normal (Statements 9, 10, 13). During an attack, people are literally running or cringing (Statements 8, 13). Blood in the urine may also occur, before or after an attack (Statement 14). Statements 2, 3, 4, and 6 are added to complete the case description (occupation is usually mentioned, as are some rather irrelevant medical history items with regard to the present condition).

AP variant:

1. Woman, 32 years, married, 2 children
2. Occupation: works as clerical staff at a large school
3. Medical history: bronchitis at age 20
4. Had her leg broken 6 years ago, as a consequence of a car accident
5. Has been taking the contraceptive pill for 9 years, interrupted for pregnancies
6. Some of her relatives are known to have coronary disease and diabetes mellitus
7. Her husband rings up, asks the physician for an immediate visit:
8. He is afraid that his wife is dying: She is rolling through the room because of the pain. She is also vomiting almost continuously
9. When the physician arrives, the pain has just subsided. The patient is sitting on the sofa and recovering a bit
10. At the moment, the patient doesn't look very ill
11. She complains about having had a convulsive abdominal pain abreast of the navel, at the left side
12. The pain is radiating to her groin
13. The pain emerges very suddenly, and then gradually subsides. During an attack she almost can't stand it
14. Earlier that day she had already seen some blood in her urine, but had had no pain at the time
15. Is this a patient with a kidney stone colic? (*yes/no*)
16. Please estimate the probability that this person has a kidney stone colic in a percentage (range, 0%–100%)

Comment:

Kidney stones are not often found in young women. In addition, nothing in her medical history points to the possible existence of this ailment. The symptoms (the same as in the previous description) are very prototypical, though.

PA variant:

1. Man, age 47, married, 3 children
2. Occupation: storekeeper
3. Medical history: bronchitis at age 30
4. Had his leg broken 6 years ago, as a consequence of a car accident
5. Four years ago: treated with medicaments for kidney stones
6. Some of his relatives are known to have coronary disease and diabetes mellitus
7. The patient appears at consulting hour with the complaint:

8. Since the other day, occasionally a sharp, stabbing pain in the abdomen, abreast of the navel, at the right side
9. The pain is radiating to his groin
10. He had already felt for a few days a somewhat nagging sensation in this part of his abdomen
11. And it is still nagging, now, in between the stabs
12. Upon inquiry, the patient admits having felt an urge to micturate more often than usual, the last few days, but the amount of urine each time was small
13. He says he doesn't feel very well
14. And reports having measured 37.8° (Centigrade) temperature
15. Is this a patient with a kidney stone colic? (*yes/no*)
16. Please estimate the probability that this person has a kidney stone colic in a percentage (range, 0%–100%)

Comment:

This pattern of symptoms and complaints is possibly, but not very likely, caused by kidney stones.

AA variant:

1. Woman, 32 years, married, 2 children
2. Occupation: works as clerical staff at a large school
3. Medical history: bronchitis at age 20
4. Had her leg broken 6 years ago, as a consequence of a car accident
5. Has been taking contraceptive pills for 9 years, interrupted for pregnancies
6. Some of her relatives are known to have coronary disease and diabetes mellitus
7. The patient appears at consulting hour with the complaint:
8. Since the other day, occasionally a sharp, stabbing pain in the abdomen, abreast of the navel, at the right side
9. The pain is radiating to her groin
10. She had already felt for a few days a somewhat nagging sensation in this part of her abdomen
11. And it is still nagging, now, in between the stabs
12. Upon inquiry, the patient admits having felt an urge to micturate more often than usual, the last few days, but the amount of urine each time was small
13. She says she doesn't feel very well
14. And reports having measured 37.8° (Centigrade) temperature
15. Is this a patient with a kidney stone colic? (*yes/no*)
16. Please estimate the probability that this person has a kidney stone colic in a percentage (range, 0%–100%)

Comment:

Upon seeing this patient and hearing her complaints, the physician will probably not immediately think of kidney stones. However, the possibility of this disease is surely not ruled out.

Note—PP, enabling conditions and consequences prototypical; AP, enabling conditions atypical, consequences prototypical; PA, enabling conditions prototypical, consequences atypical; AA, both enabling conditions and consequences atypical; NO, case description inconsistent with illness script header.

