Identifying Archaeological Knowledge using Multi-Dimensional Scaling and Multiple Constraint Satisfaction

by

Anthony Graham Tilbury

This thesis is submitted to the University of Durham in candidature for the Degree of Master of Science.

Departments of Psychology & Archaeology
Durham University
1994.

Dedicated to my Wife, Karen, for her long-suffering at the hands of my research.

Declaration

I hereby declare that the work reported in this thesis has not been previously submitted for any degree. All material in this thesis is original except where indicated by reference to other work.

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Dr Rosemary Stevenson & Mr Ian Bailiff for all of their help, support and encouragement over the length of this work. They are both knowledgeable and able tutors, and I have learnt a lot from my contact with them.

I would additionally like to thank the support I recieved from the members of both my departments, especially my appraisors who gave me an insight into my work which I had not previously seen: Dr. Crook, Dr Dooley, Dr Johnson & Dr Noel. I also received much support from the staff at the TL Lab and the members of the Cognitive Post-Grad Group. I single out Oliver Davidson from those, for putting up with me for a year and a half, and for getting my mind to work. My thanks also go to Dr Dave Robson of the Computer Science Department for much useful advice and/or the cups of coffee, and to my good friend Simon Lawrence, for his work on the project in the early stages and his tutelage in computer programming.

I wish to thank Keith and Linda & Noel and Olwen for their continued presence and offers of help, especially Keith for always being there whenever I needed him, but always letting me do things how I wanted to do them.

Finally I thank Karen for being my wife - I love her from the very depths of my heart. If this thesis were just half as good a study as she is a lady then it would be a very worthy tome indeed.

Abstract

In this thesis, I look at the current state of research in two fields: the cognitive psychology of learning and expertise & the development of Intelligent Tutoring Systems, especially their methods of modelling the users knowledge state. Within these areas I proceed to examine the way that these theories have overlapped in the past and consider their recent divergence, suggesting that this parting of the ways is premature. I then consider other relevent research so as to suggest a hypothesis where a symbolic connectionist approach to the modelling of knowledge states could be a solution to previous difficulties in the field of Intelligent Tutoring. This hypothesis is then used to construct a method for its examination and also a computer program to analyse the collected data. I then undertake experimental work to validate my hypothesis, and compare my results and methods with a pre-established technique for interpreting the data, that of multi-dimensional scaling. Finally the method now shown to be feasible is discussed to indicate the its sucess and highlight its shortcomings. Further suggestions are also made as to further research avenues.

Contents

Chapter 4 - An ITS for Archaeology

Chapter 5 - Testing the Symbolic Network

Chapter 1

Introduction

The purpose of this research work has been to examine the possibility of designing an Intelligent Tutoring System for Archaeological Science, and in particular to look at how such a system could represent the knowledge of this field. Within this thesis I examine the current state of the research literature in the fields of Cognitive Psychology and Intelligent Tutoring to enable the identification of a suitable method with which to pursue the modelling of Archaeological knowledge. This looks at the wider issues of teaching and learning as a necessary pre-cursor to the specific instances of the application of these theories in Intelligent Tutoring. In particular I look closely at the underlying pedagogy of these techniques, and criticise them for their shortcomings, offering an alternative. The work draws mainly upon the differences between expert and novice performances within a knowledge domain and how these differences may be identified and quantified.

The study is performed using a method identified as holding potential to solve the problems already identified in previous works - the technique is that of multiple constraint satisfaction, as exemplified by the symbolic connectionism approach of Holyoak & Thagard (1989, 1990). Using a sample population of novices (first year undergraduates) and a small set of experts (three University Lecturers) the ability of this method is compared and contrasted with the established methods of multi-dimensional scaling. The results and conclusions of these experiments are then discussed, and further studies identified. To start with, I begin my review of the literature with that of Learning & Expertise.

Learning and Expertise

It is difficult to restrict oneself to a single field of psychology when considering learning and expertise. This is because one cannot simply look at how people acquire expertise and become experts without firstly defining what an expert is - what about them it is that sets them apart from the novice. This consideration has been addressed and the opinion of psychologists (or at least the general consensus) is that an expert can perform *better* than a novice at the particular field, especially at *problem solving* in the field (be it novel problems or familiar problems). Thus we have to look to the field of problem solving and the different psychological theories which it has produced. Only with due consideration to how an expert is supposed to function can we then carry our discussion forwards to see how an expert is different from a novice in the light of these theories. It is my intention, therefore, to firstly look at problem solving, and to rapidly move through theory from the historic to the current. Then armed with this prior discourse I shall move on to look more particularly at the question of what *learning and expertise* are.

Problem Solving

The first theories expounded on problem solving may be accredited to the Gestalt psychologists, such as Duncker (1945) and Kohler (1969). They understood that there were information processing types of skills, but also recognised the limitations of such skills, defining Set and Functional fixedness. However, they were intrigued by *insight* or *ill-defined* problems, and the way in which the answer seemingly 'came' to one in a 'flash' with no apparent conscious thought process. An example of such a problem is that set by Maier (1933). Subjects were given a task of tying together two pieces of string which hung down from the ceiling. The distance between them was too great for you to simply reach out to both simultaneously, but an array of standard tools lay on the floor to help you. The solution to the problem was to use a pair of pliers as a bob-weight on one string and set it swinging, thus getting it close enough to the other string for the subject to grab hold of both together. In Maier's experiment, subjects who did not work out the problem in 10 minutes were given a 'subconscious' clue. An experimenter would come in to check on their progress and 'accidentally' brush the string swaying with their shoulder. Many subjects obtained insight immediately afterwards, even though they couldn't say why the insight came to them.

The way in which the solution occured to the subject by a sudden change in 'point of view' (here in using the pliers as a bob-weight) as opposed to the logical sequence of steps that characterised information processing was only documented though, not explained. We can see that somehow the functionality of the pliers is eventually overcome using analogies to prompt such a Gestalt switch, but we are offered no real explanation of how it did so. Indeed, it was this singular lack of a theory or explanation for *insight* problem solving that may have led to the monumental interest shown in the work of Newell and Simon (1972). They actually published a <u>theory</u> of an information processing <u>solution</u> to problem solving, and that alone marked it as a great breakthrough.

Using their program, GPS (General Problem Solver), Newell & Simon employed weak means-end analysis to solve problems in a fashion that could easily be applied to any field. Weak means-end analysis is the employment of an algorithm whereby a goal is firstly identified, and then intermediate and soluble sub-goals are identified which will reduce the 'distance' between the current state and the desired goal. This takes the form of identifying previously known IF - THEN rules (adopting the terminology of logic and computing) from a library whose preconditions (the IFs) match those defined in the problem. If the outcomes of one of these rules would then result in the end state being closer to the Goal state, then it would be selected and a new search would take place to bring the Goal state yet closer to the new current state. This would occur repeatedly, with the preconditions of each subsequent search equal to the outcomes of the previous resultant search. The mechanism can also work in reverse, thus employing either a forward, backward or combined method of searching the *problem space*. Table 1.1 shows how this process could work in a shopping scenario. Firstly the Goal is identified, and an Action and certain preconditions are inherent in the selected IF - THEN rule. The preconditions need to be checked before the rule can be executed, and in this instance we have assumed that the first precondition has not been met. This would result in the generation of a new Subgoal in order to fulfil the condition. This Subgoal also takes the form of a Goal with associated preconditions for action. This will be found to be true or else the conditions of it will also become Subgoals. The process will repeat itself through all conditions and all Subgoals until the problem is solved. As an account of how some of human problem solving works this theory is very useful and accurate. However, the claims of Newell and Simon that it successfully mimics human behaviour do not necessarily hold up to close scrutiny on some problems as I will shortly report: specifically the *insight* problems identified decades before by the Gestalt psychologists. This inevitably means that the work of Newell & Simon is open to attack as it does not account for human problem solving universally. Until more recently though, there has not been an alternative **theory** of problem solving offered. That state of affairs has now changed.

It was apparent that not all of problem solving could be accounted for by GPS and other information processing accounts, as their claims were more thoroughly tested. Predictions had been made about problem solving such as serial and conscious searching through the problem space. However work with expert chess players showed that this is not necessarily the case: although they appear to spend a long time in finding the correct move during a game, the solution is invariably

Table 1.1 - An Example of Means-End Problem Solving.

If you needed to go shopping to stock up on provisions for the coming week, then you could *solve* such a *problem* through decomposing it into goals and sub-goals as such.

IF		THE GOAL IS TO GO SHOPPING	GOAL
	AND	THERE IS A SHOPPING LIST	PRECONDITION - {P1}
	AND	I HAVE MY CAR KEYS	PRECONDITION - {P2}
	AND	I AM IN MY CAR	PRECONDITION - {P3}
	AND	I KNOW WHICH SUPERMARKET TO USE	PRECONDITION - {P4}
THEN		DRIVE TO THE SUPERMARKET	ACTION

If all of the preconditions are met, then the action will occur. However, if they are not all met, for instance if P1 was not true, then a search would take place for a new rule for which that precondition is a goal state. This rule would then be retrieved and acted upon to solve this sub-goal.

IF		THE GOAL IS TO HAVE A SHOPPING LIST	GOAL
	AND	I KNOW WHAT INGREDIENTS ARE NEED	PRECONDITION - {P5}
	AND	I KNOW WHAT IS MISSING FROM THE CUPBOARDS	PRECONDITION - {P6}
	AND	I HAVE A PEN AND PAPER	PRECONDITION - {P7}
THEN		WRITE THE REQUIRED ITEMS ON THE PAPER	ACTION

This procedure can be repeated with further sub-goals for unfulfilled preconditions until the problem has been decomposed to a point where the problem solving can begin. The preconditions will then be met and the actions carried out, until all of the first preconditions are met. The problem is then solved and you can go shopping. Adapted from Stevenson (1993).

available very quickly; it is just that the chess masters spend a long time checking their solution for possible flaws. The solution offered was to adopt the idea of automatic and rapid production rules, although an alternative is to follow Holyoak & Thagard (1989) using a symbolic connectionist approach. The latter method has great advantages in also accounting for insight problems as will be explained shortly. It was evidence such as this that caused people to doubt information processing as a universal panacea, and to start looking again at the findings of the Gestalt psychologists into insight problems - a set of problems whose solutions had already been documented as occurring without conscious thought and which did not apparently possess any logical process or steps to their solution. This interest began a search for a theory which could describe the insight problems and would thus fill the gaps which were coming to light in the information processing view of problem solving. The answer that was found was the process of analogical thinking.

Analogical thought was first described as a paradigm for problem solving at the start of the 1980's. A good example of what is meant by this is provided in the seminal work of Gentner & Gentner in 1983. In their experiment they used different analogues of electricity to see how this affected peoples' conceptual understanding of it. The subjects were presented with either a water_flowing_in_pipes analogy or people_moving_through_a_racetrack analogy. In the water_flowing_in_pipes analogy pipes were given as analogous to wires in electrical circuits, as both connect components together. A water pump was given as analogous to a battery as both provide a supply to their respective circuits, and a narrow pipe was given as analogous to a resistor as both restrict the flow of the supply through their respective circuits. Water pressure was given as analogous to voltage, narrowness of pipe to resistance and flow rate of water to electrical current. Also the positive relationship between flow rate and pressure in the water_flowing_in_pipes analogy was given as analogous to the positive relationship between current and voltage in electrical circuits, and the negative relationship between flow rate and pipe narrowness in the water_flowing_in_pipes analogy was given as analogous to the negative relationship between current and resistance in electrical circuits. Similar analogies were made with the people_moving_through_a_racetrack analogy, such as a turnstile being given as analogous to a resistor as it restricts the flow (of people) in its circuit.

Each analogy had strengths and weaknesses in explaining electrical theory. For resistors, the water analogy implies that two narrow pipes in parallel will allow less water flow than one narrow pipe as each constricts flow, and hence by analogy two resistors in parallel should have more resistance. Using the people analogy though, two turnstiles in parallel instead of one turnstile should allow a quicker flow of people and so, by analogy, two resistors in parallel should have less resistance than one on its own. The analogies for batteries should also give one true and one false representation, but this time the water analogy gives the true analogy. Their experiments showed that these predictions were the case, in the way the subjects solved electrical problems which they were presented with. These results successfully proved that by the selection of different *source* analogues, the subjects could indeed be influenced in the way which they subsequently understood electrical problems. This is because of the *mapping* of different *elements* from the *source* onto the *target* - the taking of different parts of the analogy and using them as examples to understand the concept of electricity.

By 1987 Metcalfe & Weibe had demonstrated that indeed, not only can analogical thought be shown to be a factor in Human problem solving, but that in some insight problems at least, there is no possibility that conscious information processing of any sort is carried out. In a series of experiments they presented subjects with logical and insight problems to solve such as those presented in Table 1.2. The subjects were also instructed that as they attempted to solve these problems they were to report how closely they felt they were to a solution as they went along (they were actually prompted every 15 seconds by the experimenter for a rating). This report was given in terms of a 'warmth feeling' rated from cold through to hot (and finally solution obtained) which was then scored on a seven point scale by the experimenter. The study showed that with the logical problems the subjects moved in ratings from 1 through to 7 in a fairly uniform fashion as they approached their solution, a finding consistent with Newell & Simon's GPS hypothesis. However, with the insight problems there was no observable shift from cold to warm prior to the solution being obtained - a direct contradiction of their work. This still did not account for the different way in which insight problems were tackled though: what it showed was that information processing in itself was not entirely adequate, or even capable of providing all of the answers.

Table 1.2 - Two problems from Metcalfe & Weibe (1987).

An example of a serial problem:-
"Given containers of 163, 14, 25 and 11 ounces, and a source of unlimited water, obtain exactly 77 ounces of water"
(Luchins, 1942).

An example of an insight problem:-
"Water lilies double in area every 24 hours. At the beginning of summer there is one water lily on the lake. It takes 60 days for the lake to become completely covered with lilies. On which day is the lake half covered?"
(Sternberg & Davidson, 1982).

One of the main points that was found in research into analogical problem solving was the importance of the structure of an analogy to its retrieval & application. The most famous example of this point is the suite of 'Fortress' analogies as studied by Gick & Holyoak in 1980. The problem to be solved is a medical situation, and is described thus.

"A patient has a tumour in his stomach which must be destroyed. A doctor has a medical device which can send a beam of radiation into the patient's body to destroy the tumour. However the tumour is surrounded by healthy tissue. If the beam of radiation is powerful enough to destroy the tumour it will also destroy the surrounding healthy tissue which is unaccepTable, but if the power of the beam is reduced such that the surrounding tissue is undamaged then the tumour will likewise be unaffected"

The appropriate solution to this dilemma is to fire in multiple beams of radiation from several points around the patient so that the beams 'converge' at the tumour. In this way the healthy tissues are only subjected to a tolerably low amount of radiation whilst the tumour receives a much higher dose and is destroyed. What Gick & Holyoak did in their experiment was to 'prime' the subjects with one of several alternative 'Fortress' analogies before presenting them with the target problem. The 'Fortress' problem is described thus.

"A small country has been overthrown by a dictator who now rules from within a strong fortress. It is surrounded by many farms and villages, and roads radiate from it like spokes on a wheel. A great general arose to oust the dictator, and gathered a large army powerful enough to capture the fortress in one concerted attack. However, the dictator had planted mines along the roads to the fortress, and they were set to prevent a large body of men attacking together, but to allow small groups to pass unhindered so he could retain control of the surrounding areas himself. Not only would a large group set of the mines and destroy itself, but it would incur the dictators wrath and he would retaliate destroying many of the surrounding villages. An attack thus seemed impossible"

The source analogy given is structurally identical to the target to allow for a successful transfer of the analogy. By this I mean that each individual part of the source analogue has an identical corresponding part in the target. Thus the <u>General has an Army</u> is analogous to the <u>Doctor has a Radiation Beam</u> and the <u>Patient has a Tumour</u> is analogous to the <u>Country has a Dictator</u>. Likewise, the tumour being surrounded by healthy tissue is analogous to the fortress surrounded by villages and farms; the destruction of the tumour is analogous to the destruction of the fortress & the destruction of the destruction of the tumour leading to the destruction of healthy surrounding tissue is analogous with the assault on the fortress leading to the destruction of the army and the retaliatory destruction of the surrounding villages. Without this structural identity the analogy would not be successfully utilised.

The source analogues given were of three different types. In one the general splits his army into small groups and sends each down a separate road into the fortress thus capturing it: this is analogous to the 'correct' solution - this is the convergence condition. In a second source analogue the general finds one unmined road and sends his entire force down it - the open passage condition', and in a third the general digs a tunnel to the fortress and attacks through it - the incision condition. In response to these different sources the subjects did indeed produce different solutions to the tumour problem. Whilst those with the convergence analogue often found the converging beams solution, those with the 'open passage' source analogue often suggested the passing of the radiation beam through an equivalent open route in the body such as the oesophagus. Similarly those with the 'incision' analogue often suggested cutting the patient open to remove a line of healthy tissue down which to fire the radiation beam.

These examples have given a flavour to the rich research of the past fifteen years into analogical problem solving which in itself is most merit worthy. By the end of the decade though (1989, 1990) Holyoak & Thagard had provided the first ever <u>theory</u> of solving insight problems with their own computer programs of Analogical Thought - ACME (Analogical Constraint Mapping Engine) & ARCS (Analogical Retrieval by Constraint Satisfaction). That a theory could be proposed was much more valuable than the descriptions of analogical problem solving alone. Drawing on the work of Gick and Holyoak (1980) & Gentner and Gentner (1983), it was already apparent that the structure of the analogies held a great deal of importance for the mapping of source problems onto the target problem. That is the source, such as the fortress problem only maps successfully if it matches the structure of the target - the tumour problem. Before discussing the ACME & ARCS work though we must consider two other pieces of evidence, since the theory of Analogical Thought which Holyoak & Thagard proposed draws additionally on the observations that analogical thought is both semantically and pragmatically constrained.

In 1986 Gentner & Toupin clearly showed how semantics plays a role. They presented children with a story including animals as actors, each with a various role. The children were asked to act out the story as it was read to them using specially provided props. They then presented a second version of the story to the children where the plot was very similar but the actors were of three different experimental conditions. The actors were either similar actors in the same roles (SS), dissimilar actors in the same roles (DS), or similar actors in dissimilar roles (SD). As predicted, the transfer of the first analogy to the second was greatest for the SS condition, whilst the SD condition caused the poorest mapping. Thus although the structure stayed the same, the semantic content also had its effect.

The issue of pragmatics has been clearly demonstrated by both Holyoak & Thagard (1989) and Keane (1990). Holyoak & Thagard cite as a 'common sense' observation that in situations such as politics, where both semantic and structural content may be the same the pragmatism of the person's own bias or opinion comes in to play. They defend this statement with a simulation in the ACME system to the way in which a decision about the CONTRA rebels in Nicaragua can be biased by analogy to either Hungarian rebels (labelled as freedom fighters in the west) or the P.L.O. (labelled as terrorists). Both are structurally and semantically identical source analogues, but the personal bias of the subject associates the label of freedom fighters or terrorists to them - a purely pragmatic constraint. They argue that depending upon which of the possible different views you take of the CONTRA rebels will affect which of the analogies is successfully transferred, and indeed the simulations they ran with the ACME program using the data above confirmed their expectations. However, the original assumption about human nature is unsubstantiated empirically for this example, as interesting as the results may appear to be. More convincing then is the work of Keane on this area, as he does indeed base his work upon experimentation.

Keane (1990) presented subjects with the Maier problem of joining two distant pieces of string, but made some changes. Firstly there was a second way of joining the strings, by using a long pole to pull one string over to the other. Secondly, all of the subjects were presented with a story before the experiment to act as the source analogy instead of receiving the experimenter hint. The analogy had two versions, and was concerned with a helicopter rescuing people trapped in a burning building. In one version the rescue is successfully performed by people swinging to an adjacent building from a rope dangling under the helicopter. In the other version the same strategy was shown as too time consuming for an effective rescue, and so a different strategy was used to save the people. Thus the same analogy of a swing rope is given as source, but the pragmatic difference is that it is seen to work for one group but not by another. Indeed, the transfer of the analogy does follow the expected pattern in that those given it as a successful story transferred it and used the pliers solution, whilst those given it as an unsuccessful story ignored it and used the pole solution.

In the ARCS program (1990), the target problem (which is inputted in a predicate calculus notation) has its predicates taken as a 'probe' for retrieval. Then where structurally similar[1] predicates are found in memory a possible mapping is established and stored. All possible mappings to sources in memory are explored, and then all of the candidates are taken and a constraint satisfaction network is created around them. Different possible mappings for the same target predicate are mutually inhibitory whilst possible mappings from the same source but for different predicates are mutually excitatory. Semantically similar mappings receive additional excitation

---

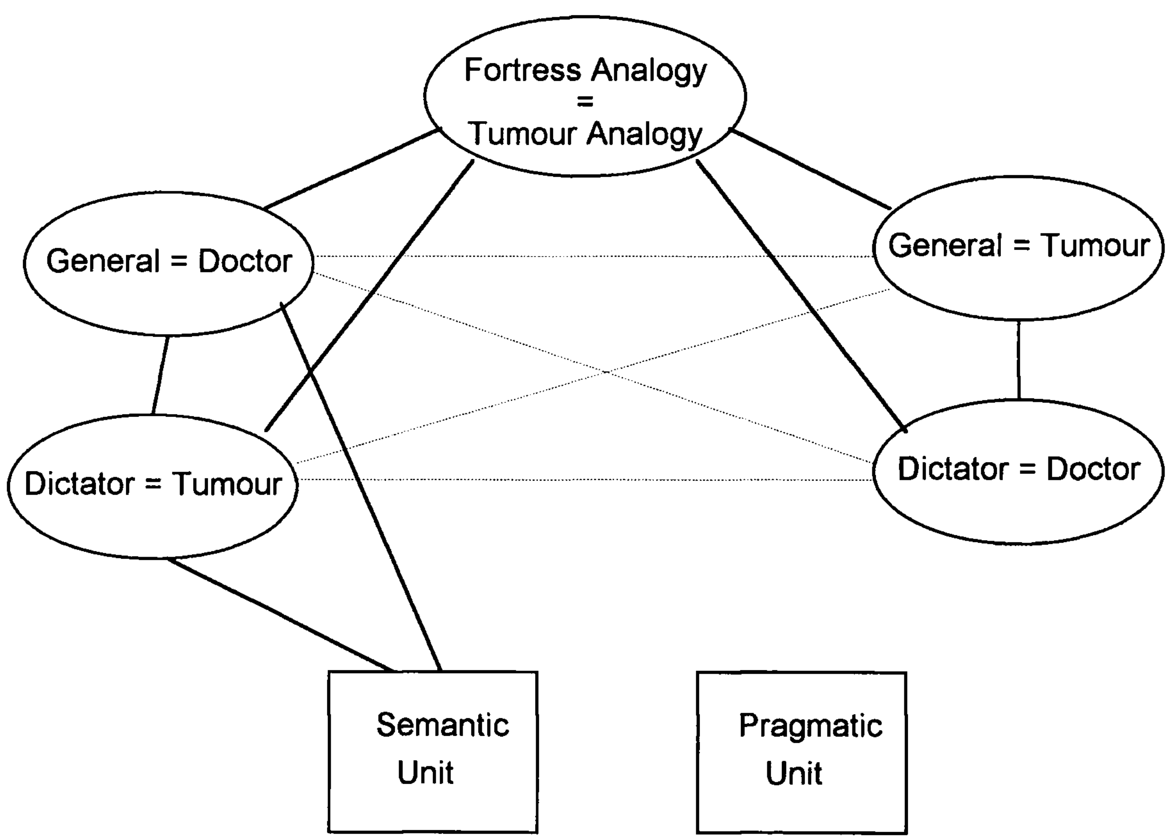
[1] - A structurally similar predicate is one that takes the same number of *arguments*. For example Loves(*John,Mary*) {which is a notation for John loves Mary} is structurally similar to Hates(*John,Mary*) though they may be semantically different. Likewise Loves(*John,Mary*) is structurally different from Loves(*John,Mary,Baby*) whilst they may be semantically similar.

from a semantic unit, and prior beliefs/assumptions receive additional excitation from a pragmatic unit. The network is then run and allowed to settle to asymptote. The mappings which then end up with the greatest activation mark the analogy for retrieval. (This has been until very recently unsubstantiated as a simulation of human behaviour, but recently published experiments do show that structural similarity does indeed play a role in retrieval too - Wharton, Holyoak, Downing, Lange, Wickens & Melz, 1994).

Once the analogy is retrieved it has to be mapped on to the target. ACME (1989) takes the source and pairs all possible mappings which are structurally similar to those in the target problem. Units also receive pragmatic and semantic weighting, but they are viewed as psychologically less important in the mapping than in the retrieval. Again a constraint satisfaction network is established with excitatory links between pairs having different target predicates and inhibitory links between pairs having the same target predicate. The network again settles to asymptote, and the best mapping from source to target is apparent in the final activation states of the possible mappings units

To make this idea clearer, consider the partial constraint satisfaction network in Figure 1.1. The possible mapping of the two analogies has been made because of the structural similarity of the analogies. However, there are several possible mappings which can be made - for instance, the General could be mapped on to the Doctor or the Tumour. However, it could not be possible for him to be both and so all possible mappings of General are mutually exclusive. In the network there is an inhibitory connection between them (represented by a dotted line). If the General were to be analogous to the Doctor, it would also therefore follow that the Doctor could not be analogous to the Dictator, and so there is also an inhibitory link between these possible mappings. Meanwhile it is fully consistent for the Dictator to be analogous to the Tumour given the General analogous to the Doctor and so these mappings are mutually excitable (as represented by a solid line). As all the mappings are structurally possible they all receive excitatory links from the Fortress Analogy = Tumour Analogy mapping. However, the mapping of the General and the Doctor are semantically similar in this story and so would receive an excitatory link from the semantic unit of ARCS. It would be this additional input to the system which would cause it to settle with higher activity on the General = Doctor mapping than the General = Dictator Mapping. In this partial example the pragmatic unit of ARCS makes no contribution, but it in general would act like the semantic unit, adding excitatory links where a possible mapping would be pragmatically favoured.

Figure 1.1 - A Partial Constraint Satisfaction Network for the Fortress Analogy = Tumour Analogy Mapping



Holyoak & Thagard have successfully used ARCS to find suitable analogues for problems such as retrieving correctly the Sour Grapes fable from amongst all of Aesops fables, and for finding the best 'convergence' analogy from different contenders including the Fortress problem (after Gick & Holyoak, 1980) to solve the Tumour problem (Thagard et al, 1990). They have also used ACME to successfully simulate the mapping of sources onto targets for the same 'convergence' problems, and also for many other analogy scenarios including Rutherford's analogy of the solar system to the atomic nucleus and the missionaries and cannibals problem to the farmers dilemma. These simulations match the observed psychological results on analogy found previously by themselves and other researchers (Holyoak & Koh, 1987; Gick & Holyoak, 1980; Gholson et al, 1988; Falkenheimer et al, 1986; Gentner & Toupin, 1986; Thagard et al, 1990; Kittay, 1987) in two ways. Firstly, that the mappings obtained were the same as with human subjects, and secondly that the percentages of subjects who retrieved an analogy formed the same rank as was made by the number of cycles which the constraint satisfaction network took to settle; i.e. the program found analogies as comparably difficult as did human subjects.

Problem Solving

We are left therefore with two competing research paradigms for problem solving. We have two apparently different types of problem - serial, logical, conscious problems and parallel, insight, unconscious problems. We likewise have apparently two hypotheses about problem solving. That all problems are solved through serial, logical, conscious (though eventually automated) information processing strategies, and that problems may be solved through unconscious, parallel constraint satisfaction methods. Thus acquainted (and armed?) with the two opposing theories of problem solving, we may move on to look at theories of expertise.

Acquiring Expertise

The acquisition of expertise was pursued by the information processors as much as possible, but as with the Gestalt Psychologists before them they were only capable of description, not explanation (DeGroot, 1965, Chase and Simon, 1973). That was until the work of J.R. Anderson. His ACT (Adaptive Control of Thought) theory (1983) is very similar to that of Newell & Simon in its use of IF ... THEN rules, or Production Rules as Anderson calls them. However, it additionally states in explicit terms how such rules may be constructed from declarative knowledge originally, and then how these rules may be changed, altered and modified so that the knowledge itself changes from that which characterises a novice, to that which characterises an expert.

ACT is a 3 stage theory. In the first stage the learner utilises pure declarative knowledge only. This is viewed as raw use of knowledge, such as being instructed to drive a car. Here the student gets a declarative input [instructions from the driving instructor] and converts them into actions [i.e. release accelerator, depress clutch, take gearstick out of first, put gearstick into second by moving it downwards, release clutch, press accelerator again]. The learner moves on to the second stage when they start to form procedures (IF ... THEN rules).

The second stage of ACT is that of proceduralization. Here the explicitly learnt declarative knowledge is compiled into procedures. Thus a procedure for changing from first gear to second gear would be created, i.e.:

P1

IF		THE GOAL IS TO CHANGE FROM FIRST GEAR TO SECOND GEAR
	AND	I STILL HAVE MY FOOT ON THE ACCELERATOR
THEN		RELEASE ACCELERATOR

P2

IF THE GOAL IS TO CHANGE FROM FIRST GEAR TO SECOND GEAR

AND I HAVE RELEASED THE ACCELERATOR

THEN DEPRESS THE CLUTCH

AND LIKEWISE FOR THE REST OF THE MOVEMENTS

The second stage of ACT is also marked by the compilation of procedures, so the Procedures P1, P2, ... for changing gear from first to second will eventually be compiled to give just one procedure, P:

P

IF THE GOAL IS TO CHANGE FROM FIRST GEAR TO SECOND GEAR

THEN RELEASE ACCELERATOR, DEPRESS CLUTCH, MOVE GEARSTICK IN TO SECOND, RELEASE CLUTCH, PRESS ACCELERATOR.

The third stage of ACT is marked by three further refinements, and is the stage of fine tuning procedures. The three refinements are strengthening, Generalization and Discrimination. Strengthening is a simple adherence that each time a procedure is used it becomes stronger: that is, more automated, quicker firing, more subconscious, more durable. Generalisation occurs when the common elements of a set of productions are noticed and form an overarching production. In our continuing car example, having acquired productions for the changes from second to third, third to fourth, and fourth to fifth gear too, our subject may now generalise from these similar productions to a new changing up a gear production P_G:

P_G

IF THE GOAL IS TO CHANGE UP A GEAR

THEN RELEASE ACCELERATOR, DEPRESS CLUTCH, MOVE GEARSTICK IN TO THE NEXT HIGHEST GEAR, RELEASE CLUTCH, PRESS ACCELERATOR

The final point is discrimination, and this is the acquisition of knowledge as to when the production is inappropriate. In the continuing example of gear changing, discrimination would add the notion that it is inappropriate to use the general form of the changing-up procedure P_G if the gearstick is already in the highest gear: thus the discriminated form is P_{GD}

P_{GD}

IF THE GOAL IS TO CHANGE UP A GEAR

AND THE GEARSTICK IS NOT ALREADY IN THE HIGHEST GEAR

THEN RELEASE ACCELERATOR, DEPRESS CLUTCH, MOVE GEARSTICK IN TO THE NEXT HIGHEST GEAR, RELEASE CLUTCH, PRESS ACCELERATOR

The ACT theory then follows the information processing line but makes clear the stages which a learner passes through as they transit from novicehood to expertise through the conscious conversion of declarative knowledge to procedures, which with practice are compiled, then generalised, discriminated, strengthened and automated. However, all is not rosy with Anderson's theory. In the same way that Newell & Simon, and more particularly Information Processing itself fell short of universal applicability, so too does ACT. The wealth of criticism against it has now reached such a proportion that it must surely be a matter of little time before an analogical account of expertise is fully expounded. The criticism levelled so far at Information Processing (IP) accounts of expertise and ACT in particular are numerous, and I will here cover in the main those highlighted by Keith Holyoak (1991) and Rosemary Stevenson (1993), as they between them look at almost all of the relevant evidence. I shall go though this evidence though case by case, as there are some 13 or more valid and distinct objections.

Case 1: I.P. accounts of expertise claim that experts in a field have superior pattern perception for their domain. However, Allard & Starkes (1991) in a study of expert volleyball players found that experts had no better pattern perception for the game than novices. This is undoubtedly due to the fact that offensive positions in volleyball are designed to be misleading, and so are not used by the opposition as a part of their game playing strategy.

Case 2: I.P. accounts of expertise claim that problems are solved both more quickly and more easily by experts than by novices. However Scardamalia & Bereiter (1991) in a study of expert writers found that they took longer over their work and felt more pain in doing their work than did novices.

Case 3: I.P. accounts of expertise argue that experts have a superior memory for knowledge of their domain. Adelson (1984) showed that in computer programmers though, novices had a better memory for code than did experts. Experts did have a good domain memory too, but for different things, such as the overall structure of the program.

Case 4: I.P. accounts imply that procedures become highly automated and inflexible with expertise. However, Anderson himself (1987) has shown in work on text editors that domain knowledge can be flexibly re-organised. So too have Lesgold, Rubinson, Feltovich, Glaser, Klopfer & Wang (1988) in a study of radiologists. In this experiment both novices and experts were shown an X-ray of a patient who although healthy, had had a lung removed previously. This causes the heart to move in the rib cage and gives the impression of an enlarged heart - indeed, this is the diagnosis all gave initially. However, after being told that the patient was also in perfect health and had previously undergone major surgery, the experts immediately reorganised their thoughts and came up with the correct diagnosis, whilst the novices remained helplessly unenlightened.

Case 5: I.P. accounts clearly state that expertise is the proceduralization of conscious declarative knowledge. Berry & Broadbent (1984) though suggest that implicit knowledge can be acquired without any conscious declarative understanding of the knowledge learnt, and yet still exhibit expertise.

Case 6: I.P. accounts most distinctly emphasise the domain specificity of expertise. Yet Ericsson & Polson (1988) have shown through the study of the waiter J.C. that expertise in one activity (memory for restaurant orders) can be transferred to memory for a different domain (abstract categories). Also Doerner & Schoelkopf (1991) demonstrated that businessmen can successfully transfer their expertise from their own domain to a complex simulation of subsistence agriculture.

Case 7: I.P. claims that expertise is a function of practice. However Chi et al (1989) clearly demonstrates with physics students that the time spent is not relevant, but that the form that the practice takes is what is crucial. Also Gentner (1983) showed that practice time is not of its own enough to achieve expertise in typing, but that a restructuring of the task is also necessary.

Case 8: I.P. accounts of expertise state that experts use domain specific productions whilst novices use weak declarative knowledge in problem solving. Yet Larkin (1981) found that expert physicists used declarative knowledge to solve problems too, but that they used it in a more expert way. They reformulated the problem first, and then applied declarative knowledge.

Case 9: I.P. accounts of expertise also state that novices initially use weak means-end analysis to solve problems and later on this is proceduralized in domain specific, expert, problem solving methods. However Sudler et al (1983,1985) has shown that on algebra-word problems, the use of means-end analysis actually impairs performance.

Case 10: I.P. ideas of expertise say that expertise is accompanied by a switch from backwards search to forwards search strategies. Anderson himself though (1984) has shown that expert computer programmers use backward search, whilst Doerner & Schoelkopf (1991) have shown that expertise can be hallmarked by flexibly switching between several different strategies.

Case 11: I.P. tells us that for the novice to be successful in attaining expertise they need feedback on performance and they need to have clearly defined goals. In direct contradiction to this however is the work of Sloboda (1991). He tells us that such feedback and clearly defined goals can actually be detrimental in the case of acquiring musical expertise, and this is backed up empirically by another study by Sloboda & Howe (1991) which links failure to achieve expertise with the pressure of such feedback and goals.

Case 12: I.P. necessarily implies that if we use productions as experts, then by finding those productions we may predict expert behaviour. Lundell (1988) tried to perform such a task, compiling productions for experts and then trying to predict the experts' solutions to problems. This however failed to be predictive to any statistical significance, and worse still one experts' model was as predictive of any other expert. By comparison though, the same process was also undertaken using a trained connectionist network which was both statistically predictive and also was specific to the individual expert upon whom it was modelled.

Case 13: I.P. similarly implies that if we may find such expert production rules we may then teach these rules to novices and hence allow them to gain expertise. Indeed this is central to Intelligent Tutoring Systems if they are to work using an Information Processing paradigm. Wenger (1987) argues against this paradigm, and lays the failure of intelligent tutoring firmly at the feet of overlay ideas of expertise. (I will discuss this more later on in this thesis, as it is a pivotal point, and one which we seek to remedy).

As you can see, there is a large body of evidence against ACT and the information processing point of view. Indeed, Holyoak (1991) sees that most of these points are already answered by the symbolic connectionism theory of analogical thought which he and Paul Thagard propose. Yet to stop here would leave much still unsaid about expertise. Again it would appear that we are not far from two competing views of expertise as we have two competing views of problem solving. Is it necessarily the case though that they are competing views? As I said earlier. there are 13 *or more* cases against pure information processing. However, there may well be room for both theories in accounting for expertise. To understand more deeply we must look at how expertise is acquired in a wider context. That of social implications and of metacognitive skills and learning strategies.

Other Avenues

To examine further the theme that the two different accounts of problem solving and expertise may co-exist I will start with the work of Giyoo Hatano (1988). This paper is of great importance for its clarity in distinguishing between two different types of expertise - Routine & Adaptive - and for doing so in real life activities. Hatano studied Japanese children acquiring expertise in the use of an abacus for mathematical computation. They start off with the declarative knowledge of its operation, and with practice they continue to increase in speed and accuracy in using the abacus. Experts can perform many complex arithmetic problems in a very short space of time (through proceduralization), and some can even do away with the physical abacus itself. Such a tale appears to closely follow the ACT account of expertise. Indeed, these children also conform to the domain specificity of ACT, finding it impossible to transfer their knowledge to a different domain: with a pencil and paper they lose their 'expert' abilities of speed and accuracy (Amaiwa, 1987). These experts Hatano defines as Routine, because they lack the ability to adapt to novel situations and problems.

To illustrate what an adaptive expert is by comparison, Hatano directs us to the studies of Carraher, Carraher & Schliemann (1985,1987). In their studies of the Brazilian children who are street vendors they found that they had acquired their expertise through the need for honest and transparent salesmanship, with constantly changing prices. Also, their normal working mathematics is vocal, not silent, and their mathematics is worked upon visibly concrete representations - the quantities of fruit they are selling. This obviously led to the children acquiring a deep conceptual understanding of arithmetic since in experiments they were able to apply (transfer) their knowledge to a different mathematical domain. This, according to Hatano, means that they are adaptive experts.

The reasons for this difference are apparently due to the way in which the street vendors learnt their mathematics. Adaptive experts have to possess a deep conceptual understanding as opposed to automatic procedures if they are to achieve cross domain transfer and the ability to tackle novel problems, and according to Hatano there are four conditions which when met lead to the acquisition of conceptual, not procedural knowledge. These conditions are: 1/ That the learner continually encounters novel problems, thus emphasising broad conceptual knowledge over proceduralized speed at a few problems; 2/ That encouragement is given to the learner to seek comprehension, thus leading to the employment of metacognitive skills during learning; 3/ That the learner is free from immediate feedback, as otherwise speed will be encouraged at the price of understanding & 4/ That learning takes place in a vocal and social exchange, as this promotes self examination of knowledge more than solitary learning will do. By these categorisations, the Japanese abacus users conforming to none are well placed as routine experts, whilst the Brazilian street vendors conforming to all 4 conditions are well defined as adaptive experts.

Hatano does also note however, that this does not imply that procedural knowledge is necessarily inferior, as it has advantages of speed and accuracy over conceptual knowledge. He merely seeks to note that they are indeed different and distinct. This notion is clearly supported by the study which Mayer and Greeno (1972) carried out. They taught two groups of learners about elementary statistics, but one group was taught procedural knowledge first (formulae) and then conceptual knowledge (relating statistics to everyday life and occurrences) at a later point while the order was reversed for the other group. The group taught procedural knowledge first performed better at subsequent testing for rote application of formulae, but floundered at questions about the formulae themselves and at insoluble questions. The conceptual knowledge first group by comparison showed the exact opposite pattern of ability. The training to routine or adaptive expertise is thus possible, and the choice may be motivated by the end need for such expertise.

Another study which makes a distinction between routine and adaptive expertise is that of John Sloboda (1991). He has studied the expert musicians at the Chetham school in Manchester to investigate the differences between expert musicians who 'make it' to the uppermost echelons of their profession (concert soloists, first violins etc.) compared to the average attainers (normal orchestra members). To do so he obtained reliable predictions of future performance from the school's teachers, and thus also managed to question the learners about their practice (learning) regime whilst it was still fresh in their minds. The most striking distinctions to be found were that those who achieved true greatness: 1/ Came from non-musical families where their efforts were judged as praise-worthy (when a musical set of parents could have been more critical, if trying to be constructive); 2/ Spent more time practising with second and third instruments whilst their less able peers concentrated more on their first instrument & 3/ Concentrated less on scales and set pieces but instead explored other styles or their own compositions/improvisations. This study too enforces the theoretical viewpoint of Hatano that there is a societal / motivational importance to the acquisition of different expertise, and that an adaptive expert obtains broader and deeper conceptual knowledge through their learning instead of concentrating on procedural routines.

It does not take a large leap of faith to see that such a broad base of experiences instead of repetitive practice at a small range of problems can lead not only to the acquisition of deep conceptual understanding, but that it will also by necessity mean that adaptive experts have a greater base of possible analogues to apply to novel situations than a routine expert. Thus we may also attach firmly the notion of adaptive expertise to analogical theories of problem solving as we did routine expertise to procedural models of problem solving and procedural models of expertise. We thus open the door even wider towards an analogical theory of expertise. To try and formulate what such a theory would need to incorporate though, we shall have to continue our characterisation of how adaptive expertise is acquired, and also think perhaps about the ground in-between pure expertise, be it routine or adaptive.

There is now an increasing body of evidence that metacognitive skills are of immense importance to the acquisition and utilisation of adaptive expertise. Two generally accepted metacognitive abilities are self-explanation and self-monitoring. I will look at both these skills in the context of different studies, starting with self-explanation. This metacognitive skill can be seen most easily in the study by Chi et al

(1989). The paper presents an analysis of students' study habits and methods when given novel learning (about physics), and it also categorises the subjects by testing of their learning into "Good" or "Poor" students (by whether they were in the more or least successful half of the subjects). Looking for correlations in this data, it was apparent to Chi et al that the two groups definitely employed different learning methods. The "Poor" students did not employ any metacognitive abilities, but instead focused on specific examples in a routine fashion: they sought to solve problems without engaging in self-monitoring or explanation. The "Good" students by comparison did engage in those metacognitive activities, not relying upon the examples but approaching the field in a conceptual and adaptive way. This shows that apparently by college age, learning strategies have already been established and fixed, which correspond to either the employment of metacognitive abilities or lack of it.

Similar results were found by Brown et al (1977, 1981, 1983) as to the use of metacognitive monitoring in acquiring adaptive expertise. The task of summarising text using comprehension strategies is a complex one with 6 strategies identified by Brown and colleagues. The difficulty of such a task may be appreciated given that even college students do not always utilise all of the strategies when summarising. However Brown et al demonstrated that the use of metacognitive monitoring can be effective in the selection of all strategies for use. In an experiment they took three subject groups and asked them to summarise various texts. One group was merely told to perform the task; a second was told also that there were 6 useful strategies which should be employed in summarising, and they were also tutored in the 6 strategies. A third group not only received the instructions of the second, but were also taught about metacognitive monitoring - they were explicitly instructed in how to check that they *had* used the 6 strategies in their summarisations. The results of the experiment showed that indeed, the third group performed better than the second group who performed better than the first group. It was also observed that the greatest improvement with the third group came when they had to summarise a difficult text or when the subject had originally been less proficient in summarisation originally. Thus metacognitive monitoring allowed the subjects to check that they were utilising a skill which others had learnt but failed to appropriately apply.

Allied to these two studies is the work of Dweck and colleagues (1978, 1980, 1985, 1988). This work was carried out on junior school children, and it was discovered that by that age children had already become fixed into one of two learning categories - Helpless or Mastery-oriented. Helpless children when confronted with a difficult problem attributes the failure to themselves; they will become avoidant, trying to change the subject to something which they are skilful at, but suffer from a decrease in performance at tasks. Mastery-oriented children by comparison do not attach failure to themselves but mastery; they work hard and employ self-explanation as they endeavour to overcome the difficulty. This results in increased task performance. Dweck sees this as the result of the children's intuitive theories of their own intelligence. These are respectively identified as Fixed or Incremental theories of intelligence. Helpless children see intelligence as fixed and given, thus if you can only ever be so intelligent then a difficulty will always be a difficulty. To this theory then, the answer to such difficulties is obviously to ignore them and go on to something which you can achieve, but it also means the learner feels inferior and stressed by their inability causing degraded performance. Mastery-oriented children see intelligence as incremental though, thus learning can continue to add to the intellect, thus by overcoming a difficulty the learner can increase their intelligence. To this theory then, the solution to a difficult problem is perseverance, as by understanding the problem better through self-explanation and repeated attempts to solve it the learner may gain mastery, solve the problem and increase their abililty.

Again, such a distinction can be looked at in terms of a difference between adaptive and routine learning strategies: Helpless children concentrate on a few problems at which they can become quick and efficient whilst mastery-oriented children employ metacognition to discover a wider set of problems and solutions and thus attain greater conceptual understanding. The study also takes on board some possibly underlying social trends too. Helpless children were not originally helpless - younger children do not exhibit such behaviour. What has happened is that society has forced the children into such intuitive models of intelligence through the experiences the child has had. If teachers only value routine expertise (rapid and accurate problem solving) then the learner will see such as the only thing worthy or to be valued. This ties in also with the work of John Sloboda (1991) mentioned earlier. The encouragement and motivation a child receives (or any learner) will affect their view of problem solving and expertise, albeit that the example may not be intentional.

These three studies then underline the importance of metacognitive skills in the acquisition of adaptive expertise, and that people use such adaptive learning strategies from the primary school through to university level. It also shows us that people get stuck in their ways easily, and usually stay stuck. However, with explicit instructions the learner can be made to employ adaptive methods when they previously did not make use of metacognitive skills. This implies that someone who has never followed the adaptive/Mastery-oriented/metacognitive path before can, through explicitly addressing such lack, be brought to employ adaptive methods of learning and problem solving and so acquire adaptive expertise. This then is almost enough for us to define what a theory of adaptive expertise would look like, but we must add one final (and perhaps to some surprising) element. This centres on the question of how pure are adaptive or routine expertise.

Stevenson & Palmer (1994) note that routine problem solving (proceduralization) has one great power - the freeing up of working memory for other tasks. Without this ability an expert could never tackle the problems commensurate to his or her level, as their working memory would be overloaded by all of the rest of their domain knowledge. This point must hold just as true for an adaptive expert as it does for a routine expert too, or else they could never tackle expert problems (unless they had a superhuman working memory!). This insight is deceptively simple yet it can all to easily be forgotten (or ignored). However, without the ability to access rapidly at least some of our prior knowledge (such as terms and concepts), new knowledge acquisition would quickly become impossible because of the limitations on our working memory. This now in place too, we may move on.

We can propose that there are 4 required components to an Analogical Theory of Expertise then in any domain.

1/ The acquisition of conceptual knowledge about the particular domain.

2/ The acquisition of a large amount of diverse knowledge both from within and without the domain.

