The Knowledge Engineering Review, Vol. 21:4, 335–344. © 2006, Cambridge University Press
doi:10.1017/S0269888906000968 Printed in the United Kingdom

How experts reason: the acquisition of experts' knowledge structures

ERWIN W. VAN GEENEN and CILIA L. M. WITTEMAN

Diagnostic Decision Making, Behavioural Science Institute, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands;
e-mail: c.witteman@socsci.ru.nl

Abstract

We were faced with a shelved knowledge system at a health insurance company. The system used Bayesian statistics. For correct computations such a system is of course well suited, but not for the people who use it: people do not easily formulate their expertise in numerical probabilities. We addressed the question of whether it is possible to find out in which format experts do represent their knowledge, and if so, what format that is. This representation might then replace the format of numerical probabilities in the knowledge system's interface, increasing its ease of use. We found one method that includes, apart from a technique to elicit knowledge contents, techniques to also acquire its format: Cognitive Structure Analysis. In this contribution we discuss this method, use it in two studies, and show that it yields valid results. It gives insight into the format of experts' knowledge structures; experts turn out to use mainly frame representations.

1 Introduction

A health insurance company in the Netherlands had asked a knowledge engineering company to develop an expert system for underwriting. The company engineered a system that used Bayesian statistics. In essence, the system lists, for each disease, a number of factors that have been identified by medical doctors as important for the underwriting process. Each insurance applicant is scored on these factors, yielding a so-called applicant profile. On the basis of this profile, the system assigns the applicant to one of three risk groups (low-risk group, medium-risk group, or high-risk group) by means of the Bayesian classification rule. The classification procedure requires numerical probabilities of the absence and presence of each factor. These probabilities were assessed by expert medical judges. They were asked to estimate for each factor (independently of the other factors) the probability of its being absent or present in the low-risk group, the medium-risk group, and the high-risk group. The experts had major difficulties transforming their knowledge into this Bayesian format, and they frequently indicated that they found it uncomfortable to use and difficult to understand. As a consequence, the whole project failed.

This business case illustrates an issue which has lately been addressed in the human–computer literature. The formalisms used in knowledge acquisition are typically determined by the computational representations implemented in the knowledge system, such as a Bayesian structure. However, such formalisms are often unfamiliar to domain experts, and of a form wholly unlike the representations they actually use (Cheng *et al.*, 2001). The experts may then be unable to communicate their knowledge properly to the knowledge engineer. It would seem proper that the tables are turned, and that knowledge engineers learn about and use the experts' knowledge structures (cf. also DuBois & Shalin, 1995; Chipman *et al.*, 2000). This is appropriate irrespective

https://doi.org/10.1017/S0269888906000968 Published online by Cambridge University Press

of whether the engineers are constructivists who build systems that perform well, cognitive modellers who aim to actually capture the experts' reasoning in a system (cf. Motta, 2001), or interactionists who think knowledge is acquired and structured incrementally through interactions (Boy, 1997).

First acquiring experts' cognitive knowledge representations to subsequently elicit their knowledge in this format is not part of mainstream knowledge engineering methods. In fact, among the vast number of documented knowledge elicitation methods (see, e.g., Neale, 1989; Olson & Biolsi, 1991; Cooke, 1994; Burge, 1998), we found only one method for acquiring both first experts' cognitive representations and then the knowledge itself: Cognitive Structure Analysis (CSA). It was articulated by Leddo and Cohen more than 15 years ago (in 1989), but it has not received much attention from the scientific community. It is not clear why this is so; it may be because it is not easy to get hold of the papers. One (Leddo *et al.*, 1988) is a technical report to the U.S. Army, the other (Leddo & Cohen, 1989) is a paper in the proceedings of a conference. Neither has been published as a journal article.