APPENDIX C
Number of Statements (Screen Images)
and Words per Case Description

Case	No. Statements	No. Words			
Practice					
Perforated otitis media	10	98			
Monilia of the mouth	9	81			
Nervous gastritis	17	177			
NO					
Hypothyroid	14	148			
Impetigo	12	132			
Iron-deficiency anemia*	15	177			
Hodgkin's disease	16	171			
Reflux oesophagitis	14	138			
		Variant			
		PP	AP	PA	AA
Experimental					
Metastatic sigmoid cancer	14	137	128	137	138
Rupturing aneurism	14	172	176	164	167
Urosepsis	14	132	131	132	131
Dermatitis peri-oralis	14	172	176	164	167
Vaginal candidiosis	14	136	165	144	164
Kidney stones colic	14	167	165	138	147
Carcinoma of the head of the pancreas	15	150	145	159	155
Stomatitis aftosa	10	94	94	85	86
Digitalis intoxication	12	110	130	123	117
Epidural hematoma	14	180	160	178	157
Nervous abdominal pain	19	193	216	201	236
Pediculosis pubis	8	90	106	93	109
Herpes zoster	11	105	94	106	96
Meningitis (as a complication of mumps)	13	121	122	115	117
Hepatitis A	17	140	140	147	140
Pre-infarct syndrome	18	160	178	185	186

Note—PP, enabling conditions and consequences prototypical; AP, enabling conditions atypical, consequences prototypical; PA, enabling conditions prototypical, consequences atypical; AA, both enabling conditions and consequences atypical; NO, case description inconsistent with illness script header. *This case was omitted from the analysis.

APPENDIX D
Experimental Design of the Sequences

1A	2A	3A	4A	1B	2B	3B	4B
pr1	pr1	pr1	pr1	pr1	pr1	pr1	pr1
pr2	pr2	pr2	pr2	pr2	pr2	pr2	pr2
pr3	pr3	pr3	pr3	pr3	pr3	pr3	pr3
4	4	4	4	14AP	14PA	14AA	14PP
5PA	5AA	5PP	5AP	15PP	15AP	15PA	15AA
6PA	6AA	6PP	6AP	16	16	16	16
7AA	7PP	7AP	7PA	17AA	17PP	17AP	17PA
8	8	8	8	18PA	18AA	18PP	18AP
9AA	9PP	9AP	9PA	19PA	19AA	19PP	19AP
10PP	10AP	10PA	10AA	20	20	20	20
11AP	11PA	11AA	11PP	21PP	21AP	21PA	21AA
12AA	12PP	12AP	12PA	22	22	22	22
13AP	13PA	13AA	13PP	23PP	23AP	23PA	23AA
				24AP	24PA	24AA	24PP
14AP	14PA	14AA	14PP				
15PP	15AP	15PA	15AA	4	4	4	4

APPENDIX D (Continued)

16	16	16	16	5PA	5AA	5PP	5AP
17AA	17PP	17AP	17PA	6PA	6AA	6PP	6AP
18PA	18AA	18PP	18AP	7AA	7PP	7AP	7PA
19PA	19AA	19PP	19AP	8	8	8	8
20	20	20	20	9AA	9PP	9AP	9PA
21PP	21AP	21PA	21AA	10PP	10AP	10PA	10AA
22	22	22	22	11AP	11PA	11AA	11PP
23PP	23AP	23PA	23AA	12AA	12PP	12AP	12PA
24AP	24PA	24AA	24PP	13AP	13PA	13AA	13PP

Note—pr, practice case; PP, enabling conditions and consequences prototypical; AP, enabling conditions atypical, consequences prototypical; PA, enabling conditions prototypical, consequences atypical; AA, both enabling conditions and consequences atypical; NO, case description inconsistent with illness script header. Each sequence starts with three practice cases, which are the same for all sequences. The experimental cases fall into one of 4 groups, with the same case variant for each member of a particular group in a given sequence. For example, Cases 10, 15, 21, and 23 belong to 1 group, and in Sequence 1, the PP variant of these cases is presented, in Sequence 2 the AP variant, in Sequence 3 the PA variant, and in Sequence 4 the AA variant. Each sequence contains 4 cases of each case variant. Cases 4, 8, 16, 20, and 22 are examples of NO cases, of which only 1 variant exists. The A and B forms of each sequence differ only in the order in which the experimental cases were presented.

(Manuscript received February 9, 1995;
revision accepted for publication July 21, 1995.)