3/ The acquisition of metacognitive abilities which will allow both

a) the formation of links between 1/ & 2/. &

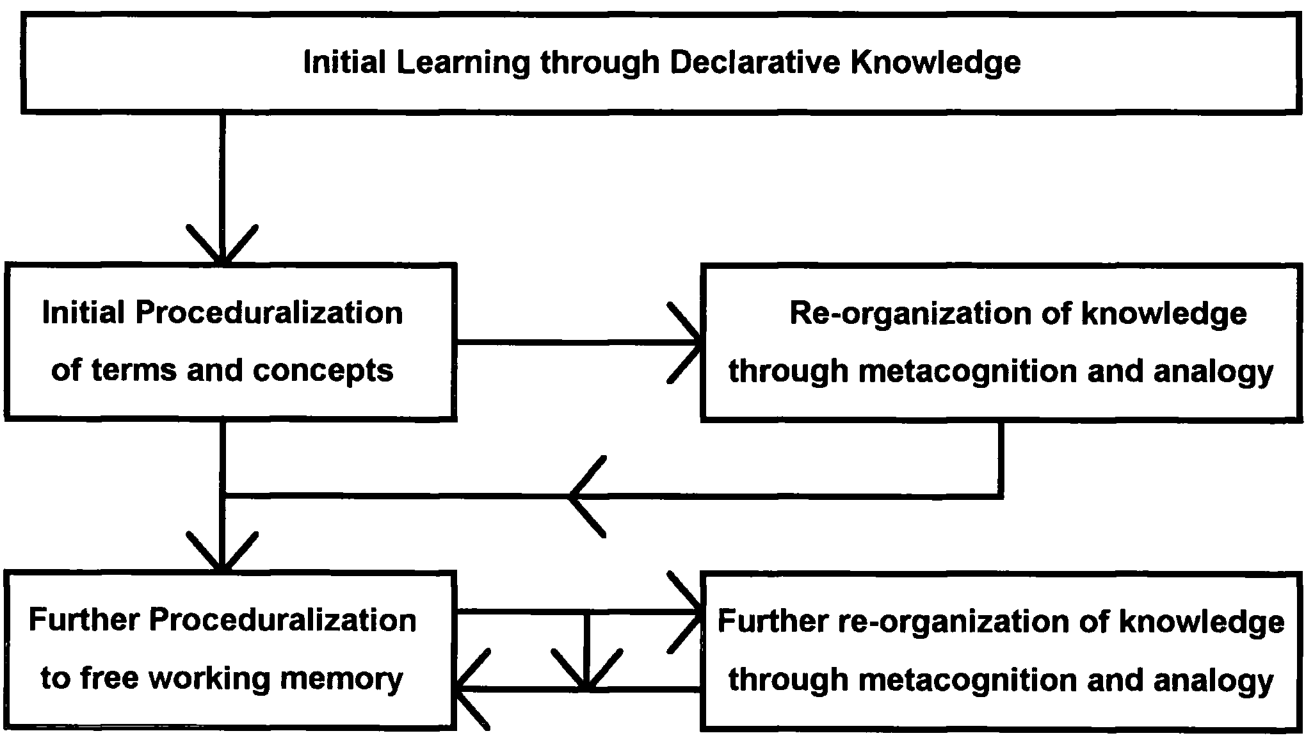
b) the acquisition of 1/ & 2/.

4/ The acquisition of procedural skills to allow for the free space required in working memory by 1/ & 2/.

Thus we may see from above that Adaptive expertise necessarily is reliant upon routine expertise. Indeed, by looking at Figure 1.2 we can see how the expertise and its acquisition can be represented as a cyclical process. Obviously in a puritanical view of adaptive expertise we would only look at the right hand side of the diagram whilst a puritanical view of routine expertise looks only at the left hand side. However it is conceivable that a routine expertise could exist independently of metacognitive reorganisation of knowledge and of analogical transfer (as indicated by the arrows by-passing the right hand side of the diagram), whilst adaptive expertise must necessarily be reliant to some extent upon the facets of routine expertise. I think that most problem solving lies in the shades of grey between both camps, but that lack of metacognitive ability can lead some people to be (almost) purely routine experts. Here I will end my review of the current state of expertise in the literature.

Figure 1.2 - An Overview of Expertise

Routine methods:	**vs**	**Adaptive Methods**
Proceduralization	vs	Conceptual Understanding
Anderson's	vs	Holyoak & Thagard's
ACT Theory		ACME/ARCS
		+ Metacognitive Skills



Chapter 2
Knowledge and its Elicitation

It is apparent from the previous chapter that there are different types of problems and different ways in which we solve them - logical problems which are arguably solved by the application of rules or procedures, and insight problems which are arguably solved by the retrieval of an analogy. In both cases, the answer to a particular type of problem is retrieved from memory, and in each case that answer is of the appropriate form to solve the problem posed. What we have not yet touched on is the way in which these answers are stored in memory - the way in which we represent knowledge. It is useful to understand this for three purposes: firstly, as it acts as a further explanation to the psychology of problem solving, and also of expertise; secondly as this will also aid us in our discussion of the way in which Intelligent Tutoring Systems may store knowledge of experts and students & thirdly because it paves the way for a discussion on how we may elicit knowledge from subjects, a point which is central both to the system and method which is finally adopted in this research.

Representing Knowledge

When considering the representation of knowledge we have to look at the work done and the theories put forward in the light of the different types of knowledge that people have studied. We have already identified that when people are trying to solve problems that they employ knowledge that is either in the form of a production rule or an analogy, and these will be discussed together shortly. Knowledge in the form of production rules may be thought of as 'knowing how', as they are used when want to know how to do something. You may know how to tie your shoe-lace for instance or drive a car, or indeed you may know how to solve a complex problem. This kind of knowledge is most often refereed to as *procedural knowledge*, as it embodies the procedures which we may follow to achieve a task. There is however another kind of knowledge, which may be thought of as 'knowing that', as it embodies your knowledge of the world in an almost 'raw' sense, and also our knowledge of concepts. You know that there is stuff called grass that you walk on, and you know that it is green. You also know that your favourite aunt is called Ethel, and that Elvis died in 1978. This 'knowing that' knowledge is most usually referred to as *declarative knowledge*, and it does indeed declare our knowledge of the world. It is essential to understand declarative knowledge and the ways in which it has been postulated we represented it. This is because declarative knowledge is a necessary precursor to procedural knowledge as will also shortly be explained. Declarative knowledge probably also underlies analogical problem solving. What appears to be retrieved are the concepts of the source analogy. Firstly then we shall look at declarative knowledge.

Declarative Knowledge

The first theory on how we represent knowledge has roots in classical and modern philosophy - that knowledge is represented as logical arguments. The way in which these Logic Based Systems are described in modern times is in the predicate calculus notation which was devised by the philosopher Frege in 1879. In this style of notation, a *predicate* is the word which acts upon other words, and these others are the *arguments* which the predicate takes. Hence the phrase 'John owns the book' would be given in predicate calculus notation as OWNS(john,book).

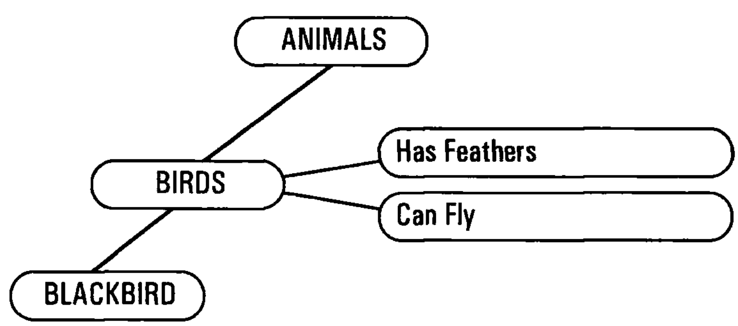
As our knowledge about the world is supposedly logically defined in these systems, we can exercise the rules of logic upon such knowledge. Hence we can represent certain parts of our knowledge not only by the structure of the predicate calculus and its contents but also by the application of logical operators upon the notation (IF, AND, NOT, OR). For example, we could logically deduce from IF OWNS(john,book) AND GIVES(john,michael,book) that it is now the case that OWNS(michael,book). We can also make statements which apply not only to individuals but also groups or classes of individuals by using a variable in our logic. For example, to represent the statement 'All grass is green' we can use the following notation:

For all x, if x is grass then x is green.

Such logic based systems ,thus can be thought of as representing all knowledge as a set of propositions (such as OWNS(john,book)) to which the rules of logic may be applied to generate new sentences and new knowledge which was previously not made known to the person. Unfortunately there are draw-backs to this style of representation of knowledge, of which the most obvious discrepancy to human behaviour is the matter of retrieval of knowledge. The way in which a logic based system can order such a set of propositions would be no more than an arbitrary list, perhaps ordered by the chronology of our acquisition, but nevertheless still a serial list of propositions. To then retrieve any information from such a list would require the exhaustive searching down the list until the appropriate knowledge was come across, which implies that we cannot easily find appropriate knowledge and that some items of knowledge (those at the top of the list) would be more quickly recalled than others (those at the end of the list) without any other factor making a difference to these relative speeds. However, this is not what we find.

The real observed case is that we can very easily access areas of our knowledge, and that related concepts are retrieved very rapidly. To try and account for such behaviour the theory of semantic networks (e.g. Quillian, 1968) was proposed. In such theories knowledge of the world is organised in a tree-like structure (in a similar way to taxonomical classifications). In this structure the joining points of the branches, the nodes, are symbolic of concepts. The branches themselves are symbolic of the relationship between the nodes. A typical example of such a semantic network is to consider the taxonomy of the blackbirds as in Figure 2.1.

Figure 2.1 - A Semantic Network for the Blackbird.



Here the 'blackbird' node is linked to the 'bird' node to represent its inclusion in the class of birds, which in turn is linked to nodes for 'has feathers' and 'can fly' to indicate these attributes of birds, and also to the node of 'animals' to indicate birds' inclusion in this class. The organisation of the network is such that a superset is always above any particular node and a subset below, whilst attributes are at the same level. In this way our knowledge of the world is rapidly accessible because of the hierarchical structure of the semantic network. Networks can also represent knowledge of individuals too, by having nodes for them. Thus we could represent the fact that we have a particular garden visitor who is a blackbird and who we call Morris by having a Morris node linked to the blackbird node in our example.

Semantic networks can represent knowledge of the world then too, not just the underlying concepts.

To be an improvement on the logic based systems semantic networks would have to incorporate the successes of those systems in addition to providing a better explanation than those systems at their identified weak points. Logic based systems are capable of performing the logical deductions and inferences on their knowledge base, producing new knowledge which had not been previously made explicit but was implicit in the knowledge. Such identification of new knowledge from the knowledge base can also be performed within a semantic network. For instance, from our example we have that a Bird has the attributes of 'has feathers' and 'can fly', and also that the relationship exists that a blackbird is a part of the class of birds. Given the fact that we have arranged our nodes in the superset/subset fashion and that a subset always inherits the attributes of the superset we can infer that a blackbird has feathers and can fly too. To stop us having a penguin in this arrangement who can fly, we would have to give our penguin a node as a a subset of Bird but with its own attribute that it cannot fly, and also make the proviso that in terms of inheriting attributes, those closest to the node have precedence over those inherited from further away. With these given, a semantic network is as capable as a logic based system to make appropriate deductive inferences, and superior because of its speed of recalling conceptually related information.

Even though such a representation is superior to logic based systems in its closeness of approximation to humanity there are still problems with semantic networks. One criticism of semantic networks, for example, is that of exemplars. It has been shown (Rosch, 1973) that some members of a class are viewed as more acceptable than others. For instance, a robin is viewed as a better example of birds than a penguin, and good examples are recalled more quickly than bad examples. To try and account for such evidence a revised version of semantic networks was put forwards (Collins & Loftus, 1975). In this view all nodes in the network are always at some base level of activity or above. When the activity level of any particular node reaches a threshold level there is then a *spreading* of *activity* through the network to connected nodes, and that part of the network is accessible to conscious attention. To then account for exemplars, they only need to have a lower than average threshold whilst a poor example has a higher than average threshold, or the connections between the exemplar and its superclass node could allow more activity to spread more quickly down it thus allowing for quicker activation than a poor example whilst both have an identical threshold.

It is still the case though that semantic networks are not truly adequate as there are enduring problems even after the revision to its operation with the addition of spreading activity. These are problems such as flexibility of retrieval and of knowledge of the real world. The answers given to such problems were to group semantic networks together as *Frames* or *Scripts*. A Frame is a re-orientation of semantic networks with the purpose of characterising knowledge about concepts, whilst a Script does a similar thing for situations. Consider a dog: given a semantic network we could understand it was a subset of animals and a superset of many breeds (such as Collie, Spaniel, Doberman ...) and that it has certain attributes such as four legs, a tail and so on. What a semantic network does not tell us is that dogs are infamous for attacking postmen, are usually found in their owners houses or at the park playing, and that their owners will often feed them, take them for walks and have them vaccinated at the vets. To try and encapsulate such a wealth of knowledge in one 'chunk' (a group of separate elements which can be manipulated as a whole) the notion of Frames was devised by Minsky (1977).

Figure 2.2 - A Frame for a Dog

DOG

Slot:		Value:
superset:		animal
number of legs:	default:	four
subset:	optional values:	collie, spaniel ...
situations:	optional values:	park, house ...
owner's actions:	optional values:	walk, feed, vaccinate ...
own actions:	optional values:	chase the postman ...

Adapted from Stevenson (1993) by the author's permission.

In a partial representation for a dog then, such as Figure 2.2, we can see that a frame has certain attributes which are defined by the slots for information and fillers for those slots. The concept is then defined by the configuration of these attributes and the values which they have taken. A superset is a slot in the frame for a dog, and it always takes the value of animal - this is because the taxonomic relationship will never be other than as it is now. The number of legs is also a slot, but this only takes a default value - we assume the default to be true, but accept that we could be told otherwise. This accounts for the fact that in most situations a dog will have four legs, but some particular individuals may have lost one leg through a misfortune. In such a case the particular subset of dog, an individual's frame, could specify that the slot for number of legs takes a value which would then over-ride the default. In such inheritances then only an unfilled slot would inherit the filler from a superset frame, working just as a semantic network does. Slots can thus be filled by a reference to another frame, or by a default value. Another type of filler is an optional filler, such as for owner's actions. Here there are many typical fillers for a slot, any of which could be appropriate at a given time. Frames thus organise the information inherent in a semantic network in order to reflect the way in which such information is known and used in the real world.

The similar theory used for situations is that of Scripts (Schank & Abelson, 1977). Where Minsky's Frame told us about a stereotypical dog, a script is a description of a stereotypical situation or event. A script will contain information such as the name of the script, the props and roles in the script, the entry and exit conditions for the script and a set of scenes which could take place in the script. To clarify this, consider the example of a restaurant. In the Restaurant script the props would be Food, tables, Cutlery, Crockery, Menus etc. and the roles would be Customer, Waiter/Waitress, Cook etc. The entry conditions would include the Customer being hungry and having money, and the exit conditions would be the Customer having less money and not being hungry. Typical scenes would be Entering the Restaurant, Ordering, Eating and Paying the Bill, and a scene would consist of many actions, e.g. for Ordering the actions may include: Customer asks for a menu, Waiter brings menu, Customer orders from menu, Waiter gives order to Cook, Cook prepares the food.

Such a representation again re-orients the knowledge it contains to conform to real world expectations of knowledge, such as knowing how to order food is related going to a restaurant because you are hungry. The computer programme which was written to test this theory by the authors seems to confirm its ability to act in a human way. When presented with information about a particular restaurant visit, the programme could infer from the information 'John was served quickly' & 'John left a large tip' that John left the large tip because he was served quickly. Likewise it could infer from 'The waiter gave John a menu' that John was given the menu to enable him to order. The system is thus capable of making the kinds of deductive inferences that the other representational systems were capable of, but of doing so in a real world setting.

One problem with the notion of such scripts was their heavy burden on human memory - to encapsulate paying a restaurant bill in a restaurant script and then paying a hotel bill in a hotel script and so on is clearly wasteful of memory. To account for this, Schank proposed MOPS - Memory Organisation Packets. These take items common to many scripts (such as the paying of bills) and form one single form of the knowledge which is then linked to any scripts which utilises it. Such an action not only makes scripts more flexible but also allows for learning. For instance, individually learnt actions could be generalised after a time in a similar way.

Unfortunately the problem with Scripts and Frames is that they are too big (even with MOPS) as is shown by further recourse to human behaviour. We do not act like a Script or Frame in as much as we do not usually recall the entire information which either would predict. We may infer a few additional things about situation or concept, but not the whole set of related information. Also they are still not flexible enough. If there was a fire in a restaurant, we would most certainly know what it was and what to do. However there is certainly no fire scenario in the restaurant script, and to make all such possible links to each individual script again takes us into problems on memory usage through the repetition of information, even if it is only stored as links to MOPS. A further problem which exists is that of retrieval of a Frame or Script. These must be organised similarly to semantic networks in that each Frame or Script is called much as a Node is - by a name. After which the associated information is retrieved. Where it otherwise (such as all parts of a Script of Frame allowing for recalling of the whole) we come back to our practical problems with memory space and also might wonder as to the point of organising knowledge into Frames and Scripts if not to allow for a common recall under one idea. However, the point is that in many cases the keyword - the script title - will not of itself be 'activated' - will not be mentioned in a sentence. For an example, consider the phrase - "The five hour journey from London to New York". It is most apparent from this phrase that the journey must have been by aeroplane, but there is certainly no information in the phrase by which we could retrieve the information about aeroplane travel by which to make the deduction that the plane must have been the method of transport.

So where does this leave the representation of declarative knowledge? We still have no definite solution to the way in which declarative knowledge may be represented. What we have done is to briefly look at the major theories which have been proposed in order, highlighting how problems with each have led to the devising of a new way to consider the representation of declarative knowledge. We have drawn attention to the ways in which it is possible that declarative knowledge may be represented, and that in itself is sufficient for us to move on to look at procedural knowledge, and the ways in which it could be represented. We will come back to the path set out by this work later though, and consider where the representation of declarative knowledge goes from here - hybrid systems. Firstly though we must look at procedural knowledge and at connectionist systems of representation.

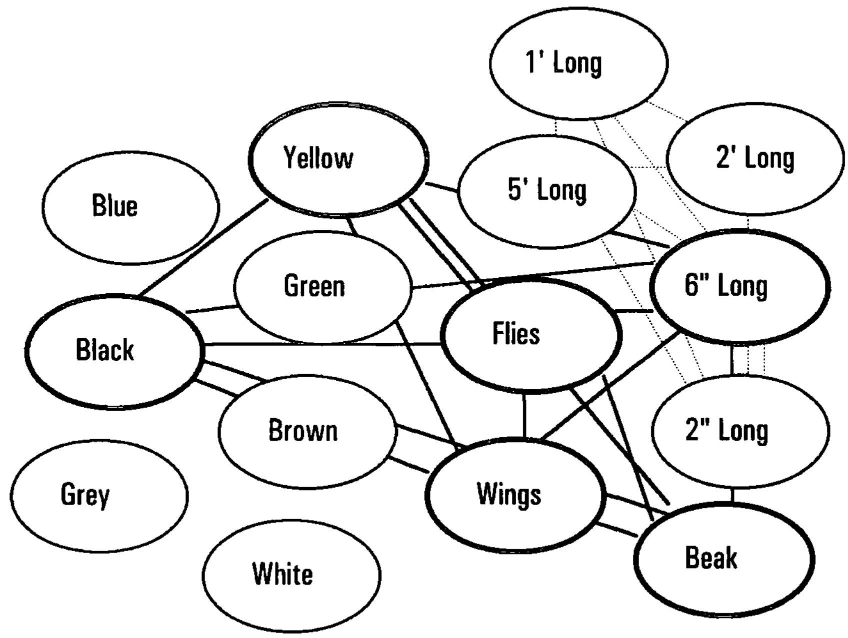
Procedural Knowledge

Procedural knowledge as I have already mentioned is the knowledge which explains how we do things. The main ways in which we might solve a problem or otherwise engage in using procedural knowledge is through the application of production rules. Production rules (Anderson, 1983) are units of knowledge which tell us how to achieve a goal. They are stored in our long term memory and consist of a condition, possibly preconditions and an action. An example of production rules and a discussion about them has already been given in earlier in this chapter when we discussed Anderson's theory of expertise. However it is worth adding here that this theory as with all of the others so far discussed this chapter is both symbolic and serial. To re-iterate this means that the thing that represents our knowledge of the world in these theories is an explicit representation of that thing. Thus in a semantic network a Blackbird is represented by a Blackbird node, and in a Frame a Dog is represented by a frame called Dog. In addition, these symbolic theories all assume that the processing of information occurs consciously in our short term or working memory, and is thus constrained by its limitations on size. Therefore, all the processing of information has to take place in a serial or linear fashion, with related information having to be processed one after the other. These two points make the serial symbolic theories common in several flaws. Firstly, these systems logically require that given the correct information a correct answer will be achieved - this is hardly the hallmark of human behaviour! Secondly, such systems also would imply that given the same conditions, it would be difficult for them to recover from an error - a point which has already been shown as false in this chapter. Thirdly, the idea that all processing of information is conscious has also been shown as unsatisfactory in the previous chapter (c.f. insight problems). Finally, it is also the case that these systems are relatively bad at coming to conclusions in cases of incomplete evidence - however, we ourselves are remarkably good at inducing a conclusion from impoverished stimuli.

Parallel Distributed Processing

To try and answer these criticisms of serial and symbolic knowledge representation, a new direction was sought, and in 1986, McClelland, Rumelhart (and the PDP research group) unveiled a solution - Parallel Distributed Processing (PDP). Instead of a serial processing of separate items of knowledge, all items would be processed in parallel, and instead of a symbolic representation of the knowledge, a distributed representation would be used. This is quite a conceptual leap from the serial and symbolic, so to try and explain more clearly consider Figure 2.2 and the case of a blackbird. In our previous symbolic representations, a blackbird would be represented as a node in a symbolic network called blackbird, or as a blackbird frame using Minsky's theory. However, it is readily apparent that in a distributed representation there is actually no 'unit' called blackbird at all.

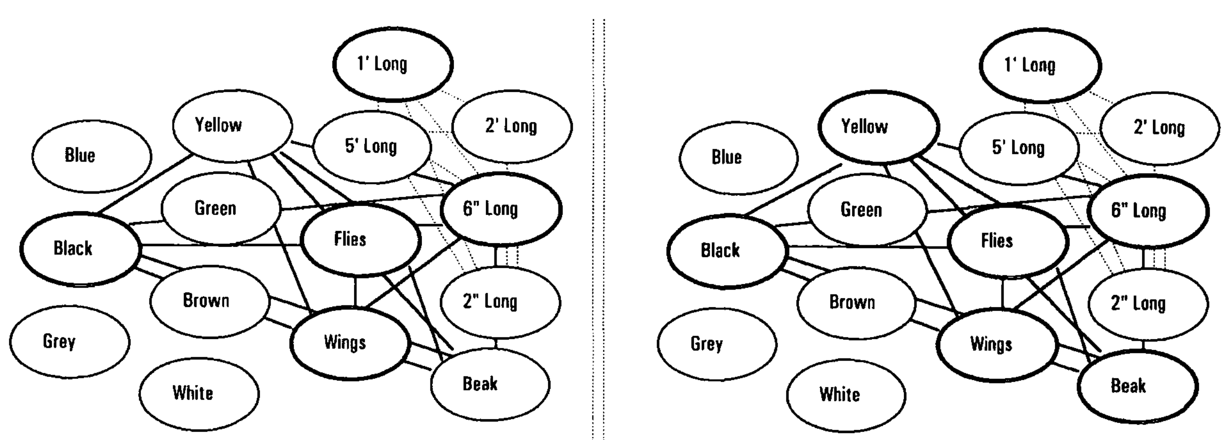
Figure 2.3 - Possible PDP Network Showing Activity Pattern for a Blackbird.



Instead there is a series of 'elements' which might be perceptual cues to many different birds (in this partial network). All of the elements would normally be connected, although some of the connections would be mutually exclusive, such as the length of the bird - these links are represented in the Figure with dotted lines. The lines connecting elements associated with a blackbird (dark circles) have been shown as solid lines, and the rest left off for clarity. Given the perceptual stimulus of a blackbird for the first time, the elements which make up the blackbird would become 'activated'. By then being told that this stimulus is a blackbird, the 'pattern of activity' can then represent the stimulus - the blackbird. In this way the representation of the blackbird is 'distributed' over several basic elements which together constitute a blackbird, without the need for a blackbird at all. The activity from any one activated node may also spread to other connected nodes, so in continuing exposures to a blackbird stimulus there will be a continual flow of activity down the links from each of the requisite elements. This flow of activity will be greatest through those links which connect two elements that are both parts of a blackbird. Using a neurological argument it is assumed that such usage will increase the size or strength of the link between these elements, so in a training period of exposures the activity pattern for a blackbird can be 'learnt'. This effect can only be effectively considered though because of its parallel nature. We can look at the activity spreading from all active nodes simultaneously.

To continue looking at this example, the parallel and distributed nature of these representations is most strikingly demonstrated in the case of an incomplete perceptual stimulus. Consider a fleeting glance at a bird as it flies overhead. Maybe you could see that it was black, but couldn't make out any other colours (such as the yellow beak), and you think its about six to twelve inches long, but couldn't say exactly. This is where the ability of a parallel and distributed system to actually process the information is most powerful. Looking at two new versions of the original network in Figure 2.4, we can see that initially (on the left) this would give us a different set of starting conditions from our perceptual stimulus. However, the activity would flow most readily down the pre-existing channels for a blackbird (lines shown), also activating the yellow and beak nodes (on the right). The two activated length nodes are mutually inhibitory, so as the six inch node receives additional activation from the other nodes it will inhibit the 1 foot node, and eventually the network will reach a settled state where the blackbird nodes are activated and recognition can occur.

Figure 2.4 - Identifying a Blackbird from Incomplete Information



This tends to confirm human behaviour in the identification of a most likely candidate from partial information. Consider also that the less the information the longer it would take for the network to 'infer' the identity of the mystery bird, as more spreading of activity would have to take place before threshold levels were reached (if at all). This could be related to uncertainty in an inference and taking a while to make a decision. As I have described it, parallel distributed processing as a model for representing knowledge explains several observable phenomena of Human memory. Firstly, the ease of retrieval of data. As I have explained, the name of any particular node is not needed to retrieve its information, as it only exists as a pattern of activity distributed over several elements. Indeed, it only requires that some of these elements be initially activated to recognise the stimulus. Thus rapid retrieval of knowledge is possible even in those situations where symbolic systems are unable to cope, and indeed beyond those situations into those even further deprived of information. This may also be expressed in terms of how the system works if damaged. If any part of a symbolic representation were 'damaged' then nothing would be known about that damaged part of the system. In a parallel distributed system, because the knowledge is not local to any part of it damage to an area would have no more effect than a lack of stimulus information did - it may take longer to achieve a decision, but there may not be an effect on the overall functioning. Only with massive destruction of the distributed system would we see a severe performance decline. Even so, with increasing damage to the system there is a 'graceful degradation' of performance. Instead of all or nothing (in symbolic systems), we have an incremental decrease in performance. This can be compared to people with brain damage, where minor cases of brain injury are not system threatening, but with increased damage performance becomes worse by stages (such as the slow decrease in patient ability from continued brain haemorrhaging in Alzheimer's disease).

PDP models can also simulate the best aspects of symbolic systems as a natural part of their operations. The idea of default assignment, so critical to inheritance in semantic networks and also in frames is a part of the spread of activation in parallel distributed systems. In the earlier blackbird example, even when we had the fleeting glance of the bird because we expect a bird to usually have a beak the activity flowed through its usual links and activated the beak (a default assignment). The only way to stop this happening would be in the presence of an over-riding stimulus, such as a blackbird in clear view with a facial injury. In such a case an element for 'no beak' would be activated, and this would by nature have an inhibitory link to the 'beak' node. Inspite of the usual predisposition for the system to activate beak, the node would receive constant inhibition from the 'no beak' node, and since this is a permanent stimulus with the bird in view it would necessarily over-ride the default for the blackbird representation. As in all other ways this activation pattern resembled a blackbird, and more closely resembled its activity pattern than any other, the bird would still be assumed to be a blackbird - just a particular special instance of a blackbird without a beak. The system thus also handles memory recall even in a situation where the stimulus directly contradicts a usual condition of the knowledge.

To finish off our examination of what a PDP model does well, consider the case of prototypes and exemplars. These again are fairly critical parts of the Frames/Scripts systems, and also of the revised semantic networks representation. Is it possible though for a distributed system to encapsulate this abstract idea? In symbolic systems prototypes such as the idea of a blackbird would be stored as a unit of knowledge, and particular instances of that knowledge, such as individual birds with slight marking changes or injuries etc. would be stored as other individual and discrete units. In McClelland & Rumelhart's work they gave a PDP system a definition of 50 instances of a conceptual type (individual cats) each of which were slightly different in their spread of activation across eight elementary units. Each activation pattern was associated to a unit for the cat's name. Then a partial stimulus was given to the system which was of general cat traits but not any individual cat already given, and the system generated a prototypical cat without activating any particular individual and without prior exposure to a prototype cat pattern. Prototypes are thus an emergent function of PDP systems.

Unfortunately, there is a very big (and quite obvious) drawback to any PDP representation - a lack of relational data. As the model takes elements of any knowledge and represents the knowledge itself as an activity pattern over these elements, there is no easy way to express the relationships between this data. For instance, consider the phrase 'John loves Mary'. To represent this as a PDP network we would necessarily have three interconnected nodes, but the problem is how to distinguish this from the phrase 'Mary loves John'. The only difference between the two phrases is the relational information contained within the linear structure of the phrase, but this information is lost if the phrase is decomposed to a set of elements. The only way to get around this is to use propositions in the representation which are by nature symbolic representations, not elements. It thus seems inevitable that as both systems offer answers to the problems which the other representation cannot accommodate that a hybrid system of knowledge representation is going to be most likely to have all the answers to Human knowledge representation.

Hybrid Systems

In the Symbolic Connectionism of Holyoak & Thagard (1989, 1990) as described in the previous chapter, the use of symbolic representations is combined with the parallel constraint satisfaction which is a part of the parallel nature of PDP systems. The other notable attempt at a hybrid model is the work of Kintsch (1988) in creating a model of Human discourse. In both these systems a set of logical symbolic rules generate a large number of 'contenders' for the solution to the initially described problem. Then a network is created with the contenders and a series of constraints (syntactic, semantic and pragmatic) are solved in parallel over the entire network, with the most highly activated 'contender' chosen as the solution (and bought into conscious thought at that time). By using symbolic representations in the system the problems of relational information in PDP systems is overcome. By using parallel constraint satisfaction the problems of purely symbolic systems of incomplete knowledge, ease of retrieval, unconscious problem solving and memory limitations are also solved. With further research a unified theory of knowledge representation appears possible. This then ends my brief look at the representation of knowledge.

Eliciting Knowledge

There are four common ways in which the knowledge of a person is elicited in psychology. These are through inference from experiments in controlled situations, by interviews or self reports, through observation or protocol analysis and by repertory grid or scaling methods. Each of these methodologies has its area of application and many associated pros and cons. It is my intention to briefly talk about the four different methodologies and then to comment at the end as to the choice of elicitation method chosen for use in this study.

Experimental Inferencing

This is perhaps the most common method of eliciting knowledge used by psychologists. Subjects are asked to perform tasks under experimentally controlled conditions. These are invariably created with a particular theory in mind, and are constructed in order to prove (or disprove) it. If the subjects behave as predicted by the theory which designed the experiment then we may infer that the subjects probably acted as they did in accordance with that theory. The use of this methodology for inferring problem solving strategies of expertise is not particularly appropriate. The main problem is that the knowledge which we are usually seeking to define in cases of expertise is large and complex. In any kind of experimentally controlled situation it is only through the successful isolation of one (or a very small number) of factors which allows us to investigate the variable(s) under consideration. At the best then we would be forced into preforming many many experiments in order to elicit the requisit amount of knowledge.

A second, and perhaps more telling problem is that as already stated this type of knowledge elicitation is completely theory driven. It is only by knowing in advance what you want the results to be that you can then set about building the experiment to isolate those elements and investigate their role in the situation. If you are seeking to elicit knowledge for which you do not have a good theory or model already created, this kind of method will again prove very inappropriate. A final restriction with this type of method is that because it relies on control over the situation, it is virtually a necessity that it be performed under laboratory conditions. This will prove to be a major obstacle with certain domains where the knowledge would not be accesible under lab conditions, either because of its complexity or the length of time involved in performing the task.

Interviews and Self-Report.

One of the most common (and perhaps obvious) methods of finding out how someone thinks when solving a problem is to ask them. This technique has been very common in the field of expert systems. An acknowledged expert in a field will be asked how they attempt to tackle a task, and then their answers are used as the raw information by which the same task could (and should) be tackled in the future. This has been critiscised as being overly simplistic, and a further refinement of the method is not to believe *carte-blanche* that the reported rules and knowledge of the expert is actually how they did accomplish the task. In this latter case the expert's responses are treated as a source of information from which the rules and knowledge may be derived or induced.

The problems with the methodology are mainly based on the naivety of the assumption in the first case - that the reported solutions from the expert are in fact those which they used. This is shown to be unlikely with evidence both from the information processors and also their critics. According to Anderson's Theory of knowledge representation (ACT), declarative knowledge becomes proceduralised as a part of the upwards progression to expertise, becoming unconscious. It is thus necessarily the case that expert knowledge (by the ACT theory) is unavailable to conscious recollection. Therefore any reported knowledge which the expert thinks they used in their solution to a task is unlikely to be a valid reflection. What is more, the work of Berry and Broadbent (1984) demonstrated that learning is not necessarily bound to pass through a declarative step. They showed how a complex rule-based game could be mastered through the acquisition of skilled (expert) knowledge without there being any declarative knowledge for them to proceduralise in the first place - only the feedback of their performance in the game. Here too though, we see again that whatever knowledge the experts may be applying, it is most unlikely that it is available to conscious recollection.

A further (and possibly fatal) flaw in the use of Self-Reports as a method of knowledge elicitation is their susceptibility to 'post hoc explanations'. Having sucessfully completed a task without any conscious attention to the process of its performance it is very possible to then attribute the performance to some unrelated phenomena. To give a facile example, consider the sub-conscious task of catching a ball. This is not a process which we learn through verbal instruction but actively through practice. However, if asked to self report on the process a subject could endeavour to describe their own actions and theorise about them, creating a misleading picture of the task's solution. Obviously such a facile example denigrates the use of Self-Reports. They can indeed be very useful at providing knowledge about a domain, and in some cases the task performed may well be open to later report. However, the direction of this kind of knowledge elicitation has moved on now to the more robust technique of protocol analysis

Protocol Analysis & Observation.

The solution to 'post hoc explanations' in Self-Reporting is to take away the opportunity to make them. This is commonly achieved now through the taking of Protocols. These are records of a subjects efforts to solve a task *at the time* of the attempt. Instead of Self-Reporting on the solution to a problem after the fact in taking a protocol the subject would be asked to verbalise their thought processes concurrently with the task itself. The form of a protocol need not necessarily be verbal, but commonly this has been the form which they have taken.

It is of course vital that the taking of Verbal Protocols does not fall foul of the problems associated with Self-Report at all. Thus protocols are not in themselves treated as knowledge of a task itself. Instead they are treated as the raw data from which knowledge of a task may be extracted. By treating the protocol as the end product of a cognitive process in the same way that the task performance itself is an end product of a cognitive process, it may be possible to work backwards to the underlying knowledge. This then brings us to the obvious argument as to what methods are applicable or appropriate to the interpretation of the protocol data to infer knowledge. The subjects own interpretation could be used, but that would fall foul of the problems inherent originally in Self-Reporting. Instead we must rely upon our own interpretations, and these will undoubtedly be coloured by the theories to which we adhere.

A common way in which protocol analysis is actually utilised is to define a problem in advance in the light of a theory, and to then use that theory to make predictions to the possible methods of problem solving which the subject could apply. These can then be used to create a scoring form which can be used in analysing the protocols. In Newell & Simon (1972) they were investigating information processing in problem solving and defined a scoring method in advance to test the theory. For instance, as they were concerned with the creation and use of goals and sub-goals by subjects they decided to looked for statements of the form of "I am trying to do ..." or "I am trying to achieve ..." as indicating the current goal state to which the subject was working.

Verbal protocols of course suffer from the problem that not all tasks are verbally expressible or have little or no verbal component to their thought process. As I have already said though, protocols need not necessarily be bound to verbal information. The style of protocol analysis may also be applied to purely observational studies. Taking a verbal protocol is merely asking a subject to 'think aloud' as they carry out a task. This is akin to inferring cognitive processes from experimentally manipulating tasks, but as the task is in no way subject to experimental variation itself, it provides a more realistic indicator of real task performance. We may thus make our own protocol of a subject's behaviour and task performance without restricting ourselves to the confines of the laboratory.

Repertory grids & Scaling Methods

The repertory grid is a method devised by Kelly (1955) originally to elicit the social beliefs and constraints of people. Subjects were initially asked to generate a list of people, and subsequently to name dimensions upon which the subjects felt that some of the people in their list showed similarities or differences. Having created a set of dimensions the subjects are then asked to rate the people on the list for each of the dimensions. Analysis of the data through multivariate statistics can then be used to find clusterings of people in these dimensions or of dimensions themselves. This technique has been taken beyond its social creation into purely cognitive knowledge elicitation by replacing people as elements in the data with concepts.

A related application is to use multi-dimensional scaling methods to visually illustrate clusterings of elements in the multi-dimensional space, and to allow for visual identification of the dimensions. This work is akin to reporatory grid workings, but is more interpretive. Subjects are given a set of concepts as a list of all possible pairwise combinations and asked to rate them on a scale for similarity. These similarity measures are assumed to be representative of the psychological association between the concepts. The ratings can then be used to plot the concepts in a multi-dimensional space and then scaled to be illustrated in two (or more) dimensions. The plot can then be used to uncover latent strucures in the data - clusters of concepts indicating commonalities; dimensions being visible in diametrically opposing concepts.

The weakness with reporatory grids and scaling methods is that they are generally very poor at eliciting procedural knowledge. Understanding the way in which a subject views the relationship between the kettle and a tea-bag is not going to give you instruction on how to make a cuppa! By contrast however, a procedural knowledge elicitation technique is unlikely to furnish you with any deep and conceptual insight into the knowledge domain - just how to solve a particular problem.

To date, Intelligent Tutoring Systems have in the vast majority of cases used procedural knowledge for their representations of expert knowledge, and as I will argue later I believe that this is a great hindrance to their ability to work well. To this end, the wethods of this research will follow a non-procedural approach, utilising a Hybrid Knowledge representation based upon the work of Holyoak & Thagard. It becomes obvious that the knowledge must be elicited in the appropriate format and so some form of repertory grid or multi-dimensional scaling technique will be used. It will later become apparent that a multi-dimensional scaling approach most closely fits the need of the knowledge representational system I chose, and so I will elicit knowledge of the domain using pairwise concept similarity ratings.

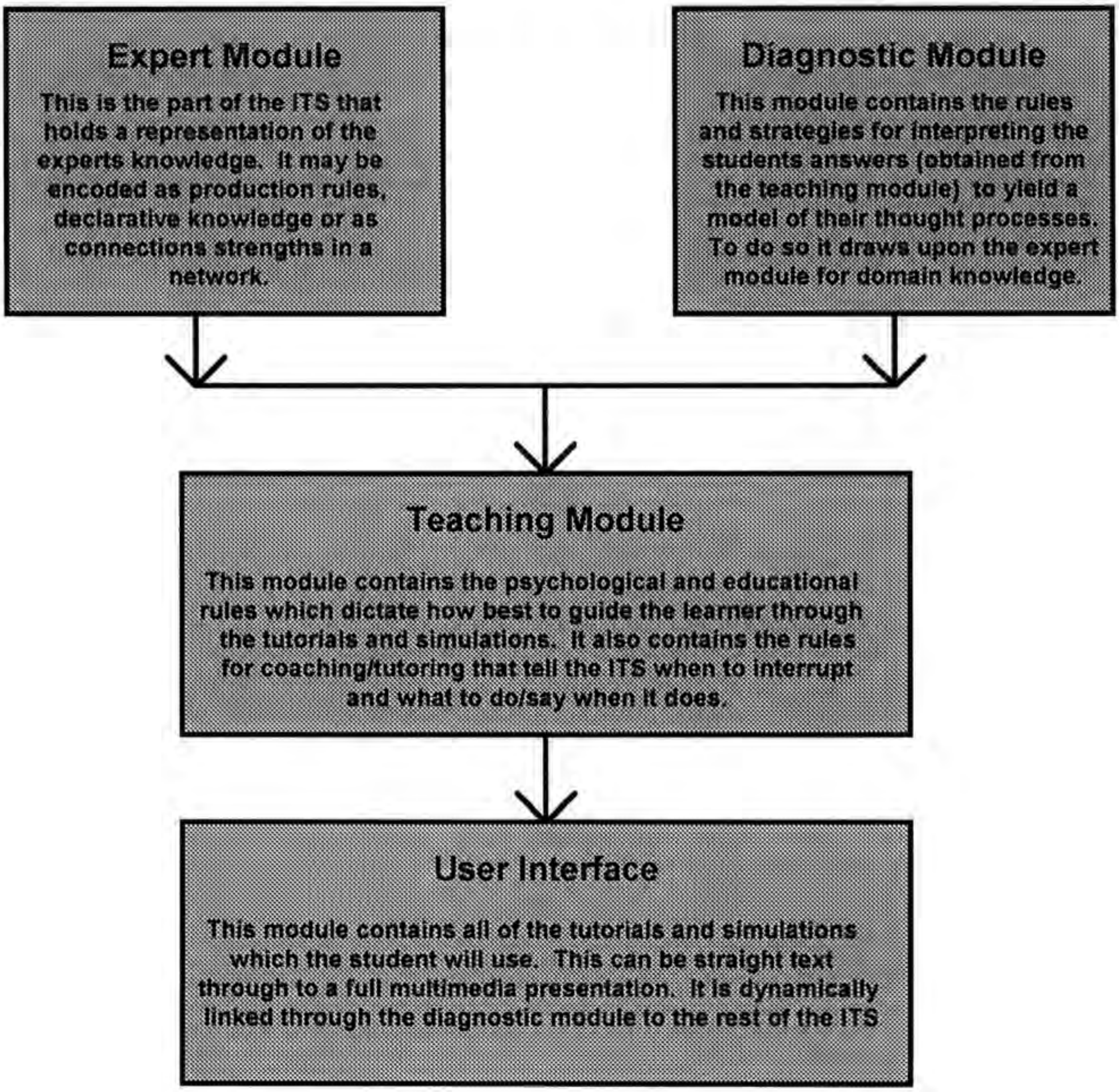
Chapter 3
Intelligent Tutoring

Intelligent Tutoring Systems

An Intelligent Tutoring System (ITS) is the upper most pinnacle of Computer Aided Instruction / Computer Aided Learning (CAI or CAL). It embodies a computer mediated session of educational benefit to the participant. This goes beyond the mere use of a computer for presentation of an expert's knowledge or for the monitoring of a student's competence at pre-prepared questions. Instead it should be able to harness the information inherent in the interaction between the user and the computer (the learner and the teacher) to be able to *Model* the student's knowledge. By extracting such information from the user the ITS can then function like an *Intelligent* Tutor, to adapt and react to the *individual* needs of the user. To do this though, the ITS has to be able to draw on the research carried out in AI, Cognitive Science and Education in addition to having the relevant domain expertise. To successfully combine all of the requisite abilities to create the ITS it has become customary to design the ITS in a modular fashion, with each part of the system performing different (though linked) functions.

Most ITSs comprise of four distinct parts or sections: the *expert module*, the *diagnostic module*, the *teacher module* and the *user interface*. These four parts interact together to provide the user or learner with the best possible environment within which to learn (as shown in Figure 3.1). Most ITSs are defined only in terms of the expert representation that they utilise, because usually it is only this first criterion that is addressed in the design. Thus we may commonly refer to WEST (Burton & Brown, 1982) as a 'Black-Box Tutor' which it truly is, but will more often neglect to mention that WEST is also an 'Intermediate States - Flat Procedural'

Figure 3.1 - A Schematic Representation of a Typical ITS



Tutor (if we refer to its diagnostic methods) which would be an equally valid and no less incomplete description of it. However, most ITSs are creations which are designed to test cognitive models, and so it is hardly surprising that authors invariably classify them by this criterion. This vast majority have only cursory teacher modules (e.g. the BUGGY Tutors, Burton, 1982) and user interfaces (e.g. PROUST, Johnson & Soloway, 1984), and frequently the diagnostic module is mis-used as merely an extension of the expert module (e.g. WUSOR, Goldstein, 1982). I shall endeavour to approach these four aspects of ITSs individually & to substantiate my criticisms at that later point. I will then concentrate on some recent developments and applications by way of a conclusion.

The Expert Module

What is supposed to put the ITS above other forms of CAI/CAL is the incorporation of an intelligence in the tutoring, especially through the *expert module*. This part of the ITS must embody the knowledge of the experts in the particular field or *domain*, as well as their skills and problem solving abilities. The way in which this information is represented is central to the way in which the whole ITS will operate, as this dictates how the modules communicate with each other, and is what really makes the ITS function intelligently. Historically the *expert module* originated from the work into expert systems done in Artificial Intelligence (AI). Since those beginnings it has moved on to adopt Cognitive Models to help it function more effectively, and so it has developed in parallel with the changing theories of Cognitive Psychology (see Chapter Two). I shall now go through these developments in more detail, starting with the first developed Intelligent Tutoring Systems.

The first ITSs employed what are called *Black Box expert modules*, because their design only allowed for an output to be associated to an input without any possibility of understanding why this association had been made. Such expert modules in actuality would usually be little more than a large database of possible steps which were worked through in an exhaustive search (e.g. WEST, Burton & Brown, 1982), or used completely non-human algorithms to deduce an answer (e.g. SOPHIE, Brown & Burton, 1975). In WEST all possible moves are calculated (moves are made by combining three numbers in any legitimate arithmetical combination) and then the consequences of the moves are computed. The move giving rise to the best completion of an overall criterion (such as creating the furthest distance between the computer and the player) is then chosen. Similarly SOPHIE, an ITS developed to analyse electronic circuits, worked by using mathematical equations to find the faults.

These expert modules clearly operated very differently to the way in which a human would attempt to solve the problem, which although not strictly a problem for the *expert module* was a difficulty for the *teaching module*. Tutoring needs to be able to help the learner, and *Black Box Experts* can only correct a wrong answer with the right answer, which will not help the learner understand *why* they were wrong.

The next stage in the development of the expert module then was *Glass Box Expert Modules*, so called for the ability of the teaching module to look within the *box* at how the expert module reached its answer. These ITSs drew upon the original work in the field of *expert systems* in AI, where much effort had been put in to the design of computerised representation of an expert's knowledge. This form of expert knowledge was gained through *knowledge engineering*, whereby the answers to questions were collected from the appropriate expert sources and collated into a form for use in the computers programming. This gave rise to programs which could accurately and precisely recall the exact answers to specific questions. This was a logical answer to the problems of the *Black Box Expert*, as it meant that the *expert module* inherently contained human knowledge.

It was *expert systems* such as MYCIN (Shortliffe, 1976) which were used as the expert modules for ITSs, which in this case was GUIDON (Clancey, 1982). Since they relied upon the knowledge of a human, there was data available to the teaching module from the expert module which could effectively be used. Indeed to use them would also be cost effective as the expert module already existed and so could literally be used 'off the shelf'. Unfortunately, although the knowledge was human in origin, for speed and efficiency it was accessed by the system through backwards search or exhaustive search algorithms. Thus the actual thought processes used became non-human, and the explanation which they could yield turned out to be of limited additional help to that which a black box expert could yield.

This led to the insight then that you cannot instruct, correct, or advise competently unless you can understand how the expert is *thinking*, and this became a major turning point in the development of ITSs. It became obvious then that the *teaching module* had to have access to a representation of the expert which truly reflects the Human thought processes. It was this need that led to the adoption of theories of cognitive psychology for building *expert modules*. These *Cognitive Models* fall into three classes, each claiming greater cognitive fidelity than the others. This mirrors the way in which cognitive theories have developed, competed and advanced (perhaps) in the last twenty years. The first set of expert modules from this cognitive tradition are those employing *Procedural Knowledge*.

Procedural Knowledge models employ the problem solving strategies epitomised by Newell & Simon (1972) and Anderson (1982) with knowledge stored as *production rules*. An incorrect solution can then be traced along a production until the point where they diverge, and this marks the point required for tutoring (as it is where the student went wrong). This is a somewhat simplified picture since the identification of the student's reasoning process, and hence the procedure that they used, is not always straight forward - this though is a problem for the diagnostic module, and is discussed under that heading. The underlying assumption however of the validity of procedural knowledge is the same, regardless of diagnosis. In some instances this is a useful way of thinking and teaching, but it is a very narrow avenue to solely pursue in tutoring. It can only allow for one correct solution to a problem, and only one (or a small set) of allowable ways of coming to that answer. Many systems have been developed along this methodology, indeed all of the systems so far mentioned utilise some form of IF ... THEN rules to operate. The best exemplars of this type of ITS though are probably Anderson's own systems (for obvious reasons) - The Geometry Tutor (1985) and the LISP Tutor (1986).

In the LISP Tutor the student is learning the programming language LISP, and the Tutor is designed to make sure that the student will successfully create a functioning programme. The computer screen is split in half, with a tutor 'window' and a code 'window'. These two text windows are the User Interface. The tutor will give the student a task to complete (selected in a sequential fashion from problems in a LISP programming book, written by Anderson Corbett and Reiser) and will then allow them to start coding. The tutor has two main interventions - firstly it will offer syntactic guidance, both by spotting spelling mistakes and also by providing 'templates'. For instance, if the student was to define a function they would have to write a piece of code with the keyword <u>defun</u> followed by three requisite arguments. Thus as soon as they type 'defun' in the code window the Tutor obligingly gives them the defun 'template' telling them what three arguments are required. The other tutor intervention is to interrupt whenever a mistake is made and instruct the student as to the nature of the mistake. It will then wait for a correction and will not allow continuation with the set problem until the mistake has been rectified (although further help on the mistake may be given).