In another method, the widely used CommonKADS methodology (Schreiber *et al.*, 2000), knowledge acquisition is also understood as a two-step process. First the knowledge type (or structures) is defined, and then instances of this type are elicited. However, contrary to CSA, in KADS the knowledge engineer determines the knowledge structure after reading and thinking about the domain, and not by listening to experts. CommonKADS does not include a method to elicit knowledge types from domain experts.

In this contribution we present the results of our endeavour to acquire experts' knowledge structures with CSA in a real domain, that of medical underwriting (for more details see van Geenen (2004)). In Section 2, Acquiring knowledge structures, we explain CSA. In Section 3, we present two empirical studies in which we applied the method and tested its validity. We end with a discussion and directions for future research.

2 Acquiring knowledge structures

CSA recognizes that different syntactical structures (formats, types) may be used to convey knowledge contents (semantics). In the first stage of CSA, the knowledge structures that are involved in the execution of a task are identified. This is the aspect of CSA that interests us most, and which is absent in other knowledge elicitation methods. In interview sessions experts are asked to talk about how they perform a task or to think aloud while actually performing it, and the knowledge structures they use are identified from a transcript of these sessions. In the second stage of CSA the task-specific contents of the knowledge structures are elicited.

2.1 Knowledge structures in CSA

With CSA, experts may be found to use any of the following five knowledge structures: Memory Organization Packets (MOPs), production rules, semantic networks, object frames, and mental models. These five structures meet two important criteria: first, that psychology has provided empirical support for their occurrence; and second, that they represent knowledge differently (Leddo & Cohen, 1989).

MOPs represent procedural knowledge, and constitute an extension of Schank and Abelson's scripts (Schank & Abelson, 1977). In essence, scripts represent single events that are described on one level of abstraction. Schank (1982) introduced MOPs for encoding events at different, interrelated levels of abstraction. A MOP represents a chain of events that are temporally ordered and causally connected to reach a goal, and includes planning knowledge.

Production rules represent procedural knowledge and take the form of if – then propositions. This structure can be expanded by allowing multiple conditions and actions, such as if A or B, then C and D. Production rules in CSA are fired probabilistically.

https://doi.org/10.1017/S0269888906000968 Published online by Cambridge University Press

Semantic networks have been defined by Quillian (1968) to represent declarative knowledge. Semantic networks are graph structures in which nodes represent concepts and arcs between these nodes represent relations among concepts. Different types of semantic networks exist (Sowa, 2000), but CSA recognizes only definitional networks which emphasize "is-a" relationships. They contain hierarchical and categorical relationships between the concepts; for example, chronic enteritis is-a disease and colitis ulcerosa is-a chronic enteritis. The inheritance of information between concepts is very important in such networks. The subclass "colitis ulcerosa" possesses, by default, the attributes associated with the class "chronic enteritis".

Frames were introduced by Minsky (1975) to represent declarative knowledge. They refer to collections of slots and slot fillers that describe stereotypical concepts. The slots can contain knowledge about, for example, the configuration of the concept and its goal or purpose. CSA adopts so-called object frames, in which the slots contain objects (e.g. the frame of a kitchen) with slots like sink, stove, etc. In our knowledge-intensive and more abstract domain, we adopt concept frames (e.g. the frame of a patient) with slots such as age, gender, weight, etc. Frames are particularly suitable for representing chunked knowledge about a concept. Research suggests that people do not have a fixed number of slots for frames, but that this depends on the context in which the frame is used (Stillings *et al*., 1995).

Mental models are conceived of as small-scale theories or models about real or imaginary events. They were introduced by Kenneth Craik (1943), who argued that the mind constructs models of reality for purposes of reasoning and explanation (see also Johnson-Laird, 1983; Johnson-Laird *et al*., 1998; Markman, 1999). Leddo *et al*. (1988) and Leddo & Cohen (1989) see mental models as structures that organize objects and that capture how changes in one object cause changes in other objects and produce events. A well-known example is the mental model of a heat exchanger used to cool a fluid in a machine (Williams *et al*., 1983). In our application we replace objects by concepts, for example a medication to cure a disease.

