6

Characterizing novice and expert knowledge: towards an intelligent tutoring system for archaeological science.

Graham Tilbury[1,2], Ian Bailiff[1] and Rosemary Stevenson[2]

[1] *Department of Archaeology*, [2] *Department of Psychology,*
Durham University, University Science Site, South Road, Durham, DH1 3LE, U.K.

6.1 Background

Teaching archaeology is becoming increasingly difficult due to the wide range of educational backgrounds of the students and to their increasing numbers. Whilst Computer Aided Learning has been and will continue to be useful in easing this burden upon teaching resources, it is limited in its uses due to the way in which its operations differ from a real human tutor. If however we can understand how students in archaeology learn and how their knowledge changes as they increase in expertise, then we may attempt to embody this knowledge in a Computer Aided Learning system that could therefore respond intelligently. Such an Intelligent Computer-Aided-Learning system, or Intelligent Tutoring System as it is more often called, is the long term goal of this research work. The aim of the present study was to take a first step towards this goal by identifying the knowledge that experts and novices have of scientific dating techniques and determining the differences between them.

6.2 Introduction

To elicit and identify the knowledge of the subjects (archaeologists) we used a technique for mapping out how a person views the relationships between pairs of concepts. The technique makes use of statistical scaling procedures to infer the cognitive organisation of a set of concepts. For example, Stevenson, Manktelow and Howard (1988) used a paired comparison task to investigate the cognitive organisation of computing concepts in expert and novice computer scientists. They took 10 concepts important to programming in PASCAL and presented these 10 concepts to the subjects in all possible pairwise combinations. The subjects were required to rate the concepts according to their similarity by assigning a number to each pair ranging from 1 (very similar) to 7 (dissimilar). These ratings gave a measure of the psychological distance between the concepts and multidimensional scaling techniques were then used on the resulting distance matrices to uncover latent structure in the data. The structure can be represented in multidimensional space that can then be used to infer the cognitive structures of the subjects.

Scaling techniques, such as the one just described, are widely used to discover the knowledge of experts. They have also been used successfully to identify some of the differences between the knowledge of experts and novices in a particular topic (Stevenson *et al.* 1988). We therefore used this technique here to identify differences in the content of experts' and novices' knowledge of scientific dating techniques. However, there are some features of a person's knowledge organisation that are not captured by these techniques. In particular, it is not possible using these methods to assess the *coherence* of a person's beliefs. Beliefs are coherent if related beliefs are consistent. For example, suppose that a person believes that the concepts Cat and Dog are highly related and also that Cat and Pig are highly related. Then, in a coherent belief system, the concepts Dog and Pig will also be highly related. If they are not, we would say that this part of the person's knowledge structure was incoherent (Smolensky 1986).

The above example is a simple one and most people's belief systems are very complex. Someone's knowledge of scientific dating, for example, would involve a set of inter-relationships between the concepts underlying a particular technique (such as mitochondrial DNA dating) and a set of higher order relationships between each of the different techniques. Given the complexity of such knowledge, it is not surprising that inconsistent beliefs are common. However, we would expect a difference between novices and experts in this respect. An expert's knowledge of scientific dating should be more coherent than a novice's. We therefore investigated this issue by supplementing the statistical scaling techniques with an alternative analysis that allowed us to estimate the coherence of each person's knowledge. We did this by constructing a network for each subject where the nodes in the network were the individual concepts associated with mitochondrial DNA dating, and the strengths of the links between the concepts were specified by the subject's rating of similarity between each pair. Having constructed such a network for each subject, we were then able to derive a measure of coherence of the network by using a modification of the formula developed by Smolensky (1986).

6.3 Method

6.3.1 Subjects

Sixteen subjects were used. One was an expert (a lecturer in scientific archaeology) and 15 were novices (archaeology students at either the end of their first or beginning of their second year of undergraduate study).

6.3.2 Materials

We selected 11 words that described individual concepts contributing to the overall idea of mitochondrial DNA (mtDNA) dating, e.g. ancestry, temporal mutation.