The teaching module thus contains some very simple procedures for deciding the nature and timing of tutoring. Advice is given instantly and progress is inhibited until the mistake is corrected. Advice is however graduated such that if a general or conceptual explanation does not provide an appropriate correction then more specific help is given, finally ending with the code itself being provided by the tutor. This more extreme intervention is reached more rapidly if the tutor recognises that the error has been repeated frequently or if the tutor cannot diagnose the error. At a much more coarse level, after each pair of lessons a quiz is administered. Failure at the quiz leads to a re-try, and further failure forces the student through the lessons again. Two quiz failures after this will then see them sent to the human tutor.

In order to make the judgements that the learner sees, the LISP tutor has to have some expert representation of these programming problems and a way to diagnose the students mistakes to allow for appropriate messages at the point of intervention. The expert module contains production rules which conform to Anderson's ACT theory (as described in chapter 2) for LISP programming, of the form:

IF	the goal is to add some numbers together
THEN	code a call to + and set goals for the arguments

Then if the student had been asked to generate a code where they had to add together some numbers the expert model tells us that they would have to start with the code '+'. If this was the action of the student then the next production would be consulted for which the IF statement was a match to the successfully carried out production - to code the arguments. If however the THEN clause was not followed by the student - the coding of '+' then the tutor will intervene. As the nature of this comparison is immediate and step by step, the assumption is that there are no stages in the students thought which are missing from the information given to the Tutor. The entire need for diagnosing at the immediate level is redundant - all the tutor has to do is see if the student action mimics the production, and if not then the student is wrong.

There are however some broader criteria for diagnosis which is described by Anderson et al as Model Tracing. That is, whether or not the path which the student's coding is going down will lead to a successful fulfilment of the problem set by the tutor. This is because there may be several ways to successfully complete any given task - for instance, to add a list of numbers you could add them all together first to last, or last to first, or add them in pairs even. To avoid 'hard-coding' all possible permutations which would appear to be a less than elegant demonstration of the ACT theory the general LISP productions are used in conjunction with a problem specification. Instead of just following the correct solution path it can attempt to model the student's steps using their code so far as refinements to the problem specification. This is made possible by taking all the legitimate LISP productions plus a set of erroneous productions (bugs) which have been experimentally identified. All possible combinations for these rules, given the initial problem and the current code entered, are calculated, yielding several possible codings for the next step. When the student then writes some more code it is compared to the calculated codings. If any of them match, then it may be assumed that the student has used the productions involved in the matched tutor's solution in their own method. The tutor can then examine those productions to determine whether the student is or is not on the correct path to the solution. If the productions used were all legitimate then the student is still on track, whereas if they included known misconceptions the student is incorrect and requires tutoring. If no codings were matched then path is considered unrecognisable. It is the comparison of the student's actions to those predicted by the model which gives the process its name of Model Tracing.

Such a methodology may well be appropriate in the case of skill based problem solving, such as learning to drive a car, where there is only one (or a few) sequential ways in which to approach the problem. The same methodology though may be singularly lacking at more general tasks where a wide body of knowledge is required and where a sequential or linear approach is not appropriate. Thus to teach general principles and conceptual understanding requires a different approach, and thus the use of *Declarative Knowledge* for the expert module is the next cognitive process to consider.

Declarative Knowledge expert modules owe much to the cognitive theories of such as Schank & Abelson (1977) and Minsky (1975). General declarative knowledge is organised into *scripts* or *frames* and the knowledge of the expert exists both in the nature of the links within and between them. It also resides in a separate set of rules of an If - Then variety to employ the general knowledge. The first use of such a methodology was in SCHOLAR (Carbonell, 1970). This ITS utilised interlinked "frames" to contain the knowledge of South American geography, and had separate rules for conducting the tutorial session. A more recent example of this style of expert model is the rainfall tutor of Stevens, Collins & Goldin (1982). Within the expert model a large body of knowledge on rainfall is stored in the form of scripts such as the one below:

EVAPORATION:	Actors	Source:	Large-body of water
		Destination:	Air-Mass
	Factors	Temperature(Source) Temperature(Destination) Proximity(Source,Destination)	
	Functional-Relationship	Positive(Temperature(Source)) Positive(Temperature(Destination)) Positive(Proximity(Source,Destination)	
	Result	Increase(Humidity(Destination))	

In this case, the evaporation schema, there are 'slots' for the actors in evaporation, the factors that influence evaporation, the functional relationship between these factors and the result of evaporation. Similar schemas also exist for all the other aspects of the water cycle, such as condensation, cooling and rain. In the course of a tutorial interaction, these 'slots' could be filled in with the students knowledge if the tutor can give the student appropriate questions to isolate the individual elements of understanding. The 'slots' to be interrogated and indeed the schemas are not necessarily set in a pre-defined order. Instead by using a Socratic dialogue the student's answers can both provide their knowledge to the system and dictate the course of the questioning. Appropriate knowledge is stored as correctly filled 'slots' whilst mal-rules are inappropriately filled 'slots'. However, in order to extract all the necessary knowledge IF ... THEN rules are needed to organise the Socratic dialogue, such as IF the answer is correct for a particular case THEN ask a question for a prior cause. As is typical of this area though, the expert module is not a part of a functional ITS in the same way that Anderson's LISP Tutor is. It is a completely separate entity with proposed methods by which it *could* work as an ITS - there is no real implementation however.

The final step (to date) has been in the use of *Qualitative Processes* in the expert module. Broadly speaking these models examine the mental behaviour underlying problem solving of dynamic systems, and the work of DeKleer and Brown (1984) is perhaps most noteworthy of them. Their work on a conceptual representation of a pressure regulator proceeds as follows. For them, a process of envisionment takes place which comprises of the construction and subsequent simulation of a causal model, which may be understood locally at any interaction along the causal chain. To them each local concept is a *confluence* of the local relationships and constraints, and the overall concept of the regulator's functioning is the set of all confluences.

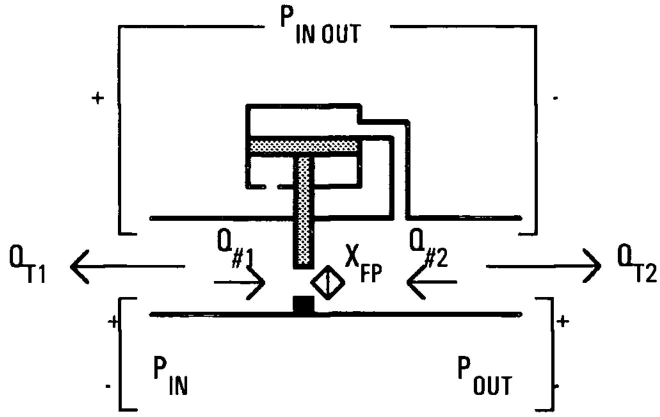
The expert representation of knowledge in this system is similar both to schemas (which encode declarative knowledge) and productions (which encode procedural knowledge). In the modelling of a pressure regulator (Figure 3.1) each part of the device is described as a set of relationships which relate to the effects that surrounding parts of the device have upon it. By way of example, consider the 'confluence' of relationships at the pressure regulators valve:

$\delta P_{in,\ out} - \delta Q_{\#1(v\ v)} + \delta X_{FP} = 0$, where $\delta P_{in,\ out}$ is the change in pressure,

$\delta Q_{\#1(v\ v)}$ is the change in flow &

δX_{FP} is the position of the valve control.

Figure 3.1 - A Representation of a Pressure Regulator



The actual knowledge which is held in this representation is a combination, not only of the individual confluences of relationships, but also of the causal links which we may make when trying to understand the system. To understand the pressure regulator locally, we must understand that IF the output pressure rises, THEN this will in turn decrease the valve opening. However, IF the valve opening is decreased THEN the output pressure will fall. Thus the confluence of these conditions is the equation given which will maintain an (dynamic) equilibrium. On a wider scale however, any individual confluence is part of a larger system. In an example of an electronic circuit (as much of DeKleer & Browns' work comes from their SPICE Tutor for solving electrical circuit problems) a similar dynamic equilibrium is investigated (Figure 3.2). Here individual confluences for the effects of a battery, a clapper and a coil are considered:

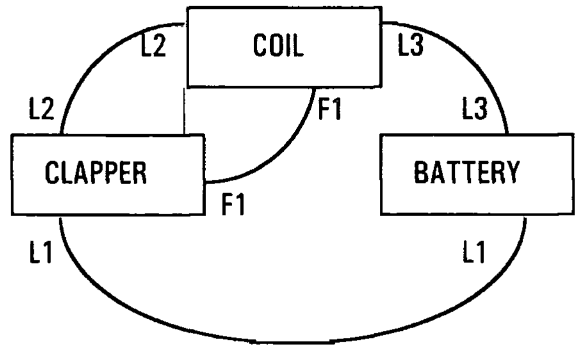
Battery: L1 <=> L3

Clapper: Open: L1 <=0, L2 <=0. IF F1=0 THEN Clapper becomes Closed.

Closed: L1 <=1, L2 <=1. IF F1=1 THEN Clapper becomes Open.

Coil: On: F1<=1. IF L2=1, THEN Coil becomes OFF.
IF L2=0, THEN Coil becomes OFF.
Off: F1<=0. IF L2=1, THEN Coil becomes ON.
IF L2=0, THEN Coil becomes ON.

Figure 3.2 - Attribute Topology of the Buzzer Circuit



The battery is simple, as its relationship is a permanent ON for current flow in the circuit. For the Clapper, if it is open, then there is no current between L1 & L2; if there is no force F1 then this will make the clapper close. If the Clapper is closed, then there is current between L1 & L2; if there is a force F1 the clapper will become opened. For the Coil, if it is On then there will exist a force F1 (an electro-magnetic attraction upon the clapper arm); If there is no current in L2 then the coil will become Off, and if there is no current in L3 then the coil will become off. If the coil is Off then there will not be a force F1; If there is current in L2 then the coil will become On, and if there is current in L3 then the coil will become off. This still only defines the individual elements though, just as the pressure regulator only defined itself & not the whole system of which it is a part. To see how a system functions, all of the confluences must be considered together in a casual model.

In our circuit example we must first assume certain initial conditions, such as the coil is off (0), there is no force (F1=0) and the clapper is closed (0). If we then start at the battery, then the current is always flowing (I1=1, I3=1). Progressing to the clapper, we find that I1=1 implies I3=1, and since there is no force the clapper is closed. Moving on to the coil, we find that it will become on (I2=1, I3=1) and that therefore this will cause the force F1 to be. If we carry on our investigation of the causal links, we find that this will now cause the clapper to become open which in turn causes the circuit to break (I1=0, I2=0). This again in turn will cause the coil to become off, and then the force F1 to disappear, and this then in turn will cause the clapper to become closed. All of which brings us back to the initial conditions with which we started, and so the whole process can cyclically repeat itself. This process thus, from the system viewpoint, causes the clapper to move continuously backwards and forwards, thus making a buzzing noise.

This work is perhaps ill-defined at the present due to the lack of research so far published. However, as an expert module it perhaps embodies both declarative and procedural knowledge, and it also has quite obvious parallels with connectionism. The individual confluences are most certainly rule based but are organised in a schema based fashion. More so, the actual system analysis can only occur when the individual elements are processed together, with a parallel exchange of causality between the individual confluences. For these reasons I include it, as I see the work as being a possible half-way-house to the symbolic connectionism expert module which is a goal of this research work. That then is a brief look at the Expert Module of an ITS. Next I shall discuss the Diagnostic Module.

The Diagnostic Module

The diagnostic module is the part of the ITS which compares the student's behaviour to the expert, and it tries to infer the student's knowledge. This results in some kind of representation of the student's knowledge called the *student model*, which stores the results of the diagnosis for use by the rest of the ITS. There are many complex ways to divide up the plethora of different student models which exist. Amongst these are the notions of the Bandwidth in diagnosis, the Cognitive Theory underlying the Knowledge Representation and whether the student model exists as an Overlay to the expert model (a difference model), or as an assembly of identified Bugs or Bug Parts. This information is then available (or should be) to the rest of the ITS, specifically through the Teaching Module, to allow for modifications to the interaction which takes place so as to continually improve the quality of the interaction with regards to the student's learning outcomes.

The information for diagnosis comes from the interaction of the learner with the ITS, be it through moves made in a game, answers to computer generated questions, commentaries on actions or commands issued to a (simulated) device. The challenge is to interpret the data garnered and thus create the student model. The classification suggested by VanLehn (1988) is worth examining because of its completeness, as it focuses on all of the different methods employed so far.

Student Models

For Student Models VanLehn breaks down their classification into three dimensions as I said earlier - Bandwidth, Representation and Difference Model. Each ITS has its Diagnostic Module classified by VanLehn by these three categories. To start with then, is the dimension of Bandwidth. This is a measure of the quality of input which the diagnostic module receives about the student. He identifies three Bandwidths - approximate mental states, intermediate states and final states. Final states are ITSs working with only the end products of cognition such as BUGGY (Burton, 1982). In Buggy an arithmetic problem is presented to the learner, and the only user input that Buggy receives is the final answer to the problem, with no idea as to the stages which the subject went through to come to their decision. This is low bandwidth because relatively little information is given as input, and means that to work out the processes which are going on in the subject's problem solving a lot of work and assumptions will have to be made by the system (and which will shortly be discussed in more detail under Diagnosis).

The Intermediate States category is used where some information is available to the Diagnostic module concerning the subject's steps in coming to their solution. This will not be all of the steps in their thought process, but obviously there is less work to do to extrapolate from this information to understand all of the steps that they made than is involved when working with Final State information. As there is more information elicited from the subject this is classified as Medium Bandwidth. An example of where such a Bandwidth is used in an ITS is in WEST (Burton & Brown, 1982). Here an arithmetic game is played by the subject, where the calculation of sums with random numbers is used to progress along a playing board. The player can use any valid arithmetic operator (+,-,/,*) and the higher the number made the further along the board the subject can move. There are also short cuts you can take, and moves which will send the opposition (the computer) backwards, so the highest possible number is not always the best move. Thus, the input of each turn is not in itself the final solution as the playing of the entire game represents the entire problem. Likewise the input is clearly not the entire thought process, as the calculation and choice of a number are not directly available to the ITS. Thus these 'mid game' moves are typical of the use of Intermediate States of information.

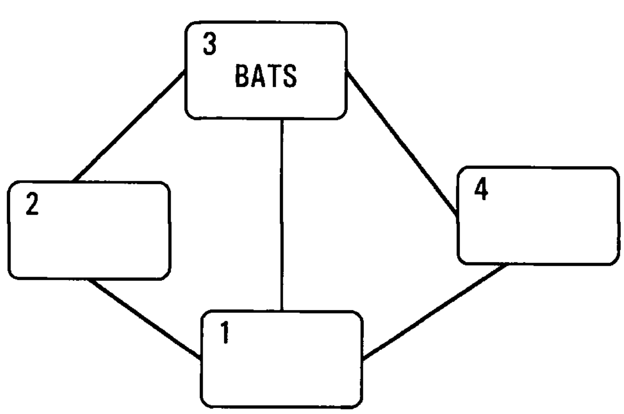
Approximate mental states are the highest bandwidth, and are where the user is questioned at every point along possible decision trees, such as in GUIDON (Clancey, 1982) or the LISP tutor (Anderson, 1984 - described earlier). Such interrogations are deemed to show virtually every mental state during cognition, and thus no further extrapolation will be required by the diagnostic module in order to understand the users thought processes. The whole idea of the Bandwidth Dimension though seems dis-important to me. It is granted a useful categorisation for descriptive purposes, but VanLehn seems to be making much more of it - he implies that High Bandwidth solutions are somehow more valid by virtue of their closer scrutiny of our thought process.

The problem is however that we cannot make this claim universally as VanLehn appears to - quality of input cannot be ascribed on this Bandwidth criterion. Some problems are solved in a linear, information processing style such as mathematical problems and I will not argue with that; but surely neither can we argue with the fact that some problems are not, but are solved in some analogical fashion such as insight problems (see Chapter 2). Thus it is ludicrous to say that having the Final State solution only (Low Bandwidth) will be less valid or less true than Approximate Mental States (High Bandwidth). We must study the two different processes of problem solving in the most appropriate way for each of them (i.e. examining intermediate states or final states for analogical problem solving). We must not condemn *a priori* a diagnosis to be less valid than another purely in terms of using a lower Bandwidth input. Were we to do so we would have to guarantee that the approximate mental states are indeed much better representations than intermediate or final states are, and clearly as the idea of approximate mental states in analogical problem solving is impossible, I feel that such a guarantee cannot be made. If we are therefore to keep a Bandwidth element, it will only be useful to us if applied less strongly. If we are aware of its descriptive meaning and know with what kind of input it is trying to deal (something which is lacking in VanLehn's classification) then it can be of use to us. In its present form though I would consider it dangerously misleading due to its confined outlook (linear problem solving only).

The second categorisation of VanLehn's is the idea of the Representational type of the ITS. This seems a bit superfluous as it merely re-iterates the expert module's knowledge base (although VanLehn does subdivide Procedural Knowledge into systems which do or do not employ the use of sub-goals). However, as a descriptive classification it is acceptable. There is little point in here repeating the different types of representational types, as they have been covered briefly already within this chapter and have been more fully explained earlier in this thesis (Chapter 2). However I may again note that the fact VanLehn does not cover non-linear systems reflects the fact that there are no fully functional ITSs which rely purely upon a connectionist approach or which apply themselves to non-linear problem solving. His omission in this respect is thus unsurprising.

The third categorisation which is offered to us is that of the Difference between Student and Expert. This is either as an Overlay Model, or as a Bug Library. An Overlay Model is one such as that used in WUSOR (Goldstein, 1982). In this ITS a game of 'Hunt the Wumpus' is played. In this game the player enters a network of caves. In each cave there are four tunnels leading to other caves, the whole system thus being interconnected. Within each cave there may exist a hazard; a pit, giant bats or the Wumpus. If you walk into a pit you die, and if you walk into the Wumpus you are lunch! Giant bats are very unpredictable taxi cabs - they swoop down on you and pick you up, dropping you at a random location in the cavern system (possibly on a pit or the Wumpus). Fortunately you are warned of the existence of these dangers in neighbouring pits by 'feeling a draught' (Pit in an adjacent cave), 'hearing a squeak' (Bats in an adjacent cave) or 'smelling a Wumpus' (Wumpus in an adjacent cave). To kill the Wumpus and win you have to shoot an arrow into his cave from an adjacent cave, and you have a very limited supply of arrows.

Figure 3.3 - An Example of a Possible Cave Structure in the Hunt the Wumpus Game.



Success in WUSOR depends upon mastering certain skills. For instance (see Figure 3.3), if you were in cave 1 and 'heard a squeak', and then moved to cave 2 and 'heard a squeak' you can logically deduce that there are bats in cave 3. There are other more difficult rules about dangers, such as double evidence of an unconfirmed danger make a cave more risky than a single evidence of danger. It is assumed that anyone's skill level is based upon how many of these skills they have mastered, and the fact that they are sequentially learnt. Thus you have to know that single evidence of danger (i.e. 'I smell a Wumpus') means the adjacent caves are risky before you can learn the 'double evidence of danger' rule. To then model the student expertise is a simple matter. The system identifies what strategies they have adopted from amongst those previously worked out as required for expert play by the ITS's creators (i.e. did they choose to enter a risky cave with only one evidence of danger rather than two - if so they have learnt the double evidence rule). Then the student's level of expertise is simply a measure of which rules they are using from that master set.

This type of system of modelling expert / novice differences is called an Overlay Model because it treats the learner as an imperfect facsimile of the expert, and merely lacking in some of the expert's knowledge and skills. Thus we may represent the student's knowledge by *overlaying* a card with holes cut in it on to a representation of the expert's knowledge, allowing us to 'see' only selected parts of the expert's knowledge. The student's knowledge is therefore a pure subset of the experts knowledge. Bug Libraries on the other hand although still treating the learner as an imperfect approximation to the expert, do not take the view that the learner brings no prior knowledge with them to the learning situation. Instead it assumes that they will bring with them both some correct rules which are the same as the expert's knowledge but also with some mal-rules or 'Bugs' which are different from the experts'.

In the Buggy Tutors (Burton, 1982) then, the student's tasks are simple arithmetic problems, and the expert module contains a list of rules (productions) for successfully performing simple arithmetic such as addition, with separate rules for additions of elements with and without a carry, and for what to do with the carry if there is one. There is also a set of pre-identified Bugs which are rules that the students may use but are incorrect ways to solve the arithmetic problem. One example of such a Bug would be to forget the carry when adding numbers together that totalled more than 9. In this type of tutor, the diagnosis of the student's learnt rules and learnt Bugs (or mal-rules) would be performed in a similar way to that already described for Anderson's LISP Tutor - all possible rules (both correct and mal-rules) for a problem are used to generate all possible answers. The student's answer is then compared with those the system has calculated, and then if a match is found the student can be assumed to have used those rules which the system did to arrive at the same answer. The student can then be represented as a collection of identified Bugs and correctly learned rules. These Bugs and rules are either assembled from Bug Libraries or Bug Part Libraries (the difference between which I will explain in diagnosis). However it too means that the student is viewed as an imperfect simulacrum of the expert; a view which I will challenge most vehemently in discussing my model later on.

Diagnosis

The other central part to the diagnostic module other than the Student Model itself is the diagnostic techniques with which the model is assembled. Many different techniques have been tried, and VanLehn charts 9 specific types in great detail. I will however limit myself to discussing the trends within these techniques, although the interested reader can examine the classifications he offers more fully if they desire. The techniques of Model Tracing, Path Finding, Condition Induction and Expert Systems all employ a similar method to diagnose procedural knowledge, and so I will explain their general nature in the light of Model Tracing, as this was described earlier in this chapter (Anderson's LISP Tutor). Basically, if the student demonstrates knowledge sufficient to move from position A to position B, such as coming to the correct piece of coding in the creation of a LISP programme, then we may conclude that they are aware of the rule or chain of rules which account for knowing how to get from position A to position B (Model Tracing/Path Finding) The difference between path finding and model tracing is that Path Finding uses lower Bandwidth information than Model Tracing, and therefore it is a chain of rules as opposed to a single rule that is identified. An ITS which uses the Path Finding method is PIXIE (Sleeman, 1982; Sleeman et al, 1991). In PIXIE students have to solve algebraic equations and enter simplifications as they progress towards an answer (Intermediate States Bandwidth). However, as this does not necessarily contain all the steps they carried out in their solution, only a chain of rules can be assessed, not necessarily individual rules. Alternatively we can attribute the bug or bugs consistent with having erroneously gone from A to B to the student, such as they wrote an incorrect piece of code as earlier discussed. This could of course be the case with either Model Tracing or Path Finding.

Expert Systems have more complex sets of rules than other ITSs and so can give rise to responses from students which show possible partial knowledge of a rule. GUIDON (Clancey, 1982) is an ITS which utilises a medical Expert System called MYCIN. In it the student follows a prompted set of questions to carry out a medical diagnosis, but there are many medical conditions where a certain set of symptoms could be indicators. For instance, the student may come to the correct answer from considering only certain of the symptoms, not all of them. Since the correct answer could thus have been arrived at by more than one different chain of steps then we can only reasonably assume that they are aware of one or other of the possible sets of rules to make the correct (or incorrect) diagnosis. In this case a percentage probability may therefore be attached to a rule (or several rules) as to the confidence the ITS has in the student understanding of that rule (or rules). After further trials it is possible that some of these rules may receive increases to their confidence from a different diagnosis whilst others do not, and that eventually after a certain threshold the system may be confident that the learner does know a particular rule.

The second trend I identify is in bringing intermediate state knowledge up to approximate mental states through inference procedures. This is the idea behind Plan Recognition and Issue Tracing. In Plan Recognition the jumps from A to B are too big to have only one possible route (i.e. Path Finding) and so the possible routes have to be constructed by various algorithms. If the steps are only small you could employ a Model Tracing approach on the results of the Path Finding, but if the chains are long then an alternative method will prove more expedient. In the CIRRUS system (VanLehn, 1987) the user input is parsed to form a discrete set of 'visible' actions (visible because it is only these parts of the problem solving process which the ITS can 'see' as input). These 'visible' actions are treated as the broadest level of a hierarchical tree structure, and they are linked together at branch nodes by 'invisible' subgoals, going all the way back along the tree's nodes until you get to just one branch at the top which represents the entire goal state. In plan recognition, it is the inferring of the treeful of 'invisible' subgoals (the assumed unseen steps that the subject utilised) which is the diagnostic task. Once the subgoals have been identified, possible paths may be recognised by an exhaustive search of the tree 'plan'. If there is a unique solution then this will represent the course which the subject took, and it will be a depth first, left to right traversal of the tree plan (assuming a left to right parsing of the subject's solution steps). If the plan recognition finds no solution then it assumes a fault in the subject's methods, so if this approach is used with a Bug Library system the Bugs can then be used in the invisible nodes of the tree plan and a Buggy solution be found instead. If there is more than one possible solution identified, various heuristics can be used to find the more probable route, or a repeated set of measurements on similar problems could be used to try and reduce the set of solutions to a unique route.

Issue Tracing is the essentially the same process as Model Tracing, but it is applied to a more coarse level of Bandwidth. Working with only intermediate knowledge states as in WEST the tutorial aims to 'coach' the user in the issues of arithmetic; it is not interested in the actual mechanics of the arithmetic as is the case with the BUGGY Tutors. In Issue Tracing instead of identifying fine grain actions such as the various attempts in the learner's head to reach the eventual answer, it recognises coarser grained issues, such as whether the learner used 'multiply' in its arithmetic or can use short-cuts in the game. The amount of time 'issues' are used in a game is counted, as is the amount of times an issue could have been used and wasn't. If an issue is used many times and missed relatively few then the student is believed to understand that issue. The reverse would indicate that the student doesn't understand the issue. The threshold for the ratio (for indicating a lack of understanding) is set high to try and avoid a mis-diagnosis. This is caused by the fact that one issue could have been 'missed' purely because another issue could have and was used with equal or greater success. When the threshold is reached however, the Coach will issue advice to the player on how they could improve their game.

The third trend is the set of techniques used with Bug Models. These all function similarly in that the answers to many pre-prepared questions are analysed to find the fewest number of Bugs consistent with all of the error types found. An example of the type of material with which they all work (arithmetic problems) is given for ease of understanding the processes used.

60	811	Bugs or mal-rules.
- 17	- 46	
50	845	1/ 0 - N = 0
57	835	2/ $N - M = \lvert N - M \rvert$
50	835	3/ Both 1 & 2.

Here we can see how two common arithmetic mal-rules would manifest in two different subtraction tasks. The first mal-rule is that zero minus any number is zero, which is obvious in its manifestation in the first subtraction, and also occurs in the second subtraction after the 'borrowing' of the ten for the units column subtraction. The second mal-rule is that in any subtraction the smaller number is always taken away from the larger number. When both these Bugs combine though they become more difficult to follow, as in the first subtraction they give the same answer as the first mal-rule in isolation as the first bug suppresses the action of the second (Bug #1 always taking precedent). However in the other subtraction task the action of the second mal-rule means that a borrow is never made and so the first mal-rule never has a chance to be active.

The first technique used in Buggy models is that of Decision Trees, as in BUGGY (Burton, 1982). Here all possible Bugs for each question are calculated in advance by brute force, and consulted when the answers are given. A tree of all of the possible routes through Bugs to get from Question to Answer can then be constructed for all the questions a candidate will attempt. For example, if the subtractions were given to a subject in the order above, the decision tree would start with the first subtraction and branch to the answers 50 and 57. If the answer given was 57 it would be linked to only Bug #2, and a Bug would be successfully identified. If the answer given was 50 though it would not discriminate between Bug #1 & Bug #2 combined or only Bug #1. The tree in this latter condition could then point to the second subtraction. Here the decision tree would have the answers of 845 and 835 as branches. If the answer was 845 this is linked to Bug #1 only, and so the Bug would be identified. If the answer was 835 this would be linked normally to Bug #1 & Bug #2 or only Bug #2. However we have already identified earlier in the tree that Bug #2 on its own is not the case, so this branch can instead point to the conclusion both Bug #1 & Bug #2. In the real ITS 55 Bugs were under investigation - a possible 3025 Bug pairs and an even higher number of higher order combinations. To allow for more efficient diagnosis a decision tree was drawn up to try and create a set of unique paths through as small an amount of questions as possible to diagnose all possible bugs and bug combinations. Assuming that no errors are made (as opposed to the application of Bugs) then from this Decision Tree a minimum number of Bugs consistent with a high percentage of the errors (about 90-95%) are selected. To work with all the possible Bugs is computationally expensive though, especially as Bugs can be compounded with other Bugs to produce new Bugs.

It is the explosion of possible combinations caused by combining more than two Bugs that prompted the move to using Bug Parts instead and dynamically assembling the parts into specific Bugs. Thus Generate and Test is used in DEBUGGY(1982) to select a small number of fundamental Bug Parts consistent with a majority of the errors in a student's performance. What this means is that initially only sole Bugs are sought out by the ITS in an initial trial period. Assuming that the learner possessed both of the Bugs discussed earlier, then either of the subtraction problems already given would indicate the possible use by the learner of either or both of these Bugs. Having thus successfully identified a small number of possible primary Bugs, a set of all Bug pairs could then be generated (which is much smaller than the set of all possible Bug pairs). This set is then used and subtraction tasks are used at this stage which will identify some Bugs and discard others. The then smaller still set of unconfirmed primary bugs would then be combined into a higher order combination and the process repeated incrementally until all the primary Bugs are identified or until there is no improvement in the match. This Generate and Test method is a very general technique, however, and is more effective if heuristics are used to home in on specific Bugs sooner.

The final technique developed is that of Interactive Diagnosis, where the Generate and Test Method is carried out dynamically <u>during</u> the student's session, and the possible Bugs identified are used to modify the interaction whilst the student is still working thorough problems. With BUGGY & DEBUGGY a set list of subtractions is given and worked with by students, and modifications and selections of new material to try and further discover Bugs or refine the opinion of possible Bugs to definites occurs 'off-line' - there is no instantaneous re-scheduling of materials. With IDEBUGGY though new questions are sought which will test the ITS's hypotheses about possible Bugs as would happen off-line with DEBUGGY to help it distinguish between similar Bug types. IDEBUGGY does this though as the student attempts more tests, actually re-defining the questions given as the student takes the test. This has been the most successful diagnosis method so far employed by Bug driven ITSs, but the process is still not rapid enough to ensure that the student isn't sometimes kept waiting for noticeable periods of time whilst the ITS searches for the next problem to present to the subject.

That concludes my look at the Diagnostic Module. Despite the ideals behind this Module, it is apparent that feedback is not always quick enough or immediate enough to be of use in instruction. In BUGGY & DEBUGGY there is just a collection of data, no modification of the user environment and no tutoring & IDEBUGGY is sometimes too slow in its modifications of the material. In WEST to avoid incorrectly assuming a person's lack of knowledge, they have to make the same omissions repeatedly before being coached. Since the program does not alter the environment to look for these issues, they could go un-noticed for a very long time. Other diagnosis methods may not necessarily tell us the whole story either. In Model Tracing and its variants any un-anticipated actions will cause the programme to assume that the learner has made an error without knowing that this is the case, and worse still an ITS like the LISP Tutor would specifically stop them from proceeding with their course of action until they changed they way back to the pre-planned routes. I especially draw attention to the way in which all of these models treat the learner's knowledge as a sub-set of the expert's.

Although the shortcomings of overlay models are many, it seems that the Buggy theories often side-step this criticism themselves, or are even offered as a solution. People have assumed that the problem with overlay modelling is that learners cannot bring prior knowledge with them which is either a mal-rule or a correct rule. Such prior knowledge is identifiable and can form an expert's library of rules. This means however that buggy models assume that there are only so many possible correct ways to solve a problem, and so many possible incorrect ways. The learner possess a finite number of these (pre-calculated) rules and the buggy rules can thus be identified and therefore corrected. The end result of such a process is that the student may be turned out as a perfect recreation of the expert. If a learner however brings with them correct knowledge or incorrect knowledge which is not identified in the library of Bug parts or the library of correct rules then the ITS will not be able to act appropriately. This is a major problem, and one to which I will frequently return. In a wider sense we can therefore say that buggy models too fall foul of overlay modelling. They are intent on turning the student into the expert, and thus although they allow for the student to begin with a different set of knowledge to the expert, they expect them to finish up with exactly the same knowledge - the ultimate overlay. This predisposition is common to overlay modelling as well as buggy modelling, and I will refer to it as the Overlay Paradigm in Student Modelling. It is a point to which I will return in somewhat greater detail in Chapter 4. For now, however, I shall move on to the third module in a typical ITS, the Teaching Module.

Teacher Module

As is said at the beginning of this section in the Schematic Representation of a Typical ITS (Figure 3.1) the Teaching Module should be responsible for three main activities. These are the selection and sequencing of materials presented, the coaching or tutoring of the user, and the modifying and updating of the computer's models. I shall thus tackle my appraisal of ITSs under these three criteria. The use of any teaching strategy has only slowly come to be adopted by the ITS field, with early projects such as SCHOLAR (Carbonell, 1970) dependant mainly on student choice to modify the interaction through help options, rather than through the intelligent adaptation of the tutoring by the machine. However more recent ITSs have started to address these issues somewhat better, with two main approaches being evident: *Tutoring* and *Coaching*.

An ITS such as WUSOR (Goldstein, 1982) or WEST (Burton & Brown, 1982) which operates at the level of issues or super-ordinate concepts are typical of the class of ITS which employs a Coaching Strategy. These ITSs have a set of heuristics based on good psychological and educational practice to try and keep the ITS session running smoothly and profitably for the user's learning outcomes. These heuristics are numerous and varied, but a few examples serve to illustrate the points:

A SESSION SHALL NOT BE INTERRUPTED UNTIL THE USER HAS HAD A CHANCE TO ACQUAINT THEMSELVES WITH THE ITS;

COACHING WILL NOT BE GIVEN UNLESS THE OUTCOME IS DRAMATICALLY BETTER THAN THE STUDENT'S MOVE.

THE LEARNER WILL NOT BE INTERRUPTED TWO MOVES CONSECUTIVELY;

COACHING WILL NOT BE GIVEN IF THE USER IS GOING TO LOSE WHATEVER;

These all serve to try and maintain interest from the student and to only coach when and where it will be most beneficial and best remembered. Then, when the coaching is finally to be given, it is subdivided into different levels of tutoring depending on the overall competence of the user. For example, in WUSOR a player of Hunt the Wumpus could be about to stumble into Giant Bats in a particular cave, for which they have already gained double evidence, whilst they could move into a cave for which they only have single evidence of Bats - a less risky option. The kind of coaching they may get at this stage would be a message such as "Multiple evidence for Giant Bats is more dangerous than single evidence for Giant Bats". However, if the player had already apparently learnt this rule in relation to pits, the intervention could be based upon this different level of competence - "We have already seen that multiple evidence for Pits is more dangerous than single evidence". This will hopefully maintain relevance and interest more effectively than starting from scratch again as the first message did and insulting the player's intelligence. The help information which is given is also accessible upon demand from the ITS. These Coaches thus make intelligent decisions upon when to make an interruption to coach, and also make a (semi) intelligent decision as to the form the coaching takes when it is administered.

In ITSs that Coach such as the gaming environments of WEST and WUSOR there is just one continuous session, so there cannot be any alteration of the materials that they present. Instead the emphasis is on the manipulation of the Tutor's parameters. In WUSOR you can change the rules of the game after one session and before the next (if the player is still the same) to make the game more challenging and so develop more difficult skills. For instance, the single evidence for a danger is quite an easy skill to master if you can only hear a danger from one cave away - if you extend that to two caves away the need to use multiple evidence and work such things out carefully becomes much more essential to success.

In ITSs that Tutor however, the emphasis is on the manipulation of the materials presented. In an ITS like Kimball's Calculus Tutor (1982) the teaching module consults the student model to discover their level of competence in the various calculus solving techniques, and then selects material to tutor with from the areas of greatest student weakness. Thus the ordering of the material is changed through the intelligence of the system. The system also offers on-line help and advice at request, with different levels of help from clues and hints through to completing the problem for the user. By reason of the nature of this interaction though, there is no need for the ITS to interrupt the student's progress, and so it does not need to be bogged down with the heuristics evident in the Coaching style of interaction. A similar adaptation also occurs during the interaction in IDEBUGGY (Burton, 1982). However, the BUGGY series acts more like a diagnostic aid such as the MYCIN expert system than as an intelligent Tutor. Although BUGGY and DEBUGGY collected data and formulated student models, it is only with IDEBUGGY that we see intelligent adaptation of the material, but only so as to better facilitate diagnosis - there is no on-line modification of *instruction* at all - indeed there is no tutoring of the student whatsoever. To thus say that these ITSs have a teaching module or tutoring strategy is generous in the extreme. By my generalised definition at the start of this chapter, they are at least one module short of an ITS.

According to Halff (1988) the teaching module should address three main issues - The nature of Learning, of Teaching and of the Domain. He espouses the view that although we understand that the user is not a "blank slate" we do not adjust our ITSs so as to weed out the old inappropriate knowledge whilst we are sowing the new. This we need to tackle, and it is a major problem for our understanding of learning. Similarly he sees fault in our teaching strategies, in that we are not always appropriate in using the expert model as a teaching representation. He thus suggests we adopt a Propaedeutic representation of the knowledge - an intermediate stage of representation which would allow the user to acquire learning skills initially instead of skilled performance in the domain, but that would allow for this to be learned later through practice. This suggestion comes from a production system point of view comparable to Anderson's PUPS but can equally well be applied to creating short term production style knowledge bases from which to then explore more conceptual knowledge: a possibility which I have already considered in my review of Expertise, especially noting the suggestion of Stevenson & Palmer (1994). Such a method is a worthy suggestion, however it may be implemented.

Halff also identifies three issues which a Teaching module should address in the selection of materials: firstly, that each session should be solvable and comprehensible to the learner in the light of their curriculum so far; secondly, that the instruction should be 'transparent' so that the exercises or structure of the tutor reflects the underlying knowledge, & thirdly, that the material is individualisable to the learner. The first of these criteria is perhaps that most frequently addressed in ITSs. In Anderson's LISP Tutor the teaching strategy is based on a few very simple heuristics. Within a lesson there are heuristics such as 'always interrupt when a student is wrong' and 'if the student is repeatedly incorrect on the same topic give them the solution'. At the broader level it tests after every other lesson, and if the test is failed progress is not allowed to the next lessons, but revision of the last lessons is forced instead. Thus Anderson fulfils the first of Halff's selection criteria, but does not succeed on the other two points.

Structural transparency can be related to a schema approach in ITS structure. In SCHOLAR, Carbonell's ITS for South American Geography, the information is internally organised into schemas for each country with slots filled for appropriate country information or pointing to other schemas. In Scholar this structure dictates the method of teaching, as the interrogative style of tutoring is based on the underlying schema representations. A typical session might start with the question 'Did you know that Argentina is a Country in South America?' (emphasising the Continent Country Super/Sub Ordinate relationship). SCHOLAR might then enquire if they wish to know more of Argentina, and then offer other information which relates to Argentina (such as Population, Location, Cities, Borders). When it then is time to move on to another situation it could ask if the student knows another country on the same latitude as Argentina, or whether it remembers which country borders Argentina. It also emphasises these relationships when testing the subject at the end of a session. Typical summary questions would be of the type 'Is Buenos Aries a City in Argentina'.

Halff's third point is that of Individuation. This can somewhat be seen in the re-selection of materials in IDEBUGGY, though as this is purely for the aid of diagnosis and not for the student's direct learning enhancement it is a poor example. If the re-sequencing was to make a point in the tutorial interaction to the learner, then it would be a very good example of individuation. In WUSOR there is a certain degree of individuation, in that players of different competences receive different playing advice and conditions, but the level of individuation is very coarse as the tutor only operates at an issue based level of instruction. Therefore the individuation will be common to many people, not user specific. A similar form of individuation appears to happen in the LISP Tutor, where there is graduated advice and only competent users can move on to the next stage of learning, but this is a phantom - the graduated advice is a constant gradation, not one based on the user's skills, and the requirement of competence to move on does not change the sequence of presentation beyond recovering old material until they are ready for the new - the next stage of novel learning will always be the same for all students. The only real individuation in the LISP Tutor comes from the individual Bugs which define an individual student's model. These are wholly individual representations for a particular learner and can be used to individuate the *diagnosis* stage. The teaching materials however are never re-ordered to account for the diagnosis. It merely indicates whether the learner's solution is correct or incorrect.

Halff also indicates that within the learning experience the knowledge should be sequenced for relatedness and generality. That is, novel concepts should be tackled in an order of priority where concepts closely related to existing knowledge are taught before more distantly related topics. Then, when new concepts are tackled, they should be taught as a generalised concept before moving on to specific instantiations of that concept. Going back to SCHOLAR we can see that there is some ordering for relatedness. The ITS doesn't just teach South American Geography at random or a set of unrelated countries. It moves to other related areas to teach new information. Having learnt of Argentina and its border with Brazil progressing to learn of Brazil. Also the distance of relationships are tackled from closer relations first - thus going from South America to Argentina and then Argentina to Buenos Aires, not straight from Argentina to Buenos Aires.

As can be seen, there are many desirable elements in the teaching strategy of an ITS which should be addressed by the Teaching Module, and that to date many ITSs have only addressed some or other of these points. To successfully utilise all these strategies would require the blending of two or more of the ITSs described, with each from a different school of representing and interpreting knowledge. As I have described, current models do not necessarily address the topic of teaching at all (BUGGY Tutors), let alone fully. Indeed drawing upon the Psychology of Learning and the Acquisition of Expertise, it is easy to see that some hybrid representation (employing a combination of both approaches) may indeed be necessary to allow for all teaching strategies to be instantiated in the same ITS. The point must be though, that whatever is appropriate to the situation must be used to try and ensure three things - The Relevance of the Instruction, the Memorability of the Instruction and the Interest of the Instruction, and these three must all be assured for each individual user. This is obviously an area open to much research, as the major ITSs have not adequately addressed the topic. What has been done so far is bitty and incomplete, and is something about which there is little feedback, and on which more studies will have to be carried out. It is also of note that these same goals can also be shared by the user interface as will shortly be discussed. That then completes a brief look at the work of the Teaching Module to date.

The User Interface

This final section of the ITS should be the place wherein the tutorial materials reside, the simulations are run etc., and so could contain a large diversity of elements. In many current ITSs though, this module has largely been confined to text processing, be it through a sophisticated natural language processor or a set of multi-choice style pre-prepared questions. The BUGGY series for instance employs little more than standard text methods for presenting its arithmetic questions, as does Kimball's Calculus Tutor. Even a simulation such as WUSOR fails to take advantage of most of the other possibilities and sticks to plain text and a planning screen for the student to record their suspicions (and only then in later versions). Looking more closely at Figure 3.4 we can see that the entire Tutorial session is carried out through standard text from the Tutor and single letter responses from the leaner. The planning screen may force them into thinking about the gaming process & therefore the underlying concepts, but it is the textual interaction which is more typical of the other named Tutors. In this there is nothing to stimulate or interest the learner, and nothing to encourage thinking about the concepts or knowledge encapsulated within the ITS. They are thus missing out on a lot of learning potential which can be obtained through (carefully) utilising a multimedia user interface.

The point of using multimedia presentation is two-fold. Firstly, the presentation of information in several different modes is much more likely to aid learning. This is because it will both promote greater memorability due to the complexity of the information (by which I mean the number of connections the new knowledge can make to relevant prior knowledge that the learner possesses) and likewise is more likely to engender deeper conceptual learning for the same reason. Secondly, and just as importantly, such a presentation of the information is more likely to engage and sustain the motivation of the subject.

Figure 3.4 - Sample Tutorial Screens from WUSOR: Game Screen & Planning Screen.

Game Screen.

We are now in cave D. The neighbours are caves C, G, and B. What now? [Tutor]

> B [Student] {Command moves student to Cave B}

We are now in cave B. The neighbours are caves A, D, and F. What a stench! The Wumpus must be in one of the neighbouring caves. Squeak! I hear bats. They must be in of the neighbouring caves. What now? [Tutor]

>X+ [Student] {Command marks cave as hazardous in planning advice screen}

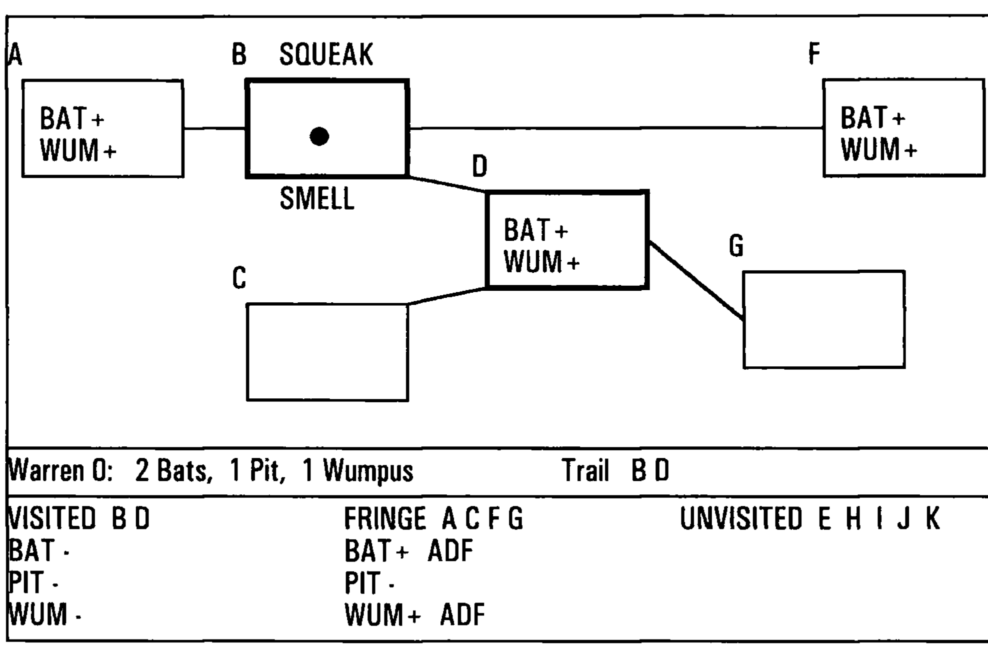
Which danger (Bats, Pits or Wumpus)? [Tutor]

>BW [Student]

Which Caves? [Tutor]

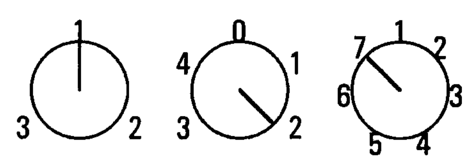
>ADF [Student]

Planning Screen.



Simple devices have been used in some of the ITSs developed already, such as the graphic display of the gaming board and coach in WEST. The display consists of a graphical gaming board covering about one third of the screen. The board has a twisting trail along which stage-coaches (the counters in this game) may be moved. The stages are on the trail as graphical images, as are the towns (although they are more representational than detailed). Similarly the shortcuts are marked as direct routes between twisting parts of the main trail. The three random numbers which make up the turns are displayed both as numerals and as points on three 'clock face' types of indicators. There is an on screen area for the player to type in a calculation using buttons for the allowable numbers and operations. There is also a hint box for them to gain advice on their calculation, and they can perform many different calculations before click an 'ok' button to enter their final number. Lastly there is a coaching display box which is seen as a squared 'speech bubble' coming from the mouth of a childish cartoon of a sports coach (with coach written across his chest). In his spoken words he gives advice, such as suggesting a different calculation they could have performed which would advance the learner's stage-coach much further, and offering them to take their move again (with a yes/no clock box).

The advantages of the WEST board are several. Firstly there are the motivational factors. It is designed to be fun for the intended users (primary school children) with moving stage-coaches, buttons to 'click' on screen and a friendly cartoon character offering them the advice (as opposed to the 'computer' itself). Also it is taking advantage of several different ways of presenting the information to the learner, i.e. the numbers and the mathematics. Firstly the numbers are randomly assigned on screen both as numerals, (i.e. 1, 2 and 7) and then also as in three number indicators:



Then finally the numbers are also represented by the movements of the stage-coaches along the trail - counting along a number line.

More complex designs have occurred recently too, such as the graphical interface in STEAMER (Holan, Hutchins & Weitzman, 1984) and in RBT (Woolf et al, 1987; Woolf, 1988a). The approach of these two systems (STEAMER & RBT) is even more appropriate, with the Tutorial Session (the running of a steam powered plant) displayed in a number of user chosen fashions. Taking RBT which is a simulation of an industrial boiler used in paper milling, if we look at the computer display we are presented with a wealth of graphical and textual information available on several different screen layouts. We can display simulations of a number of vital instruments, switches, warning lights or meters. The information can also be displayed as time traces, or as a physical representation of the power plant itself. In the main screen there is a large graphical representation of the actual boiler, with simulated actions occurring within the graphic. For instance, you can see the fuel being squirted onto the flue bed, can see the position of valves and their opening and closing & the flow rate of the liquor to the boiler. The main screen also has a section of basic dials showing Fuel composition and feed rate, water feed rate, steam temperature, pressure and flow rate and the composition of the flue gas (in important elements). There are also four Tutor guides to the overall Safety, Efficiency, Emissions and Reliability of the RBT operations. To find out more information in RBT you can look more closely at particular parts of the boiler, view the entire control panel or look at graphs for trends in the boiler's operations.