2.2 Discriminating the knowledge structures

The same knowledge content may be represented in any of the five different structures that Leddo & Cohen (1989) distinguish, and which we consider in our studies. Translating one representation into another is, in theory, simple (cf. Markman, 1999; Sowa, 2000). For example, a concept frame with different features in its slots may be rewritten using the following production rule: *if* it is a concept *then* these features may be expected. Recognizing this interchangeability in theory, we want to know which structure people naturally use in practice.

To identify the knowledge structures that experts use in the transcripts of the knowledge elicitation sessions (see below), discriminators are used. Leddo & Cohen (1989) propose that each knowledge structure can be identified by three key discriminators: the structure's Basic unit (Section 2.2.1), its Organizing principle (Section 2.2.2), and its Inferential use (Section 2.2.3). We present an overview in Table 1, and discuss each in turn.

2.2.1 Basic unit

The "basic unit" in a knowledge structure is an event, a concept, or a combination of both. Frames relate to collections of concepts, rather than to single concepts as used in semantic networks. Both describe declarative knowledge. Production rules and MOPs refer to events, since they capture procedural knowledge. Mental models use both concepts and events; they emphasize the causal relationships between concepts to produce events.

2.2.2 Organizing principle

The "Organizing principle" refers to "what drives the overall organization of the knowledge" (Leddo & Cohen, 1989, p. 276). Slightly modifying the authors' original terminology to suit our

https://doi.org/10.1017/S0269888906000968 Published online by Cambridge University Press

Table 1 The three levels of discriminators for the five knowledge structures in our adapted version of CSA

Basic unit	Organizing principle	Inferential use	Knowledge structure
Concept	Chunking	Concept completion	Frame
Concept	Taxonomy	Inheritance	Semantic network
Event	Data-conclusion mapping	1. Forward chaining 2. Backward chaining	Production rule
Event	Goal	1. Forward chaining 2. Backward chaining	MOP
1. Concept 2. Event	Causality	1. Prediction 2. Explanation 3. Conclusion 4. Hypothesis	Mental model

knowledge-intensive domain, we distinguish the following: chunking, taxonomy, data-conclusion mapping, goals, and causality.

Frames represent implicit goals or purposes. To more precisely distinguish frames, we use "chunking" as their organizing principle. Frames have been introduced particularly to represent chunked knowledge about concepts (Minsky, 1975); chunking refers to the process of grouping and linking elements to a single concept.

If knowledge is organized as temporal or spatial co-occurrences, without a goal or purpose, then CSA suggests a semantic network. The role of such networks is to represent taxonomies. Consequently, "taxonomy" is the organizing principle of semantic networks.

Production rules represent temporal co-occurrences. Leddo & Cohen (1989) do not go into the specifics of this organizing principle. Personal communication with one of the authors revealed that this may be explained by the special status of rule-based representations: "The criteria for scripts, frames, and models are more stringent than the criteria for rules. Thus, if a particular case does not fit into any of the former representation types [. . .] the fall-back representation is rules" (Cohen, 2002). We argue for a clearer organizing principle for production rules. For decades production rules have constituted a dominant paradigm of human cognition, in the production system model for human performance (Newell & Simon, 1972). The paradigm's fundamental claim is that cognitive skills involve performing sequences of condition–action pairs. The model works by mapping data onto the condition part of rules until a conclusion, the action part, is reached. We decided to use this principle, "data conclusion mapping" for short, to identify the organization of rule-based representations.

In MOPs, the knowledge is organized around goals. Causality is involved in MOPs too, since the events in a MOP are causally connected to each other. However, MOPs are specifically meant to capture goal-directed knowledge, and causality is subordinate to goal-directedness.

Causality serves as the leading and unique principle in mental models, since this principle is used in reasoning and explanation. Mental models do not necessarily include explicit or implicit goals, as one can just seek to explain why, for example, increasing medication would decrease pain.