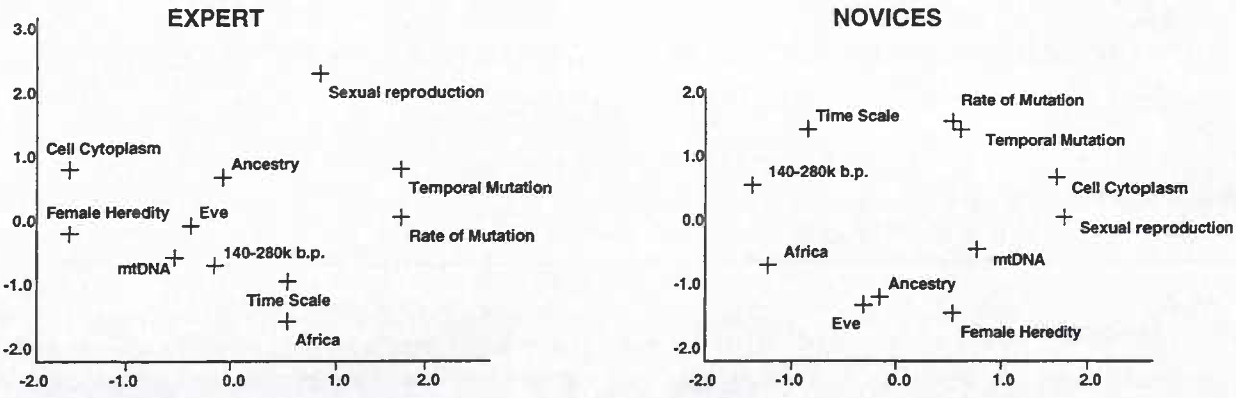

Figure 6.1: The location of the 11 concepts used in the paired comparison task in two dimensional space.

These words were taken from lectures on mtDNA dating previously given to the students. Each word was paired with every other word making total of 55 pairs. The pairs were presented in a single booklet and were in a different random order for each subject.

6.3.3 Design and Procedure

Subjects were presented with all 55 word pairs and asked to rate the similarity of each pair on a scale from 0 to 10, where 0 indicated dissimilar and 10 indicated very similar. A 'don't know' category was also provided for cases where subjects were unfamiliar with the concepts in a pair. The resulting similarity matrix for each subject was analysed using multidimensional scaling. The data were also used to construct a network of concepts and links using a program called NetG which was specially written for this purpose. The program also calculated the coherence of the networks. Each link between pairs of nodes was assigned a strength that reflected the subject's rating of that pair of concepts. Each node was then assigned an arbitrary initial 'activity' which was allowed to spread through the network according to a formula that was modified and developed from Holyoak and Thagard (1989). This firstly diminished the activity of a node by a certain amount, and then increased it in proportion to the activity of other nodes to which it is connected. To perform this calculation once for each node of the network takes a period of time, called a cycle. After several cycles the network will reach a stable state, where any further cycles will not change the activity of any node. Coherence is measured at this point.

6.4 Results

First, we determined whether the subjects used the full range of numbers when rating each pair of words. We found that the novices did use the full range, but the expert was much more likely to use the numbers 0 and 10.

6.4.1 Multi-Dimensional Scaling

A two dimensional scaling solution was produced for the 11 concept names in the comparison task. This yielded similarity matrices of these concepts in two dimensional space, one for the expert and one for the novices (see Figure 6.1). Observation of the novices' matrix in Figure 6.1 suggests that the novices distinguish between biological concepts (on the right hand side of the space) and what might best be called historical concepts (on the left hand side). Observation of the expert's matrix in Figure 6.1 presents a more complex picture. The concepts that the novices grouped together do not form discrete categories in the expert's matrix. We will have more to say about the expert's groupings in the Discussion (below). For the moment we will simply note that the organisation of the same concepts is qualitatively different in novices and experts.

6.4.2 Network Data

The calculation of coherence yields a values between 0 and 1, where 0 is incoherent and 1 is completely coherent. The average value of this figure for the novices was .63 while the expert's value was .89. Thus, as we predicted, the expert's knowledge structure was more coherent than the novices'. We also measured the number of cycles needed for the networks of the novices and of the expert to reach stable states. The expert's network settled after 39 cycles, while the number of cycles needed for the novices ranged from 37 to 46, with a mean of 40.