In close up the on screen graphic is enlarged to provide a better 'picture' of localised events, but otherwise the other on screen graphics (the Tutor guides and the partial control panel) are unchanged. The graphic can also be completely turned off to allow for the rest of the control panel to be fitted on screen. This gives a complete break down of Fuel rates, temperatures, composition, pressure and state, and similarly full accounts of the Steam, Flue Gases, Combustion and Water. On the final screen, the trends screen, the user can request to see how any of the factors in RBT (such as fuel temperature, Steam flow, O2 content in exhaust fumes) have varied over time , allowing the learner to see how their actions have altered them.

The RBT Tutor then allows the learner to attempt to run the boiler, altering various functions such as fuel rate into the boiler, or performing external actions on the boiler such as 'rodding' the fuel vents to let air flow in more easily (assuming they were clogged). When the user gains competence at 'day-today' operations, problems can then be induced in the boiler for the student to react too. All changes made to the function of the boiler through RBT are made by selecting actions from a pull-down menu of all allowable user actions (such as 'rod the air ports', or 'put in liquor'). The screen even has an area for an alarm to flash (the area is reserved on all screens) and is accompanied by a noise. This prompts the user to see what is going wrong in the boiler, or they can consult the alarm menu to be told what function has tripped the alarm (such as increased exhaust emissions). After use of the tutor it is intended that people will be capable of running (under supervision) a real boiler.

In RBT simple numerical information, such as a flow rate, can be displayed visually in up to four different modes. This has the added advantage over the more usual alphanumeric displays, that it supports the use of both rapid proceduralized skills and also deeper conceptual development. For instance, by observing a numerical readout from a meter or gauge you could use a piece of procedural knowledge (such as IF flow X > lmn THEN shut valve Y) whilst you could obtain / use conceptual knowledge from the time traces (such as observing over time that a trend in one part of the system CAUSES a particular trend in a different part of the system). The system is interactive, allowing many complex functions to be bought together in a real life task, and for the student to learn how they inter-relate in a non-threatening environment. True mastery (conceptual) as well as knowledge of how to cope with emergencies appropriately and quickly enough (procedural) are taught. It is probably the best application of a Tutorial system I have come across.

Another good example of the way in which the User Interface should be designed is the work of Price & Hobbs as described by Price (1993) and Soper & MacDonald (1994). Price & Hobbs are the executive producers of a set of computer aided learning materials for the teaching of basic University level Economics called

WinEcon. Within their programs there is a consistency of screen design which helps the learner to easily attempt any of the different parts of the package, both in uses of areas and also use of symbols and icons. The learning sessions are performed through the presentation of text and graphics, as well as utilising many different forms of student interactions and computer simulations. Graphs and models can be 'tweaked' and the result of varying conditions dynamically observed. For instance, in investigating an economic phenomenon called the Pareto effect (to do with the allocation of goods) the user can alter the initial conditions of a calculation and then see the effects this has on the outcome of the effect both in comparison to previous conditions in a scatter-plot and also for the particular instance in a pie chart. Consistency in design also means that summaries and important points are easy to locate and become more memorable. Likewise the factual knowledge which is presented as rules and is thus of the style of productions is complemented by the conceptual learning encouraged by the interactive parts of the tutorials. In addition, the actual design is modular, allowing wholesale changes to the operations and design of the packages to be instituted regardless of who authored any one particular teaching module, and without necessarily altering the User Interface. Indeed, any modern authoring package or even spreadsheet or office suite is capable of creating the most aesthetic of graphical user interfaces (GUIs), as the courseware for teaching kinetics amply demonstrates (Young & Heath, 1994).

Recent Directions

From the papers presented at the World Conference on Artificial Intelligence in Education in the late Autumn of 1993, it was clear that the research paradigm within Intelligent Tutoring Systems had shifted. In the keynote address itself, Kurt VanLehn pointed out the changes. Progress in modelling was seen to have come to a halt. Mostly people pursued the Anderson ideals of productions or similar information processing (IF - THEN) expert models, with an overlay student model (be it overt overlay or implied overlay through bug libraries). What ITS developers had failed to properly come to terms with though was the pedagogy of the interactions and this was where we were now to turn. He drew upon examples of self-explanation (a metacognitive skill discussed in the previous chapter) to show how we could best learn. He then went on to suggest (most forcefully) that without real progress in our understanding of student learning and the subsequent application of this knowledge to the design of ITSs we would not be able to improve their performance in any significant way. Indeed, it was even indicated that production style modelling such as based on Anderson's ACT theory be sacrificed too and replaced with less cognitively realistic but more robust models which will actually work 'in the field'. ITS development in the Expert module (and Diagnostic module) should stop, and we should instead focus on the teaching module and the Student interface.

This kind of view is echoed by many other researchers. Warren, Goodman & Maciorowski (1993) argued successfully for increased re-use of previous work. Using an executive program to co-ordinate various 'off the shelf' packages and old systems as modules within an object oriented environment saves much valuable resources and development time. Meanwhile a small amount of new programming is sufficient to tailor the package to the specific task it will perform. The modularity is stressed even further allowing for future 'upgrades' of different components of the ITS, even future developments in the modelling segments. For now however, no progress is deemed necessary or expedient in this direction. Orey, Young & Trent (1993) also commented on the prohibitively large amount of time that the creation of successful ITSs takes, and advocated the adoption of intermediary development aids, allowing for Experts to swiftly create courseware through enhanced authoring tools.

Michael Orey (in Orey & Nelson, 1993) goes on to blame some of the failure of ITSs not only on development times but on the cognitive models employed.

He sees that some form of hybridised representation is required to succeed in the creation of an ideal ITS; one where truly intelligent machines may communicate with the learner and perform the complex task of teaching. This belief is due to the problem that the form of knowledge representation is dependent upon the nature of the domain and the nature of the tutoring required. One could easily teach strictly algebraic topics (such as geometry) by production rules because of their similar nature, whilst in the teaching of troubleshooting where causal reasoning is the key to success a semantic network or frame based solution is more appropriate. To thus have an ideal ITS which conforms to Orey's desires it must be capable of dealing with all types of representations in order to deal with all types of problems - therefore hybridisation will be a necessity. Of course, one can teach something apparently based on serial logical knowledge through causal networked models (such as Brown and DeKleer, 1984) but the ideal system will also need to address the pedagogical fact that some people will learn better with one form of presentation or the other (or indeed both) regardless of the material presented (Stenning, Cox & Oberlander, Forthcoming). Similarly the ITS may need access to separate models of the learner for the same purpose: i.e. structuring the presentation of learning material may well be more easily based on a production rule style of model for a specific task but for a wider context be best served by a networked model. Obviously, if one model could hold both sets of information, much could be gained in terms of speed and efficiency, not to say our own understanding.

In line with this kind of approach the study of HyperMedia systems for Computer Based Learning and Intelligent Tutoring holds many promises. The structure of HyperMedia can be thought of as being both networked and symbolic - it is inherently a hybrid form of knowledge representation. Each nodal point can represent symbolic information or rules, whilst the inter-links between the nodes contain additional information about the particular domain (or domains) of knowledge. It must be pointed out however, that in Human use of such structures we can only follow one link at a time and only view one node of information at a time, but these are limitations we impose upon the structure by our use of it. The potential exists within it for a computer to make use of the representational form through parallel access of nodes and parallel travelling of links.

Indeed, Hybrid Models are pointed to as a possible saviour from the overlay problem (described in brief earlier) of standard ITSs by Mitchell & Grogono (1993). They espouse the view that in Intelligent Tutoring there exists an area of knowledge which we should like the learner to become proficient in (called the knowledge base) and a set of interconnections which relate all of this knowledge within the knowledge base (called the concept map). By then basing our modelling of the learner on these representations, we can have an understanding of not only what they have learnt from the knowledge base but how they see the knowledge relate to itself within the concept map. We are not therefore bound by a simple discrepancy comparison (or overlay paradigm) to deem the learner right or wrong, but can look at how they obtained their answer - to see whether the learner has developed a qualitatively different perspective whilst retaining a logically consistent network of their own.

HyperMedia CALs are not new, but the cross-over to adding some form of Intelligent Tutoring within them is. Some experimental work has been carried out on diagnosis within HyperMedia systems by Viau & Larivee (1993). In their research Viau & Larivee created an 'interactive textbook' to teach first year college students basic computer ideas and processes. The system also logged all interactions for each subject, generating a use path for time spent on any particular area of information (page) and amount of times travelled between any pair of pages. By applying pre & post use tests to the subjects (70 students) they could determine the factors of the interaction which accompanied enhanced learning. This in turn allows for a diagnostic technique where either learners who follow one or another strategy can be directed in the most educationally profitable fashion through the material or can be instructed as to the most efficient were of learning for that particular student style. Indeed, such developments was speculated upon in 1990 by Fabrice Florin, who envisioned an Information Landscape, where a Guide (The Intelligent Tutor) could accompany you through the land, journeying along roads (Tutorials), stopping to play (simulations and interactives), following a river (presentations) or simply wandering (HyperMedia) but always with a friend by your side with guidance and answers to hand.

Another example shows how the HyperMedia principle is again applied to create a domain free intelligence in a piece of CAL. The Linctus PB project (Briggs, Tompsett & Oates) utilises a networked system again, but the expert information is encoded purely as statements of medical symptoms within a node related to the medical condition which that node represents. The system then automatically creates the links between nodes, for instance where any node with a condition is linked to any node describing a symptom of that condition. Nodes also represent medications for certain conditions and symptoms. The system as a whole operates as an intelligent repository of knowledge for the pharmacist - a kind of 'memory jogger'. Details can be entered of a patient to elicit a 'diagnosis', whilst browsing through the HyperMedia system allows the pharmacist to refresh his or her memory. The system's knowledge as a whole was derived from experts in consultation with the designers, but the style of creation of the package is universally applicable.

To recapitulate the current position of ITS research, we can look to several distinct threads. Firstly there is the position of the orthodox cognitive psychologists who have viewed ITSs as a means to an end. As Orey makes clear (1993) an ITS is an excellent tool for learning about human cognition and thus they make very good experiments for proving the feasibility of different cognitive models of thought, learning and expertise. This camp has come to a point where the overheads of cognitive fidelity have still not produced a workable Intelligent Tutor with any applicability except in the narrowest of fields, and so they have now turned their attention from modelling the knowledge of learners and experts to modelling the learning process itself. Such ITSs have rarely crossed over though from the laboratory into the classroom. A second thread is that of computer based learning. This is the school of people who take the pragmatic approach that a computer that is a useful tool in learning is a lot better than one which mimics humanity but can't actually teach anyone anything. They have recently been making much better use of the resources in computer based instruction (WinEcon, Price & Hobbs). This is purely a piece of CAL software with no claims or pretensions to an expert module, teaching module or diagnostic module. Yet within the screen layout they have incorporated tasks as previously described which allow the learner to partake in interactions with the knowledge to encourage conceptual understanding. The screen design is created to ease learning through consistent presentations and interesting images and assignments: thus still incorporated learning theories in their design and therefore creating effective tutoring material.

The third thread is the emerging union between these two camps. The recognition of the Psychology and AI communities that a deliverable at the end of the day is important and of the Computer Aided Learning community that an intelligence in tutoring is desirable is opening up a new area. Work such as that of Viau & Larivee (1993) is moving the psychologists into the modelling of a practical and deliverable ITS by investigating modelling on a HyperMedia authoring system that is widely used already in CAL. In their work they took an interactive textbook designed as a piece of HyperText (in HyperCard) with other HyperMedia representations such as images available. To this standard authoring package, which is used in many CAL based applications, Viau & Larivee incorporated a diagnostic element to discover how much time was spent by learners when they used the interactive textbook on any particular topic, and to trace the routes which they took. From this data they established patterns of use of the CAL, and then performed testing on the subjects to see who had learnt most of the contents of the interactive textbook. This they proceeded to correlate with their identified strategies from the raw data and thus could now predict how use of the CAL would most effectively lead to increased learning. They are thus bringing the psychology behind Intelligent Tutoring to ready made CAL applications which could after further developments be used to 'upgrade' the CAL software into some form of ITS.

On the other hand, CAL designers such as Warren, Goodman & Maciorowski (1993) have investigated the transition of laboratory ITSs to used applications through the adoption of a modular yet communicative system which allows Intelligent Tutoring modules to be combined with proprietary authoring packages and pre-existent CAL. Instead of working within the CAL packages (proprietary authoring software such as HyperCard, ToolBook and Authorware) they are creating their own programmes to perform a similar function to Viau and Larivee - to perform the functions of an ITS. However, each module is completely independent of the others, but they 'communicate' between each other under the aegis of a piece of control software. They can thus use a CAL package as the User Interface and design whatever 'front end' to the overall ITS they desire (and not be confined just to HyperCard as Viau and Larivee are). Then they add a program to the authoring package to communicate with the control software and the ITS can function with any proprietary software. Similarly, the modelling or diagnostic modules can be inter-changed at will - perhaps in response to a breakthrough in student modelling, or perhaps to facilitate swapping between specifically production based or schema based or analogical based learning.

This together with my previous discourses into the current state of cognitive psychology with regards to modelling, eliciting and representing knowledge is sufficient background to isolate a research field. However, the field of application for such research as well as the specific system of implementation to be investigated must also be considered. In the next Chapter I therefore proceed to discuss these matters, explaining the choices of field, architecture and methodology before embarking upon a description of the experimental work.

Chapter 4
Modelling for an Archaeology ITS

Archaeological Need

A Very Brief History

My experiences have given me a very personal viewpoint on the teaching of Archaeology at University. Students in the subject tend to split into two camps from their first day at University - scientists and non-scientists, of whom the latter are traditionally the much larger proportion of students. The problem is two-fold; firstly teaching to a mixed background of students makes 'pitching the level' of the contents somewhere between very difficult and virtually impossible; secondly the actual contents themselves will be either too scientific to interest one half of the group, or too humanist for the other. Before fleshing out these points though we have to understand that the teaching of science to Archaeologists is in itself an essential objective of modern Archaeology. There are many scientific techniques available to the archaeologist to aid and enhance their work now, through the finding of sites and artefacts, to their subsequent mapping, recovery and interpretation. However, to successfully employ these scientific techniques requires conceptual understanding of their underlying science or else the techniques may be unhelpful at best, or at worst invalid and misleading as I will now explain.

By way of an example, consider the case of radiocarbon dating, perhaps one of the more popular scientific methods for helping the archaeologist identify the scientific age of a given sample. The problem is, that there are many factors relating to radiocarbon dating which will tell the scientist and the archaeologist whether the sample is useful or indeed valid. For instance, there are certain periods of history for which the scientists cannot calibrate the technique as accurately as for others. By way of an example, a Roman find may well require very accurate dating for it to be of use to the archaeologist (say within 10 years either side, perhaps less) whilst radiocarbon dating cannot actually provide a better estimate of its age than +/- 50 years. In this instance then combined Archaeological and Scientific knowledge are needed to understand that radiocarbon dating would not be an appropriate technique to use in this circumstance. Likewise the technique can only be used on finds within a certain age range and requires a certain minimum size of sample dependent upon its age to allow for a measurement of its age to be made. More importantly perhaps, there are certain preconditions upon the handling of the object for it to be dated correctly, such as complete isolation from any modern sources of carbon to avoid contamination with younger material. This also requires the archaeologist to understand what materials actually contain carbon (a good many students are surprised to learn that this means a polythene bag cannot be used in place of a paper one - tin foil is the appropriate material in this instance).

Having thus established that there is a need for the teaching of Scientific Techniques to archaeologists we return to our previous problem - the two distinct groups of archaeologists. The scientists are perhaps a lesser problem as they already possess a good deal of the requisite knowledge, albeit that they have not yet related it to the field of archaeology. They would genuinely like to learn some archaeology though (a good motivational factor in learning) and so could do well given a chance. However the non-science based archaeologists (as a general rule) have done little or no science at any previous level, only taking the minimum required 'O' levels to matriculate. They do not enjoy science, were glad to see the back of it, and are deeply offended when they reach university to be informed that they must start to learn enough science for them to appreciate the techniques that will help them in the future.

In addition to the problem of their antipathy to the subject is the class size. With there being relatively few scientists taking archaeology either as a main or subsidiary subject the lectures on scientific techniques are given to a mixed audience of predominately antipathetic non-scientists and a few science students. Whatever the level then pitched at the class, the non-scientists will resent the science whilst the scientists are unlikely to be motivated by science devoid of archaeology and pitched way below their own understanding. We thus have two problems - the student's antipathy to the subject matter and the teaching of the subject matter itself.

An ITS for Archaeological Science

The situation thus described was the prompt for investigating the development of an ITS for archaeological science. The individuation which can be gained from a one to one interaction with an interactive and responsive piece of software (by which I mean the ability of the tutor to respond to & adapt to a particular individual student in a way which is educationally advantageous for the student) could alleviate the two problems identified earlier. This could be achieved by allowing for the various different levels of scientific and archaeological understanding within the learner group. Then an ITS could individuate for any student, by altering the curriculum presented to the learner. An archaeologist could be gently bridged from their prior knowledge of archaeology through to the scientific applications which Archaeologists utilise, and then to the underlying scientific principles themselves. This would cushion the Archaeologists from the shock to the system caused by throwing them into the science head first. Likewise, the scientists to whom ancient history is most probably a closed book can be gently led from the science through the scientific applications which Archaeologists utilise (though this time focusing on the science first instead of the Archaeology), and then into the actual archaeology itself. Also, just as obviously, this allows for total re-arrangement of the material presented. Thus students of whatever background can be directed through the learning experience in whatever way is most educationally profitable given their own individual prior knowledge. This of course is given the proviso that there is an intelligence to diagnose the learners prior knowledge state first, or that the students themselves can choose their own curriculum, or both.

This would make for more effective learning possibilities, especially if the software employed additional multimedia techniques for complex displays of the information. By this it is meant that the information is presented in different forms such as plain text, graphics, or in an interactive simulation. Such complexity makes for greater memorability of the learning event. In addition, interest is more likely to be promoted by the use of animations, sounds and graphics which a multimedia presentation entails. Within the context of the tutoring the motivation of the student could thus be more easily promoted and maintained. An ITS for Archaeology is thus a possible solution which could bring excellent learning outcomes to a troubled area of higher education.

Scientific Dating in Archaeology

A General Description

The subject matter for the knowledge domains which I intend to use is drawn from Scientific Archaeology - Dating Techniques. As a part of their first year lecture courses Archaeology undergraduates learn of many different dating techniques and studying some of the common ones (such as radiocarbon dating or carbon 14 dating). This makes it an ideal area to study novice expert differences, as first years will hopefully be knowledgeable enough in the subject area to comprehend the terminology of the techniques without necessarily yet having an in-depth understanding of the processes. They would thus be a good novice group, more so than a group of people with no experience of the domain. The latter would only furnish us with the knowledge that they understood nothing, or did not know the relevant terminology. By looking at fledgling knowledge however we have something to compare with the knowledge representation of the experts - lecturers in Scientific Archaeology. To help the reader understand the domain too I offer a brief insight into Scientific Dating techniques below.

Carbon 14 Dating

Carbon 14 Dating (which is also called Radiocarbon Dating) is perhaps the most commonly known of the scientific dating techniques, being responsible for establishing the age of items such as the East Anglian Bog Man (Pete Marsh), the Sutton Hoo Long Ship and recently the Turin Shroud. Carbon 14 (or C14) is a radioactive isotope of Carbon. This means that it is subject to radioactive decay. In a period of time called the Half Life, half of the number of C14 atoms will have decayed. C14 is a radioactive nuclide, and may combine with other elements just as other isotopes of Carbon do to form molecules. This means that C14 can enter into the carbon cycle. Radioactive C14 is produced originally in the upper atmosphere due to the affects of cosmic radiation from space. This produces a constant source of new C14. Given that there is also a constant decay of C14, a dynamic equilibrium can be set up, allowing a constant level of C14 to exist. The C14 will be photosynthesised by plant life and thus gets into the food chain, and so into all organisms. The way in which it is photosynthesised depends upon the type of plant. Different photosynthetic pathways exist in different plant types which affects the relative fractions of the different Carbon isotopes which the plant absorbs. In the seas a lot of C14 which is absorbed into the food chain in plankton, and a lot then becomes a part of the shells of crustaceans. This means that the Oceans are a great reservoir of C14.

Since there is deemed to be a constant level of C14 in the atmosphere, all the time that any living organism lives there will be a constant renewal of the C14 in their body through the food chain. However, with the death of that living organism the exchange of new C14 in the organism ceases. From that point onwards there can only be decay of C14 in the organism. By measuring the amounts of C14 and the other isotopes of Carbon at a later date, the proportion of C14 compared to the others will be found to have diminished compared to the equilibrium state in the environment. Libby measured the decay of radioactive carbon and found that the half life of C14 was 5700 years (approximately). Using this figure it is thus possible to determine the period since death of any organic matter.

Advanced Carbon 14 Dating

Carbon 14 Dating has some further refinements. Firstly it is not the case that the environmental equilibrium of radiocarbon has been constant through time. It is therefore necessary to calibrate the age of any sample that has been C14 dated. This is achieved through the application of tree ring dating (or dendrochronology). Using tree rings it has been possible to construct a chronology for many thousands of years by looking at growth patterns and which we can fit in to our present calendrical system. We can also analyse the carbon content of each ring of the trees and so observe fluctuations in the proportion of C14 in the environment at any time in the tree's rings' history. This can be used to adjust the dates calculated with the assumption of a constant and unchanging equilibrium.

In C14 Dating, the calculation of the proportion of C14 in a sample is performed by measuring the radioactivity of the sample. The activity of the sample is in proportion to the number of atoms of C14 and so calculations can be made to determine the amount of the different isotopes present. Unfortunately as the sample gets older the radioactivity of the sample becomes less, and after 6 or more half lives there is so little radioactive carbon left undecayed that it becomes impossible to differentiate the activity from the sample and the activity that naturally occurs on the planet (background radiation). This Decay method of calculating radiocarbon dates is thus limited to an upper range of 30 000 - 40 000 years, beyond which age a sample cannot be dated

There is however a solution to the problem of measuring radioactive decay, and that is to use Accelerator Mass Spectroscopy (or Accelerator Dating). When the activity is so low that it is virtually indistinguishable from background radiation there are still astronomically large numbers of actual radioactive nuclides in the material. Using a particle accelerator a technique called high energy mass spectroscopy can actually count the number of atoms of each Carbon isotope in a sample directly. By so doing an older sample can still yield a radiocarbon date up to a range of 50 000 - 80 000 years old. The use of such an Accelerator Dating technique has the added advantage that a smaller sample can also be used for dating as again the size of the sample activity is no longer a problem.

A practical point also exists, that the Archaeological sample must be kept clear of any Modern sources of carbon to avoid contaminating the sample with new C14. If this were to happen then any calculations on the age of the sample would produce a falsely young age for it. Samples should always therefore be stored in non-carbon compounds

Potassium Argon Dating.

Potassium Argon Dating (or K-Ar Dating) can be used to date much older objects than radiocarbon dating. The half life of radioactive Potassium (K40) is 1 000

000 000 years (approximately). When it decays Potassium 40 yields two different possible products (or Daughter Nuclides), and does so in a constant ratio. One of these Daughter elements is the gas, Argon. Potassium occurs naturally in many minerals and rocks, one of which is the mineral Feldspar. Feldspar forms a crystalline structure, and as the Potassium 40 in it slowly decays Argon gas is formed and trapped within the crystal. If the Feldspar crystals are subjected to massive heating such as being involved in some kind of Volcanic Event then the crystal lattice is weakened such that the Argon gas is driven off.

To use K-Ar Dating as a technique all that needs be done is measure the amount of Argon gas trapped in a Feldspar crystal. This will tell you how much Argon has been produced since the last massive heating of the rock, and knowing the rate of Argon production (by virtue of the half life of Potassium 40) an age can be ascertained for the sample. Since the half life is so incredibly long this technique's range has a lower limit at 400 000 years old.

Uranium Series Dating

Uranium Series Dating (or Uranium Thorium Dating) uses the decay of radioactive Uranium to make a dating technique. Uranium has many different isotopes, more than one of which is radioactive. The decay of Uranium 234 (U234) produces several daughter products which are themselves radioactive. Fortunately the half life of U234 is relatively short (350 000 years) compared to the half life of some of the other products and Uranium isotopes (1 000 000 000 000 years approximately). Thus these other products and isotopes can be effectively ignored from our calculations. We had to be certain though as some of the other Uranium isotopes decay to the same daughter as U234 by a different path. The decay product which we are interested in is Thorium. By calculating the build up of Thorium we can determine the age of a sample.

Uranium is a naturally occuring element and is found in many rocks including sedimentary rocks. As water flows over the rocks Uranium can become absorbed into the water, however Thorium is insoluble and so cannot enter the groundwater. Underground when the water emerges into caves it can form stalactites as the Calcium in the water precipitates out. The Uranium in the groundwater can also be precipitated out and become a part of the stalactite too. As the Uranium continues to decay Thorium continues to be produced and also becomes a part of the stalactite. By determining the proportions of Uranium and Thorium in the stalactite it is thus possible to calculate the stalactite's age.

General Dating Ideas

There are several ideas which are common to the dating techniques which we have already mentioned. The actual calculation of the ages of the samples in all cases requires the application of a radioactive decay formula. All of the techniques are applied to an Archaeological material to provide a date, and each technique has an associated range to it within it is applicable. There are additionally some operating limits on the ranges, such as the size of samples and the sensitivity of equipment.

For each technique there is some clock mechanism - a process which occurs at a regular and measureable rate (such as the decay of radioactive nuclide). The rate itself must be constant, as in the constant half life of radioactive materials. This all establishes a chronometer which can be investigated to see how long it has been running. To be used as a dating technique with any practical value though it is essential to know when the chronometer started. All of the dating techniques thus have a clearly definable Zeroing Event - an action which started the chronometer running. For instance, in C14 Dating the proportion of C14 is stable until the organism dies and carbon exchange cessates. After this point (the Zeroing Event) there is a gradual decline in the amount of C14 in the organism because of the radioactive decay of C14. All of the techniques must also have a way of measuring the proportions of various elements in the samples to make the data for the calculating of the samples age, such as high energy mass spectroscopy.

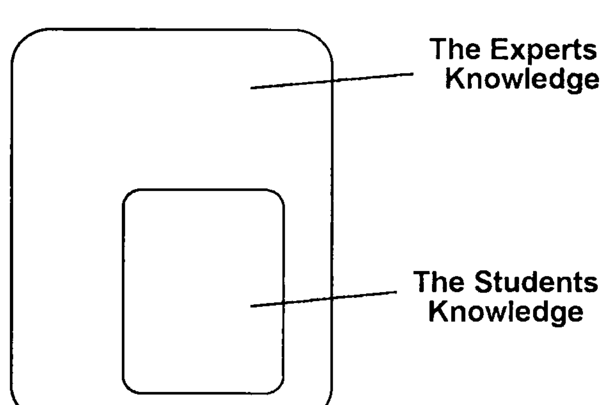
These five areas are the areas which I used in the design of the experimental material for this study. They include three completely different radiometric dating techniques, one of which, Carbon 14 Dating, was subdivided into basic dating procedures and advanced techniques. In addition, we have seen that there are some general principles that are common to all three techniques. This concludes my brief look at the radiometric techniques used. In the first study, a biological (and also nuclear) technique was investigated. That is mitochondrial DNA dating (or mtDNA Dating)

In all human cells there exists mitochondria with but one exception - the male sperm. This is because mitochondria live in the cell cytoplasm which normally surrounds cells, but which sperms do not have. This means that although our genes are jointly inherited from both of our parents, our mitochondria all come from the maternal line. Over time, as with all cells, the mitochondria may undergo mutations, and this gives us the basis of a dating technique. Since there is a temporal mutation of the mitochondria, as long as we know the rate of that mutation we can calculate the divergence of different people's mitochondria by virtue of their genetic difference. Thus the female heredity can be ascertained for all women. By calculating the divergence between women of different ethnic origins we can chart the times at which different cultures split off from a common ancestry, right back to the first ever woman to have lived who shares our own genetic make up. mtDNA Dating has shown that this common ancestor was born between 140 000 and 280 000 years ago, and that she lived in Africa. Archaeologists have nick-named her 'Eve'. That completes my brief discussion of the Archaeological materials. I shall now consider the problems in modelling knowledge.

Designing a System

Early developers of Intelligent Tutoring Systems (ITSs) made extensive use of simple overlay modelling to keep track of the student's progress as explained in Chapter 4. In these systems the expert's knowledge, however it has been encoded, is considered to be the entire sum of the knowledge concerning the particular domain. The student is considered to have acquired only a small fraction of this knowledge (if any) before learning begins as can be seen in Figure 4.1. The purpose of the Intelligent Tutoring System (ITS) in this case is to fill in the 'blank slate' that is the learners knowledge with all of the expert's knowledge as represented within the system. As the students knowledge at any one time could thus be represented by *laying* a card with holes in it *over* a card containing the experts knowledge, the method gained the name of overlay modelling.

Figure 4.1 - An Overlay Model view of Student and Expert Knowledge



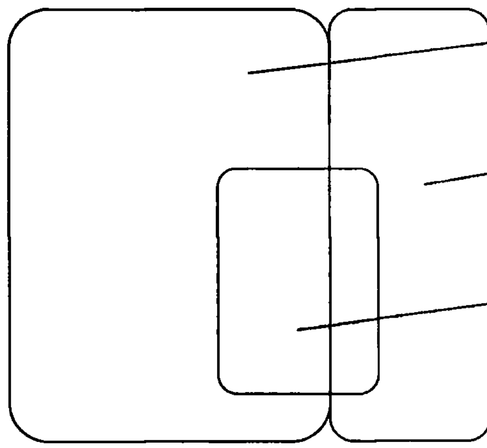
This clearly is an incomplete assessment of the student's actual knowledge, as they obviously do bring some prior experiences and ideas with them into any novel learning situation as was explained in Chapters 1 & 2. This meant that there was a gaping hole in student modelling methods, and led to a realisation that the mis-conceptions which learners bring with them must be addressed for effective remedial action to be taken and to effect appropriate new learning. This need led to the creation of 'Buggy' models to encapsulate the knowledge state of a learner within an ITS including misconceptions.

Buggy Models

In 'Buggy' or mis-conception models of student's knowledge it is assumed that the learner's lack of expertise is not only due to a lack of expert knowledge, but also parts of their own knowledge being erroneous, as is shown in Figure 4.2. The ITS either has a library of possible 'Bugs' pre-generated or else has a smaller library of 'Bug' parts which can be used to create a larger number of 'Bugs' out of several compounded 'Bug' parts. Then when the ITS notices a discrepancy in performance between the learner and the recommendation from the expert model, it can consult its libraries to see whether the failing is due to a lack of knowledge or the application of an erroneous piece of knowledge. By the application of many different problems related to the knowledge domain it is possible to calculate which different 'Bugs' or 'Bug' parts are responsible for a student's performance, and thus allow for effective tutoring to occur which can specifically address the mal-rules which the student has developed before teaching them the actual correct rules for the problems encountered.

This is where commentaries end on the development of student models. You either have an Overlay model or a 'Buggy' model. However, it is my contention that both of these types of modelling are examples of Overlay modelling, and that ITSs

Figure 4.2 - A 'Buggy' Model view of Student and Expert Knowledge



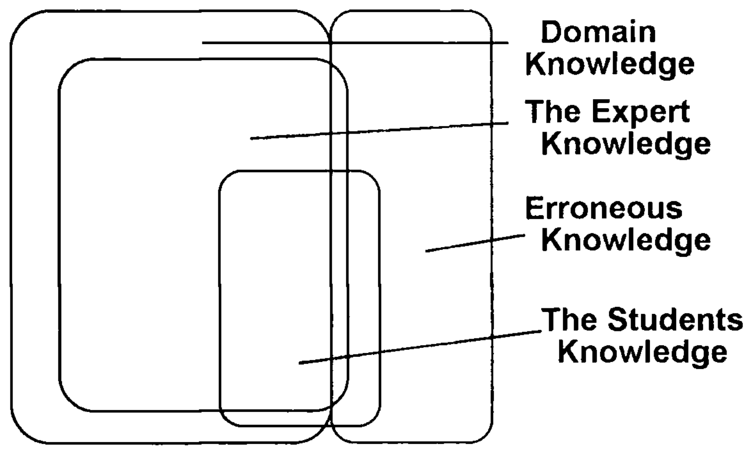
have yet to break out of an Overlay Paradigm.

Critique of the Overlay Paradigm

The problem can be represented most easily as in Figure 4.3. The domain knowledge is no longer to be seen as the expert's knowledge: they are independent areas. This should come as no surprise, since any two experts will have different views and opinions on some topics, although they may well agree on a good many more. The point is that an expert will not necessarily represent the entire domain knowledge. Likewise, it is also conceivable that an expert may in fact have some erroneous ideas themselves, maybe a few, perhaps none, but still possible. The student then may have knowledge which is also shared by the expert (although it is not all guaranteed to be 'correct') and also knowledge both appropriate and inappropriate to the domain which is not shared by the expert.

If we allow ourselves this view of knowledge then we are allowing the student not only the freedom to have inappropriate knowledge of the domain as

Figure 4.3 - Beyond the Overlay Paradigm view of Student and Expert Knowledge

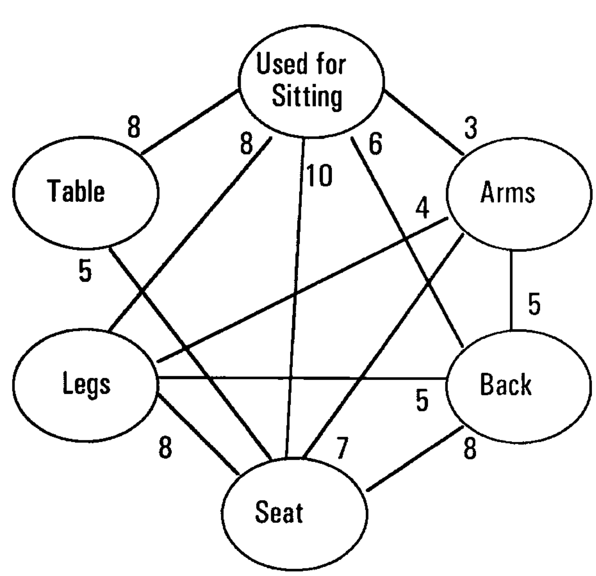


'Buggy' models allowed, but also to have or attain knowledge of the domain that is not necessarily shared by the expert. By this view I classify both Overlay modelling and 'Buggy' modelling as methods which fall foul of the 'Overlay Paradigm' in Intelligent Tutoring; that *the student's knowledge, when correct, must be a facsimile of the expert's knowledge*. This would imply that the end point of all teaching is to turn out 'clones' of our current experts both in terms of their knowledge and its application. As I discussed earlier (Chapter 2), such Routine Expertise, is not always desirable or appropriate as their are many situations which may call for Adaptive Expertise.

What we should be teaching students instead is how to think like an expert, not what an expert already understands. In this respect what is required is a system that is capable of deriving a measure of expertise which is free from the content of the domain but can be used nevertheless to discriminate between levels of expert performance and behaviour within a domain. In such a system, expert thinking and application could be rewarded rather than rote learning of the expert's knowledge as is often the case in other systems. By not shackling ourselves to the Overlay Paradigm we can still teach expertise for and within a domain, but we do not necessarily teach purely the expert's knowledge of the domain. To be able to achieve this we need a system to encapsulate the knowledge of the learner that is not a difference model from the expert model in the ITS and one that is additionally capable of storing information about the student's understanding of the subject and its concepts.

Having shown how most current ITSs fail to do this and from my discussion of the emerging Hybrid modelling approach in Cognitive Science, it should be apparent that a Hybrid Symbolic Connectionist Model would satisfy the requirements of both a student and an expert model within an ITS. The actual system which seems most appropriate to this study is the Symbolic Connectionism of Holyoak & Thagard. Within their system, the knowledge of any person (student or expert) is represented by the interconnections between individual concepts for a given overall concept. Thus for the overall concept of chair we can see in Figure 4.4 that the various components and functions of it are all linked together. Some of these links are stronger than others, and represent the subject's belief that they are more closely related.

Figure 4.4 - A Symbolic Connectionist Network for a Chair.



This model can be 'interrogated' in two ways. Firstly, we can look at the relationship judged between any two components, such as the 10 (maximum) relationship between 'Seat' and 'Used for Sitting'. In this way the symbolic nature of the model can mimic the way that a normal Overlay Paradigm system would function, as a simple comparison can be made between Student and expert answer. However, a second and more useful answer can be obtained from employing the connectionist part of the model. In Holyoak & Thagard's work the model was subject to a mathematical device to propagate *activation* throughout the network. To perform this feat, each concept (or node in the network) is initially assigned an arbritary *level of activity*. Then each node would have its activity altered by the activities of the nodes to which it is connected. Hence, with 'Seat' and 'Used for Sitting' maximally connected, Seat would receive an increase in activity equal to the activity of the node 'Used for Sitting' scaled by the strength of the connection, 10. This is obviously a greater contribution to the activity gained by the node 'Seat' than that which would come from its connection to the node 'Table' where the connection strength is only half the maximum at 5. Of course this assumes that both of the contributing nodes have the same activity, which would usually be the initial condition for the system. The process of gaining activity from the connections to other nodes is further modified by the activity of the node itself. For 'Seat' the contribution of its own activity and that of the connected nodes are weighted: the greater the activity of a node the less it is influenced by its connections to other nodes. Also, there is a decrease in activity due to decay of *activity* that is in proportion to the amount of activity in the node. This process occurs for each and every node, and when all of the calculations have been completed once, a cycle of time is deemed to have passed in the spread of activity. This propagation may be represented mathematically thus:

$$A_N(t+1) = A_N(t)*(1-d) + \frac{\sum(A_i * C_{Ni})}{\sum A_i} * (1-A_N(t))$$

where $A_N(t)$ = The Activity of the node N at a time t,

d = The decay constant,

A_i = The Activity of a node i,

$\sum$ = The sum over all nodes i, where i = 1 to the no. of nodes and $i \neq N$,

& C_{Ni} = The connection strength between nodes N and i.

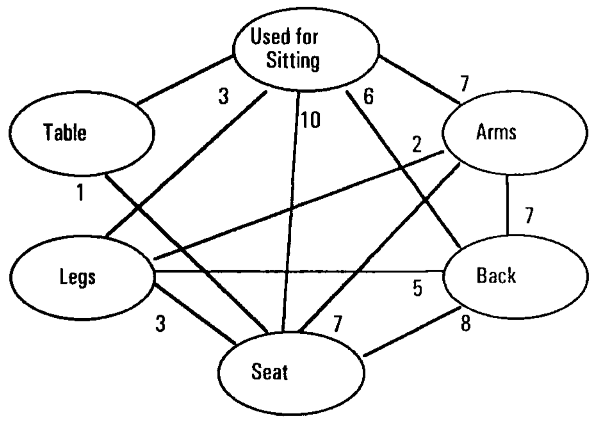
This process is repeated until such time as no further cycles of the network (applications of the formula to the data) produce a change in the activities of any node. At this point the network is deemed to have reached a stable state. The activity of any node in comparison to the others then gives a relation of the node to the overall concept, with the node having the highest activity deemed to be most closely related to the overall concept. So, in our example for instance, the activities we find are as shown in Table 4.1.

Table 4.1 - Activities for the Concepts in Chair

	Table	Legs	Seat	Back	Arms	Used for Sitting
Activity	.404	.575	.614	.537	.484	.611

This allow us to see that in the eyes of the subject the item Seat and the function of Used for Sitting are most important concepts to the overall concept of a chair, whilst Table is least important. In the mind of this subject (the author) I was clearly thinking of a desk chair, however it is easy to see that a completely different idea of a chair could be held by another person which would yield very different relationships between the individual concepts. Thus if you were to think of a comfy upholstered chair instead you may well give no relationship to table and very little to legs from any of the other nodes. Instead you may give much more weight to the arms, ending up with a network like Figure 4.5 instead.

Figure 4.5 - A different Symbolic Connectionist Network for a Chair.



Now to compare the relationships with the previous 'Expert' definition of a chair, there are only three relationships still judged identically - grounds to say that this person has a different idea of the concept Chair, and therefore according to the Overlay Paradigm the wrong concept. However, if we repeat our network analysis of the stable activation levels and compare with the previous result we can see clearly there is little difference.

Table 4.2 - Activities for the Concepts in Chair, both Desk and Comfy

	Table	Legs	Seat	Back	Arms	Used for Sitting
Comfy	.200	.436	.618	.594	.574	.608
Desk	.404	.575	.614	.537	.484	.611

As can be seen from Table 4.2 although the activities have changed somewhat, the positions of importance given to each concept are still the same, with Seat most important, followed by Used for Sitting. However, the Comfy chair sees legs in 5th most important position rather than third, but otherwise there are no rank changes. What we do not have though is the independent measure to show that this second view of chairs is just as valid. Obviously we know that both are acceptable, and also that most people are experts in the field of chairs, being able to recognise instantaneously whether an object is or is not a chair. What we need still is the measure of expertise.

Measuring Expertise

The idea of 'dynamically interrogating' the network would be difficult to use in practice due to the volume of information it would produce. If we were to continually take the activity of each and every node, even a small network would produce many items of data. If we then had to do this for both a student and an expert model to enable a comparison we would need to handle twice the data again, and we would still need to find some criteria on which to compare the information. Also, not only can we interrogate the network with an equal initial distribution of activity, but also we can interrogate it under different constraints: keeping one (or more) node(s) at a maximum activity for instance. This is called 'clamping' by Holyoak & Thagard, and allows us to investigate the effects of a given concept in relation to all of the other. It is akin to saying that the one concept is immutably true and then seeing how the relationships between the other possible concepts are altered because of it. By allowing any one or more node to be clamped we thus have another way of producing yet more information from the network on which to compare student and expert models. We are almost overwhelmed with possible comparisons.

What is needed instead is something similar to the 'goodness' criteria used by PDP researchers (such as Rumelhart & McClelland), where a statistical or mathematical function could yield an interpretation of the entire network data in a readily comparable and simple numerical form. Their method though (Smolensky, 1986) is not directly applicable to a symbolic system so an alternative technique was sought. The current research literature however failed to provide a suitable statistical test for data created by such a network. As the original work of Smolensky with

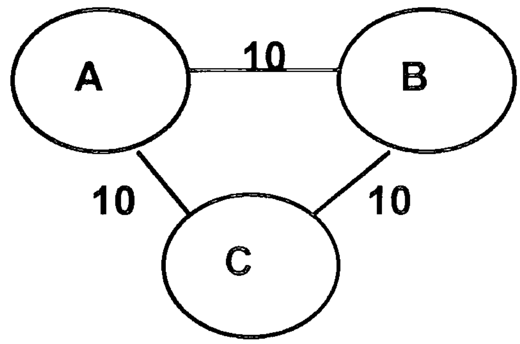
Harmony sounded plausible as a method if it could be applied to a different type of network I returned to it. A more recent study was also identified which appeared to use Smolensky's method with a symbolic network (Briton & Eisenhart, 1993) and so I studied both these works to identify a method for creating some form of 'goodness' value for a hybrid (symbolic & connectionist) network appropriate for this research. This 'goodness value' has become an important part of this thesis, and so a description and appraisal of the original works is a necessary fore-runner to my own research. I will start off with a quick look at the idea of network coherence which leads on to Smolensky's definitions of Harmony. Then I identify how he and also Briton & Eisenhart calculated their respective values for a network's Harmony, and finally what I myself learnt from these works and how I then applied this to create my own function for a network's coherence.

Coherence

To understand coherence, let us look at an example of a simple network with three inter-connected nodes: A, B & C. These nodes represent three concepts, and the connections between node pairs stand for the 'similarity' between them as perceived by the subject. A subject would be asked to rate the similarity of each pair of these three concepts on a scale of 0 to 10, where higher values represent greater similarity. If that subject judged all three concepts to be highly similar they would rate all three possible pairs (A&B, A&C and B&C) as 10, and the network representing this would be as shown in Figure 4.6.

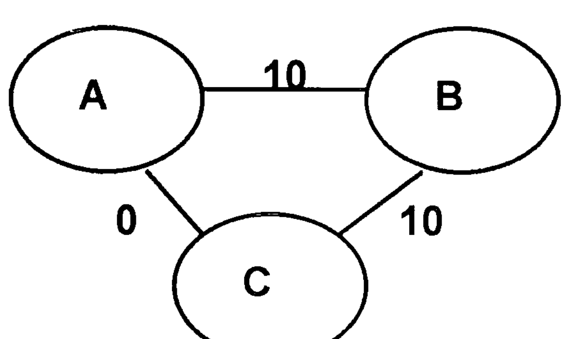
Now this network is logically self consistent if we examine it. To paraphrase the relationships we could say that A is similar to B, B is similar to C & C

Figure 4.6 - A Three Node Network with Three Highly Related Concepts.



is similar to A. Indeed we could borrow from logic and say that A=B=C is logically sensible since it expresses a transitive relationship. We can therefore say that this network is coherent. The same would also be true of a network where A-B was scored at 10 whilst B-C and C-A were scored with a zero. In this second case we could say A=B≠C, which too is logically sensible (and implies through logical transitivity that A≠ C). However, were the case to be that A-B and B-C where scored 10 but C-A were scored with a zero then we have a difficulty. On one reading of this network (see Figure 4.7) A=B=C, which implies that A=C; however, A is not equal to C and so transitivity is violated. We therefore would say that this network is incoherent. We should also say the same of a network where C-A scored 1 instead of 0 and A-B and B-C still scored 10. This is however a different case from the previous one, since the strengths of the relationships are no longer all-or-none. We therefore have to quantify coherence instead of using an all-or-none binary measure.

Figure 4.7 - A Three Node Network with Three Highly Related Concepts.



A quantitative measure of coherence can be calculated from the difference between the activation of each pair of nodes and the strength of connection between each pair. What coherence shows is the way in which the supposed relationships of the individual pairs is consistent with the overall pattern indicated by the network's inter-relationships in all other possible paths. The activation states represent for us the 'global' importance of each concept to the overall concept in a stable and unconstrained network[1]. We can then compare these overall activation levels (the activation levels of the nodes) with the ascribed similarities (the subject's ratings, which are represented by the connection strengths between nodes) by taking the nodes in pairs. We can thus look at the deeper inter-relationships which the overall activation levels express for a pair of activation levels in comparison to the predicted similarity of the concept pair (from its ascribed connection strength). This can be given as a numerical relationship between the two measures; the difference between activity levels for a pair of nodes & the connection strength between that pair. This comparison can be made mathematically to give a similarity score between the two measures, which we may call the coherence of the node pair. This coherence value can then be summated and scaled over all of the concept pairs to give a single index: a coherence value for the entire network.

Harmony

" The contribution to H of an inactive atom is zero. The contribution of an active atom a is the product of its strength and the consistency between its

---

[1] Though we could also constrain one concept and look at the changed activity pattern in the light of this, in which case we could look at the importance of all concepts to one particular concept or more, as opposed to the overall concept.

knowledge vector k_a and the representational vector r; this is measured by the function $h_K(r,k_a)$" [2]

Smolensky thus defines a value for the Harmony of a two-tier network where a sub-layer of knowledge atoms are linked with unitary strengths to a layer of representational features. Harmony is the original value calculated for the 'goodness' of a network by the PDP research group, and is a particular way of measuring the coherence of a network as previously explained. The value for Harmony, H is calculated as follows:

$$\mathbf{H}(r,a) = s_a a_a h(r,k_a), \qquad \text{where } h_K(r,k_a) = \frac{r.k_a}{|k_a|} - K$$

$$\text{and } r.k_a = \sum r_i\,(k_a)_i\,, \text{ and } |k_a| = \sum |(k_a)_i|$$

and where r = representational feature vector

(i.e. the activated nodes from the knowledge base)

a = activation vector of a_a for all a.

s_a = strength of atom vector

(i.e. the frequency of an activation pattern appearing

a_a = activation of knowledge atom a (-1,0,1)

k_a = knowledge vector (weights of connections from a to r_i)

in the discrete range of (-1,0,1).