2.2.3 Inferential use

"Inferential use" refers to the inference mechanism that is applied in the knowledge structure.

In Frames, new knowledge is inferred through the use of "concept completion", which refers to expanding the frame by inferring additional attributes. In Semantic networks, the main inference mechanism is inheritance. Inheritance can be bottom-up (from subclass to class) or top-down (from

https://doi.org/10.1017/S0269888906000968 Published online by Cambridge University Press

super class to class). Again, this mechanism allows the inference of additional attributes and/or objects.

In Leddo and Cohen's version of CSA, production rules implement forward chaining only. However, production rules are generally considered to allow backward chaining too (Winston, 1992; Luger & Stubblefield, 1998). Indeed, even Leddo *et al.* (1988, p. 32) confirm that "production rules can produce an inference chain that can serve to make inferences, by forward chaining, or deductions, by backward chaining'. We therefore add backward chaining to production rules" inferential use.

In MOPs there can be both forward and backward chaining, as in production rules. The essence is that one can make inferences from MOP-based actions to (sub)goals and *vice versa*. According to Leddo & Cohen (1989), in mental models both types of chaining are applied to causal knowledge, to generate predictions or explanations. These are not the only possible results of drawing inferences in mental models (Markman, 1999); hypotheses and conclusions are other options. We add these discriminators to operationally define the inferential use of mental models.

2.3 The combination of basic unit, organizing principle, and inferential use

Ideally, discriminators for one and the same knowledge structure are found on all three levels. For example, a clinical neuropsychologist who talks about their work may say: "I use neuropsychological tests to assess whether patients who display unclear symptoms suffer from dementia, a CVA, or other mental disorders (concepts and events). While testing the patient, I formulate hypotheses about the underlying causes of the patient's complaints (causality, and formulating hypotheses). These hypotheses may be rejected or confirmed while testing (reaching conclusions)." This is clearly a case of a mental model.

However, of course not all cases are equally ideal, and we will conclude that a knowledge structure has been used when we find discriminators on three or two levels. For example, a knowledge structure need not be used to make inferences. In this case, the knowledge structure can be identified by means of the discriminators on the levels of basic unit and organizing principle only. Also, as Cohen (2002) acknowledges, more than one knowledge structure can be applied within a single episode of reasoning. A task might involve making predictions about an event from a concept represented in a frame. Therefore, discriminators for all knowledge structures should be tested concurrently.

3 Applying Cognitive Structure Analysis

We applied CSA in the domain of medical underwriting for private health insurance in the Netherlands. Underwriters for private health insurance first assess whether an applicant meets the statutory eligibility criteria (technical underwriting). Then, the medical costs and risks that are associated with the applicant are assessed (medical underwriting). The result is a decision of whether an applicant is accepted for private insurance and, if so, under what conditions.

Underwriting is conducted on the basis of an application form, which is submitted to the insurer by the applicant. The form contains questions about the applicant's current insurance and health. The forms are processed by medical underwriters who have received intensive training for about six months and who have had to work under the supervision of experienced underwriters for about one year. For complex medical issues, medical underwriters are assisted by medical doctors.

3.1 Method

We performed two consecutive studies to acquire the underwriters' knowledge structure as well as content. In the first study, we used the knowledge elicitation technique suggested by Leddo & Cohen (1989): a structured interview technique with embedded problem solving. Open questions

https://doi.org/10.1017/S0269888906000968 Published online by Cambridge University Press

such as ''Can you describe how you generally carry out the underwriting process?'' were used to start the interview and get the expert talking. While listening to the expert, the structures' discriminators (see Table 1) were used to establish hypotheses about which knowledge structure the expert used. Then closed questions, drawn up in advance, were posed to test these hypotheses and to detect the organizing principles they applied (e.g. with a MOP: ''Do you have an overall goal in mind when thinking of these events?''), and the inferential use they made of the structure (e.g. with a production rule: ''Do you use the data to come to a conclusion?''). Small-scale problems, taken from real applications, were introduced to have the experts illustrate what they were talking about. Finally, we summarized our understanding of the structure and content of the experts' knowledge, and asked them to verify whether we had understood them correctly. This procedure was repeated until all the knowledge structures used by the expert had been identified.