6.5 Discussion

Both of our measures revealed differences between the expert and the novices. The two dimensional scaling revealed that the novices and the expert had qualitatively different cognitive organisations. The novices appeared to group the concepts around a single dimension concerned with whether the concepts were biological or chronological. This dimension, therefore, is based on similarities between the concepts that presumably existed prior to any training in scientific dating. That is, it seems to be based on previously learned biological ideas and ideas about the origins of humans. The expert, on the other hand, revealed a much more complex organisation of concepts, one based on principles of scientific dating and not discernible to a non-expert. The expert seemed to be using three main groupings. One, at the bottom centre, consists of archaeological concepts (140-280k b.p., Time Scale, and Africa), the second, at the left hand side of the space, consists of a cluster of concepts all associated with mtDNA inheritance and its restriction to the female line, and the third, at the right hand side, consists of the two more technical mutation concepts. Sexual reproduction is isolated in the space, perhaps reflecting the fact that inheritance from both parents is not relevant to mtDNA dating. Overlaid on this basic pattern is the location of four concepts close to the centre of the space (Ancestry, Eve, mtDNA and 140-280k b.p.). These are the four key concepts needed to understand the evidence claiming that all modern humans descend from Eve and they appear to be crucial in linking the purely Archaeological concepts with those concerning mtDNA inheritance. Overall, therefore, the richly organised categorisation of the expert stands in stark contrast to the simple pre-theoretical organisation of the novices.

These results confirm those observed by Chi, Feltovich and Glaser (1981) on expert and novice physicists. They asked experts and novices to categorise physics problems according to their similarity, and found the groupings of novices were based on surface similarities between the problems (e.g. problems involving inclined planes were grouped together) while experts grouped according to deep theoretical principles (such as Newton's third law). What is more, as was the case with our data, the principles underlying the groupings of the experts were only discernible to other experts.

We also examined the coherence of the networks and found that coherence was considerably higher for the expert than for the novices. Incoherent knowledge systems are likely to be a persistent feature of learners as they move from a pre-existing organisation based on previously learned knowledge (such as biological and historical principles) to a new organisation based on the subject matter being learned (such as principles of mtDNA inheritance and scientific dating). Our data suggest, therefore, that the novices were showing the beginnings of a shift in understanding by moving from an organisation based on reproduction to one based on principles of scientific dating.

Some caution must be exercised in drawing firm conclusions from these results because only one expert has been tested. Tests on additional experts are therefore needed to consolidate our findings. However, previous research has found that experts are usually in close agreement with each other on rating tasks like the one used here (Stevenson *et al.* 1988). We have shown, therefore, two ways in which the cognitive organisation of relevant concepts differs between an expert and novices. These differences have important implications for teaching and learning. For example, they highlight the importance of taking pre-existing knowledge into account when assessing students and for developing teaching techniques that point out how the new subject (like Scientific Archaeology) organises things differently from previously learned ones that are not in the primary area, such as Biology and History, as well as how they are similar. In this way, new learning can build on earlier learning rather than being in conflict with it. Furthermore, if we are to exploit the full potential of computer aided learning, then the construction of Intelligent Tutoring Systems will need to incorporate these techniques. They will also need to infer a model of the learner's knowledge if the techniques are to be used successfully. Our research shows how these learner's models can be inferred.

Acknowledgements:

We thank Peter Rowley-Conwy for identifying the expert's organisation of concepts and Simon Lawrence for writing the basic NetG program.

Bibliography

CHI, M. T. H, FELTOVICH, P., AND GLASER, R. 1981. 'Categorisation and representation of physics problems by experts and novices', *Cognitive Science*, 5, 121–152.

HOLYOAK, K. AND THAGARD, P. 1989. 'Analogical Mapping by Constraint Satisfaction' *Cognitive Science*, 13, 295–355.

LAWRENCE, S. H. 1994. *Developing Tools for an Intelligent Tutoring System*, Unpublished B.Sc. Dissertation, Department of Computer Science, Durham University.

SMOLENSKY, P. 1986 'Harmony Theory', in J. L. McClelland, D. E. Rumelhart and The PDP Research Group (eds.) *Parallel Distributed Processing*, MIT Press, chapter 4.

STEVENSON, R. J., HOWARD, M. J. and MANKTELOW, K. I. 1988 'Knowledge Elicitation: Dissociating Conscious Reflections from Automatic processes' in D. M. Jones and R. Winder (eds.) *People and Computers IV*, Cambridge University Press, 565–580.