Looking at the main equation for the value of **H**, the part of it which is $s_a a_a$ is used to scale the similarity rating, so that it is bounded in the range [-1,+1]. It also means that any node not activated from the knowledge base is not included in the calculation of a value for Harmony. The other part of the equation is the basic

---

[2] p 222. "Information Processing in Dynamic Systems: Foundations of Harmony Theory", Smolensky, P. Chapter 6 in *Parallel Distributed Processing*, McClelland, J.L., Rumelhart, D.E. and the PDP Research Group., MIT Press, 1986.

similarity rating, $h(r.k_a)$. Since $|k_a|$ is purely a scaling device h is also bounded in the range [-1,+1], with each non-zero connection of k_a for each activated knowledge atom a contributing a scaled binary relative to its own state. This is on its own a simple rating scheme to yield a Harmony value for a two-tier network. It is modified however by the addition of the constant K, also bounded in the range [-1,+1]. If K=0 then **H** remains the simple Harmony equation, being the concordance of those features which agree less those that disagree. However, Smolensky identifies that this means if over 50% of k_a agrees with r, then an increase in the activity of a will necessarily mean an increase in **H**, which is not in itself necessarily a good move. If this were the case, then an increase to the activity of any node would cause an increase in Harmony, regardless of the node's relationship to the rest of the network. Thus in a marginal case where only just half of all the input of activated nodes in the knowledge base agree with the connection weights there are almost as many nodes which do not agree and yet still have the ability to raise the Harmony of the network through an increase in their own activity. However, by ranging the value of the constant K, we may alter this criteria of percentage agreement through the range of 0-100%, thus giving us greater control over the calculation of **H**.

Essentially then, Smolensky's work on Harmony may be paraphrased thus: If the connection strength and the activation concur, then $r_i (k_a)_i$ = 1.1 or -1.-1 =1, whilst if they differ then $r_i (k_a)_i$ = 1.-1 or -1.1 = -1. This kind of data manipulation and calculation is indeed well served by the matrix mathematics which Smolensky employs, as it quickly and efficiently deals with the large number of computations involved in any reasonably sized network, and since it is very easy to express in its notations. However, this form of calculation will only work because the states are represented in a Binary form. The probability of a particular pattern of activity occurring is not explicit in the connection strengths in this representation but is instead inherent in the outcomes of all possible input values of the network. This also means that any scaling has to be applied after the matrix multiplication stage of the calculations has occurred. To use similarity rating on an analogue scale within the bounds (-1,+1) could not be treated similarly, since to use a reduction ad absurdia, $\cdot 7 \times \cdot 7 \neq 1$!!! As Smolensky himself puts it, " It will turn out to be convenient to denote present and absent respectively by +1 and -1 .... Other values could be used if *corresponding modifications* [my emphasis] were made in the equations ... The use of continuous numerical feature variables, while introducing some additional technical complexity, would not affect the basic character of the theory". [3]

Symbolic Networks

The application of such a measure of the internal coherence of a network to a symbolic network as used in this research would be most useful as I previously stated, and at the 15th Annual Conference of the Cognitive Science Society a paper was presented by Briton & Eisenhart (1993) which claimed to have done just that. In their experimental work, a questionnaire was produced which elicited conceptual knowledge in the same fashion as in our project's data collection, and which was apparently used to create symbolic networks just as we had done. They claimed that this showed results which distinguished between experts and novices in two ways - firstly that the settling rate of their networks was decreased by expertise, and secondly that Harmony was increased by expertise. The only difference between their methods and ours is that they used a 7 point scale for their data collection where we used 11, and that they used the Kintsch and McClelland & Rumelhart system of bounding to [-1,+1] for the range whereas I followed Holyoak & Thagard in using [0,+1]. Briton & Eisenhart expressed their equation for Harmony much as Smolensky did, putting

---

[3] p 214. "Information Processing in Dynamic Systems: Foundations of Harmony Theory", Smolensky, P. Chapter 6 in *Parallel Distributed Processing*, McClelland, J.L., Rumelhart, D.E. and the PDP Research Group., MIT Press, 1986.

forward their own modified version. This is not, however, spelled out formally in notation in the conference proceedings, merely in common English. By referring back to Kintsch's original work (Kintsch, 1988) though it is not impossible to make a formalisation, and as such this is what their formula appears to be:

$$H(r) = \frac{[(a_a . a_a^T).W]}{\sum |(k_a)_i|} . 1$$

where r is the set of features,

k_a is the weight vector for a,

a_a is the activation vector for r (+1,-1) &

a_a^T is its transpose,

and where W is the weight matrix. This, from Kintsch, is a matrix where the connection weights for the nodes are repeated on both sides of the diagonal, but has a null value placed in the diagonal itself. i.e. ., for a 3-D matrix,

$$W = \begin{pmatrix} - & w_{21} & w_{31} \\ w_{12} & - & w_{32} \\ w_{13} & w_{23} & - \end{pmatrix}$$

In this situation, the similarity rating for any 2 nodes is represented by their product multiplied by the connection weight between them (and later summated and scaled). Essentially then, this appears to be a barely altered version of Smolensky's work cut down for use in a homogenous network: $a_a.a_a^T$ replaces $r.k_a$ to get around the change from a two-tier network to a homogeneous network; the constant K is permanently set at zero, and the scaling factor of $\sum |(k_a)_i|$ is identical. Briton & Eisenhart even re-scaled their original data from the range (0,+1) to (-1,+1) to make use of Smolensky's formula. However, to account for the change-over to continuous activation energies (even though they actually used a small set of discrete numbers within the range) Briton & Eisenhart have introduced a term to multiply by the weight matrix, and then to summate over all nodes before scaling. This appears to somewhat blur an individual node's contribution to the overall Harmony, but I have worked through the equations with algebraic variables when I found myself unable to understand their methods (see Appendix A for the full workings). Simplification of the smallest non-trivial network (a 3 node network) demonstrates they are actually using a generic formula thus:

$$H(N) = a_N \{W_{AN} + W_{BN} + \dots + W_{NM}\}$$, where there are M nodes and where $W_{NN} = 0$.

This leaves us with an underlying method which is completely un-sound. It is clear now that they are trying to replicate Smolensky's work in a homogeneous network utilising continuous activity levels, which as I demonstrated earlier is impossible to perform using the matrix multiplication techniques which Smolensky successfully employs. However, in their avoidance of this pit-fall they have chosen to ignore the similarity rating between nodes completely! Instead they merely compare the node's activity to the weights of the surrounding connections, a calculation hardly worthy of being called Harmony!

Implications

In the light of this discussion it is apparent that a simple measure of internal coherence for a homogeneous continuous network is desirable, and achievable. We therefore propose to use a simple correlation between each pair of nodes summated over the whole network. We express this as

$$H(r) = a_a \times \left(\frac{\sum_{ij} h(a_i,a_j)}{N} \right) \quad [i \neq j]$$

where a_i = the activity of node i & a_j is the activity of node j;

w_{ij} = the connection weight between nodes i & j,

N = the number of non-zero connections (w_{ij}),

a_a = the percentage of non-zero activity nodes,

& a_a, w_{ij}, a_i & a_j are bounded in the range (0,1).

We also define h(ai,aj) to take the forms of a subtraction difference. That is, the similarity between nodes i & j is |ai-aj|, and then h is the smaller of the similarity (sim_{ij}) / W_{ij} or W_{ij} / (sim_{ij}). Thus h(ai,aj) = sim_{ij} / W_{ij} ($sim_{ij} < W_{ij}$) or W_{ij} / sim_{ij} ($sim_{ij} > W_{ij}$) where sim_{IJ} = |ai-aj|. We also define $h(a_i,0)$ & $h(0,a_j)$ as 0.

However to do this loses some of the rigidity of Smolensky's work, since the continuous data is experimentally collected from subjects. We thus have no guarantee that there is any inherent relationship between the scale as used by the subjects, and thus the subtraction comparison is flawed. It will probably be forgivable to assume that the scale is (approximately) linear (or at least to assume that it is a more defensible approximation than a non-linear scale) but it is no longer the tight mathematics that Smolensky's equations were. Also because the mathematics are used on a homogeneous and not two-tiered network, the decomposability is not assured, especially under the circumstances of the continuous data used. Therefore it does not seem appropriate to lay claim to the title of Harmony for this measure of internal coherence, as it implies the strictness of Smolensky's work. Instead we shall choose to refer to our measure simply as the coherence of the network.

This then gives us the independent measure of expertise which we required to overcome the Overlay Paradigm. Indeed, for the prior example about the concept of Chair, although the Comfy Chair example is perhaps less prototypical we can still see that the representation has a certain amount of coherence to it, as did the Desk Chair (see Figures 4.4 & 4.5). Their respective coherence values using the formula just given are 0.572 (Comfy) & 0.672 (Desk). By comparison, if a person really did not understand what a chair was, only knowing it had a seat for sitting on, then the coherence drops to only 0.333 (connection strengths: Seat & Used for Sitting - 10; all other pairs, 0). As a Hypothesis for testing, I believe that the theoretical measure of expertise which is derived from the Symbolic Connectionism of Holyoak and Thagard and the work on coherence of Smolensky and also Briton & Eisenhart is empirically observable and able to discriminate between a group of novices and a group of experts within the chosen domain of study, archaeological science.

In summary, there is an imperative need to aid the teaching of Scientific Archaeology, caused mainly by the diversity in student background. We could possibly achieve an improvement in teaching by using a computer aided learning system (capable of making intelligent judgements as to the learner's level of expertise in the domain of Scientific Archaeology). To create such an intelligent tutor it is a pre-requisite that we can successfully identify novice and expert knowledge in that domain and successfully discriminate between them. We have seen that to truly claim that such a system accounts for the prior knowledge of the learners coming into the ITS environment we must address the problems of the Overlay Paradigm. By using a system which is not constrained by a set of expert rules or student mal-rules in making that decision but which can identify expert thinking, we can endeavour to tutor intelligently. We have shown that by looking at relational judgements it is indeed possible that a measure independent of the actual rules of a domain can be taken. If this measure does indeed discriminate effectively between experts and novices then we can avoid the problems of the Overlay Paradigm, and would have taken the first step on the road to an Intelligent Tutoring System for Archaeological Science. We must now proceed to confirm or refute this supposition empirically.

Chapter 5
Testing the Symbolic Network

Introduction

The purpose of the empirical work is to establish whether a symbolic connectionist system can be used to represent knowledge in the field of Scientific Archaeology and discriminate effectively between the knowledge of experts and novices using a coherence measure (and thus avoiding the problems inherent in the Overlay Paradigm). To test our belief that this will indeed be the case we have to elicit knowledge from both novices and experts within the domain and endeavour to see if our method is indeed capable of performing the distinction. We will also be comparing and contrasting its performance with a tried and tested way of interpreting the elicited knowledge as a further test of the validity of our methods.

To elicit the knowledge structure of the subjects in the format required for the creation of a symbolic connectionist network, a paired comparison procedure was chosen. The data collected can then be represented on a diagonal matrix showing the responses to all the possible concept combinations. Previous work has concentrated on the interpretation of these similarity matrices through the use of a multidimensional scaling technique (or MDS). These techniques compress the data into a two dimensional representation of the 'concept space', and have been used both to look at the internal cognitive structure of a particular concept area, and to compare the structure of novices and experts. This successful use of the similarity matrices to distinguish novices from experts will be used as a comparison for the constructed network data. The symbolic networks created should be able to differentiate between novices and experts through the internal structure of the 'concept space' as the MDS techniques do, but also through the identification of the *coherence* of the knowledge organisation. I shall explore these claims in the light of theory and past work to show how this new application is expected to function.

The relational data present in the similarity matrix can be represented visually as MDS plots through the application of a scaling formula. The data is scaled from a large number of dimensions (as many as 24 in this work) into only two dimensions for ease of visualisation. This will almost certainly lead to a large loss of data in the final plot unless the original data represented a two dimensional space. An example of the technique can illustrate how it can be used. Given a list of the major cities of the USA and a list of the air travel distances between all of these cities we may construct a similarity matrix, as shown in Figure 5.1. The matrix is diagonal since the distance from City A to City B is necessarily the same as the reverse journey, making the upper right hand side of the matrix redundant.

Figure 5.1 - A Similarity Matrix for the air distance between 12 major USA Cities

City in the USA	Atl	Bos	Cin	Col	Dal	Ind	Lit	Los	Mis	StL	Spo	Tpa
Atlanta	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Boston	1068	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Cincinatti	461	867	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Columbus	549	769	107	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Dallas	805	1819	943	1050	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Indianapolis	508	941	108	172	882	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Little Rock	505	1494	618	725	325	562	n/a	n/a	n/a	n/a	n/a	n/a
Los Angeles	2197	3052	2186	2245	1403	2080	1701	n/a	n/a	n/a	n/a	n/a
Memphis	366	1355	502	586	464	436	137	1831	n/a	n/a	n/a	n/a
St Louis	558	1178	338	409	645	234	353	1848	294	n/a	n/a	n/a
Spokane	2467	2747	2067	2131	1891	1959	1988	1227	2042	1820	n/a	n/a
Tampa	467	1379	928	985	1077	975	912	2480	779	1016	2821	n/a

By applying the MDS technique, we are given a set of co-ordinates to place each of the list members (Cities) in two dimensional space. By charting all of the list members together on a plot as in Figure 5.2 (the MDS plot) we can investigate their inter-relationships.

Figure 5.2 - The MDS Plot for the 12 Major US Cities.

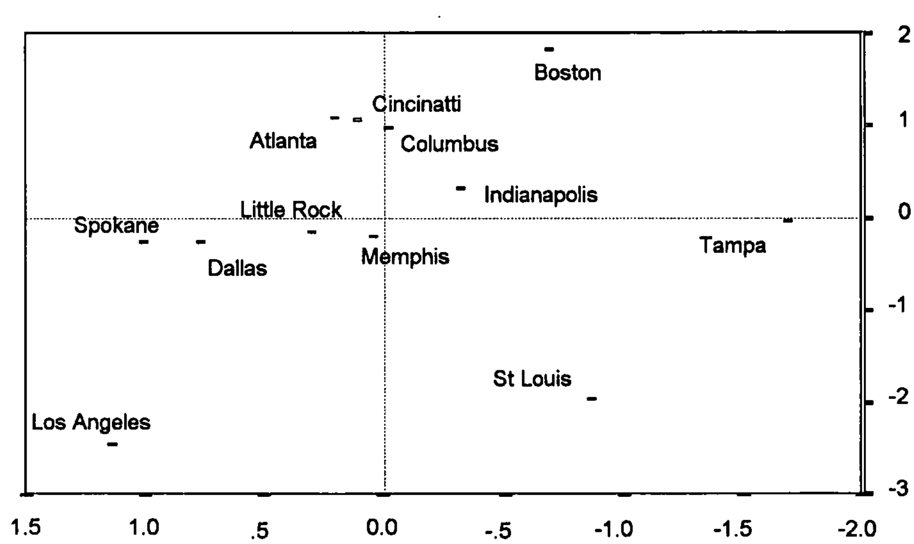
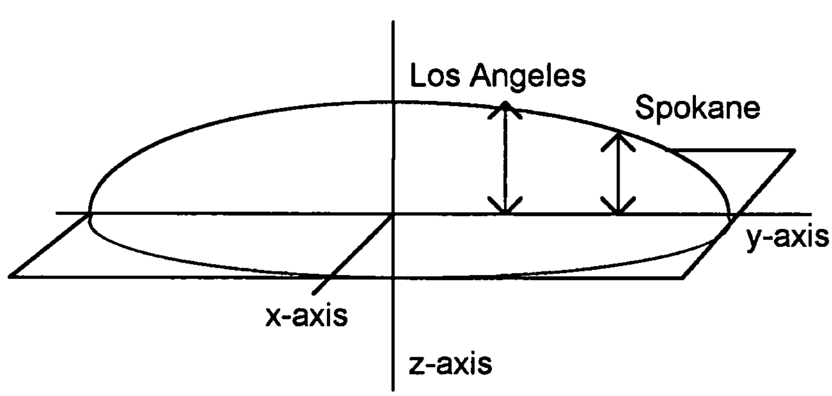


Figure 5.3 - A US Map for comparison with the MDS plot.

We can attempt to discern any common features by attempting to label one or more axes on the plot. In this example that is an easy task, as the plot shows the geographical positions of the Cities as if on a map of the USA. By orienting the plot we can then resolve the x and y axes as longitudes and latitudes or Northings and Eastings. In this case, as Spokane lies due north of Los Angeles we can find that our North South axis and then rotate the MDS plot to line this axis vertically up the page (as I have done). By then lining up cities on the map (such as L.A., Spokane & Boston) we can see how the other cities almost match up. This gives us a very useful tool for making a qualitative judgement about the content of the knowledge investigated. We can quickly see how close any two cities are by air. What we cannot see is any data lost in the scaling process. In this example, the air distances over such a large area of the globe are going to be heavily dependant upon the curvature of the Earth. We are thus missing out on a third dimension, z, being the distance each city is displaced from an imaginary section cut through the Earth as demonstrated in Figure 5.3.

Figure 5.3 - A section through the Earth, showing the effect of Curvature upon the USA.



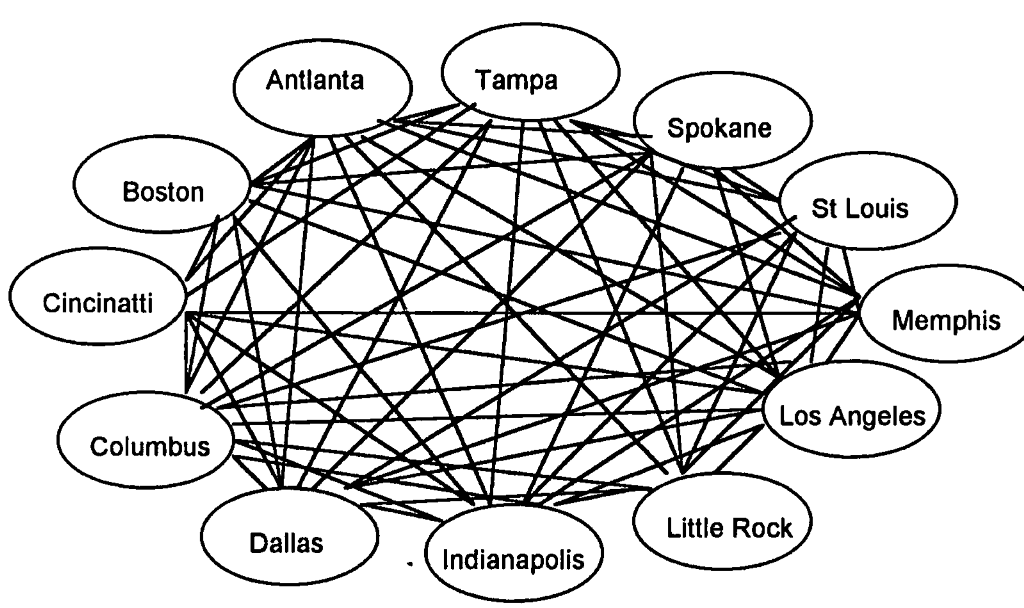
This data has been lost in the scaling process, with no axis left in the MDS plot which could show us that data. Indeed, the reason that the MDS plot and the map of the USA are not in exact alignment is because the map is drawn to show the effects of the Earth's curvature (as is indicated by the latitude and longitude lines) and has not been 'flattened' out. As you can imagine, the loss of data is therefore entirely dependent upon the number of different inter-relationships which exist between the points in the data set. In this case only one dimension has been lost and so the majority of cities do still appear in their correct location (allowing for the variation in curvature), although some, notably St Louis & Atlanta, are harder to fit into place. In some instances this loss of information can be quite acceptable: imagine having to read a road map rendered to show the Earth's curvature as well as the M25! MDS plots are thus good for identifying the two major relationships in a visualisable way, which will allow us to make non-exact judgements about the data.

A previous experimental use of the MDS technique should also be considered, to show how this scaling of a physical relationship can be applied to scaling a belief relationship, and then to the creation of a 'cognitive space' which represents a subject's organisation of knowledge. Such a study was performed in 1988 by Stevenson, Manktelow & Howard. Concepts important to the area of investigation are taken, and all possible pairwise combinations of them are presented to the subjects in a booklet or questionnaire. In this case, the field of study was computer programming, specifically procedures in PASCAL. The key words used in the questionnaire were single words representing important concepts to the topic drawn from videos of first year lectures. The subjects were asked to rate the similarity of the paired concepts on a scale. The scale used was 1 - 7, where a rating of 1 would indicate that the subject knew of no relationship between the pair of concepts, whilst a 7 would indicate that they believed the two concepts to be intimately related. The rest of the range was to represent all intermediate strengths of relationships

In the study, 8 Novices (1st Year Undergraduates) and 2 Experts (Computer Science Lecturers) took part. The similarity matrices produced for the subjects were interpreted using the MDS technique to produce a plot of the 'concept space' of the key words used. This showed a qualitative difference in the novice and expert responses. Both had two prominent groupings of concepts in their concept spaces, but the grouped concepts were different for novices and experts. The technique was also employed to interpret the similarity of the novices' and experts' responses by interpreting the different use of the two dimensions by the subjects. In this plot the novices were seen to place much more weight in one of the identified dimensions than the other, whilst the experts were much more consistent in placing even weight on both dimensions. Each dimension was also identified in terms of some higher order strategy in programming (the exact nature of which is not relevant here).

However, the data that is needed in multi-dimensional scaling can also be used to create a symbolic connectionist network. To do so, we firstly take the same similarity matrix we had before. Each concept is then the label for a node in a network, and each and every node is connected to each and every other node. The connections between nodes are also labelled, with the strength of relationship as shown in the similarity matrix. For our previous example of US cities then, a symbolic connectionist network would look like Figure 5.4.

Figure 5.4 - A Symbolic Connectionist Network for the 12 US Cities.



As the data in the symbolic network is unchanged from its state in the similarity matrix, it is possible to extract it again to create the MDS plots: the connection strengths simply being re-interpreted as the original data. This is an advantage over the MDS technique as the raw data can be retrieved from the network at any time, whilst the MDS plot is a static and irreversible process - you could not retrieve the initial data simply from the MDS plot itself. However, the advantage of the symbolic network is not in mimicking the MDS technique, but in its ability to yield a quantitative description of the novice expert difference through a measure of the coherence of the network (see also Chapter 4).

As we saw in the last chapter, the idea of coherence within a network is derived from the work of Smolensky (1986) on Harmony. If a person holds inconsistent beliefs about a concept area, such as this paradoxical set of logical statements, then we can say that their beliefs are incoherent.

$$A = B, \qquad B = C, \qquad C \neq A.$$

Within a connectionist structure it is possible to make such comparisons between concepts throughout the structure, and to ascribe a quantitative value to each component part. The sum of these parts is the Harmony of the network (Smolensky). This measure of harmony is taken from 0 to 1, where 0 is a completely incoherent set of beliefs and 1 is completely coherent. Tied to this measure is a second, that of the speed with which the constructed network reaches a stable state. Smolensky's work though was concerned with pure connectionist networks and with two-layer (Input-Output) networks. To make the work suitable for our purposes requires a change to the calculation of the coherence, and with deference to Smolensky I chose not to use the label of Harmony for our measure. This is because Smolensky's measure is very rigorous as a mathematical entity, and this cannot be guaranteed in the same way for a symbolic network. Previous work has been carried out though in making this step

(Briton and Eisenhart, 1993) and a measure of network coherence for a symbolic connectionist network has been produced. Briton and Eisenhart's research indicated that both the coherence of a network and also its settling speed are discriminators of expertise.

In the work of Briton and Eisenhart, subjects were presented with paired comparison tasks and again asked to rate the material for similarity. The scale they used was from 0 - 7, and they too created a similarity matrix for each subject. They then scaled their data to fit boundary values of [-1,+1] after the work of Kintsch (as discussed in Chapter 4) and followed his method for the propagation of activity through the network to a final stable state. That is,

$$A_{N(T+1)} = \left(A_{N(T)} + \Sigma \left(A_i * C_{Ni}\right) \right) \quad / \quad \Sigma A_i \qquad \text{(Eqn. 1)}$$

where $A_{N(T)}$ = the activity of a node N at a time T,
A_i = the activity of a node i,
& C_{Ni} = the connection strength between nodes N and i,

and the sum of all activities for all nodes is scaled to total 1 each cycle.

Briton & Eisenhart expressed their equation for Harmony much as Smolensky did, putting forward their own modified version. Essentially they barely altered Smolensky's work, but because of the differences in the networks each used they ended up with what I earlier argued to be a poor measure of coherence. Each node's activity is correlated with the connection strengths between itself and all the connected nodes, but no notice is given to the activity of the nodes to which it is connected. Thus:

$$H(N) = \Sigma_{N=1\text{-}M} \left(a_N \{W_{AN} + W_{BN} + \ldots + W_{NM}\}\right), \text{ where there are M nodes and where } W_{NN} = 0.$$

To facilitate the research, a computer program, called NETG, was written to construct and run symbolic networks and to calculate their coherence. The program was written in PASCAL on a Hewlet Packard Apollo Workstation. The core program was created by Mr Simon Lawrence, and was then customised and extended by myself. The program in its current form takes the concepts as an input and then prompts for the relationship between each pair. This information may be stored to a file and retrieved at a later date. The program also allows for the selection of different methods of operation - either allowing the network to be run using the Kintsch / McClelland & Rumelhart formula (mentioned previously and labelled as Eqn.1) for propagation or one based on the formula used by Holyoak&Thagard which is:

$$A_{N(T+1)} = (A_{N(T)} * (1\text{-}d) + (\Sigma (A_i * C_{Ni})) * (1\text{-}A_{N(T)})$$

In Holyoak and Thagard's formula the differences to the Kintsch / McClelland & Rumelhart formula are minor theoretical ideas. Instead of controlling the size of activity in each node through scaling of the network to a constant activity, the node's activity at a time T+1 is equal to its activity at a time T decreased by a decay constant d. The contribution of the other nodes is also scaled by the previous activity of the node N, thus making it harder for a node which has already achieved a high activity to then lose that activation.

The NETG programme also allows the selection of different ways to calculate the network's coherence, and other functions such as automatic reduction of network size and 'clamping' of concepts may be selected. The network may then be run, and an output produced detailing the node's final activity states and the network's settling rate and coherence. The network may either be run automatically through to a stable state, or it may be manually instructed to perform any integer number of cycles of the network.

Through experiments we intend to collect data for evaluation by the standard method of MDS and through my own network program (NETG) which can assess the coherence and settling speed of the symbolic network as well as yield the relational data inherent in the MDS technique. It is my hypothesis that my program will be an effective novice/expert discriminator.

Experiment 1

The first data collection was intended to investigate the feasibility of the NETG program. A small group of second and third year Archaeology undergraduates were given a simple questionnaire on mitochondrial DNA (or mtDNA) theory. The aim of this study was to elicit the knowledge which novices and experts have of mtDNA theory and to identify it using both the MDS technique and the NETG programme. Then the two methods can be compared.

Method

Subjects:

Sixteen subjects were used. One was an expert (a lecturer in scientific archaeology) and 15 were novices (Archaeology undergraduates in their second or third year of study).

Materials:

11 concepts (single words or two word phrases) were selected from the lectures given on the subject as contributing to the overall concept of mtDNA theory; e.g. mtDNA, Temporal Mutation, 'Eve' and Africa. Each concept was paired with every other concept to create 55 pairs. These were presented in a random order to each subject in a single questionnaire. The questionnaire (an example of which is in

Appendix B) firstly gave a small example of the format of the questionnaire. The subjects were requested to indicate on a scale of 0 - 10 how closely related they felt each pair of concepts to be, by circling the appropriate number on the scale provided. A typical line of the questionnaire was like this:

Africa & mtDNA - 0 1 2 3 4 5 6 7 8 9 10.

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. This task was given to one lecture group and their lecturer in the second half of a normal teaching session. The resulting similarity matrix for each subject was then analysed using the multidimensional scaling technique and also by constructing a symbolic connectionist network. In the network, each concept becomes a node, and each link between a pair of nodes has a strength attributed to it equal to the subject's rating of the pair in the questionnaire. An initial arbitrary 'activity' is assigned to the nodes and allowed to spread though the network according to a formula developed from that used by Holyoak and Thagard (1989) and described in the introduction of this chapter. When the network reaches a stable state, the coherence and settling rates are produced by the programme.

Results

First, we examined the raw data of the similarity matrix. The distribution of strength of relationships as shown in Figure 5.5 shows that the students were more likely to stick to the middle of the range. The expert by comparison used the extreme values (0 & 10) much more frequently. Both groups did however make full use of the range, but where the novices scored greater than 7 it was only for very obvious archaeological associations. Additionally, the expert's distribution was equally high across the range of concepts, whereas the learners tended to concentrate their high ratings over a small group of concepts.

Multi-Dimensional Scaling:

A two dimensional scaling solution was produced for the 11 concepts in the questionnaires. As Figures 5.6 & 5.7 show the novices possess a very different organisation to the expert in their knowledge. The concepts for novices are distributed fairly evenly around the edges of the plots (with some small localised grouping) while the expert plot has a central left grouping and two other bunches; one in the middle bottom and one to the right.

Figure 5.5 - Distributions of Strengths of Relationships in the Questionnaire, by group.

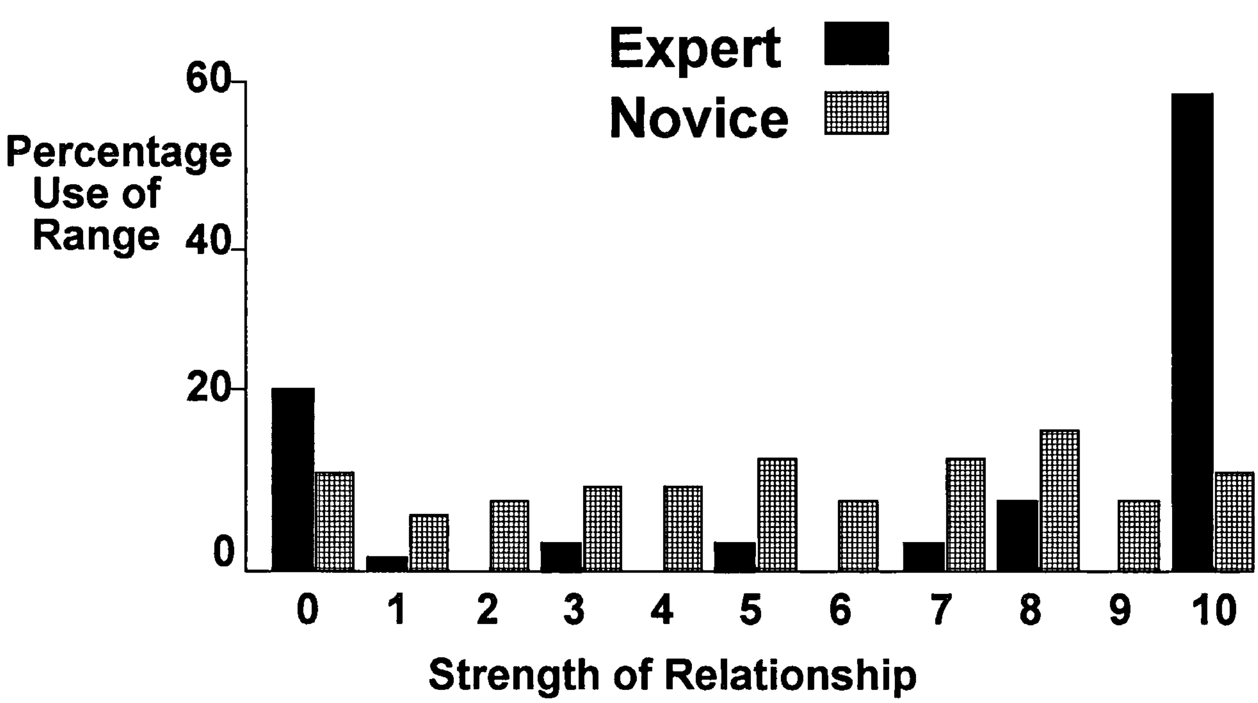


Figure 5.6 - MDS Plot of the 15 Novices. A 'concept space' for mtDNA Theory.

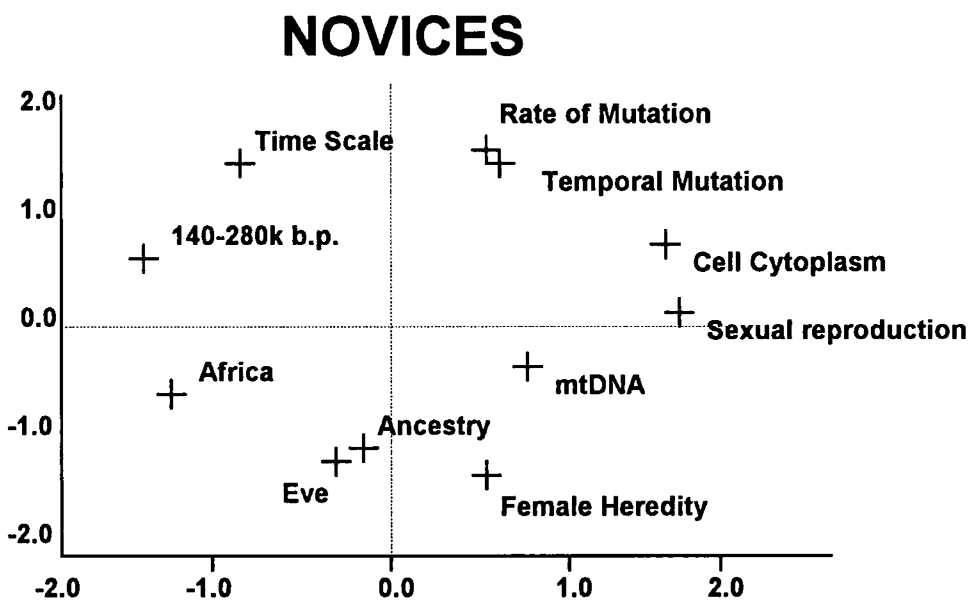
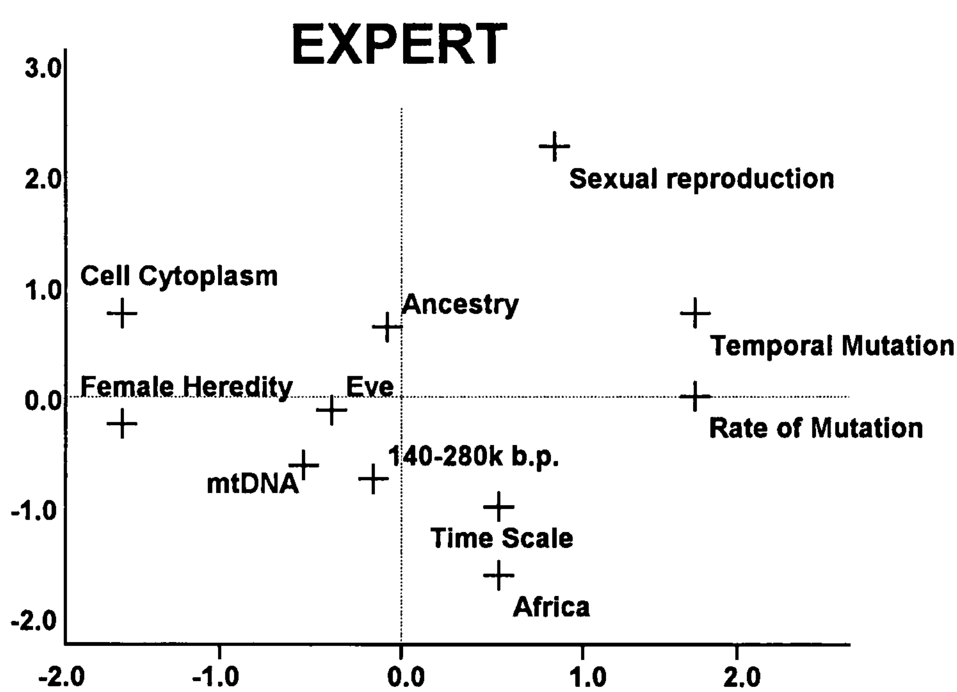


Figure 5.7 - MDS Plot of the Expert. A 'concept space' for mtDNA Theory.



Network Data:

The connection strengths which were elicited from the subjects as paired concept similarity ratings were then used to construct a network as described in the earlier chapters. For each of these networks a calculation of the network's coherence was then made using the formula which I constructed and which was detailed at the end of Chapter 4. The calculation of coherence for the subjects in the novice group yields values between .384 and .799 with an average of .629, while for the expert this value was .889 as shown in Figure 5.8. We also measured the number of cycles needed for the networks of the novices and of the expert to reach stable states (see Figure 5.9). 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.

Figure 5.8 - - Box and Whisker Plot of Coherence, by group.

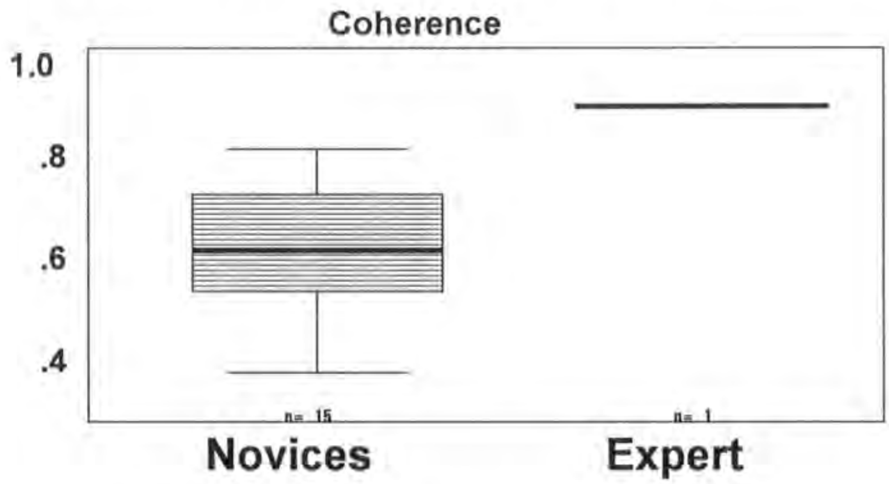
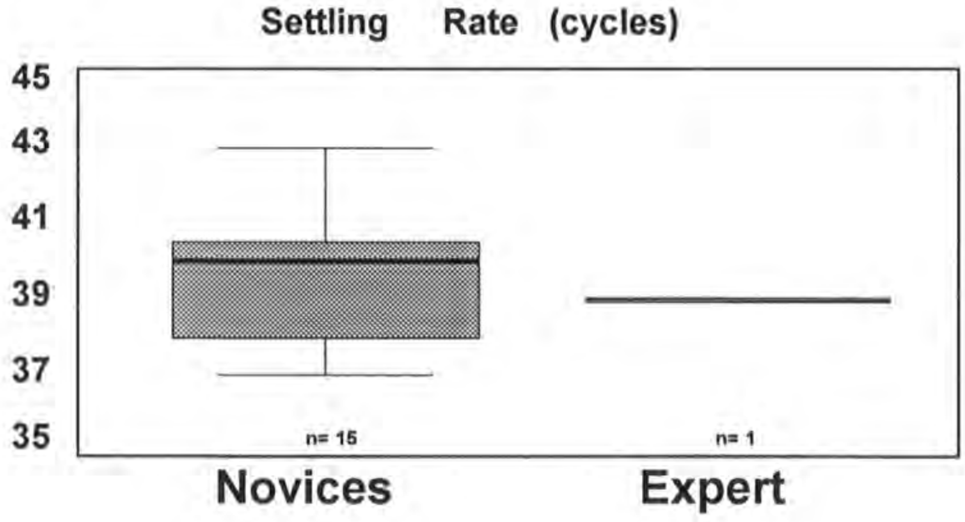


Figure 5.9 - Box and Whisker Plot of Settling Rate, by group.



Comparisons:

The scaled data of the MDS plots can be compared directly with the data derived from the NETG program. Where the MDS plot shows a scaled representation of 'cognitive distance' in two dimensions as pure physical distance between any of the features, a comparable measure can be drawn from the NETG programme by *clamping* a feature (Holyoak and Thagard, 1986) and then observing the settled activities of the other concepts (Table 5.1). By clamping, a single concept has its activity set to the maximum value, and then the activity within the network is allowed to propagate through it as before. However, on each subsequent cycle, the 'clamped' concept retains its maximum activity. Thus, when a stable state is reached, it will reflect a vastly increased influence from the clamped concept. It therefore makes the final activities for the network data 'centred' upon that particular concept.

Table 5.1 - A comparison of the NETG and MDS data for the concepts relative to mtDNA (Expert's data).

Network Feature	Network Activity	MDS Distance
Ancestry	0.517	.39
Eve	0.510	.14
140-280K b.p.	0.507	.16
Time Scale	0.490	.58
Rate of Mutation	0.458	.86
Temporal Mutation	0.434	.91
Female Heredity	0.432	.49
Africa	0.406	.50
Cell Cytoplasm	0.406	.65
Sexual Reproduction	0.222	.97

In table 5.1 the 10 concepts are listed in column 1, and the second and third columns contain the original data derived from the technique. In column 2 we see the activities of the other 10 concepts after the network has settled with the 11th concept, mtDNA clamped. This is a measure of network activity, and higher activity represents greater cognitive similarity and hence smaller cognitive distance. The third column shows us the distance between mtDNA and the other 10 concepts on the MDS plot; the cognitive distance. To compare the two different techniques we perform a correlation calculation between them[1]. This gives us a correlation of -0.6793, which with 9 degrees of freedom is significant at less than the .05 level.

Table 5.2 - A comparison of the NETG and MDS data for the concepts to the overall concept of mtDNA Theory (Expert's data).

Network Feature	Network Activity	MDS Distance
Eve	0.112	0.426
140-280K b.p.	0.111	0.681
mtDNA	0.109	0.783
Time Scale	0.105	1.159
Ancestry	0.104	0.624
Rate of Mutation	0.092	1.427
Temporal Mutation	0.086	1.568
Female Heredity	0.083	1.681
Africa	0.078	1.823
Cell Cytoplasm	0.074	1.915
Sexual Reproduction	0.036	2.274

[1] - This and all subsequent correlations use the Pearson correlation.

Similarly, we can compare the unclamped data from the NETG program to the distance of the features in the MDS plots from the 'central point' of the plot, showing the 'cognitive distance' in both cases of the features from the overall concept. This can be seen in table 5.2, where again the features are listed in the first column and the next two columns show the same measures as before. This time though, the network activity is for an even distribution of activity with no concepts clamped and the distances measured off of the MDS plot are from the central point (0,0), not from any of the other concepts. Again we can correlate the two techniques, and this gives us a correlation value of -0.9909 which for 14 degrees of freedom is significant at less than the .001 level.

Table 5.3 - A comparison of the NETG and MDS data for the concepts to the overall concept of mtDNA Theory (Mean Novice data).

Network Feature	Network Activity	MDS Distance
Eve	0.102	1.12
140-280K b.p.	0.063	2.05
mtDNA	0.101	1.15
Time Scale	0.085	1.39
Ancestry	0.103	0.81
Rate of Mutation	0.094	0.97
Temporal Mutation	0.090	1.19
Female Heredity	0.104	0.70
Africa	0.092	1.07
Cell Cytoplasm	0.076	1.19
Sexual Reproduction	0.078	1.52

We can also perform these correlations for all of the novice subjects. The correlation is performed on the mean of the MDS & NETG data for for the 15 novice subjects as shown in Table 5.3. This yields a correlation significant to the 0.01 level of -0.8759.

Discussion

In the Multi-Dimensional Scaling data, the novice plot (Figure 5.6) can easily be resolved into a higher order conceptual grouping. The concepts move from the left hand extreme of the plot being Archaeologically related through to the right hand side of the plot and being Biologically related. A second axis running from top to bottom of the plot can also be resolved as running from chronological events through to non-chronological events. This axis however is not as obvious a distinction as is the case of the first.

For the expert plot shown in Figure 5.7 the organisation is very different and was actually interpreted by the expert himself. This is because the work of Chi, Feltovich & Glaser (1981) clearly shows us that it requires an expert in any particular field to make sense of the underlying principles which dictate an expert's organisation of knowledge. For this plot then, there are three distinct groupings of concepts. The main (and central) group is the set of concepts which relate to the process of mtDNA inheritance {'Eve', mtDNA, Ancestry, Female heredity & Cell Cytoplasm}. A second and also central grouping is related to the Archaeology which accompanies the theory {Africa, Time Scale & 140-280 ka b.p.}. The final grouping relates to the underlying genetics of the theory {Temporal Mutation & Rate of Mutation}, though it is placed further away from the central groupings. The final concept, Sexual Reproduction, is viewed as inconsequential in mtDNA dating by the expert, and so is left isolated at a greater distance from the other groupings and from the centre of the plot. The MDS technique thus clearly shows a difference between the organisation of concepts in novices and the expert.

The network data yielded two measures of comparison between expert and novices. The coherence value derived from their networks shows quite clearly that the expert's thinking is logically self consistent to a greater extent than the novices as seen in Figure 5.8, and thus is a good measure of expertise. The novice group also showed a good distribution of coherence values indicating (possibly) a spread in knowledge of the subject ranging from similar levels of 'understanding' to the expert through to values only half that of the expert. The settling rate of the network was the other measure, and in this instance proved to be less effective as a discriminator. Although the expert's rate was below the mean value for the novices, we can see in Figure 5.9 that the spread of novice settling rates surrounds the expert's value.

The attempt to compare the two techniques empirically is shown in the correlation test applied between the two sets of data which the techniques yielded. This clearly shows in the case of unclamped network data and MDS distances from centre of a distribution that for the group of novices as a whole there is a significant correlation between the two techniques. The expert data also yields a significant correlation. In the light of these correlations I am confident to state that the workings of the MDS scaling technique operate in the same fashion as the NETG program. However, as indicated earlier in the chapter, because of the nature of scaling to two dimensions there is necessarily a loss of information where more than two discrete dimensions are under examination. It appears, however, that in the case of this first experiment where the data size of the knowledge domain (11 nodes only) is small that we can easily resolve into two prominent axes with little loss of information.

Experiment 2

Following on from the first experiment, we undertook a larger collection of data in an attempt to provide more valid conclusions. To elicit and identify the knowledge of the subjects (Archaeologists), we again used the statistical scaling technique of Multi-Dimensional Scaling along side the network programme NETG. In this collection of data however, the length of the questionnaires was increased and the area of interrogation widened to five topics; all concerned with radiometric dating techniques in archaeology. It was also our intention to gain an interpretation of conceptual understanding both within and between the five areas of radiometric dating used :- General Dating Techniques, Uranium Series Dating, Potassium-Argon Dating, Advanced Radiocarbon Dating & Normal Radiocarbon Dating (as described in more detail in Chapter 4).

Method

Subjects:

36 subjects were used. 3 were experts (lecturers in scientific archaeology) and 33 were novices (Students at the end of their first year of studying Archaeology). The expert subjects were all university lecturers in England. One was from Durham University; the other two were taken from a list produced by the expert from the first experiment indicating other lecturers in the country with the appropriate expertise in scientific dating. Many were approached by telephone, of whom 5 agreed to help out. They were all sent out questionnaires anonymously - of these, only two returned , all of whom were suggested by the expert from the first study as knowledgeable in the appropriate areas of Archaeology.

Table 5.4 - The five topic areas and their associated concepts.