In the second study, the knowledge elicitation technique we used was protocol analysis, which involves asking experts to perform a task while thinking aloud. Although this technique has its critics (see, e.g., Baron, 2000), it has been widely acknowledged to provide rich sources of knowledge (Neale, 1989; Ericsson & Simon, 1993; Cooke, 1994; Schreiber *et al*., 2000). The experts conducted health insurance underwriting for 30 real applications, which had been processed two or three years earlier, while thinking aloud. Their verbalizations were tape recorded and transcribed into hard copy protocols. Two judges used the structures' discriminators (in Table 1) to encode the protocols and thus to identify the experts' knowledge structures. Both interviews with problem solving and think aloud are well-known techniques, also used successfully in cognitive task analysis; for example, they have been used in the design of intelligent tutoring systems (Corbett *et al*., 1997).

3.2 Participants

There were 16 medical underwriters and four medical doctors who participated in both our studies. There were 11 underwriters with a background in administration, three male and eight female, with a mean age of 28.4 years; five underwriters with a medical training, three male and one female, with a mean age of 31.2 years; and four medical doctors, three male and one female, with a mean age of 51.0 years.

3.3 Results study 1

All knowledge structures were used by the majority of participants. All used frames, 80% used semantic networks, 90% used production rules and MOPs, and 85% used mental models.

The discriminators of the knowledge structures were used on the three levels of discrimination (basic unit, organizing principle, and inferential use), except with the MOPs, for which the third level of discrimination was not found. On the level of inferential use, not every variant was used; for example, the participants only used ''forward chaining'' with production rules and no ''backward chaining''.

Frame representations were used by all the participants to represent applicant, partner, child, current insurance, requested insurance, product, medication, treatment, and medical aid. The underwriters with a medical background and the medical doctors used an additional frame for the representation of ''disease''. Semantic networks were used by most participants to classify insurance products in a taxonomy. The underwriters with a medical background and the medical doctors also applied the networks for taxonomies of diseases.

The inferential use made of mental models was associated with the background of the participants. Of the medical underwriters with a background in administration, half used no mental models at all, and the others made no predictions and formulated no hypotheses. The underwriters with a medical background and the medical doctors all used sophisticated mental models, while the medical doctors were the only participants who used backward-chaining hypotheses.

https://doi.org/10.1017/S0269888906000968 Published online by Cambridge University Press

3.4 Results study 2

To calculate, from the codes given by the two judges to the protocol segments, the frequency with which different knowledge structures were used, we used counting rules. In frames, we counted the number of concepts or attribute clusters, in semantic networks the number of taxonomies, in production rules the number of data-conclusion mappings, in MOPs the number of event chains or goals, and in mental models the number of concept-event pairs with a causal relationship. To give an example: if we found all three discriminators of a production rule representation, with four events identified, described in data-conclusion pairs and used in a forward-chaining scheme, we counted this as four production rules. We only counted a knowledge structure as such if it was identified on at least two levels of discrimination.

Medical underwriting was conducted by all underwriters ($n=20$) while thinking aloud. All participants used frames, 85% used semantic networks, 90% used production rules and MOPs, and 80% used mental models. We found that the discriminators of the knowledge structures were used on all three levels (basic unit, organizing principle, and inferential use) except for MOPs, for which the discriminators were only used on the first two levels. With production rules, the participants only used forward-chaining inferences, and no backward chaining.

The inferential use made of mental models was associated with the background of the participants. Medical underwriters with a background in administration made no predictions and formulated no hypotheses, while the medical doctors were the only participants who used backward-chaining hypotheses. For example, one doctor said "I suspect she suffers from incontinence (hypothesis); she uses diapers as medical aids (data), this confirms her incontinence (tested hypothesis)".

The frames, used by all the participants, represented the same concepts as in study 1. The underwriters with a medical background and the medical doctors used additional frames for the representation of diseases. Some frame attributes were represented as frames themselves. For example, the concept "medication" was assigned attributes such as the name of the medication, its insurance coverage, its strength, its dose, its duration, its price per unit, and its costs.

Semantic networks were used to classify insurance products, and the underwriters with a medical background and the medical doctors also used them for taxonomies of diseases. For example, one doctor said "a transient ischemic attack is a kind of cerebrovascular disorder". MOPs were used to capture the three-stage process of calculating the underwriting norm, calculating the medical costs, and relating the costs to the norm. The use of mental models appeared related to the background of the participants. Almost half (five) of the 11 medical underwriters with a background in administration used no mental models at all, the other half used extremely simple models. On the other hand, the underwriters with a medical background and the medical doctors all used sophisticated mental models.

3.5 Comparing the results of both studies

With both knowledge elicitation techniques, interviews with problem solving in study 1 and the "think aloud" approach in study 2, we found quite similar results. All five knowledge structures were used in medical underwriting, with different sub-groups of participants using different structures more or less often. We conclude that the two techniques elicit the knowledge structures and discriminators for the medical underwriting task consistently.

We also found identical contents represented by the different knowledge structures. Frames were found to represent the concept's applicant, partner, child, requested insurance, product, medication, treatment, and medical aid. With each elicitation technique we were also able to identify the additional frame representation for the concept "disease" used by medical doctors. Furthermore, using each elicitation technique, we found that medical underwriters use semantic networks for the representation of a taxonomy of insurance products and that medical doctors apply semantic

https://doi.org/10.1017/S0269888906000968 Published online by Cambridge University Press

networks for taxonomies of diseases. Finally, we were able to identify the same MOPs, and the same set of production rules and mental models with both knowledge elicitation techniques.

Two discrepancies between the results with the two elicitation techniques were found. First, fewer frame attributes were elicited during the ''think aloud'' approach (mean per participant per frame=4.8) than during the structured interview (mean per participant per frame=5.2). Second, the semantic networks were composed of fewer concepts during think aloud (mean per participant per semantic network=2.2) than during the structured interview (mean per participant per semantic network=4.2). The explanation is that knowledge that remains implicit and is not verbalized during think aloud can be elicited in interviews with additional probes. Consider the example of a semantic network that was verbalized by medical doctors during the ''think aloud'' approach. They talked about a transient ischemic attack (TIA) as a child concept of a cerebrovascular disorder. The equivalent semantic network elicited in the interview included another child concept of a cerebrovascular disorder: a cerebrovascular accident (CVA). By using probes such as ''what, if any, other concepts are involved in the taxonomy?'' we have been able to elicit a more elaborate semantic network during the interviews.

These results do not yet tell us which structure is most often applied by underwriters; all are used by a majority or all of the participants. To establish which structure is the most popular, we calculated, from the codes assigned in study 2, the average number of each of the knowledge structures the participants used. We found that frames were by far the most frequently used: on the average, for each application 8.25 frames were used. Production rules, the next most frequently used rule, were used far less often with a mean of 0.83 per application, followed MOPs with 0.68 per application, mental models with 0.50 per application, and finally semantic networks with 0.05 per application.