<table>
<tr><th>Type</th><th>Topic</th><th colspan="2">Concepts</th></tr>
<tr><td>A</td><td>General
Radiometric Terms
(GEN)</td><td>Zeroing Event
Dating Range
High Energy Mass Spectroscopy
Formulae
Operating Limits on Range</td><td>Clock Mechanism
Radioactive Nuclide
Constant Decay Rate
Archaeological Material</td></tr>
<tr><td>B</td><td>Uranium Series
Dating
(USD)</td><td>Uranium Series Dating
Sedimentary Rocks
Uranium Present in Rock
Thousand Trillion Year Half Life
U Included from Groundwater
Associated Calcite Deposit</td><td>Uranium 234
Several Daughters
Daughter (Th) Insoluble
Stalactite Formation
350 ka</td></tr>
<tr><td>C</td><td>Potassium Argon
Dating
(KAr)</td><td>Potassium Argon Dating
Potassium Present in Rock
Thousand Million Year Half Life
Drives Argon from Rock
Feldspar Crystals in Rock</td><td>Potassium 40
Two Daughters
Volcanic Events
400 ka or older</td></tr>
<tr><td>D</td><td>Advanced
Radiocarbon Dating
(ARC)</td><td>Isolate Find from Modern C14
Count Atoms instead of Activity
Range = 50 - 80 ka
High Energy Mass Spectroscopy
Range = 30-40 ka
Smaller Sample Required</td><td>Dendrochronology
Calibration
Accelerator Dating
Accelerator Mass
Spectroscopy.
The Decay Method</td></tr>
<tr><td>E</td><td>Normal
Radiocarbon Dating
(NRC)</td><td>Death of Living Organism
C-14 Produced in Atmosphere
C-14 Fixed in Organic Matter
C-14 Reservoir in Oceans
Libby's Half Life, 5568 Years
Cessation of CO_2 Exchange</td><td>Carbon Cycle
C-14
Organic Matter
Fractionation of Carbon
Photosynthetic Pathway
Carbon 14 Dating</td></tr>
</table>

Materials:

We selected 5 areas of knowledge from within the field of Radiometric Dating Techniques. For each topic, words or short sentences were selected that described individual concepts contributing to the overall idea of the area; e.g. for Radiocarbon Dating, Carbon 14 and Libby's Half-life of 5568 Years were two of the concepts chosen. Table 5.4 show the full set of chosen concepts. These words were taken from lectures previously given on the topics to the students and from the recommended course text. Ten questionnaires were then produced, each one pairing one knowledge topic with another. For each of these questionnaires, each concept was paired with every other concept from both topics. Table 5.5 shows how the different topics (column 1) had certain numbers of concepts (column 2). To make a questionnaire two topics were combined, and since the combination is commutative the mirror questionnaire is not necessary. Thus, a questionnaire combining topics C & D combines 11 concepts from C with 9 concepts from D (as seen in columns 1 and 2). This combined questionnaire, CD, is described at the intersection of row 4 and column 5, where questionnaire CD has 20 features equalling 190 concept pairs. The reverse combination CD is unnecessary, and so row 5 column 4 is left with n/a for both number of features and number of questions. The pairs were presented in a single booklet and were in a different random order for each subject. The undergraduates each received one questionnaire type (of the 10 possible) selected at random. Each questionnaire type was completed by 2, 3 or 4 subjects. One lecturer (from Durham) completed all 10 of the different questionnaires. This however proved the impossibility of collecting such a large bulk of data from busy lecturers, and so a single questionnaire type was selected to be given to other experts. From a preliminary analysis of the novice data it was clear that many of the concept pairs were completely opaque to them, and so the 'easiest' questionnaire for novices was selected for the experts to facilitate comparison with the novice data. This was the GEN & NRC questionnaire.

Table 5.5 - The make-up of the ten different questionnaire types.

Topic	Number of Concepts	GEN		USD		KAr		ARC		NRC	
		f	q	f	q	f	q	f	q	f	q
GEN	9	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
USD	11	20	190	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
KAr	9	18	153	20	190	n/a	n/a	n/a	n/a	n/a	n/a
ARC	11	20	190	22	231	20	190	n/a	n/a	n/a	n/a
NRC	12	21	210	23	253	21	210	23	253	n/a	n/a

f= Number of Features, q= Number of Questions in Questionnaire.

Design and Procedure:

Subjects were presented with all possible word pairs for their particular questionnaire (ranging from 153 to 253 pairs as indicated above). They were presented in the form of a booklet and were given to the novice group (students) as a part of a normally time-tabled lecture. The experts were sent their questionnaire by post, with a covering letter asking for their help in completing my experiment (except for the one in Durham). They were additionally encouraged by Dr Stevenson beforehand by telephone to lend their time to this work. Within the booklet each subject was asked to rate the similarity of each concept pair on a scale from 0 to 10, where 0 indicated dissimilar and 10 indicated very similar. A 'don't know' category was introduced in this study to help students who were unfamiliar with some of the technical terminology. This change came from a couple of sessions where I gave different preliminary versions of the new questionnaires to undergraduates for completion. After they had attempted them I asked for their reaction to the style and presentation of the questionnaire, and some changes were instituted because of the feedback. Specifically the 'don't know' category was thought to be better than having to say a zero relationship when the subjects knew nothing about the concept or concepts involved. The zero relationship was kept for those cases where the subjects were familiar with both concepts but thought they were unrelated. Also a breaking up of the questionnaires visually into groups of questions for ease of completion was requested. Those taking part in these informal session did not take part in the experiment proper.

The resulting similarity matrix for each subject was then analysed using both the multidimensional scaling technique and by constructing symbolic connectionist networks as before.

Results

First, we examined the raw data of the similarity matrix to see whether the subjects had made answers to the majority of the paired feature words. We found that on average 60% of the responses made in the novice (undergraduate) data were 'don't knows', though there were tremendous variations across individuals and across questionnaire types. These can be seen in table 5.6 to range from 23% - 99% for individuals and from an average of 37% (on what had been judged before the data gathering to be the *easiest* questionnaire for the undergraduates to answer) up to 88% on a questionnaire looking at much less familiar terms to them (judging from their course lectures). Overall, the novices answered 'Don't Know' to 56% of the concept pairs. By comparison the experts produced no 'don't knows' at all. It was on the basis of this analysis that the GEN & NRC questionnaire was chosen to be sent to the experts outside of Durham.

Multi-Dimensional Scaling:

A two dimensional scaling solution was produced for all of the concepts in every completed questionnaire. These could be compared only with the same questionnaire type which had been completed by other subjects. By comparing MDS

Table 5.6 - Percentage of 'Don't Knows' scored by the Novice Subjects: type & group.

Questionnaire Type	Novices		
	Subject	% DK	Means
GEN & USD	1 2 3 4	53 47 82 58	60
GEN & KAr	5 6 7	44 56 61	54
GEN & ARC	8 9 10	43 47 65	52
GEN & NRC	11 12 13 14	31 24 44 50	37
USD & KAr	15 16 17	80 32 33	48
USD & ARC	18 19 20	72 82 97	84
USD & NRC	21 22 23 24	23 47 91 97	65
KAr & ARC	25 26 27	86 63 56	68
KAr & NRC	28 29	68 65	67
ARC & NRC	30 31	54 25	40
MEAN		56	

Legend: The Questionnaire Type is the two Archaeological Scientific Dating Techniques which are combined to make it. For example, GEN & USD implies (using the notation from Table 5.4) that the questionnaire is a combination of General Dating Concepts and Concepts relating to Uranium Series Dating.

plots for the Expert in the questionnaire types with the group plots for the novices it can be seen that the organisation of the same concepts is qualitatively different between the novice and expert groups. The expert subject from Durham completed an additional set of 5 individual questionnaires for the five scientific dating topics which in combination produced the main questionnaires. This study was considered having observed how the combinations of dating domains affected the relationships within the domains themselves. This data gives us valuable information as to how the expert has grouped the concepts without combination with another topic. It is presented and discussed first, as it allows us to make sense of the combined questionnaires.

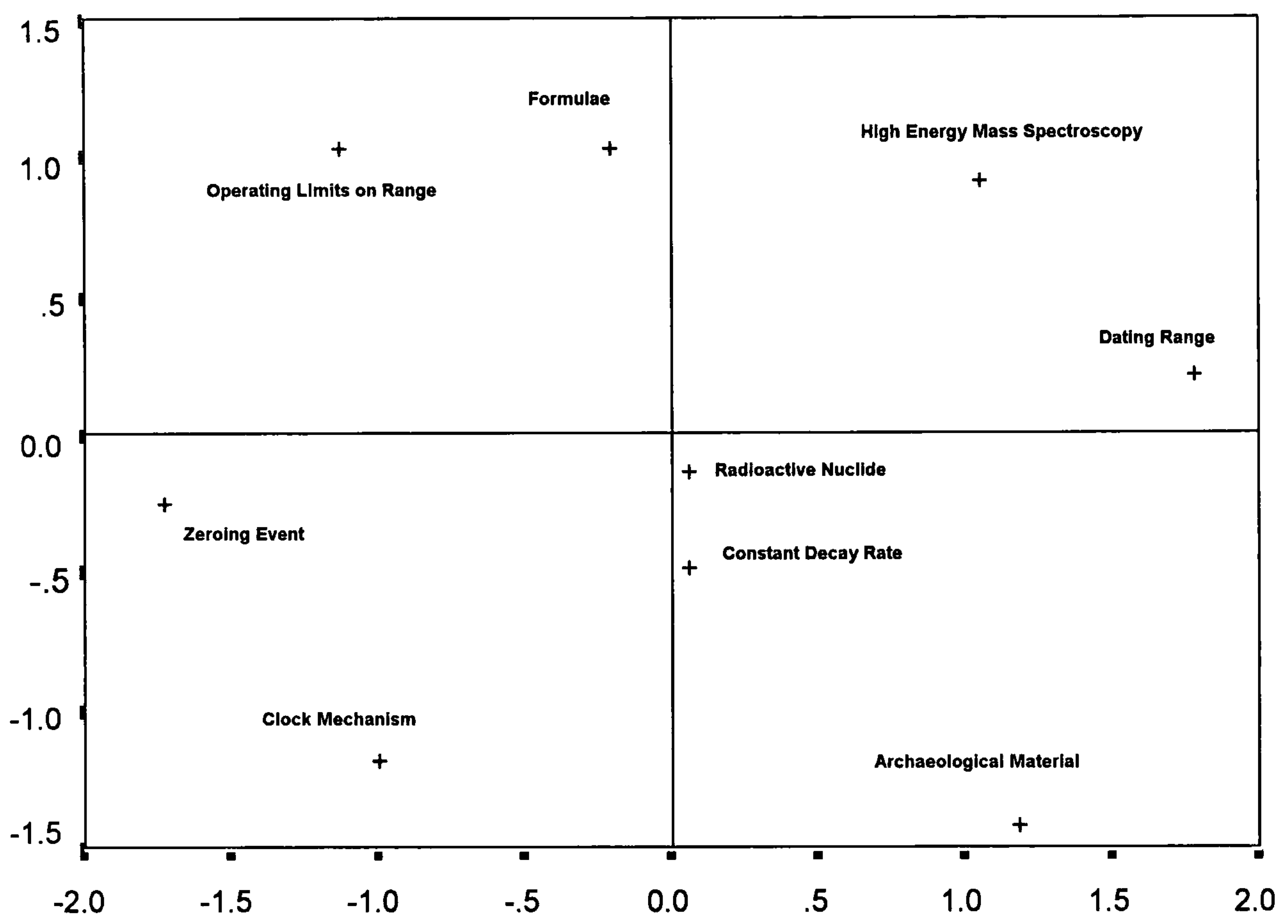
For the one questionnaire, GEN & NRC we have two groups - three experts & four novices who completed it. We examine these with the hindsight of the individual questionnaires, and in the hope of observing a difference between the expert and novice groups. For the other nine questionnaire types we have only one expert and a group of novices. This condition are less valid since there is no expert group and so are placed in Appendix C with a brief discussion of the concept groupings. They are interesting in support of the experiment, but not of sufficient weight to place in the main body of this text.

Individual Questionnaires

The Durham expert (who completed all ten of the combined questionnaires) was also asked to complete five smaller questionnaires looking at the five parts of the composite questionnaires individually; thus GEN, USD, KAr, NRC & ARC. These are discussed here.

GEN

Figure 5.10 - MDS Plot of the 'Concept Space' for the Expert answers to Questionnaire type GEN: General Radiometric Theory.



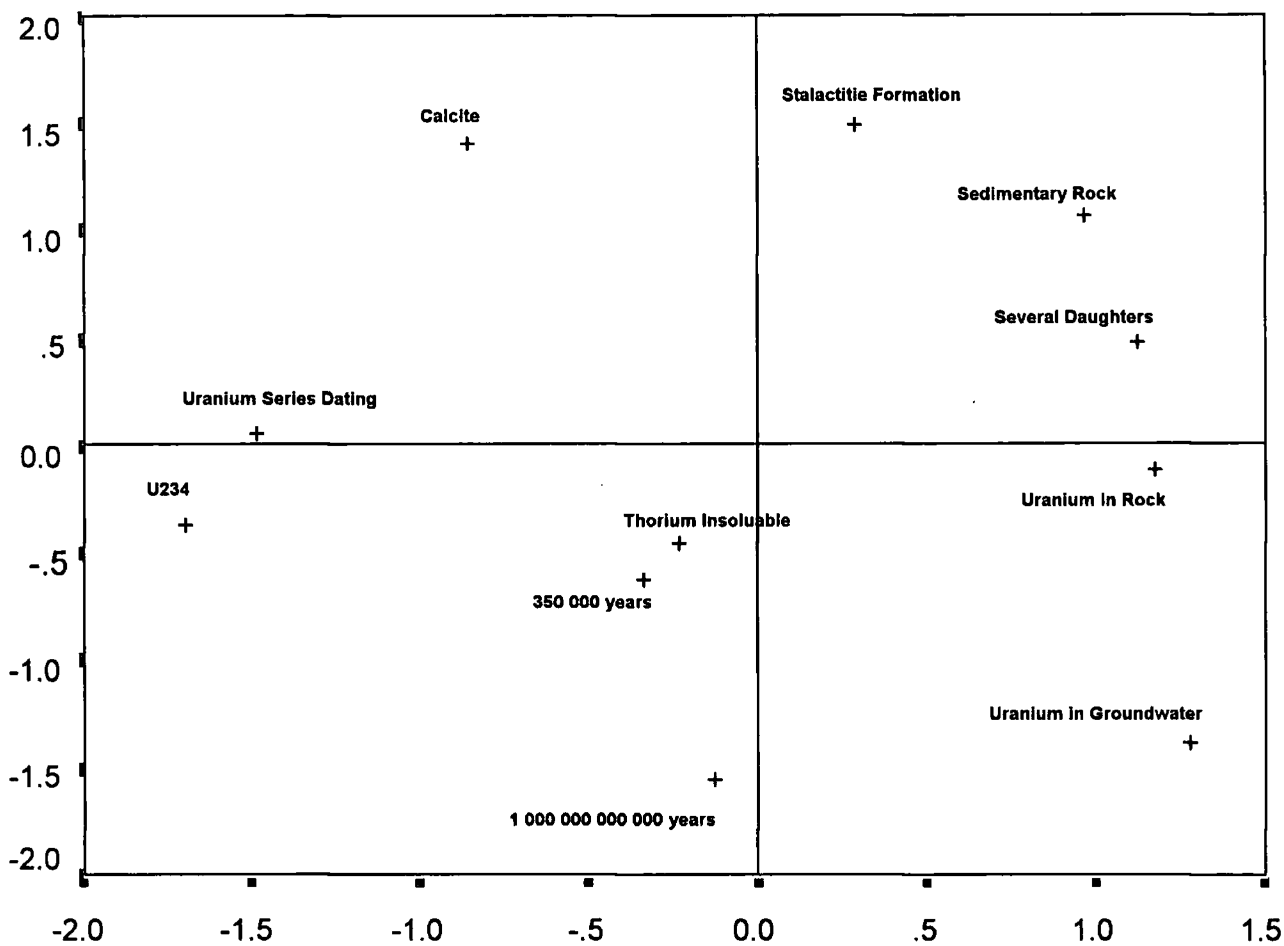
In this plot, according to the expert, the concepts have become distributed and grouped according to two criteria. The first grouping is from the centre bottom right to the bottom left of concepts which defines the general mechanisms of radiometric dating techniques. There always has to be some kind of clock mechanism for a radiometric technique to exist, be it charge building up (as in Thermo-Luminescence) or the rate of decay of a large number of radioactive particles (as in nuclear methods such as C14 & Uranium Series Dating). Likewise there must always follow some zeroing event which starts the clock 'running'. In all of the cases presented, the mechanism of the techniques is nuclear, and so there must necessarily be in each instance a radioactive nuclide whose decay is measured. Finally, it must always be the case that the decay rate is assumed to be constant with regards to each technique to allow for a time scale to be associated to the measured decay.

The other grouping which the expert makes is of the general application of radiometric dating technique in Archaeology. This revolves entirely around the ability of the technique to yield a date to the Archaeologist, and comprises of the Archaeological Material, the Dating Range & the Operating Limits on the Range. Firstly the actual material itself will constrain the technique which may be applied to it, as will the nature of the question which the Archaeologist is asking of it. This must be married to the applicable Dating Range of the techniques, as they all operate in discrete ranges of time. That on its own though is not sufficient, as there are other constraints on the operating limits of the range rather than the purely theoretical limitations. There are physical conditions on sample sizes and purities for making any dating at all. There are also considerations of the utility of the techniques, as near the 'top end' of a technique's range there may well be a drop off in the technique's precision.

Of the two other concepts, Formulae are an inescapable part of all techniques in the calculation of ages and the defining of limitations, but is viewed by the expert as unrelated to the other concepts or groupings. Similarly, High Energy Mass Spectroscopy is only viewed as relevant to the particular case of Advanced radiocarbon dating, and so although related to dating ranges again, it is not seen as a part of that grouping by the expert.

USD

Figure 5.11 - MDS Plot of the 'Concept Space' for the Expert answers to Questionnaire type USD: Uranium Series Dating.

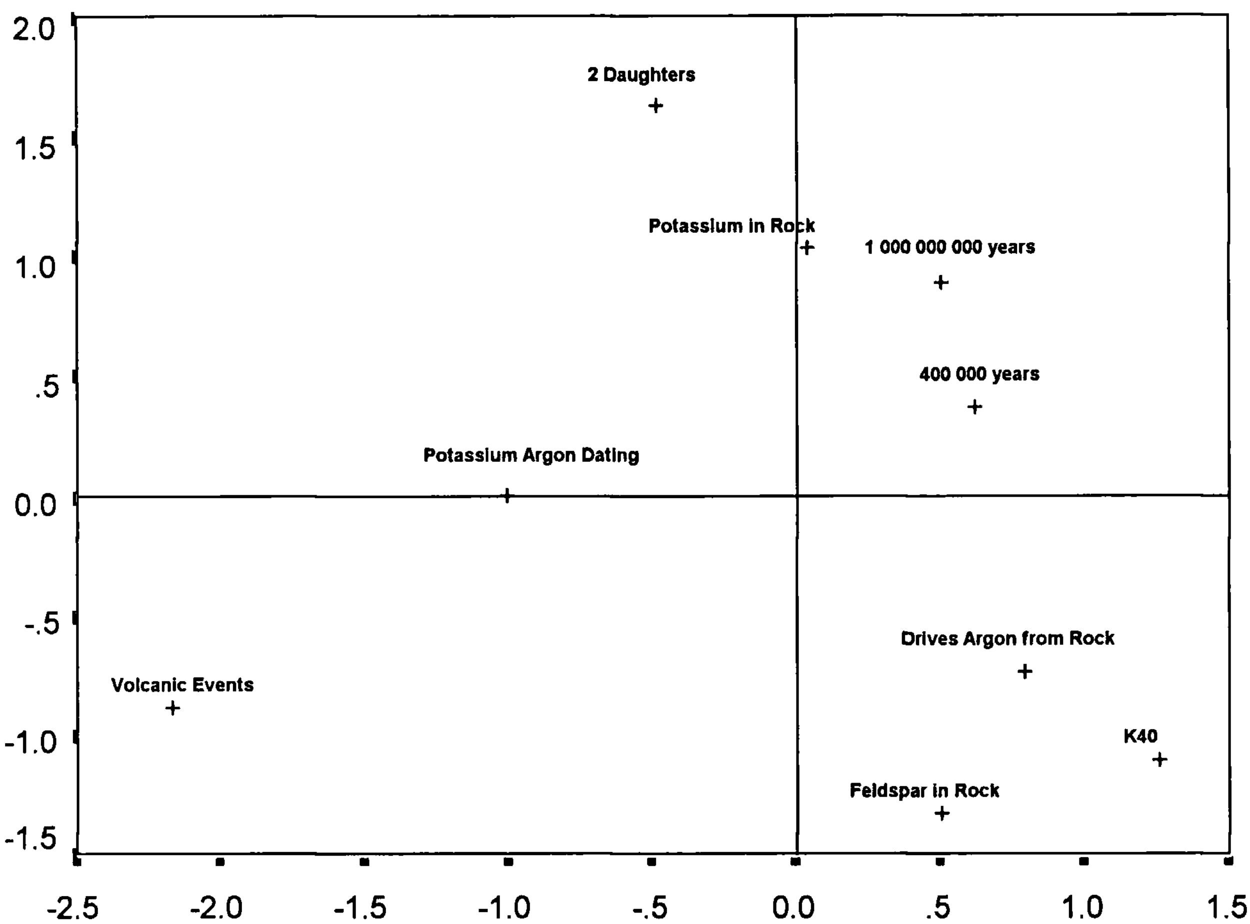


In the USD plot, the distribution was explained in the following terms. The main grouping identified by the expert was taken to include Calcite, Stalactite Formation, Thorium Insoluble & Uranium in Groundwater. These run from left to right and top to bottom and are grouped for defining the mechanism and zeroing event behind Uranium Series Dating. The mechanism functions for two reasons - firstly that there is radioactive nuclides of Uranium in groundwater, and that Thorium (a radioactive by-product of Uranium) is insoluble. Stalactites, a Calcite compound, form a crystalline structure which can trap other elements from within the groundwater. As a stalactite forms it can thus incorporate Uranium into it, but not Thorium (as it is not present in the groundwater, being insoluble). Any Thorium subsequently discovered in the stalactites is therefore a direct result of radioactive decay of the Uranium in the stalactite since the stalactite's formation. Adjoined to this spread in the top right corner are Sedimentary Rock, Several Daughters & Uranium in Rock. It is due to the fact that Uranium exists in these Sedimentary Rocks, and that water can run over them which leaches the Uranium out of the Rocks and into the groundwater. This is a necessary condition for the technique, but not critical enough for the expert to associate it to the grouping (with Uranium in the Groundwater already given). Likewise, the fact that Uranium decays into several different 'daughter' elements is not viewed as an essential to the mechanism, given we are concentrating purely on the chain of events leading to the production of Radioactive Thorium only.

The other grouping made by the expert is on the left hand side - Uranium Series Dating and U234. These are grouped as they are the fundamental concepts to this technique - being the technique's name, and the particular radioactive nuclide which is the starting point for the decay mechanism. The dates, although together, are not grouped specifically by the expert. They represent to the expert the half-life of Uranium - 350 000 years is the half life of U234 (and is also the operating limit of the technique); 1 000 000 000 000 years is the half life of another radioactive Uranium Element, U238. It is the vast difference in these half-lives which allows us to ignore some of the other daughter products of Uranium from our calculations.

KAr

Figure 5.12 - MDS Plot of the 'Concept Space' for the Expert answers to Questionnaire type KAr: Potassium Argon Dating.

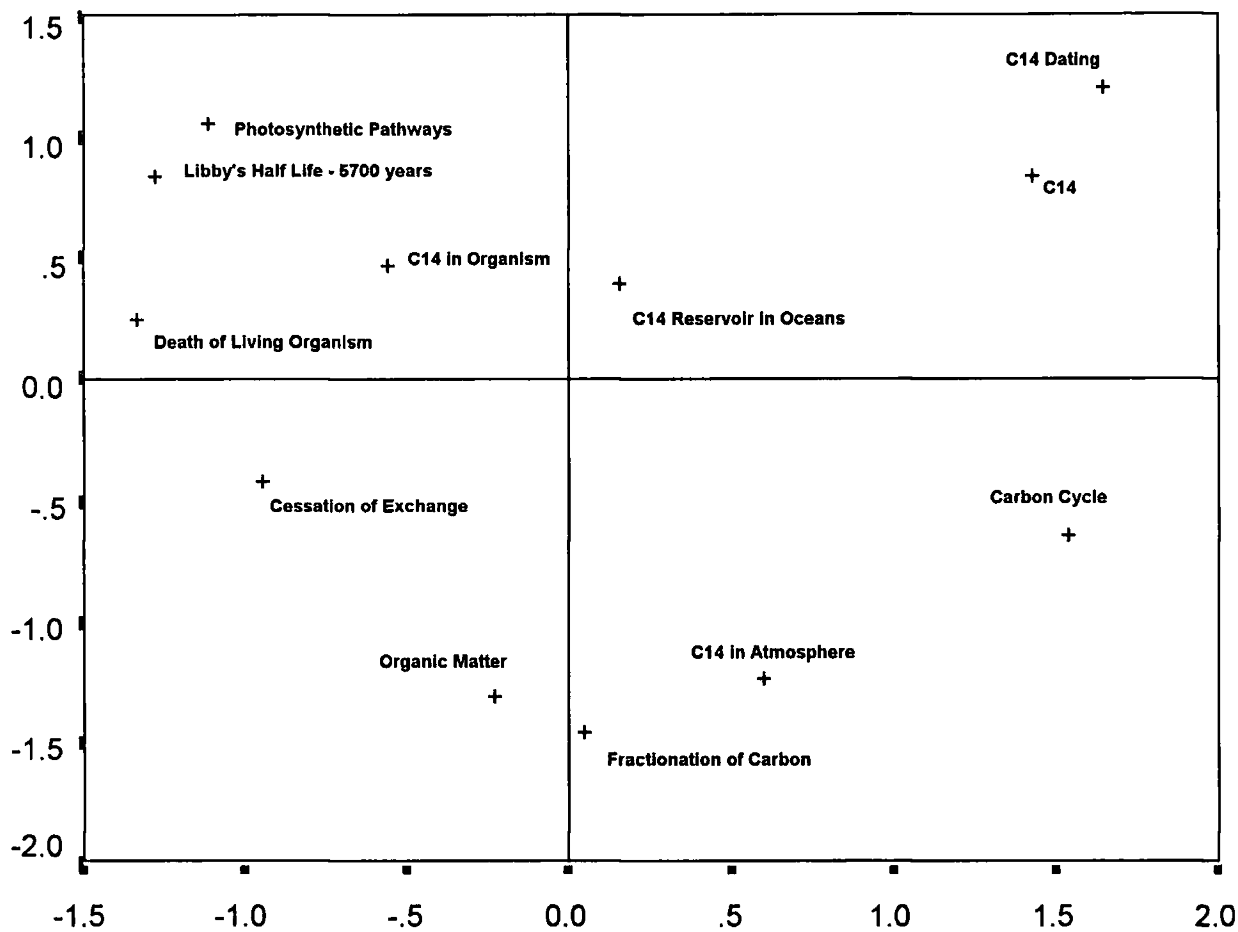


The expert's grouping of the KAr concepts identified two main bodies. The first was the set of Potassium Argon Dating, 2 Daughters, 1 000 000 000 years and K40. Potassium Argon Dating is obviously the technique's name, and the actual radioactive K40 nuclide is the basis of the technique - it has a half-life of 1 000 000 000 years. It also decays into two different daughter elements, one of which is the gas Argon. The other identified grouping is of Potassium in Rock and Feldspar in Rock. These together form the necessary conditions for the technique. In rock there is Potassium, within the actual crystals of Feldspar itself. Over time the radioactive nuclide of Potassium, K40, decays and the gaseous daughter element Argon is trapped in the crystal structure of the Feldspar. This build up gives us our clock mechanism. This is then related to Volcanic Events and Drives Argon from Rock (although not grouped with them) by the expert. This forms the zeroing event of the mechanism, as the intense heat of the Volcanic Event drives the Argon from within the Feldspar crystals, allowing for a new build up of trapped Argon to begin.

The other concept is 400 000 years. This date is the lower end of the dating range for the Potassium Argon dating technique. It appears closely next to the other date, but is not grouped by the expert.

NRC

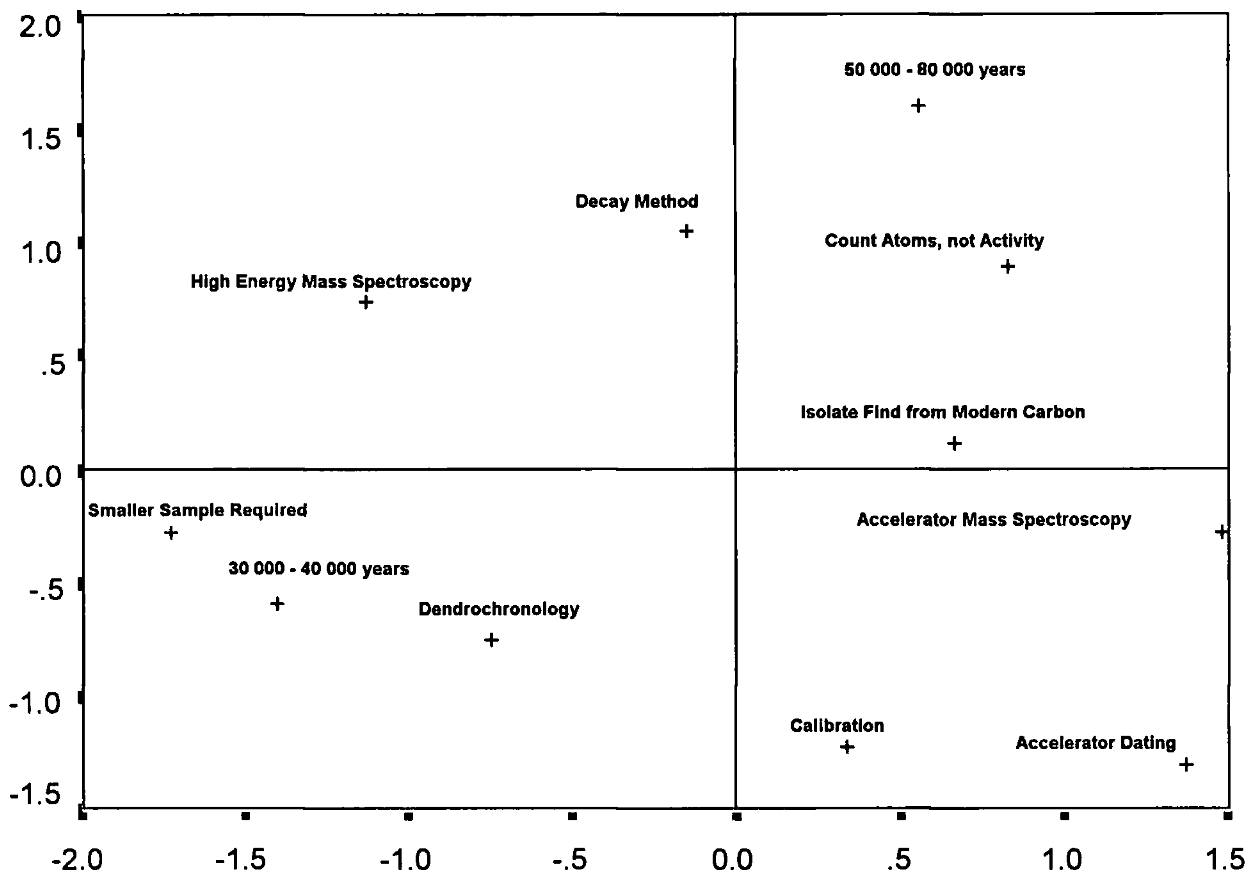
Figure 5.13 - MDS Plot of the 'Concept Space' for the Expert answers to Questionnaire type NRC: Normal Radio-Carbon Dating.



This dating technique was grouped into three main areas by the expert: necessary factors for the zeroing, the basics of the technique & the essential conditions for the mechanism. Those concepts vital to the zeroing event are grouped to the left and the bottom of the plot: Libby's Half Life - 5700 years, Death of Living Organism, Cessation of Exchange & Organic Matter. A second grouping is the basic part of the technique again - the name, C14 Dating and the C14 nuclide itself upon which the technique is built. The third grouping runs from top left to bottom right and covers all the remaining concepts. These all relate to the movement of C14 in the environment through the Carbon Cycle.

ARC

Figure 5.14 - MDS Plot of the 'Concept Space' for the Expert answers to Questionnaire type ARC: Advanced Radio-Carbon Dating.



This final technique was the most complex in terms of the expert's attempt to group the concepts meaningfully. The main link in the plot was taken to be High Energy Mass Spectroscopy. This is the process by which the physical number of carbon atoms can be counted, allowing for a more accurate C14/C12 ratio to be determined. In normal radio-carbon dating the C14 volume has to be inferred from the radioactive count which is much less sensitive and so cannot measure smaller quantities of C14 in older samples reliably against the background radioactivity of the Earth. This increases the range of carbon dating to 50 000 - 80 000 years. The process of High Energy Mass Spectroscopy requires the use of a particle accelerator, and so the technique is often referred to as Accelerator Dating or Accelerator Mass Spectroscopy. This grouping thus contained High Energy Mass Spectroscopy, Accelerator Mass Spectroscopy, Accelerator Dating, Count Atoms, not Activity, 50 000 - 80 000 years & Smaller Sample Required.

A second grouping was the concepts Decay Method and 30 000 - 40 000 years. These represent the normal radio-carbon dating technique and it's range. The Decay Method concept though appears to be crossed over into the sweep from top left to bottom right which mostly defines the first grouping. A third grouping in the central bottom part of the plot is Dendrochronology and Calibration. These are grouped as the dating technique of dendrochronology has been used to calculate fluctuations in the Earth's C14 level and thus calibrated C14 dating where it had been previously assumed as a constant level. Both this small grouping and the other concept, Isolate Find from Modern Carbon have not been integrated with the other groups. As the expert explained however, these are applicable to both normal radiocarbon dating and accelerator dating.

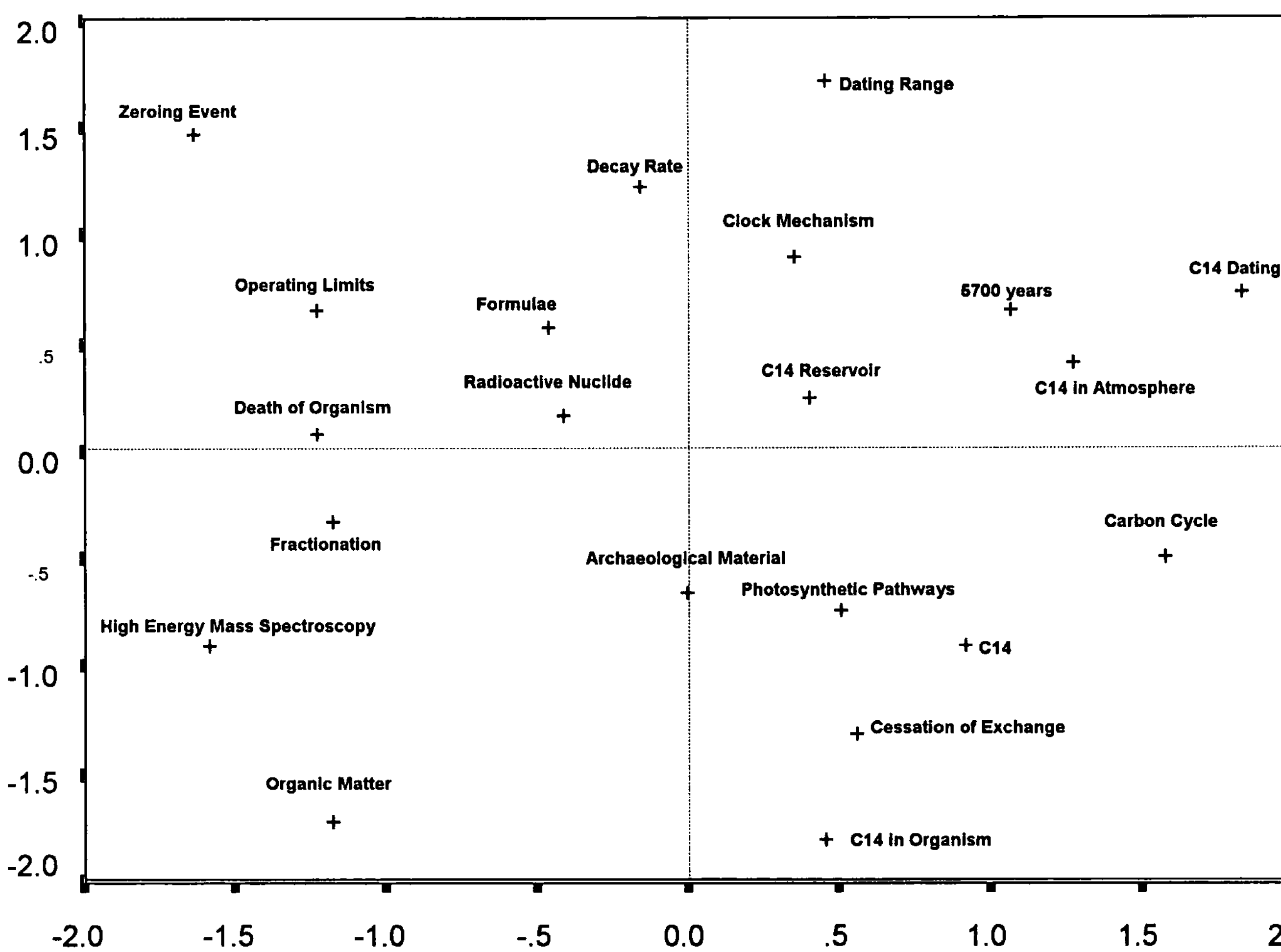
Interpretations

The explanations given by the expert as to the groupings identified in the MDS technique indicate that the general identification of mechanisms and applications seen in the General Dating Methods is extended throughout the other particular techniques. This is in agreement with the findings of previous research where experts group concepts according to deep conceptual principles rather than surface similarities (Chi, Feltovich & Glaser, 1981). In this particular case the concepts were grouped according to several deeper conceptual principles - Basis of Technique (Such as name and fundamental radioactive nuclide); Mechanism of the Chronometer (Such as Zeroing Event, Rate of Decay & the Archaeological Material through which this happens) and Utility of the Technique (Such as its Range and necessary preconditions).

In the investigation of the combined questionnaires and of the differences between the expert and novice distribution of concepts, it is these deeper conceptual principles which we shall expect to see making the difference.

GEN & NRC

<u>Figure 5.15</u> - MDS Plot of the 'Concept Space' for composite Novice answers to Questionnaire type GEN & NRC: General Radiometric Theory & Normal Radiocarbon Dating

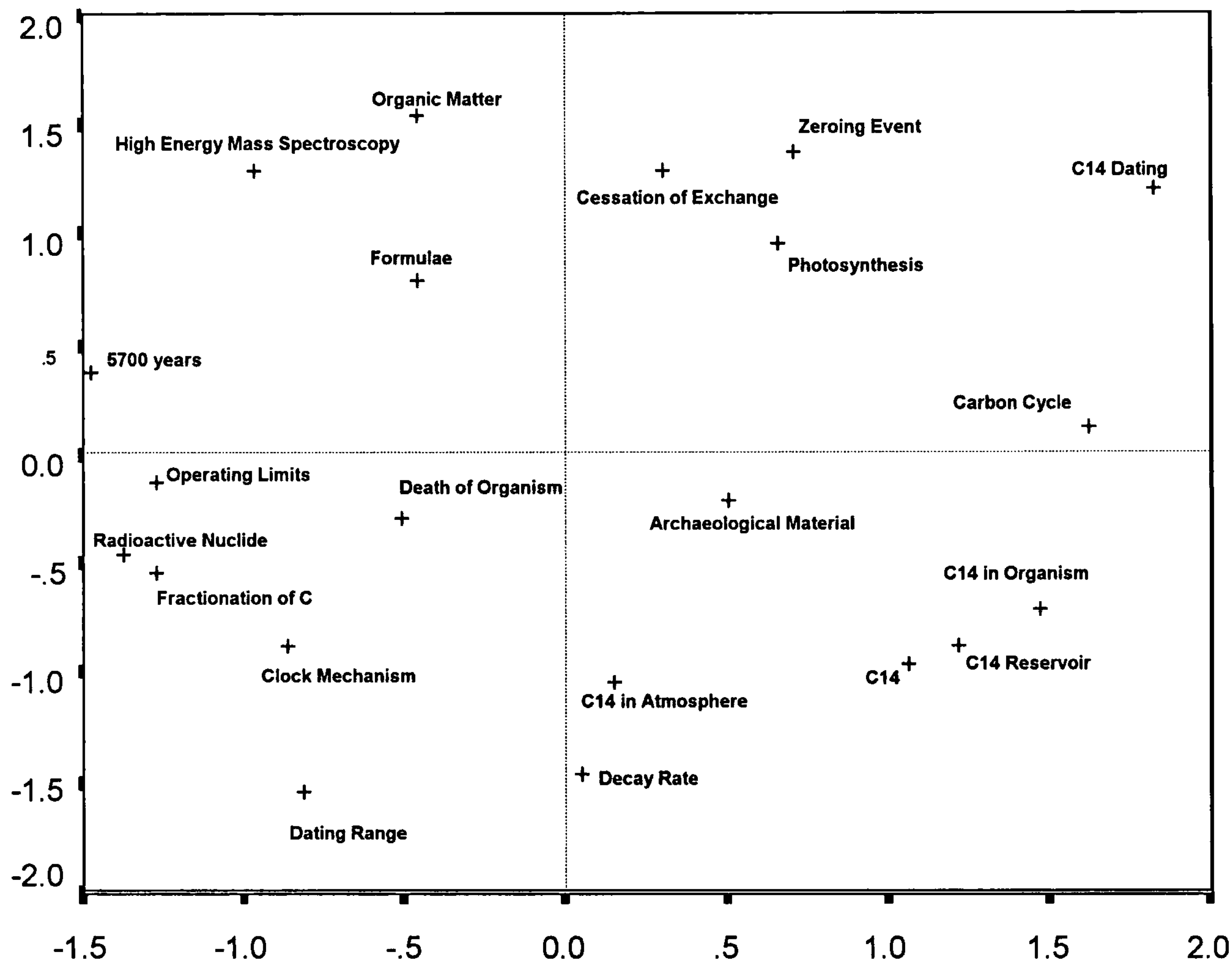


KEY	Topic	Concepts	
A	General Radiometric Terms (GEN)	Zeroing Event Dating Range High Energy Mass Spectroscopy Formulae Operating Limits on Range	Clock Mechanism Radioactive Nuclide Constant Decay Rate Archaeological Material
E	Normal Radiocarbon Dating (NRC)	Death of Living Organism C-14 Produced in Atmosphere C-14 Fixed in Organic Matter C-14 Reservoir in Oceans Libby's Half Life, 5568 Years Cessation of CO_2 Exchange	Carbon Cycle C-14 Organic Matter Fractionation of Carbon Photosynthetic Pathway Carbon 14 Dating

In the novices plot (Figure 5.15) we can see the NRC concepts being grouped together in the bottom right quadrant with almost no interconnection between it and the GEN concepts. The novices group 9 of the 11 radiocarbon ideas together: C14 Dating, 5.7 ka, C14 Reservoir, C14 in Atmosphere, Carbon Cycle, C14, Photosynthetic Pathways, Cessation of Exchange & C14 in Organism. The other two features seem discarded amongst the rest of the features rather than integrated. They (Death of a Living Organism & Fractionation of Carbon) are grouped in an area with Operating Limits on Range and High Energy Mass Spectroscopy

In the NRC part of the plot (right hand side mainly) it is hard to define any groupings. Death of the Organism and Fractionation are clearly together and separate from the rest of the group. There appears to be a huddle with C14 Dating, 5700 years, C14 Reservoir and C14 in atmosphere in the top right, and another with Photosynthetic Pathways, C14, Cessation of Exchange & C14 in Organism at the bottom right. In the GEN part of the plot (left hand side mainly) High Energy Mass Spectroscopy and Organic Matter appear to be stranded together in the bottom left. The other concepts appear fairly evenly spread. The main point is that there is little differentiation within a grouping, only between the two dating topics.

Figure 5.16 - MDS Plot of the 'Concept Space' for composite Expert answers to Questionnaire type GEN & NRC: General Radiometric Theory & Normal Radiocarbon Dating



KEY	Topic	Concepts	
A	General Radiometric Terms (GEN)	Zeroing Event Dating Range High Energy Mass Spectroscopy Formulae Operating Limits on Range	Clock Mechanism Radioactive Nuclide Constant Decay Rate Archaeological Material
E	Normal Radiocarbon Dating (NRC)	Death of Living Organism C-14 Produced in Atmosphere C-14 Fixed in Organic Matter C-14 Reservoir in Oceans Libby's Half Life, 5568 Years Cessation of CO_2 Exchange	Carbon Cycle C-14 Organic Matter Fractionation of Carbon Photosynthetic Pathway Carbon 14 Dating

In the Expert's composite MDS plot above we can see that there is a much greater interplay between the two parts of the questionnaire - GEN and NRC. Both do still retain their own integrity, with GEN coming in from the left of the plot and NRC from the right. Within the groupings themselves (allowing for the distortion of the boundaries to interlock the two techniques) we can see quite obvious groupings. In GEN we have the grouping of Radioactive Nuclide Clock Mechanism and Decay Rate at the boundary with NRC. We also have a grouping of High Energy Mass Spectroscopy and Formulae away from the border. In NRC we can see Fractionation of Carbon, C14 in Atmosphere, C14 Reservoir and C14 in Organism along the GEN border at the bottom of the plot. In the two interlocking fingers of the NRC plot we have groupings of Organic Matter & Cessation of exchange (top finger) and Libby's Half Life - 5700 years & Death of Organism (bottom finger).

However, they concepts do not stay rigidly in their own topic areas but slot together with interlocking fingers. Across these groups there are some locally close concepts, such as Zeroing Event (from GEN) and Cessation of Exchange (from NRC). There are also some locally close groups, such as Clock Rate, Decay Rate and Radioactive Nuclide (from GEN) with Fractionation, C14 in Atmosphere, C14 Reservoir and C14 in Organism (from NRC). Here it is apparent that there is a significant interaction between the two dating topics.

Network Data:

The coherence and settling rates were calculated using NETG for all the data. For the combined questionnaire GEN & NRC the coherence values and settling rates for the novice and expert groups is shown in Table 5.7. The mean figures for coherence are 0.750 for the novices and 0.607 for the experts. For settling rate the means are 56 cycles for the novices & 54 cycles for the experts.

Table 5.7 - Network Data for Expert Individual Questionnaires.

	Coherence		Settling Rate	
Subject	Novice	Expert	Novice	Expert
1	0.937	0.651	56 cycles	53 cycles
2	0.692	0.509	54 cycles	52 cycles
3	0.699	0.622	59 cycles	57 cycles
4	0.671	-	56 cycles	-
Means	0.750	0.607	56 cycles	54 cycles

The calculation of coherence for the all of the combination questionnaires subjects in the 1st group (Novices) are shown in Figure 5.16, and yield values of between 0.074 and 0.937 with an average of .489. For Group 2 (Expert) this value was between 0.509 and 0.662 with an average of 0.632. We also measured the number of cycles needed for the networks of the novices and of the expert to reach stable states. As can be seen from figure 5.18 the networks of the expert settled after 63 cycles on average, ranging from 52 to 80 cycles, while the number of cycles needed for the novices ranged from 62 to 343, with a mean of 94.

Figure 5.17 - Box and Whisker Plot of Coherence for all Combination Questionnaires.

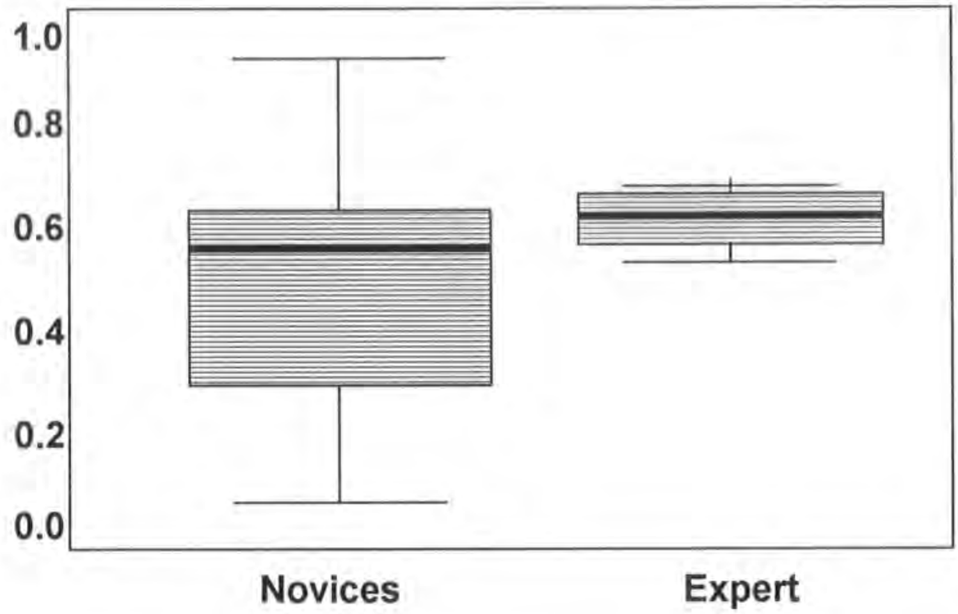
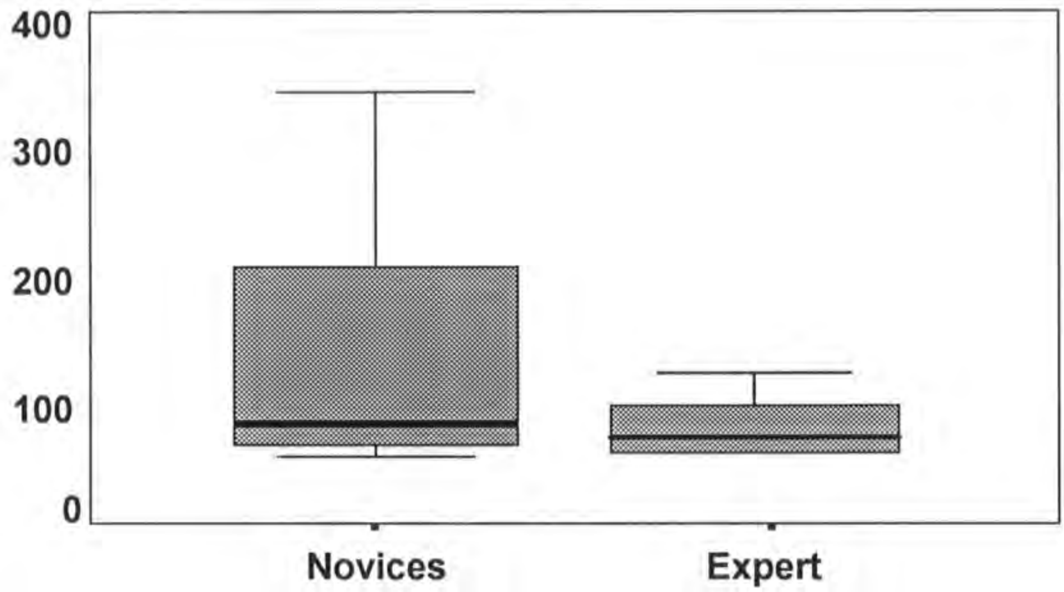


Figure 5.18 - Box and Whisker Plot of Settling Rate for all Combination Questionnaires



The Coherence Information is also shown as a plot for each of the ten questionnaire types in Figure 5.19, and in Figure 5.20 it is compiled for all types belonging to one of the basic technique types.