4 Conclusions and further research

We have assessed CSA as a method to elicit the format in which experts represent their knowledge, as well as the content of this knowledge. We found CSA to be a valid method to acquire experts' knowledge structures, and applicable in a real domain. In both studies, we were able to elicit the structures and the content of experts' knowledge. The two elicitation techniques were reliable. They both revealed that, for medical underwriting, the experts' cognitive knowledge representations were of all five types, and study 2 showed that frame representations were by far the most often used. Our findings thus support the idea that underwriting, or probably any assessment task, is a process that is not necessarily best described as just the application of rules (Chipman *et al*., 2000). Indeed, our studies suggest that this type of assessment task is probably best described as reasoning with frames.

It may be remarked that a Bayesian representation is not mentioned by Leddo and Cohen. They may have assumed that people are not good at working with numerical conditional probabilities (see, e.g., Ghosh *et al*., 2004) and will thus not be likely to represent their knowledge in this format. People seem naturally to use a less confusing representation, such as natural frequencies (Gigerenzer & Edwards, 2003), or, as we found, frames. Admittedly, we did not include a Bayesian representation in our list of five that we checked for, but the experience with the shelved knowledge system was that our experts did not, and could hardly, formulate their knowledge as numerical conditional probabilities.

We think the most valuable asset of CSA is that it allows experts to give their knowledge in their own format instead of performing mental gymnastics to explain their reasoning and knowledge. Compared to how hard it had been for the engineers of the company that developed the Bayesian system to acquire experts' knowledge, to us this was quite easy. Admittedly, the task for the knowledge engineer does get harder, because the elicitation sessions require much more preparation. However, with this investment, they easily elicit the experts' knowledge, in a way that is quite easy on the experts. A comparable conclusion was reached about the elicitation of the probabilities needed for belief networks (Renooij, 2001). This also is often seen as a major obstacle

https://doi.org/10.1017/S0269888906000968 Published online by Cambridge University Press

in building such networks, and the ease of understanding and use of the elicitation method appear to be one of the critical issues in choosing which method to apply.

Our conclusion, that people may prefer to use frames, was reached for an assessment task. Future research would need to establish whether it also extends to other types of task. However, we are convinced that the central message, namely, that knowledge engineers would be wise to implement their users' and not their systems' representations in the interfaces of their systems, holds for all types of task and all domains.

The next step would be to equip knowledge systems (e.g. the system for medical underwriting) with a user interface with frame representations, instead of with forms in which to fill in numerical conditional probabilities. We predict that the usability and acceptance of such systems will increase significantly, compared to systems in which no effort is made to translate the Bayesian computations that the system performs into an interface with a representation that appeals to the users' cognitive make-up.

References

Baron, J, 2000. *Thinking and Deciding* (3rd ed.). Cambridge, UK: Cambridge University Press.

Boy, GA, 1997, Knowledge elicitation for the design of software agents. In Helander, M, Landauer, TK and Prabhue, P (eds.), *Handbook of Human–Computer Interaction*. Amsterdam: Elsevier.

Burge, JE, 1998, Knowledge elicitation for design task sequencing knowledge. MSc thesis, Worcester Polytechnic Institute.

Cheng, BC-H, Cupit, J and Shadbolt, NR, 2001, Supporting diagrammatic knowledge acquisition: an ontological analysis of cartesian graphs. *International Journal of Human–Computer Studies* **54**, 457–494.

Chipman, SF, Schraagen, JM and Shalin, VL, 2000, Introduction to cognitive task analysis. In Schraagen, JM, Chipman, SF and Shalin, VL (eds.), *Cognitive Task Analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Cohen, 2002, personal communication.

Cooke, N. J. 1994 Varieties of knowledge elicitation techniques. *International Journal of Human–Computer Studies* **41**, 801–849.

Corbett, AT, Koedinger, KR and Anderson, JR, 1995, Intelligent tutoring systems. In Helander, M, Landauer, TK and Prabhue, P (eds.), *Handbook of Human–Computer Interaction*. Amsterdam: Elsevier.

Craik, K, 1943, *The Nature of Explanation*. Cambridge: Cambridge University Press.