Figure 5.19 - Box and Whisker Plot of Coherence for Individual Questionnaire Types.

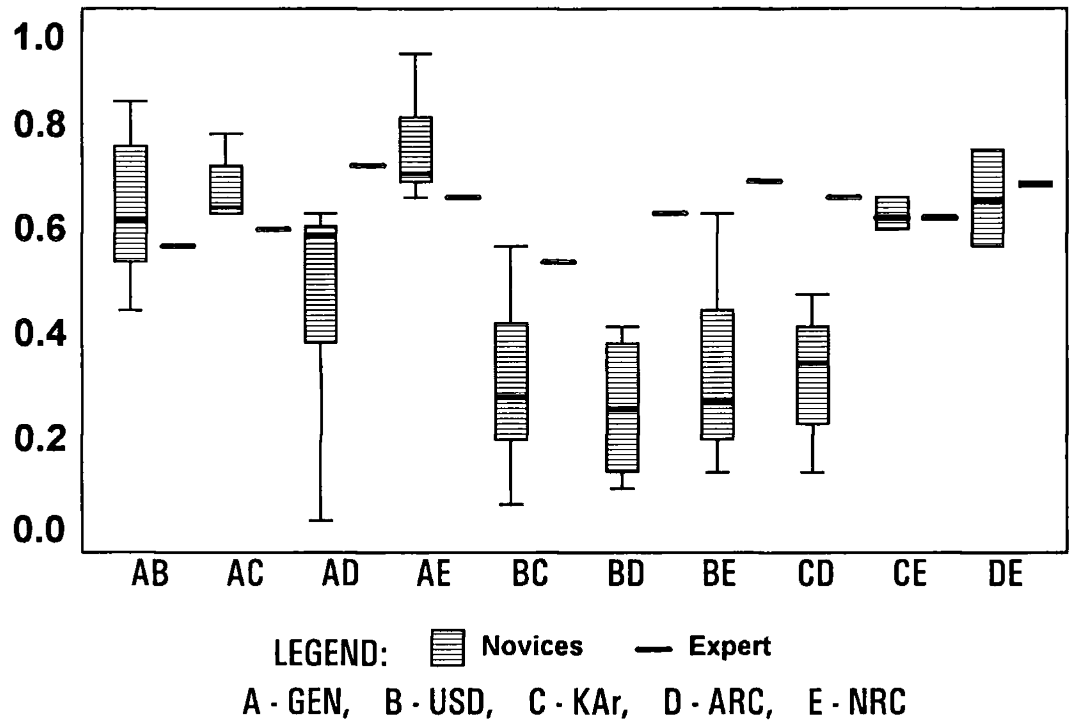
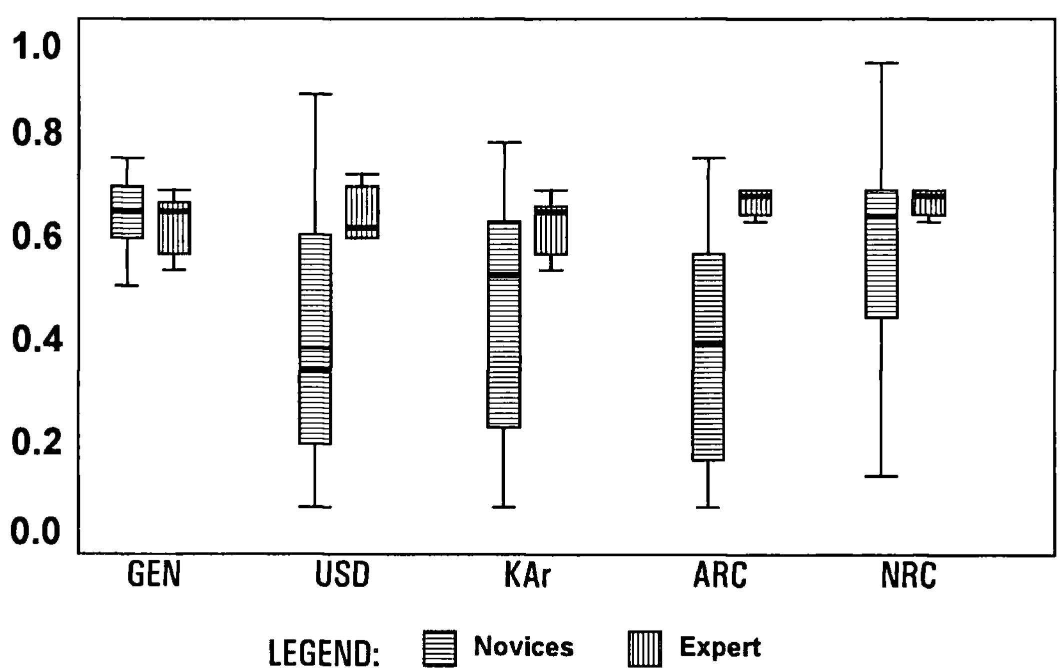


Figure 5.20 - Box and Whisker Plot of Coherence for Composite Techniques.



The coherence and settling rates calculated using NETG for the Expert on the individual plots are also shown in Table 5.8. The mean figures are 0.716 for coherence and 30 cycles for the settling rate.

Table 5.8 - Network Data for Expert Individual Questionnaires.

	Coherence	Settling Rate
GEN	0.587	27 cycles
USD	0.681	29 cycles
KAr	0.724	28 cycles
ARC	0.772	35 cycles
NRC	0.815	32 cycles
Means	0.716	30 cycles

Discussion

Multi-Dimensional Scaling

I wish to focus upon the organisation of Figures 5.15 & 5.16 - the Group plots for the Novices & Experts for questionnaire type GEN & NRC. In the novice's plot (Figure 5.15) we can see the NRC concepts being grouped together in the bottom right quadrant with almost no interconnection between it and the GEN concepts. 9 of the 11 NRC concepts occupy this area, with no obvious groupings. Those which we might expect are hard to find or justify: C14 Dating & C14 (Basis of Technique) are spatially separated; The Zeroing Event concepts area scattered with Death of Living Organism in the top left, Organic Matter the bottom left, Cessation of Exchange bottom right and Libby's Half Life - 5700 years top right. The remaining concepts which grouped to make the Techniques Mechanism are all in the bottom right hand area (bar Fractionation) but even then cannot be resolved into any pattern or arrangement in that area. The GEN concepts similarly show a failure to be organised according to deeper conceptual principles. Zeroing and the Mechanism of the Technique are closely associated according to the expert, but the three related Mechanism concepts could only just be grouped together given they fall no further from each other than any other concepts. Such a lack of organisation is more apparent in the Dating Concepts, which are even further removed from each other within the GEN half of the plot than from most of the other GEN concepts.

As to any links between the two techniques, we could look at the rogue NRC concepts in the GEN half of the plot for meaningful groupings. Death of Living Organism is clearly a Zeroing Event, but falls closer to 5 other concepts before Zeroing Event itself. The other rogue, Fractionation, is a part of the Techniques Mechanism and so could be associated to Clock Mechanism, Constant Decay Rate or Radioactive Nuclide. It is actually closest to High Energy Mass Spectroscopy, although Radioactive Nuclide is a similar distance off. There is definitely though no obvious signs of deep conceptual understanding organising the distribution and grouping of concepts in this plot. Basically, two main groupings exist for the two dating topics involved: GEN & NRC. These are the two most familiar of the 5 to the novices as indicated by the low number of DK responses these concepts elicited from the subjects, and by observing the groupings of their other questionnaire responses (see Appendix C).

In the Expert's representation (Figure 5.15), there is great distortion of the two techniques, GEN & NRC, because of the interplay between them, although both do still retain their own integrity: GEN comes in from the left of the plot and NRC from the right. However, they do not stay rigidly in their own areas but slot together with interlocking fingers. Within the groupings themselves (allowing for the distortion of the boundaries to interlock the two techniques) we can still see quite obvious groupings too. In GEN we have the grouping of Radioactive Nuclide, Clock Mechanism and Decay Rate (Technique Mechanism concepts) at the boundary with NRC above. We also have a grouping of High Energy Mass Spectroscopy and Formulae away from the border (lone concepts). In NRC we can see Fractionation of Carbon, C14 in Atmosphere, C14 Reservoir and C14 in Organism (Technique Mechanism concepts) along the GEN border at the bottom of the plot, where in fact the other grouping is the same deep concept. In the two interlocking fingers of the NRC plot we have groupings of Organic Matter & Cessation of exchange (top finger) and Libby's Half Life - 5700 years & Death of Organism (bottom finger). These together are the Zeroing Event concepts, and locally we can see them grouping across the techniques too - such as Zeroing Event (from GEN) and Cessation of Exchange (from NRC). It is thus obvious how the Expert plot conforms to the expected meaningful groupings and distributions of concepts based on the deeper underlying principles of radiometric dating techniques.

The main point being for the expert plot that there is indeed a great integrity within the two dating topics, but that in addition there is some interplay between the two topics which is both meaningful and valid.

Network Data

The Coherence values and for the GEN & NRC condition show the reason why the coherence values should be higher for the 'novices' on the combination of these two techniques is readily accountable, since they are the two topics with which they are most familiar. They are certainly the areas where novices would be expected to perform best, but to show higher values than experts is surprising. There are several possible explanations for this unexpected result, and these will be tackled in the last chapter

Over all of the questionnaires the results tend to indicate that the Expert does indeed have a higher coherence value than the Novices (means 0.632 & 0.489 respectively), and indeed the same appears to be the case for settling rates, with quicker rates for the Expert (means 80 cycles & 94 cycles respectively). This is only true to look at the means of these values though. If we look at the changes in these values over different questionnaires (Figure 5.19) and different techniques (Figure 5.20) we begin to see more of a pattern to the spread of results. For the experts there is a tendency to have a high coherence value and a low settling rate over all the techniques with little variation around the mean (s.d. 0.052). This demonstrates the consistency of the expert's coherence values and is in sharp contrast to the novices. For them the results show very high results on the GEN & NRC techniques and very low scores on the others (s.d. 0.239). This reflects the greater lack of knowledge they have in these other techniques, and also reflects the spread of abilities within the novice group.

Additionally, the plots in Appendix C gives us some further results. These contained the MDS results for the other 9 questionnaire types for the expert and the novice groups. They demonstrate further that there is little evidence of groupings within the novice plots, except in those cases where NRC (and sometimes also GEN) are a constituent of the questionnaire. This adds additional weight to the conclusion that the novices are particularly lacking in knowledge of these dating topics, most especially USD, KAr & ARC. This also amplifies the results showing the high percentage of DK responses they made to the questionnaires (Table 5.6). The experts' plots by contrast produce groupings for the two topics in each questionnaire every time, with occasional meaningful cross-topic groupings too.

A further point to be made from these expert plots is that the groupings within them are not necessarily reflections of the individual plots he made. This observation indicates how the similarity ratings for these same concepts appear to have been influenced quite specifically by the context within which they were set. This context effect was not anticipated and clearly warrants further study and investigation. I shall now move on to the final chapter, chapter 6: the Discussion of all this research work.

Chapter 6
Discussion

The Empirical Work

Both of our measures revealed differences between the experts and the novices. In the first experiment the network data showed that the expert was clearly above all of the novice values in the coherence measure, and was marginally below the mean for the settling rate. The MDS data also showed that the expert had a qualitatively different cognitive organisation of the concepts involved than the novices. In addition, we compared the activity values obtained from a stable network of the subjects with the conceptual similarity of the concepts to the overall topic of the study through the physical displacement of the concepts in the MDS plot from their origin. These two measures correlated very highly across the subjects.

The expert had clearly organised his knowledge along the lines of higher order concepts about the knowledge domain, grouping as either Archaeological Concepts , mtDNA Inheritance, or Genetic Mutation. One concept, Sexual Reproduction is isolated in space and ungrouped by the expert, reflecting the expert's belief that this is not relevant to the process of mtDNA dating. Additionally, four of the concepts (Ancestry, Eve, mtDNA and 140-280k b.p.) fall very close to the plot centre indicating a greater relevance to the overall concept of mtDNA Dating. According to the expert, these represent the crucial elements in understanding the application of the technique in Archaeology - that all modern humans descend from Eve

The novices by comparison appeared to group the concepts around a single dimension only, concerned with whether the concepts were biological or chronological. This dimension, therefore, appears to be based on similarities between the concepts that could have 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 based on the underlying principles of scientific dating. Overall, therefore, the richly organised categorisation of the expert stands in stark contrast to the simple pre-theoretical organisation of the novices.

In the second experiment we again looked at the network data for the subjects. Over all of the different questionnaires it transpired that the expert had a discernibly higher mean coherence than the novices (0.632 compared to 0.489) and lower settling rate (80 cycles compared to 94). In the more closely studied condition (GEN & NRC) however the results seemed to be turned on their heads. Here the novice mean coherence was much higher than their mean over all of the questionnaires (0.750) whilst the expert mean (0.607) was marginally under the mean value for all questionnaires.

The GEN & NRC questionnaire had been specifically chosen for the comparison to a group of experts on the grounds that it was the topic combination on which the novices would know most (because of the course work they had so far undertaken), and so the comparison might be more productive. Indeed the former is definitely true. We saw in the breakdown of coherence values by topics and by questionnaires that GEN & NRC scored much higher coherence values for the novices than the other three topics (KAr, USD & ARC). It may not seem surprising in the light of their backgrounds that this was the case. What is more at odds is the fact that the Expert's coherence scores are lower than the novice's for GEN & NRC, albeit that they are relatively consistent across all of the other questionnaire types & higher than the novices. I would suggest that the effect we are seeing is perhaps a false representation based upon the beliefs of the subjects rather than true knowledge, and will expand upon this view after talking about the MDS plots.

The MDS technique again showed us a similar pattern to that found in the first Experiment. The expert was clearly grouping the concepts in any one of the 5 dating topics according to deep conceptual understanding of the knowledge domain as shown in the description of individual MDS plots (for which the expert mean coherence was 0.716, somewhat above the mean for the combined plots - 0.632). However, the groupings on the individual plots (from the individual questionnaires) were not always completely apparent in the combined questionnaires, or at least not in exactly the same groupings. The expert combined plots showed a distinction between the two sets of dating topics, but additionally there were identifiable instances of interactions between the concepts from the two different topics. However, the main in-topic groupings were not as complex or detailed as they were in the individual questionnaire cases. This finer distinction was clearly not achieved in the combined plots and indicates the importance of the context in which the similarity ratings are taken. This is an area highlighted by this study which warrants further study.

The novice plots however show little grouping at all, even within the individual dating topics in the questionnaires. The exception to this is when NRC makes up one part of the questionnaire (and occasionally when GEN does so). It was therefore not surprising that it was the NRC & GEN condition on which the novices performed best. This brings me back to the anomalously high novice coherence values for the novice GEN & NRC questionnaire. It is my consideration that the complexity of linking the two different dating topics can cause a drop in coherence. Therefore where the expert's individual scores average above 0.7, when looking at the more the complicated situation represented by the interaction of the two techniques there is a drop off in coherence representing the more uncertain nature of this conceptual area. In the case of the novices no attempt appears to be made to understand this interaction, and so consequently there is no drop off. To see if this is indeed the case a further study could be performed to isolate these aspects and see whether the expert performance was indeed greater in the interaction between the dating topics. If so, then either the calculations of coherence need to be modified to take into greater account the cases where a subject offers no relationship or else we must have another measure with which to aid our interpretation of the coherence value.

Overall our data suggest that the experts had an observably different knowledge representation to the novices, both in terms of the network measures and also the MDS plots. In addition, the wide spread of novice scores in the network measures indicates the spread of understanding they possess. We may infer from this that some novices were beginning to shift from an organisation of knowledge characterised by prior learning to an understanding based on the deeper conceptual basis of the principles of scientific dating. Our results also appear to confirm those observed by Chi et al (1981) on expert and novice physicists: that experts group concepts according to deep theoretical principles while novices' groupings were based on surface similarities. Additionally, in Chi et al the principles underlying the groupings of the experts were only discernible to other experts; a fact which was also true of this study.

Some caution must be exercised in drawing firm conclusions from these results because only one expert was tested in the first experiment and only three were examined for parts of the second experiment. 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). It is therefore not unreasonable to expect that the results gained so far would indeed be replicated with further subjects, and that with a full set of expert data we could indeed validate the conclusions of this study. What has been shown is that we have two similar methods by which the cognitive organisation of relevant concepts can be shown to differ between experts and novices. Both are complementary; one (MDS) showing a more detailed account of the conceptual groupings visually; the other (NETG) yielding a single index measure (coherence) for expertise but containing all of the information within the MDS plot ready for dynamic interrogation.

Future Directions

There are, I believe, several ways in which this study could usefully be extended to help provide more information on the nature of novice expert differences and their application in computer aided learning situations. In so far as the actual planning and preparation of the materials are concerned there are several changes which would make any future study worth while. Firstly, having identified the application by the experts of a deeper level of conceptual organisation to the materials any new study would be well served to explicitly draw upon the further aid of other experts in designing the concepts to be used for the knowledge elicitation (instead of drawing appropriate concepts from the lecture materials without expert advice). This would allow concepts which specifically grouped into deeper conceptual sets to be identified at the outset. Then with the elicitation of materials based on them the conceptual groupings could much more easily be interpreted for a dating topic. Likewise, it would also be easier to make a more direct comparison of these groupings across more than one particular topic area in a domain, or even across domains.

A point of great concern with the data collected is the actual sample sizes. Even with the help of both my supervisors in encouraging and cajoling colleagues over a long period, it was still only possible to find three expert subjects to complete one or more questionnaire. There are two ways in which a future study could approach this vital problem. The first approach would be to treat the experts subjects in a similar way to the novices, by actually sitting with them as they do the questionnaire. Most of the possible experts approached took one look at the questionnaire (a task requiring some 30 minutes or more to complete) and said they would never find the time to do it. By making an appointment to actually go and elicit their knowledge instead as of giving them the questionnaire there is a much greater likelihood of their filling it out. Also, by making an appointment it gives the time to explain the purpose of the study and try to actually motivate the expert into wanting to help out.

The other alternative is to shorten the length of the questionnaires. This could be achieved by decomposing them into smaller parts, and indeed this could serve the purpose of examining separately the individual topics as well as the area of interaction between them. If such was done then the questionnaire could be completed in only 10 minutes perhaps (although there would be more questionnaires to fill in eventually), and this may be an acceptable amount of time for a busy lecturer to find. Getting three lots of ten minutes of a lecturer's time as separate appointments may also be feasible.

In order to handle the split data an expansion of the current NETG program would be required. At the moment it only works with an inputted network on which it can perform a coherence run to a stable state using different equations for the spread of activity, and then different equations to calculate coherence values. It also measures settling rates and is capable of automatically reducing the size of a network to speed up the calculations on a large network (discarding nodes below a threshold activity after a few cycles, and then continuing with the reduced network to a normal conclusion). Looking at individual topics and the interaction separately would be within its present grasp, but it would be useful to be able to combine them within the program to examine the overall patterns too. Similarly, it would be useful to be able to split the networks into smaller sub-networks of specific nodes at the user's request. Another useful feature would be to extract sub-networks from the body of data which had locally low or high coherence levels (or quick / slow settling rates). These could then be used to infer particular areas in a domain where a subject had need of further study or where learning had already been successful.

Perhaps the most profitable extension of this work would be in a longitudinal study of novices to see how their knowledge of certain concepts changed over a period of study such as a first or higher degree. Pairwise similarity ratings could be taken at critical times or on a regular basis over several years, and the way in which the data changed could be studied. I believe that a scenario similar to the following would be observed if such a study were undertaken. Initially there would be a small set of concepts upon which the novice had some structure from their prior learning, but between which there would be little organisation according to the new domain principles. As learning occurred these concepts would be re-organised in the framework of the new discipline, and this would be reflected in a rise in coherence values (NETG) and a shift in concept organisation and groupings (MDS). After learning had reached a certain level, new concepts would again be introduced to the learner, and the lack of understanding in them would force the coherence of an expanded network down (NETG) and place many ungrouped and displaced new concepts in amongst the previously existing organisation (MDS). As these concepts become learnt there would be an integration of them with the previous knowledge (ideally) as well as amongst themselves, resulting in an increase in the coherence of the network (NETG) and a further grouping and re-organisation of concepts (MDS). This process would repeat itself with new learning forcing a temporary reduction in coherence and disorganisation to appear in the MDS plots.

This model would not mean that an expert learning a new topic necessarily will have a total fall-off in their coherence value (or their expertise). The way that coherence works means that as the subject's knowledge base grows so to will their resilience to the affects of new learning - the proportionate fall in coherence becomes less and less. If this model were borne out, and I believe my theoretical and experimental considerations point to it being so, then it would have great implications in terms of learning and teaching. For instance, it would be the case that by the value of the coherence, certain thresholds could be identified. Below a certain coherence it may be that the subject is not yet ready to undertake new learning of a conceptual nature, being preoccupied with reorganising their existing knowledge. There may also be an upper value beyond which the subject is so sure of their new knowledge that they either become bored by their learning or entrenched in their beliefs. A receptive area could then be used to define when learning could most profitably be entertained.

We also have some pedagogical advice based on the findings of our study, not just upon predictions for the future. These novice expert differences identified highlight the importance of taking the pre-existing knowledge of students into account. The novices clearly have some pre-defined conceptual organisation for some of the concepts. To develop teaching techniques that are effective they will necessarily have to point out how the new domain (like Scientific Archaeology) organises concepts in contrast to previously learnt domains (such as Biology, History, Chemistry & Physics), where the organisation may well have been different (although that is not necessarily going to always be the case). Otherwise it is possible, according to the literature (Stevenson & Palmer, 1994), for these previous conceptual organisations to persist in spite of the new learning. By pointing this out explicitly however and addressing the issue, new learning can be built on prior knowledge rather than being in conflict with it.

The final extension of this work is into the field of Intelligent Tutoring Systems, and this was the original inspiration and motivation for the study. If ITSs are indeed to live up to their names then it is vital that they are able to address the pedagogical issues already raised both in this discussion and in the earlier reviews of the literature (Chapters 1 & 3). This can be achieved if the knowledge modelling is capable of making the distinctions identified and it is my belief that the NETG program in a future form is more than capable of doing so, and that the idea behind it is certainly well grounded now. What would be essential to make it functional in an ITS though is the ability for the Tutoring System to make inferences from the learner's actions within the learning environment which could either be used as or interpreted to yield similarity ratings for concept pairs. From my review of the literature I am certain that this is not an impossible task - the work of researchers such as Viau and Larivee (1993) point to possible mechanisms for this interpretation: for instance they studied the path followed through a HyperText system and the time spent at each of the nodes to predict scholastic performance, where the nodes represent concepts in the NETG network.

In final conclusion, I am confident to state that the multiple constraint satisfaction network, realised through the NETG program, and the application of multi-dimensional scaling techniques are both able to effectively discriminate between novice and expert conceptual organisation. The way in which these methods work allows us to address prior knowledge in the learning process, and the comparison between the NETG program and the established MDS technique lends additional credence to the methods established in the NETG software. The advantage of the NETG technique is in the ability for it to work dynamically with the data and to extract different networks from within the data set (a feature not implemented in the current version). Moreover, the fact that the network is dynamic allows it to possibly be incorporated it into an ITS as a knowledge modelling module, which could provide a possible solution to the problems outlined in my criticism of the Overlay Paradigm in ITSs.

Unfortunately, due to the lack of expert subjects willing to partake in the study, I cannot make a firm statement as to the results obtained, but am certain that they will be replicable in the light of the consistency amongst experts in these kinds of concept rating tasks as indicated earlier. I do believe though that I have been successful in making a comparison of the two different methods for interpreting paired similarity data, and have made a strong enough case to warrant further exploration of the NETG solution.

References

Adelson, B. (1984), *When Novices Surpass Experts: The difficulty of a task may increase with expertise*, Journal of Experimental Psychology: Learning, Memory & Cognition, 10, 483-495.

Allard, F. & Starkes, J.L. (1991), *Motor Skill Experts in Sport, Dance and other domains*, in Towards a General Theory of Expertise. Prospects and Limits, Ericsson, K.A. & Smith, J. (Eds.), Cambridge University Press, Cambridge.

Amaiwa, S. (1987), *Transfer of Subtraction Procedures from Abacus to Paper and Pencil*, Japanese Journal of Educational Psychology, 35, 41-48 [Japanese text with English Summary].

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

Anderson, J.R., Farrell, R. & Sauers, R. (1984), *Learning to Program in LISP*, Cognitive Science, 8, 87-129.

Anderson, J.R. (1987), *Skill Acquisition: Compilation of weak-method problem solutions*, Psychological Review, 94, 192-210.

Anderson, J.R., Boyle, C.F., and Yost, G. (1985). *The geometry tutor.* In A. Joshi (Ed.), Proceedings of the Ninth International Joint Conference on Artificial Intelligence (pp. 1-7) Los Altos, CA: Morgan Kaufmann.

Anderson, J.R., Corbett, A.T., and Reiser, B.J. (1986). Essential LISP. Reading, MA: Addison-Wesley.

Bandura, M. & Dweck, C.S. (1985), *The Relation of Conception of Intelligence and Achievement Goals to Achievement-related Cognition, Affect and Behaviour*, Manuscript submitted for publication, cited in Dweck & Leggett, 1988.

Berry, D.C. & Broadbent, D.E. (1984), *On the Relationship between Task Performance and Associated Verbalizable Knowledge*, Quarterly Journal of Experimental Psychology, 35A, 39-49.

Briggs, J.H., Tompsett, C & Oates, N (1993). *Using Rules to Guide Learners through HyperText*, Computers and Education, 20, 1, pp 105 - 110.

Briton, K.B. & Eisenhart, F.J. (1993). 'Expertise, text coherence and constraint satisfaction: effects on harmony and settling rate.' pp 266-271 in the Proceedings of the fifteenth annual conference of the cognitive science society, LEA., Hillsdale, N.J.

Brown, A.L. (1977), *Development, Schooling and the Acquisition of Knowledge about Knowledge*, in Schooling & the Acquisition of Knowledge, Anderson, R.C., Spiro, R.J. & Montague, W.E. (Eds.), Lawrence Earlbaum Associates, Hillsdale, N.J..

Brown, A.L., Campione, J.C. & Day, J.D. (1981), *Learning to Learn: On Training Students to Learn from Texts*, Educational Researcher, 10, 14-21.

Brown, A.L. & Day, J.D. (1983), *Macrorules for Summarizing Text: The Development of Expertise*, Technical Report No. 270, Champaign, IL: Centre for the Study of Reading.

Brown, J.S. and Burton, R.R. (1975). *Multiple representations of knowledge for tutorial reasoning*. In D. Bobrow and A. Collins (Eds.), Representation and understanding: Studies in cognitive science (pp. 311 -349). New York: Academic Press.

Burton, R. R. (1982). *Diagnosing bugs in a simple procedural skill*. In D. Sleeman and J. S. Brown (Eds.), Intelligent Tutoring Systems (pp. 157-183). New York: Academic Press.

Burton, R. R. and Brown, J. S. (1982). *An investigation of computer coaching for informal learning activities*. In D. Sleeman and J. S. Brown (Eds.), Intelligent Tutoring Systems (pp. 79-98). New York: Academic Press.

Carbonell, J.R. (1970). AI in CAI: *An artificial intelligence approach to computer aided instruction*. IEEE Transactions on Man-Machine Systems, 11, 190-202.

Carraher, T.N., Carraher, D.W. & Schliemann, A.D. (1985), *Mathematics in the Streets and in the Schools*, British Journal of Experimental Psychology, 3, 21-29.

Carraher, T.N., Carraher, D.W. & Schliemann, A.D. (1987), *Written and Oral Mathematics*, Journal of Research in Mathematics Education, 18, 83-97.

Chase, W.G. & Simon, H.A. (1973), *Perception in Chess*, Cognitive Psychology, 1, 55-81.

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P. & Glaser, R. (1989), *Self-Explanation: How Students Study and use Example in Learning to Solve Problems*, Cognitive Science, 13, 145-182.

Clancey, W.J. (1982). *Tutoring rules for guiding a case method dialogue.* In D. Sleeman and J.S. Brown (Eds.), Intelligent Tutoring Systems (pp.201-225). New York: Academic Press.

DeGroot, A.D. (1965), Thought and Choice in Chess, Paris: Moutor.

DeKleer, J., and Brown, J.S. (1984). *A physics based on confluences.* AI Journal, 24, 85-168.

Diener, C.I. & Dweck, C.S. (1978), *An Analysis of Learned Helplessness: Continuous Changes in Performance, Strategy & Achievement Cognitions following Failure*, Journal of Personality & Social Psychology, 36, 451-462.

Diener, C.I. & Dweck, C.S. (1980), *An Analysis of Learned Helplessness II: The Processing of Success*, Journal of Personality & Social Psychology, 39, 940-952.

Doerner, D. & Schoelkopf, J.H. (1991), Controlling Complex Systems; or, expertise as "grandmother know-how", in Towards a General Theory of Expertise. Prospects and Limits, Ericsson, K.A. & Smith, J. (Eds.), Cambridge University Press, Cambridge.

Dunkler, (1945), *On Problem Solving*, Psychological Monographs, 58 (5, Whole No. 270).

Dweck, C.S. & Leggett, E.L. (1988), *A Socio-Cognitive Approach to motivation and Personality*, Psychological Review, 95, 256-273.

Ericsson, K.A. & Polson, P.G. (1988), *A Cognitive Analysis of Exceptional Memory for restaurant Orders*, pp 23-70 in The Nature of Expertise, Chi, M.T.H., Glaser, R. & Farr, M.J. (Eds.), Lawrence Earlbaum Associates, Hillsdale, N.J..

Evans, J. St B.T. (1988), T*he knowledge Elicitation Problem: a Psychological Perspective*, Behaviour and Information Technology, 7, 2, 111-130.

Gentner, D. (1983), *Structure Mapping: A theoretical framework for analogy*, Cognitive Science, 7, 155-170.

Gentner, D. & Gentner, D.R. (1983), *Flowing Water and Teeming Crowds; Mental Models of Electricity*, in Mental Models, Gentner, D. & Stevens, A.L. (Eds.), Lawrence Earlbaum Associates, Hillsdale, N.J..

Gentner, D. & Toupin, C. (1986), *Systematicity and Surface Similarity in the Development of Analogy*, Cognitive Science, 10, 277-300.

Gick, M.L. & Holyoak, K.J. (1980), *Analogical Problem Solving*, Cognitive Psychology, 12, 306-355.

Goldstien, I.P. (1982). *The genetic graph: a representation for the evolution of procedural knowledge*. In D. Sleeman and J. S. Brown (Eds.), Intelligent Tutoring Systems (pp. 51-78). New York: Academic Press.

Halff, H.M. (1988). *Curriculum and Instruction in Automated Tutors.* In M.C. Polson, and J.J. Richardson (Eds.), Foundations of Intelligent Tutoring Systems (pp. 80-108). Lawrence Earlbaum Associates.

Hatano, G. (1988), *Social & Motivational Bases for Mathematical Understanding*, in Childrens Mathematics, Saxe, G.B. & Gearhart, M. (Eds.), Jossey-Bass, San Fransisco.

Hollan, J.D., Hutchins, E.L., and Weitzman, L. (1984). *Steamer: An interactive inspectable simulation-based training system.* AI Magazine., 5, 15-27.

Holyoak, K.J. (1991), *Symbollic Connectionism: Towards Third Generation Theories of Expertise*, in Towards a General Theory of Expertise. Prospects and Limits, Ericsson, K.A. & Smith, J. (Eds.), Cambridge University Press, Cambridge.

Holyoak, K.J. & Thagard, P. (1989), *Analogical Mapping by Constraint Satisfaction*, Cognitive Science,13, 295-355.

Howard, M. (1988), 'Knowledge Elicitation: How to Dissociate Conscious Reflections from the Automatic Processes', Unpublished B.Sc. Dissertation, Department of Psychology, Durham University.

Johnson, L. & Soloway, E. (1984), *PROUST: Knoweldge-based program de-bugging*, Proceedings of the seventh International Software Engineering Conference (pp. 369-380)

Keane, M.T.G. (1990), *Incremental Analogizing: Theory and Model*, in Lines of Thinking: Reflections on the Psychology of Thought, Vol 1: Representation, Reasoning, Analogy and Decision Making, Gilhooly, K.J., Keane, M.T.G., Logie, K.H. & Erdos, G. (Eds.), Wiley, Chichester.

Kelly, G.A. (1955), The Psychology of Personal Constructs, Norton, NY.

Kimball, R. (1982). *A self-improving tutor for symbolic integration.* In D. Sleeman and J. S. Brown (Eds.), Intelligent Tutoring Systems (pp. 283-307). New York: Academic Press.

Kintsch, W. (1988). 'The role of knowledge in discourse comprehension: a construction-integration model', *Psychological Review,* 95, 2, 163-182.

Kohler, W. (1969), The Task of Gestalt Psychology, Princeton University Press, Princeton, N.J..

Larkin, J.H. (1981), *Enriching Formal Knowledge; A Model for Learning to Solve Textbook Problems*, in Cognitive Skills and their Acquisition, Anderson, J.R. (Ed.), Lawrence Earlbaum Associates, Hillsdale, N.J..

Lawrence, S.H. (1994), 'Developing Tools for an Intelligent Tutoring System', Unpublished B.Sc. Dissertation, Department of Computer Science, Durham University.

Lesgold, M., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D. & Wang, Y. (1988), *Expertise in Complex Skills*, in The Nature of Expertise, Chi, M.T.H., Glaser, R. & Farr, M.J. (Eds.), Lawrence Earlbaum Associates, Hillsdale, N.J..

Lundell, J.W. (1988), Knowledge Elicitation and the Modelling of Expertise in a Diagnostic Task, Unpublished Ph.D. dissertation, University of Washington.

Maier, N.R.F. (1933), *Reasoning in Humans II: The Solution of a Problem and its Appearances in Consciousness*, Journal of Cognitive Psychology, 12, 181-194.

Mayer, R.E. & Greeno, J.G. (1972), *Structural Differences Between Learning Outcomes Produced by Different Instructional Methods*, Journal of Educational Psychology, 63, 165-173.

McClelland, J.L., Rumelhart, D.E. & The PDP Research Group (1986). 'Parallel Distributed Processing', MIT Press.

Metcalfe, J. & Weibe, D. (1987), *Intuition in Insight and Noninsight problem solving*, Memory and Cognition, 15, 238-246.

Minsky, M. (1975). *A framework for representing knowledge.* In P.H. Winston (Ed.), The psychology of computer vision (pp. 211-277). New York: McGraw-Hill.

Mitchell, D.M. & Grogono, P.D. (1993), *Modelling Techniques for Tutoring Systems*, Computers and Education, 20, 1, pp 55-61.

Newell, A. & Simon, H.A. (1972), Human Problem Solving, Prentice Hall, Engelwood Cliffs N.J..

Orey, M.A. & Nelson, W.A. (1993). *Development Principles for Intelligent Tutoring Systems: Integrating Cognitive Theory into the Development of Computer Based Instruction*, Education Technology Research & Design, 41,1,pp 59-72.

Orey, M., Trent, A. & Young, J. (1993). *Development of Efficiency and Effectiveness of Alternative Platforms for Intelligent Tutoring.* In P. Burna, S. Ohlsson & H.Pain (Eds) Proceedings of AI-ED '93 - World Conference on Artificial Intelligence in Education (pp. 42 - 49). Association for the Advancement of Computing in Education, Charlottesvill, VA.

Owen, E. & Sweller, J. (1985), *What do Students Learn While Solving Mathematical Problems*, Journal of Educational Psychology, 77, 272-284.

Price, S. (1993), Seminar at Bristol University on the Application of the WinEcon Tutor and its Underlying Structure, Unpublished.

Scardamalia, M. & Bereiter, C. (1991), *Literate Expertise*, in Towards a General Theory of Expertise. Prospects and Limits, Ericsson, K.A. & Smith, J. (Eds.), Cambridge University Press, Cambridge.

Schank, R.C., & Abelson, R.P. (1977). Scripts, plans, goals and understanding. Hillsdale, NJ: Lawrence Earlbaum Associates.

Shortliffe, E.H. (1976). Computer-based medical consultations: MYCIN. New York: American Elsivier.

Sleeman, D., Ward, R.D., Kelly, E., Martinak, R. & Moore, J. (1991), *An Overview of Recent Studies with PIXIE*, in P. Goodyear (Ed.) Teaching Knowledge and Intelligent Tutoring, Ablex Publishing Co., Norwood, NJ.

Sloboda, J.A. (1991), *Musical Expertise*, in Towards a General Theory of Expertise. Prospects and Limits, Ericsson, K.A. & Smith, J. (Eds.), Cambridge University Press, Cambridge.

Sloboda, J.A. & Howe, M.J.A. (1991), *Biographical Precursors of Musical Excellence: An interview study*, Psychology of Music, 19, 3-21.

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

Soper, J.B. & MacDonald, A.B. (1994). *An Interactive Approach to Learning Economics*, Association for Learning Technology Journal (ALT-J), 2, 1, 14-29.

Stenning, K., Cox, R. & Oberlander, J. (Forthcoming), *Contrasting the Cognitive Effects of Graphical and Sentential Logic Teaching: Reasoning, Representation and Individual Differences*, Language and Cognitive Processes.

Stevens, A., Collins, A. & Goldin, S.E. (1982). 'Misconceptions in Student's Understanding' in D.Sleeman & J.S.Brown (Eds) *Intelligent Tutoring Systems*, pp.13-24, Academic Press.

Stevenson, R.J. (1993), Language, Thought & Representation, John Wiley & Co., Chichester.

Stevenson, R.J. (Forthcoming), Human Learning, A Cognitive Perspective, Routledge.

Stevenson, R.J., Howard M.J. & Manktelow, K.I. (1988). 'Knowledge elicitation: dissociating conscious reflections from automatic processes' in D.M. Jones & R. Winder (Eds.) *People and Computers IV*, pp. 565-580. Cambridge University Press.

Stevenson, R.J. & Palmer, J.A. (1994). 'Learning: Principles, Processes & Practices', Cassell.

Sweller, J., Mawer, R.F. & Ward, M.R. (1983), *Development of Expertise in Mathematical Problem Solving*, Journal of Experimental Psychology: General, 112, 639-661.

Thagard, P., Holyoak, K.J., Nelson, G. & Goch, D. (1990), *Analogical Retrieval by Constraint Satisfaction*, Artificial Intelligence, 46, 259-310.

Tilbury, A.G., Bailiff, I.K. & Stevenson, R.J. (in Press) *Characterizing Novice and Expert Knowledge: Towards an Intelligent Tutoring System for Archaeological Science.* In J. Huggett (Ed) Proceedings of CAA '94 - Annual Computer Applications in Archaeology Conference.

VanLehn, K. (1993). *Cascade: A Simulation of Human Learning and Its Applications.* In P. Burna, S. Ohlsson & H.Pain (Eds) Proceedings of AI-ED '93 - World Conference on Artificial Intelligence in Education (pp. 1 - 3). Association for the Advancement of Computing in Education, Charlottesvill, VA.

VanLehn, K. (1993). Keynote Conference Address, given to the AI-ED '93 Conference, August 23rd, Edinburgh, Scotland. This address is unpublished to a great extent as it deviated very significantly from the actual paper in the proceedings.

VanLehn, K. (1988). *Student Modelling.* In M.C. Polson, and J.J. Richardson (Eds.), Foundations of Intelligent Tutoring Systems (pp. 56 - 78). Lawrence Earlbaum Associates.

Viau, R. & Larivee, J. (1993), *Learning Tools with HyperText: An Experiment*, Computers and Education, 20, 1, pp 11-16.

Warren, K.C., Goodman, B.A. & Maciorowski, S.M. (1993). *A Software Architecture for Intelligent Tutoring Systems.* In P. Burna, S. Ohlsson & H.Pain (Eds) Proceedings of AI-ED '93 - World Conference on Artificial Intelligence in Education (pp. 50 - 57). Association for the Advancement of Computing in Education, Charlottesvill, VA.

Wenger, E. (1987), Artificial Intelligence and tutoring Systems. Computational and Cognitive approaches to the Communication of Knowledge, Los Altos, C.A. Morgan Kaufmann.

Wharton, C.M., Holyoak, K.J., Downing, P.E., Lange, T.E., Wickens, D.W. & Melz, E.R. (1994), *Below the surface: Analogical Similarity and Retrieval Competition in Reminding*, Cognitive Psychology, 26, pp 64-101.

Woolf, B.P. (1993). *Toward a Computational Model of Tutoring*, Educational Technology Research and Development, 40, 4, pp 49-64.

Woolf, B.P. (1988a). *Representing complex knowledge in an intelligent machine tutor.* In J. Self (Ed.), Artificial intelligence and human learning: intelligent computer-aided instruction (pp. 3-27). Chapman and Hall Computing.

Woolf, B.P. (1988b). *Representing, Acquiring, and Reasoning about Tutoring Knowledge.* In H. Burns, J.W. Parlett & C.L. Redfield (Eds.), Intelligent Tutoring Systems: Evolutions in Design (pp. 127-149). LEA, Hillsdale, NJ.

Woolf, B.P., Blegen, D., Jansen, J.H. & Verloop, A. (1987). *Teaching a Complex Industrial Process.* Proceedings of the Artificial Intelligence National Conference, Philadelphia (pp. 722-728). Morgan Kaufmann, Los Altos, CA.

Young, C. & Heath, S. (1994). *The Spreadsheet Microsoft Excel for Effective CAL Courseware*, The Computer in Teaching Initiative Support Service File (The CTISS Files), 17, pp 54 -57.

Appendices

Appendix A

Simplifications of the Briton & Eisenhart Method for the Calculation of 'Harmony'

$$H(r) = \frac{\left((a_A, a_B) \cdot \begin{bmatrix} a_A \\ a_B \end{bmatrix} \cdot \begin{bmatrix} - & W_{BA} \\ W_{AB} & - \end{bmatrix}\right)}{\sum |(ka)i|} \cdot 1$$

$$H(r) = \frac{\left(\begin{bmatrix} a_A a_A & a_B a_A \\ a_A a_B & a_B a_B \end{bmatrix} \begin{bmatrix} - & W_{BA} \\ W_{AB} & - \end{bmatrix}\right)}{|a_A| + |a_B|} \cdot 1$$

$$H(r) = \frac{(a_A a_A) W_{AB} + (a_B a_A) W_{BA} + (a_A a_B) W_{BA} + (a_B a_B) W_{AB}}{|a_A| + |a_B|}$$

& since $W_{AB} = W_{BA} = W$, then

$$H(r) = \frac{W (a_A + a_B)^2}{|a_A| + |a_B|}$$

But, from the Kintsch model $\sum a_i = 1$, and so

$$H(r) = \frac{W}{|a_A| + |a_B|}$$

Also, for only two nodes and one connection, it follows that without a decay function $a_A = a_B$,and so $a_A = a_B = 0.5$.

Therefore,

$$H(r) = W.$$

This is as we should expect from the given example for 2 nodes, but it is only an artefact of such a simplistic case. I use it as a demonstration of the workings however since those for three nodes is that much more complex and tedious. However, if we look at the reduction for three nodes we find that;

$$H(r) = \frac{W_{AB}\,[(a_A+a_B)^2 + a_C a_A + a_C a_B] + W_{AC}\,[(a_A+a_C)^2 + a_B a_A + a_B a_C] + W_{BC}\,[(a_B+a_C)^2 + a_A a_C + a_A a_B]\}}{\{|a_A|\,.\,|a_B|\,.\,|a_C|\}}$$

Then by re-arrangement;

$$H(r) = a_A\,\{\,W_{AB}\,(a_A+a_B+a_C) + W_{AC}\,(a_A+a_B+a_C)\} + a_B\,\{\,W_{AB}\,(a_A+a_B+a_C) + W_{BC}\,(a_A+a_B+a_C)\} + a_C\,\{\,W_{AC}\,(a_A+a_B+a_C) + W_{BC}\,(a_A+a_B+a_C)\}$$

and since $a_A+a_B+a_C = 1$, then

$$H(r) = a_A\,\{\,W_{AB} + W_{AC}\,\} + a_B\,\{\,W_{AB} + W_{BC}\,\} + a_C\,\{\,W_{AC} + W_{BC}\,\}$$

So we finally get to the generic formula inherent in the process. The actual calculation of Harmony used by Britton & Eisenhart is that the Harmony of any node n is a simple sum of the activation of it multiplied by the connections to it, thus

$$H(N) = a_N\ \{W_{AN} + W_{BN} + \ldots + W_{NX}\},$$ where there are X nodes and where $W_{NN} = 0$.

Which gives us the completed formula,

$$H(r) = \sum\ (a_N\ \{W_{AN} + W_{BN} + \ldots + W_{NX}\}),$$ for N=1 to x.

Appendix B

Raw Data from Experiment 1.

Example Questionnaire

A Questionnaire on Mitochondrial DNA Dating

On the following pages you will be presented with a list of paired ideas. Each pair consists of two ideas which are both associated with Mitochondrial DNA Dating in the field of Archaeology. In total there are 55 paired items for you to look at.

What I would like you to do is to read each pair of ideas, and then indicate on the scale provided how closely related to each other you believe them to be. To help explain this to you I have included a possible example of a similar task below. If you complete it, it should help to clarify in your mind the way in which this survey works. Afterwards please do go on to fill in this questionnaire.

Thank-you for your help.

Example

In locating an archaeological site there are many different possible approaches. These include the following four techniques: Field walking, Aerial Photography, Resistivity and Magnetometry. In total this gives us 6 paired items.

Please indicate on a scale of 0-10 how closely related you believe each of the following pairs of ideas/concepts to be by circling the appropriate number on the scale below.

Field Walking	&	Aerial Photography	0	1	2	3	4	5	6	7	8	9	10
Resistivity	&	Magnetometry	0	1	2	3	4	5	6	7	8	9	10
Magnetometry	&	Field Walking	0	1	2	3	4	5	6	7	8	9	10
Magnetometry	&	Aerial Photography	0	1	2	3	4	5	6	7	8	9	10
Field Walking	&	Resistivity	0	1	2	3	4	5	6	7	8	9	10
Resistivity	&	Aerial Photography	0	1	2	3	4	5	6	7	8	9	10

Could you please use the space below to indicate any salient ideas or concepts which you believe are related to locating an archaeological site which were not included in the original questionnaire.

If you have any problems or questions, please ask for help.

Please indicate on a scale of 0-10 how closely related you believe each of the following pairs of ideas/concepts to be by circling the appropriate number on the scale given.