Dubois, D and Shalin, VL, 1995, Adapting cognitive methods to real-world objectives: an application to job knowledge testing. In Nichols, PD, Chipman, SF and Brennan, RL (eds.), *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Lawrence Erlbaum.

Ericsson & Simon, 1993.

Ghosh, AK, Ghosh, K and Erwin, PJ, 2004, Do medical students and physicians understand probability? *QJM: An International Journal of Medicine* **97**, 53–55.

Gigerenzer, G and Edwards, A, 2003, Simple tools for understanding risks: from innumeracy to insight. *British Medical Journal* **327**, 741–744.

Johnson-Laird, PN, 1983, *Mental Models*. Cambridge, MA: Harvard University Press.

Johnson-Laird, PN, Girotto, V and Legrenzi, P, 1998, Mental models: a gentle guide for outsiders, `http://www.si.umich.edu/ICOS/gentleintro.html`.

Leddo, J and Cohen, MS, 1989, Cognitive structure analysis: a technique for eliciting the content and structure of expert knowledge. *Proceedings of the 1989 Conference on AI Systems in Government, McLean, VA*.

Leddo, JM, Mullin, TM, Cohen, MS, Bresnick, TA, Marvin, FF and O'Connor, MF, 1988, Knowledge elicitation: phase I final report. Technical Report No. 87–15, Army Research Institute, Alexandra, VA.

Luger, GF and Stubblefield, WA, 1998, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. London: Addison-Wesley Longman.

Markman, AB, 1999, *Knowledge Representation*. Hillsdale, NJ: Lawrence Erlbaum.

Minsky, M, 1975, A framework for representing knowledge. In Winston, PH (ed.), *The Psychology of Computer Vision*. New York: McGraw-Hill.

Motta, E, 2001, The knowledge modelling paradigm in knowledge engineering. In Chang, SK (ed.), *Handbook of Software Engineering and Knowledge Engineering*, vol. 1. Singapore: World Scientific.

Neale, IM, 1989, First generation expert systems: a review of knowledge acquisition methodologies. *The Knowledge Engineering Review* **2**, 105–145.

Newell, A and Simon, H, 1972, *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Olson, JR and Biolsi, KJ, 1991, Techniques for representing expert knowledge. In Ericsson, KA and Smith, J (eds.), *Toward a General Theory of Expertise*. Cambridge: Cambridge University Press.

https://doi.org/10.1017/S0269888906000968 Published online by Cambridge University Press

Quillian, M, 1968, Semantic memory. In Minsky, M (ed.), *Semantic Information Processing*. Cambridge, MA: MIT Press.

Renooij, S, 2001, Probability elicitation for belief networks: issues to consider. *The Knowledge Engineering Review* **16**(3), 255–269.

Schank, RC, 1982, *Dynamic Memory*. Cambridge: Cambridge University Press.

Schank, RC and Abelson, R, 1977, *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Lawrence Erlbaum.

Schreiber, G, Akkermans, H, Anjewierden, A, De Hoog, R, Shadbolt, N, Van de Velde, W and Wielinga, B, 2000, *Knowledge Engineering and Management: The CommonKADS Methodology*. Cambridge, MA: MIT Press.

Sowa, JF, 2000, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks Cole Publishing Co.

Stillings, NA, Weisler, SE, Chase, CH, Feinstein, MH, Garfield, JL and Rissland, EL 1995, *Cognitive Science: An Introduction*. Cambridge, MA: MIT Press.

Van Geenen, EW, 2004, Knowledge structures and the usability of knowledge systems: a study in health insurance underwriting. PhD thesis, Nijenrode University, the Netherlands.

Williams, MD, Hollan, JD and Stevens, AL, 1983, Human reasoning about a simple physical system. In Gentner, D and Stevens, AL (eds.), *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum.

Winston, PH, 1992, *Artificial Intelligence*. Reading, MA: Addison-Wesley.

https://doi.org/10.1017/S0269888906000968 Published online by Cambridge University Press