Time Scale	&	140-280 ka b.p.	- 0 1 2 3 4 5 6 7 8 9 10
Rate of Mutation	&	'Eve'	- 0 1 2 3 4 5 6 7 8 9 10
Time Scale	&	Female Heredity	- 0 1 2 3 4 5 6 7 8 9 10
Rate of Mutation	&	mtDNA	- 0 1 2 3 4 5 6 7 8 9 10
Africa	&	Female Heredity	- 0 1 2 3 4 5 6 7 8 9 10
Sexual reproduction	&	140-280 ka b.p.	- 0 1 2 3 4 5 6 7 8 9 10
Female Heredity	&	'Eve'	- 0 1 2 3 4 5 6 7 8 9 10
Female Heredity	&	Cell Cytoplasm	- 0 1 2 3 4 5 6 7 8 9 10
Cell Cytoplasm	&	Time Scale	- 0 1 2 3 4 5 6 7 8 9 10
Ancestry	&	140-280 ka b.p.	- 0 1 2 3 4 5 6 7 8 9 10
140-280 ka b.p.	&	Temporal Mutation	- 0 1 2 3 4 5 6 7 8 9 10
'Eve'	&	Time Scale	- 0 1 2 3 4 5 6 7 8 9 10
140-280 ka b.p.	&	Rate of Mutation	- 0 1 2 3 4 5 6 7 8 9 10
140-280 ka b.p.	&	Africa	- 0 1 2 3 4 5 6 7 8 9 10
Ancestry	&	mtDNA	- 0 1 2 3 4 5 6 7 8 9 10

Ancestry	&	Time Scale	- 0 1 2 3 4 5 6 7 8 9 10
Temporal Mutation	&	Sexual reproduction	- 0 1 2 3 4 5 6 7 8 9 10
Ancestry	&	'Eve'	- 0 1 2 3 4 5 6 7 8 9 10
mtDNA	&	Female Heredity	- 0 1 2 3 4 5 6 7 8 9 10
mtDNA	&	140-280 ka b.p.	- 0 1 2 3 4 5 6 7 8 9 10
Sexual reproduction	&	Ancestry	- 0 1 2 3 4 5 6 7 8 9 10
Cell Cytoplasm	&	'Eve'	- 0 1 2 3 4 5 6 7 8 9 10
Rate of Mutation	&	Time Scale	- 0 1 2 3 4 5 6 7 8 9 10
Temporal Mutation	&	Female Heredity	- 0 1 2 3 4 5 6 7 8 9 10
Sexual reproduction	&	Time Scale	- 0 1 2 3 4 5 6 7 8 9 10
Africa	&	Cell Cytoplasm	- 0 1 2 3 4 5 6 7 8 9 10
'Eve'	&	mtDNA	- 0 1 2 3 4 5 6 7 8 9 10
Cell Cytoplasm	&	Ancestry	- 0 1 2 3 4 5 6 7 8 9 10
mtDNA	&	Temporal Mutation	- 0 1 2 3 4 5 6 7 8 9 10
Africa	&	Rate of Mutation	- 0 1 2 3 4 5 6 7 8 9 10
Sexual reproduction	&	'Eve'	- 0 1 2 3 4 5 6 7 8 9 10

Cell Cytoplasm	&	Sexual reproduction	- 0 1 2 3 4 5 6 7 8 9 10
Temporal Mutation	&	Rate of Mutation	- 0 1 2 3 4 5 6 7 8 9 10
Time Scale	&	Temporal Mutation	- 0 1 2 3 4 5 6 7 8 9 10
Sexual reproduction	&	mtDNA	- 0 1 2 3 4 5 6 7 8 9 10
Ancestry	&	Africa	- 0 1 2 3 4 5 6 7 8 9 10
Africa	&	Temporal Mutation	- 0 1 2 3 4 5 6 7 8 9 10
'Eve'	&	140-280 ka b.p.	- 0 1 2 3 4 5 6 7 8 9 10
mtDNA	&	Africa	- 0 1 2 3 4 5 6 7 8 9 10
Female Heredity	&	Rate of Mutation	- 0 1 2 3 4 5 6 7 8 9 10
Temporal Mutation	&	Cell Cytoplasm	- 0 1 2 3 4 5 6 7 8 9 10
'Eve'	&	Africa	- 0 1 2 3 4 5 6 7 8 9 10
Female Heredity	&	Ancestry	- 0 1 2 3 4 5 6 7 8 9 10
Cell Cytoplasm	&	Rate of Mutation	- 0 1 2 3 4 5 6 7 8 9 10
'Eve'	&	Temporal Mutation	- 0 1 2 3 4 5 6 7 8 9 10
Rate of Mutation	&	Sexual reproduction	- 0 1 2 3 4 5 6 7 8 9 10
Rate of Mutation	&	Ancestry	- 0 1 2 3 4 5 6 7 8 9 10

mtDNA	&	Cell Cytoplasm	- 0 1 2 3 4 5 6 7 8 9 10
Time Scale	&	mtDNA	- 0 1 2 3 4 5 6 7 8 9 10
Time Scale	&	Africa	- 0 1 2 3 4 5 6 7 8 9 10
Temporal Mutation	&	Ancestry	- 0 1 2 3 4 5 6 7 8 9 10
140-280 ka b.p.	&	Cell Cytoplasm	- 0 1 2 3 4 5 6 7 8 9 10
Africa	&	Sexual reproduction	- 0 1 2 3 4 5 6 7 8 9 10
Female Heredity	&	Sexual reproduction	- 0 1 2 3 4 5 6 7 8 9 10
140-280 ka b.p.	&	Female Heredity	- 0 1 2 3 4 5 6 7 8 9 10

Please indicate below how you have acquired your knowledge of mitochondrial DNA dating:-

(tick as many as are appropriate)

Lectures in Archaeology	-	[ ]
Tutorials in Archaeology	-	[ ]
Essays in Archaeology	-	[ ]
Anthropology Courses	-	[ ]
Biology Courses	-	[ ]
Others {Please Specify}		

Could you please also use the space below to indicate any salient ideas or concepts which you believe are related to Mitochondrial DNA dating which were not included in the original questionnaire.

Thank you.

Appendix C

The Remaining MDS Plots from Experiment 2.

Description

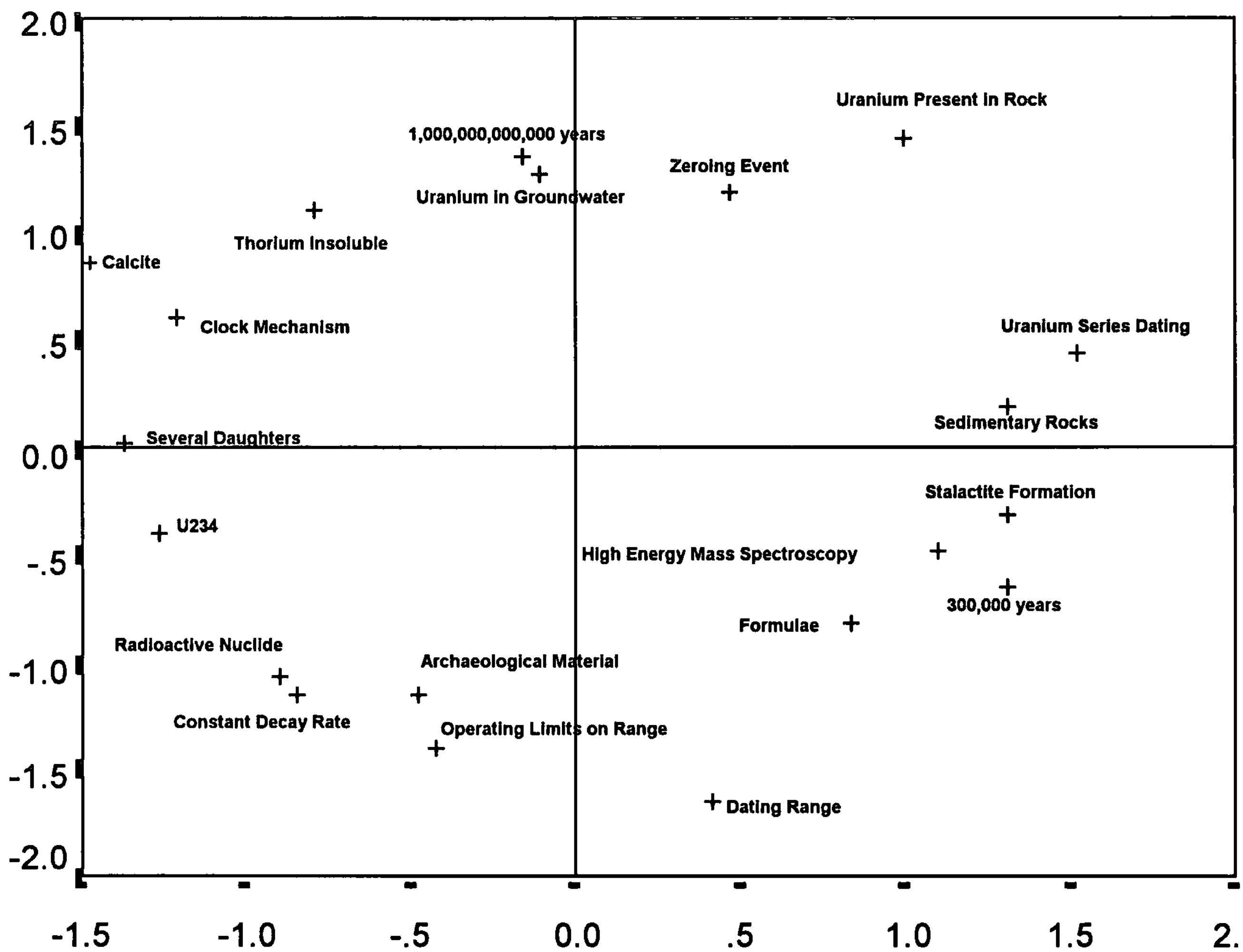
Looking at the original descriptions which the Expert gave on the individual techniques it was clear that he was organising concepts by deep conceptual principles. In the expert combination plots we can see that this grouping style is still very much in evidence, and more we can see that where there is an interaction between the two techniques it is often occurring at a point of deeper organisation. For example, in Figure 16 we can see that the distribution of concepts within the KAr and NRC halves is stretched out of shape to allow for a very specific grouping of Death of Living Organism & Drives Argon from Rock. These concepts clearly represent to the Expert the actual Zeroing Event of the chronometer for each of the two techniques. Even though such cross technique groupings pull the rest of the MDS plot out of shape, we can still see recognisable groupings within each technique that conform to the deeper conceptual organisation. Both sides show groupings for the Basis of the Technique (as described in Chapter 5) and reasonable groupings by the Mechanism of the Technique.

By comparison, the Novice plot for the same questionnaire type (Figure 15) in no way reflects the same pattern of organisation even without meaningful cross-technique groupings to distort the individual technique groupings. There is no clear organisation to the two techniques as a whole, but instead there are many sporadic localised pockets. For instance, the Fractionation of Carbon & Photosynthesis concepts form a part of the larger underlying Mechanism concept identified by the Expert, and are at a more superficial level linked by their closeness in that particular part of the Mechanism of the Technique. However, they are as far apart from the other members of that deeper grouping as is possible, and scattered too. Similar comparisons can be made throughout the rest of the combination questionnaire plots (Figures 1 - 18) to a greater or lesser degree.

The main difference between the two in realistic and easily observed terms, is that the novice plots tend to have no organisation with the concepts equally spaced and at a distances from the centre. The experts plots however tend to show more deliberate organisation with clusters and groupings, and with more tendency to have concepts or groupings nearer the centre of the plots.

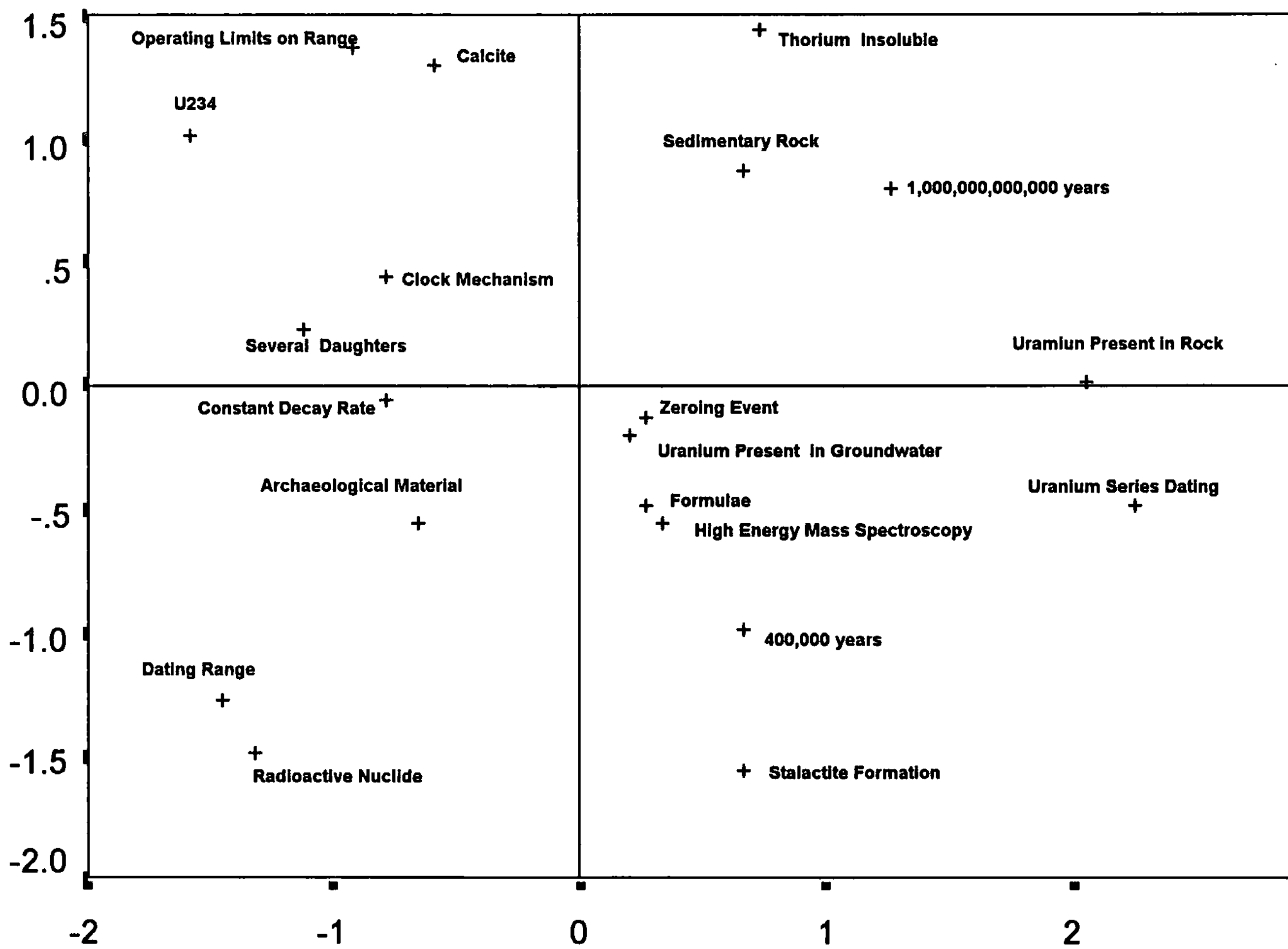
GEN & USD

Figure 1 - MDS Group Plot for all USD & GEN Novice Responses.



In this plot the USD concepts are bounded mainly to the top and also the left and right, but at the edge of the plot. The GEN concepts are all closer to the middle of the plat and spread mainly at the bottom, but also left and right. The distinguishable groupings for GEN are Radioactive Nuclide, Constant Decay Rate, Archaeological Material and Operating Limits on Range (bottom left) and High Energy Mass Spectroscopy and Formulae (bottom right). For USD they are U234 & Several Daughters (left), 1 000 000 000 000 years & Uranium in Groundwater (top) and Uranium Series Dating, Sedimentary Rock, Stalactite Formation & 400 000 years (right). There are no obvious cross-technique groupings.

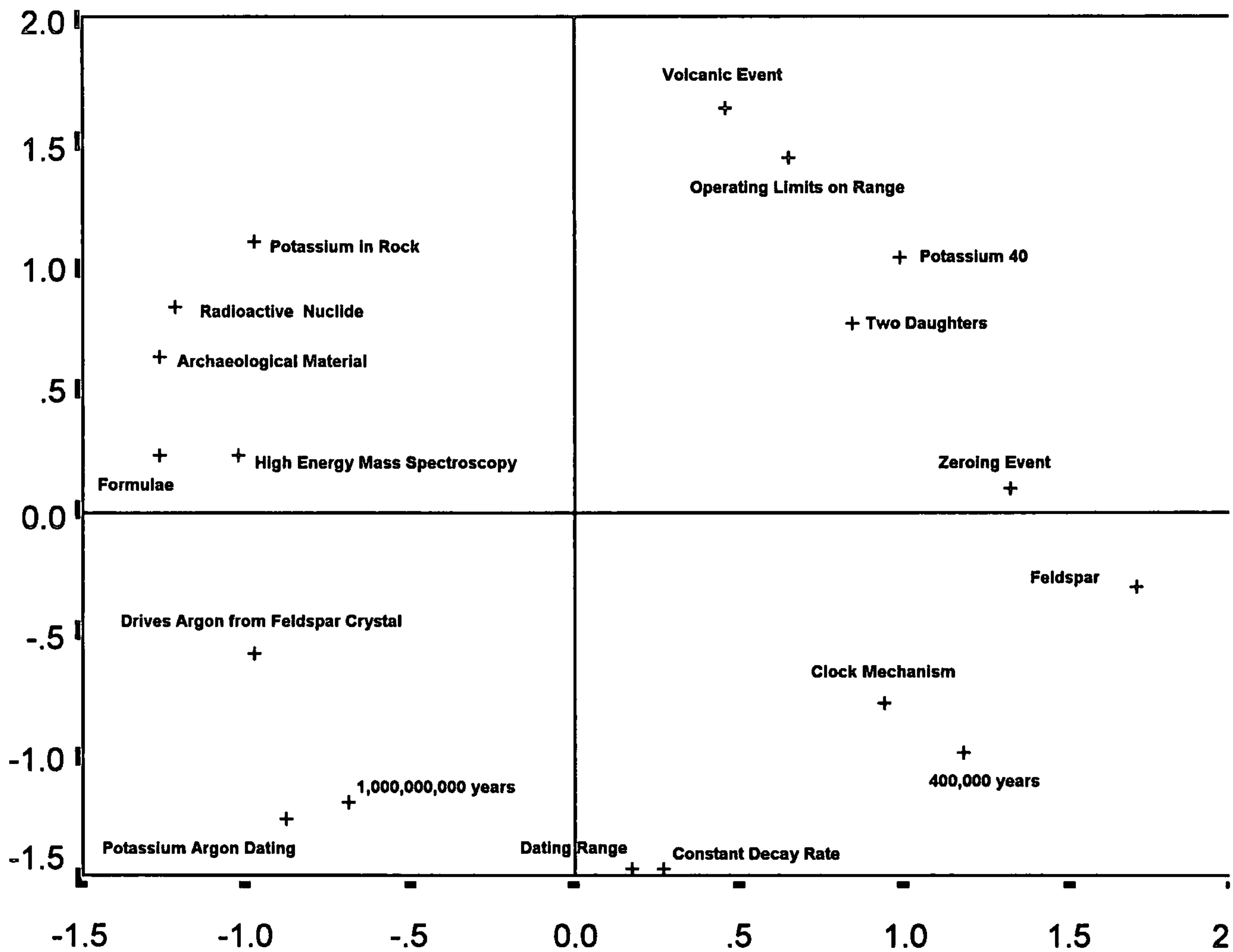
Figure 2 - MDS Plot for USD & GEN Expert Response.



In the expert plot the boundaries between GEN and USD are quite interlocked, with USD going from top left through the centre and round to bottom right and GEN from bottom right up to top left and centre right. To the left half of GEN we have Operating Limit, Dating Range, Archaeological Material, Clock Mechanism, Constant Decay Rate & Dating Range; to the right we have Zeroing Event, Formulae and High Energy Mass Spectroscopy. In USD we have Calcite, Uranium in Groundwater, 400 000 years and Stalactite Formation along the GEN boundary.

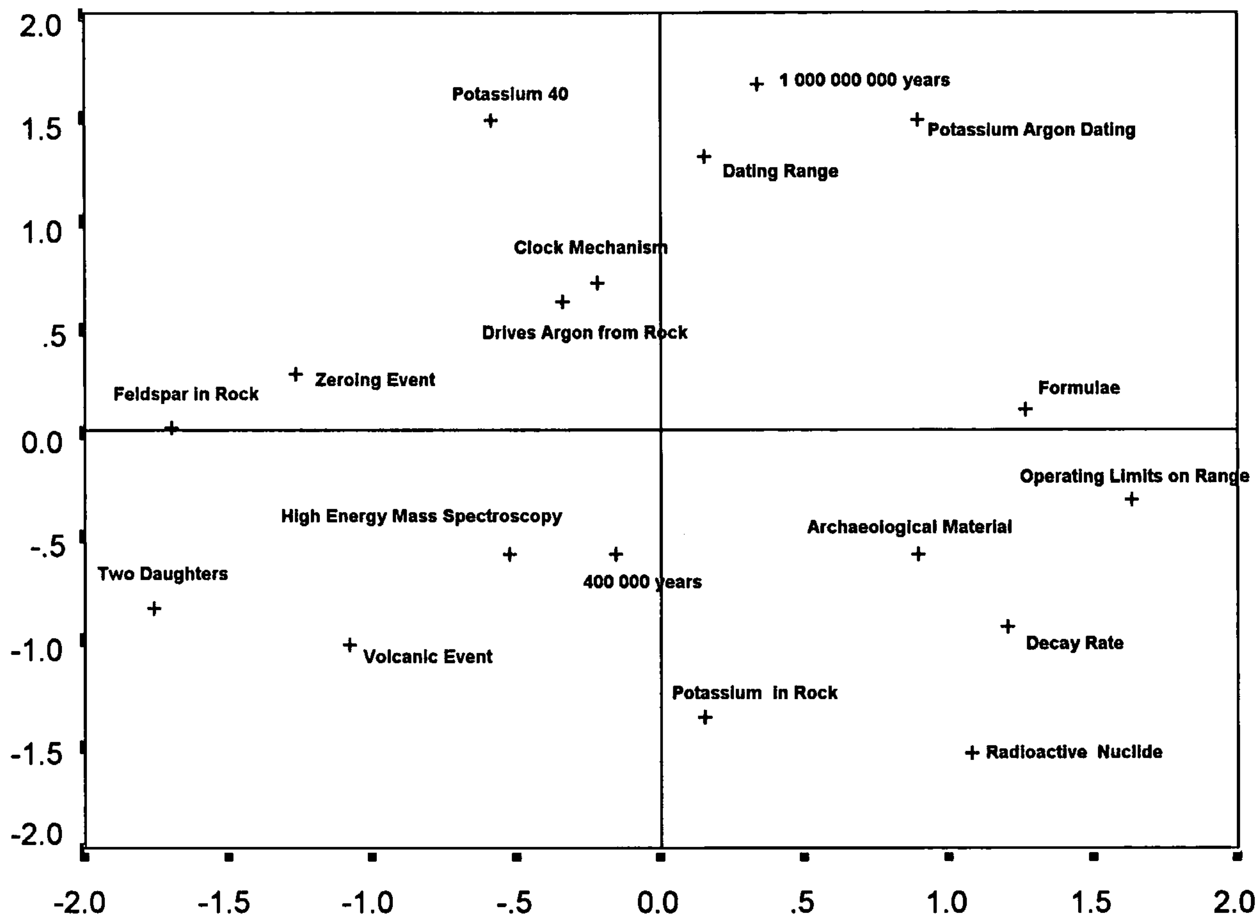
GEN & KAr

Figure 3 - MDS Group Plot for all KAr & GEN Novice Responses.



In this plot the GEN concepts are grouped to the left (Radioactive Nuclide, High Energy Mass Spectroscopy, Archaeological Material & Formulae) and centre bottom (Dating Range & Decay Rate) and right (Zeroing Event & Clock Mechanism). The KAr concepts are grouped top (Volcanic Events, Operating Limits on Range, Potassium 40 & 2 Daughters) and right (Feldspar in Rock & 400 000 years) and also bottom left (Drives Argon from Rock, Potassium Argon Dating & 1 000 000 000 000 years) and top left (Potassium in Rock). Of these, Zeroing Event & Clock Mechanism (GEN) are grouped near 400 000 years and Feldspar, whilst Potassium in Rock (KAr) is by Radioactive Nuclide & Archaeological Material.

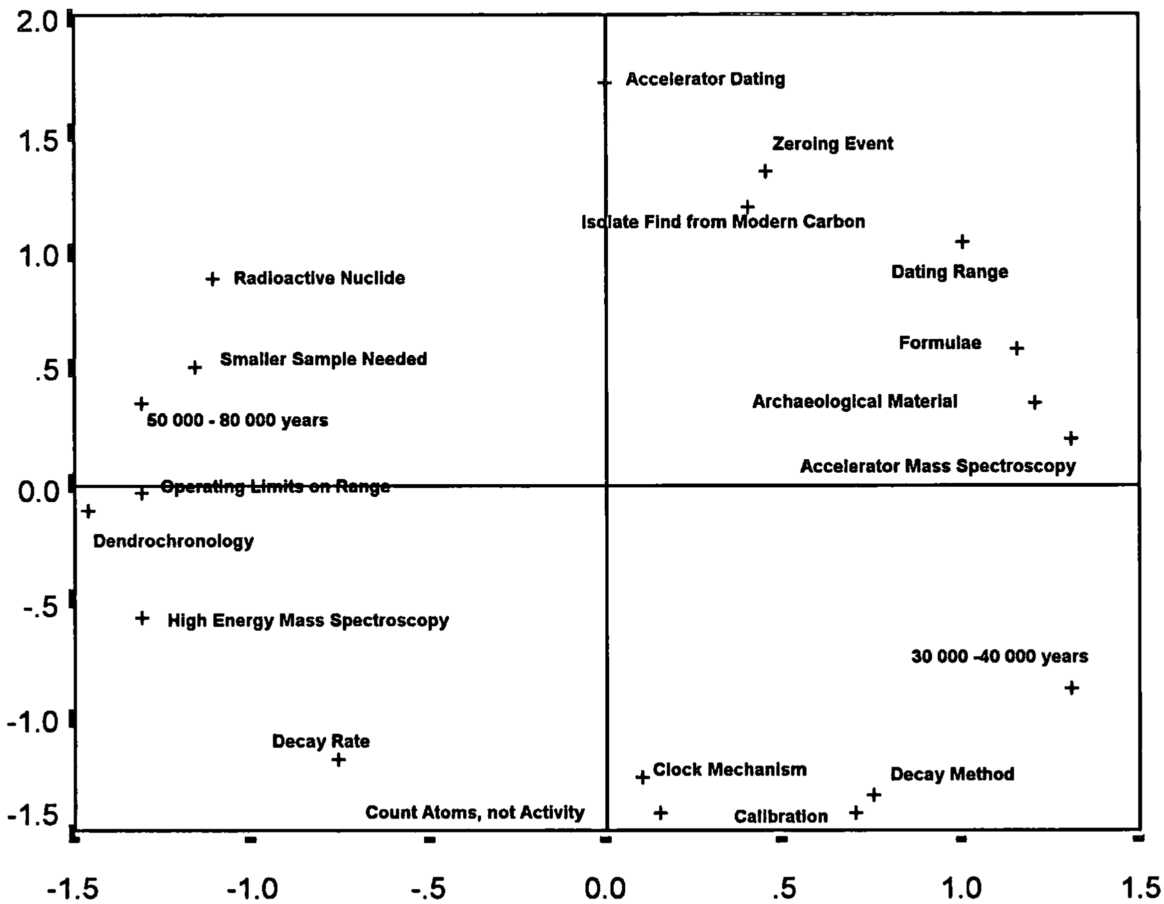
Figure 4 - MDS Plot for KAr & GEN Expert Response.



In the Expert's plot we have a rough divide with KAr on the left and GEN on the right, but there is a finger of KAr top right and of GEN going centre left. On the KAr finger Potassium Argon Dating and 1 000 000 000 years group next to Dating Range (GEN) whilst the GEN finger groups Zeroing Event and High Energy Mass Spectroscopy. This GEN group also sandwiches a KAr grouping of Drives Argon from Feldspar and 400 000 years between itself and Clock Mechanism. Radioactive Nuclide, Decay Rate and Operating Limits on Range group in the bottom right (GEN). Potassium in Rock, Volcanic Events and Feldspar are along the bottom of the GEN finger.

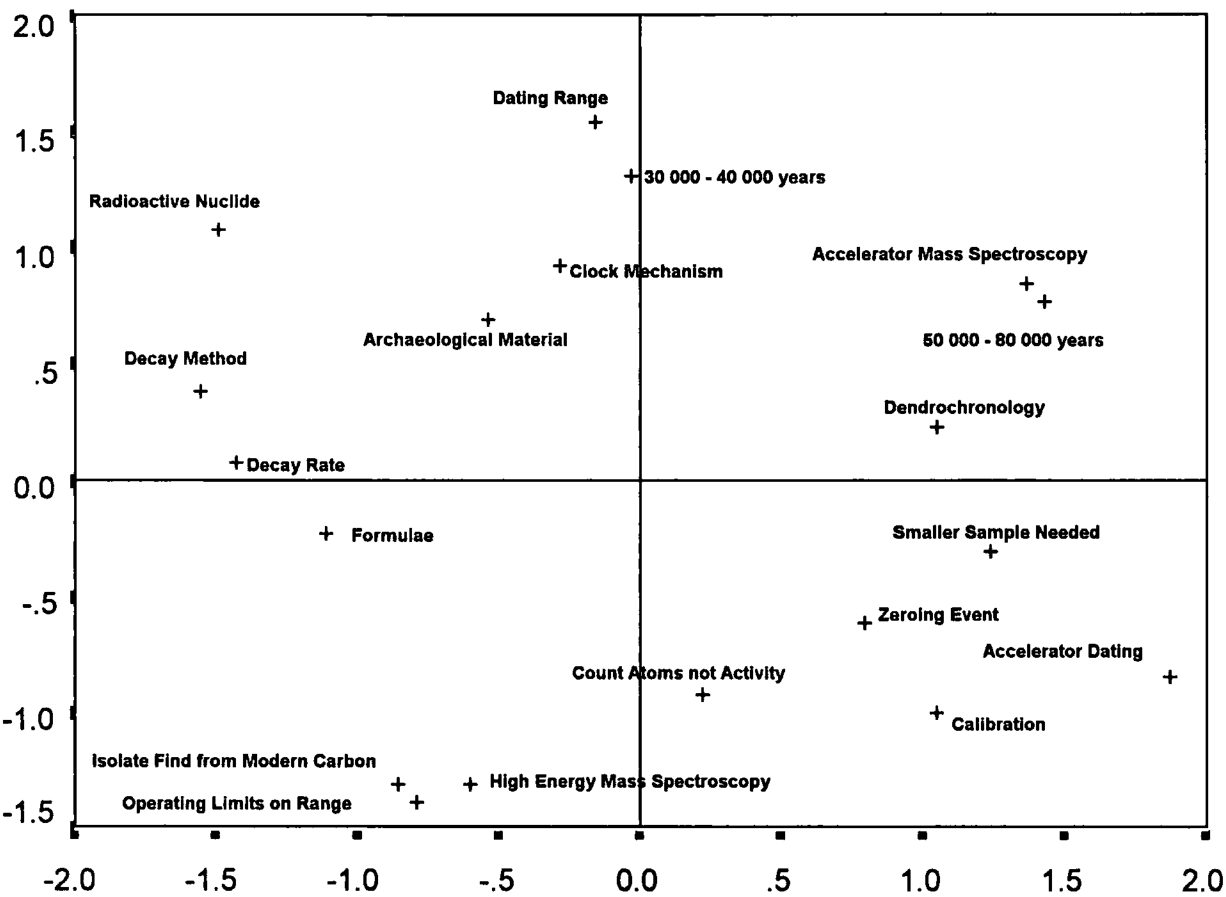
GEN & ARC

Figure 5 - MDS Group Plot for all ARC & GEN Novice Responses.



In the novice group plot we have groupings of GEN bottom left (Operating Limits on Range, Decay Rate & Clock Mechanism) and top right (Zeroing Event, Dating Range, Formulae & Archaeological Material). We also have ARC groups left (Smaller Sample Needed, 50 000 - 80 000 years & Dendrochronology) and bottom right (Count Atoms not Activity, Calibration, Decay Method, 30 000 - 40 000 years & Accelerator Mass Spectroscopy). Cross technique groupings are Zeroing Event (GEN) and Isolate Find from Modern Carbon at the top, Clock Mechanism & Count Atoms not Activity (ARC) at the bottom, Accelerator Mass Spectroscopy (ARC) & Archaeological Material and Formula on the right and Operating Limit on Range (GEN) with Dendrochronology, 50 000 - 80 000 years to the left.

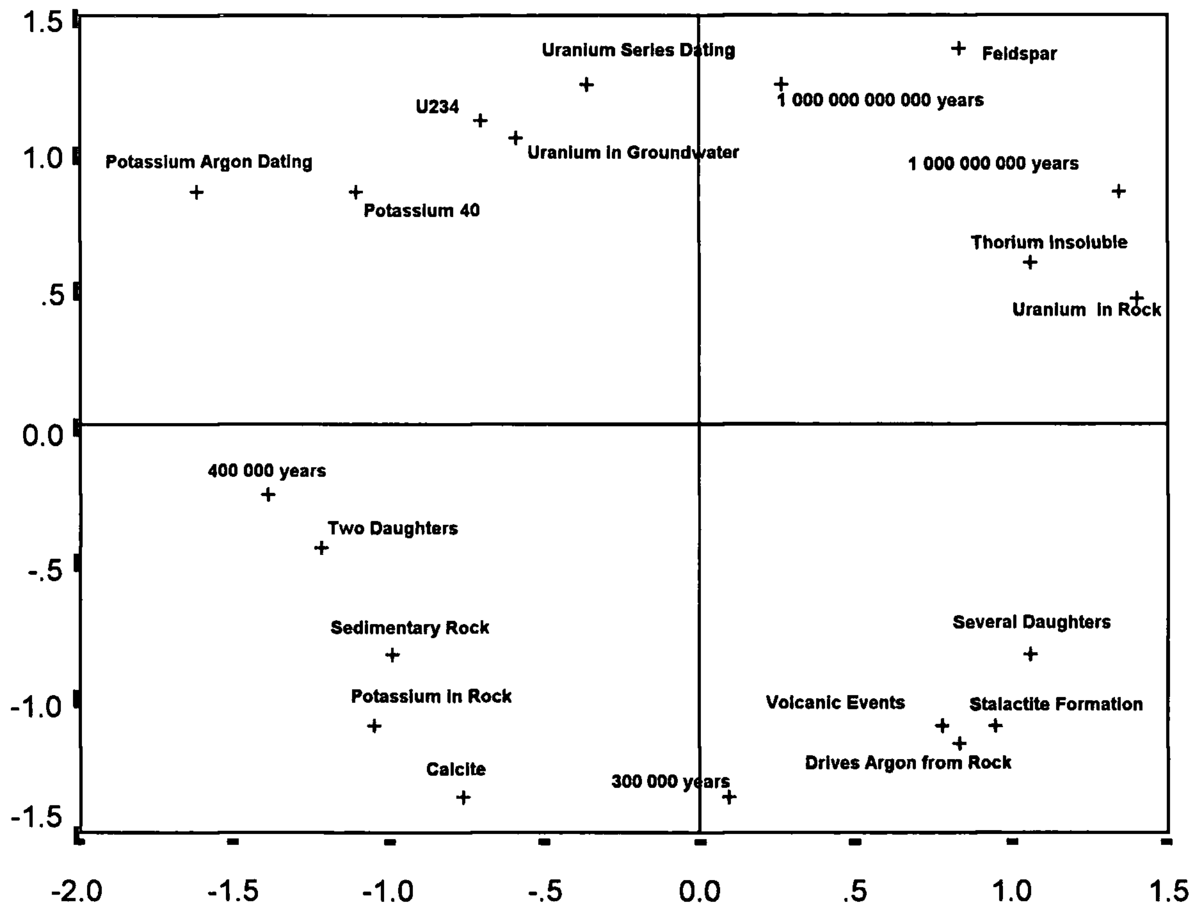
Figure 6 - MDS Plot for ARC & GEN Expert Response.



In the Expert plot we have ARC predominantly to the right hand side and GEN to the left. The exceptions are that Decay Method and Isolate Find from Modern Carbon (ARC) are very well into the left and isolated from the rest of ARC. Similarly 30 000 - 40 000 years is very isolated at the top and left of the ARC right hand side. Isolate Find from Modern Carbon is grouped by Operating Limits on Range; Decay Method is near Decay Rate; 30 000 - 40 000 years is grouped by Dating Range. In ARC, Accelerator Mass Spectroscopy is grouped with 50 000 - 80 000 years.

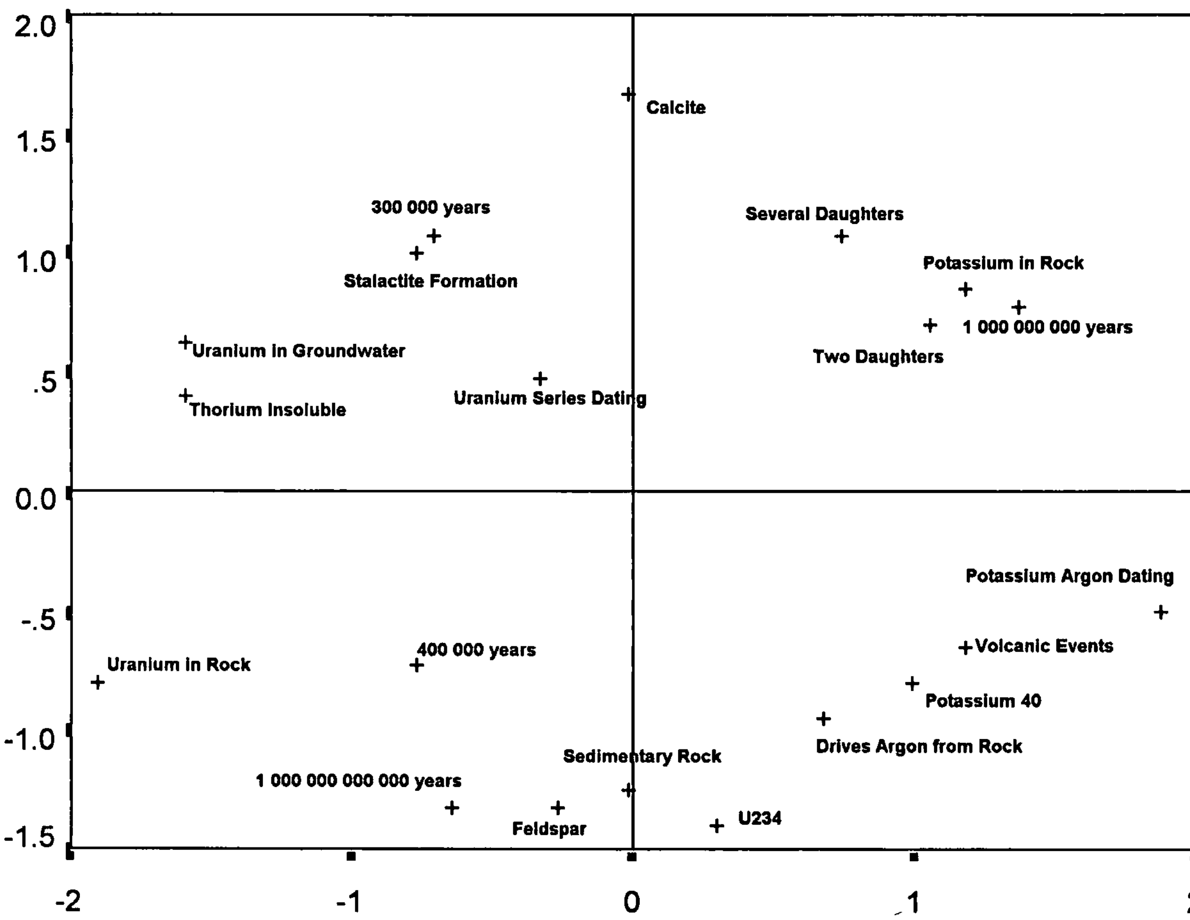
USD & KAr

Figure 7 - MDS Group Plot for all KAr & USD Novice Responses.



The groupings in the novice plot is predominantly left for KAr: Potassium 40 & Potassium Argon Dating (top left), 400 000 years, 2 Daughters & Potassium in Rock (bottom left). There is also 1 000 000 000 years & Feldspar (top right) and Volcanic Events & Drives Argon from Rock (bottom right). USD is predominantly right: 1 000 000 000 000 years, Uranium in Rock & Thorium Insoluble (top right), Several Daughters & Stalactite Formation (bottom right). There is also Sedimentary Rock, Calcite & 300 000 years (centre bottom) and Uranium Series Dating, U234 & Uranium in Groundwater (centre top). Crossover groupings are Several Daughters & Stalactite Formation (USD) and Volcanic Events & Drives Argon from Rock (KAr), and Sedimentary Rocks (USD) & Potassium in Rock (KAr).

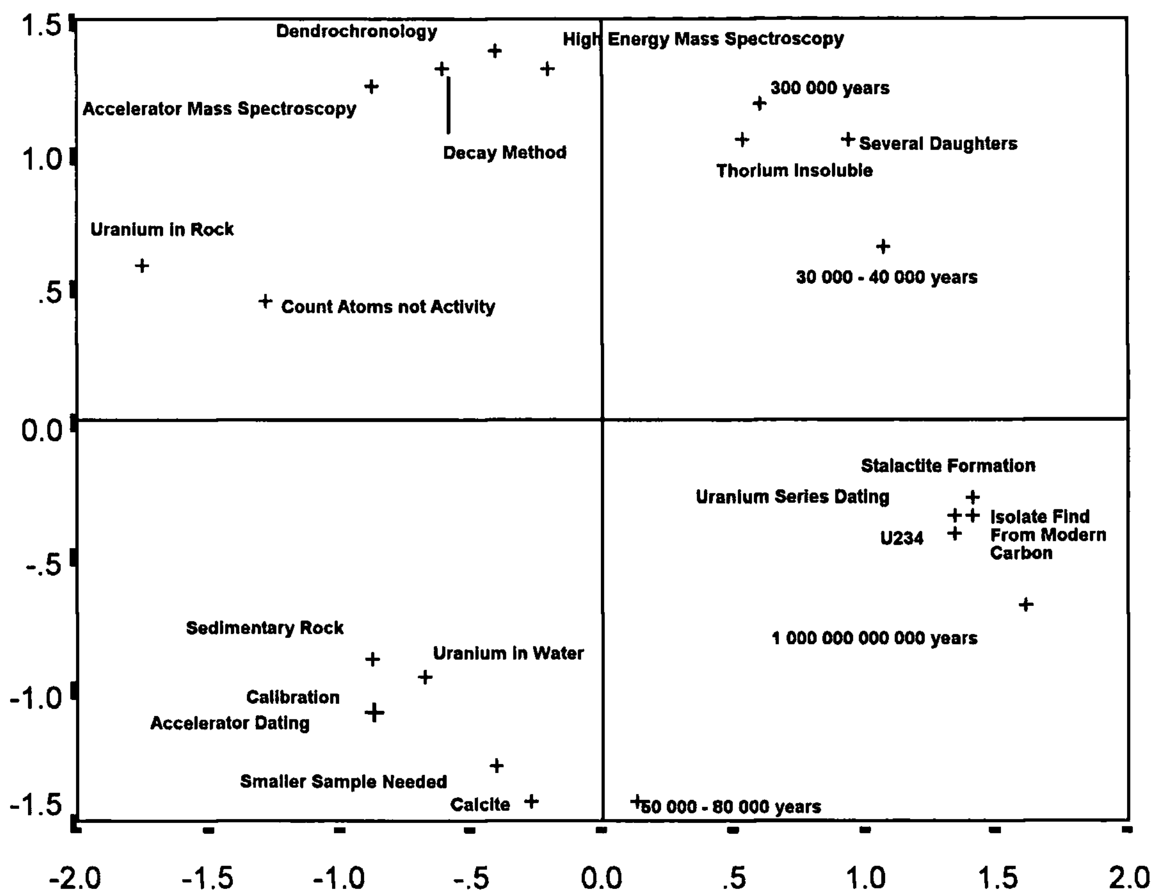
Figure 8 - MDS Plot for KAr & USD Expert Response.



The Expert here has roughly a split of USD left and KAr right. The grouping of 400 000 years & Feldspar (KAr) encroaches to the left, and 1 000 000 000 000 years, U234 and Sedimentary Rock (USD) encroaches to the left. There is otherwise no merging of the two techniques. Cross technique groupings do exist, with Several Daughters (USD) and Two Daughters (KAr) together, aswell as the overlap between the two encroaching fingers. Main groupings within techniques are: Two Daughters, 1 000 000 000 years & Potassium in Rock (KAr) and Thorium Insoluble, Uranium in Groundwater, Stalactite Formation & 300 000 years (USD).

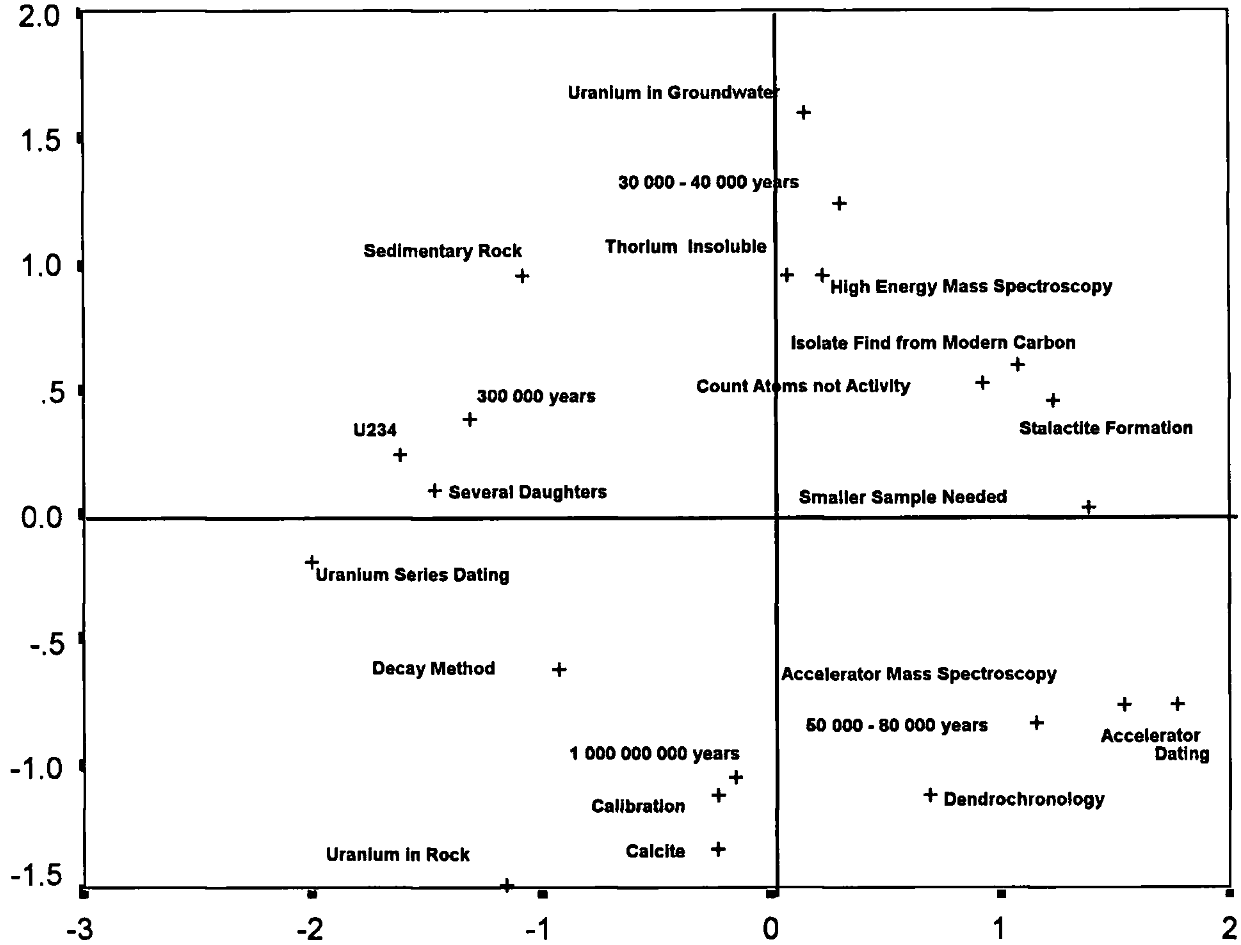
USD & ARC

Figure 9 - MDS Group Plot for all ARC & USD Novice Responses.



For this Novice Group plot there are several local groupings. For USD Sedimentary Rock, Uranium in Groundwater & Calcite (bottom left) with Calibration, Accelerator Dating & Smaller Sample Needed for ARC. There is also a USD grouping top left; 300 000 years, Thorium Insoluble, Several Daughters, with an adjacent ARC concept - 30 000 - 40 000 years. Left there is a USD grouping too - Stalactite Formation, Uranium Series Dating & U234 with an ARC concept - Isolate Find from Modern Carbon. There is additionally an ARC grouping top and left: Accelerator Mass Spectroscopy, Dendrochronology, Decay Method and High Energy Mass Spectroscopy.

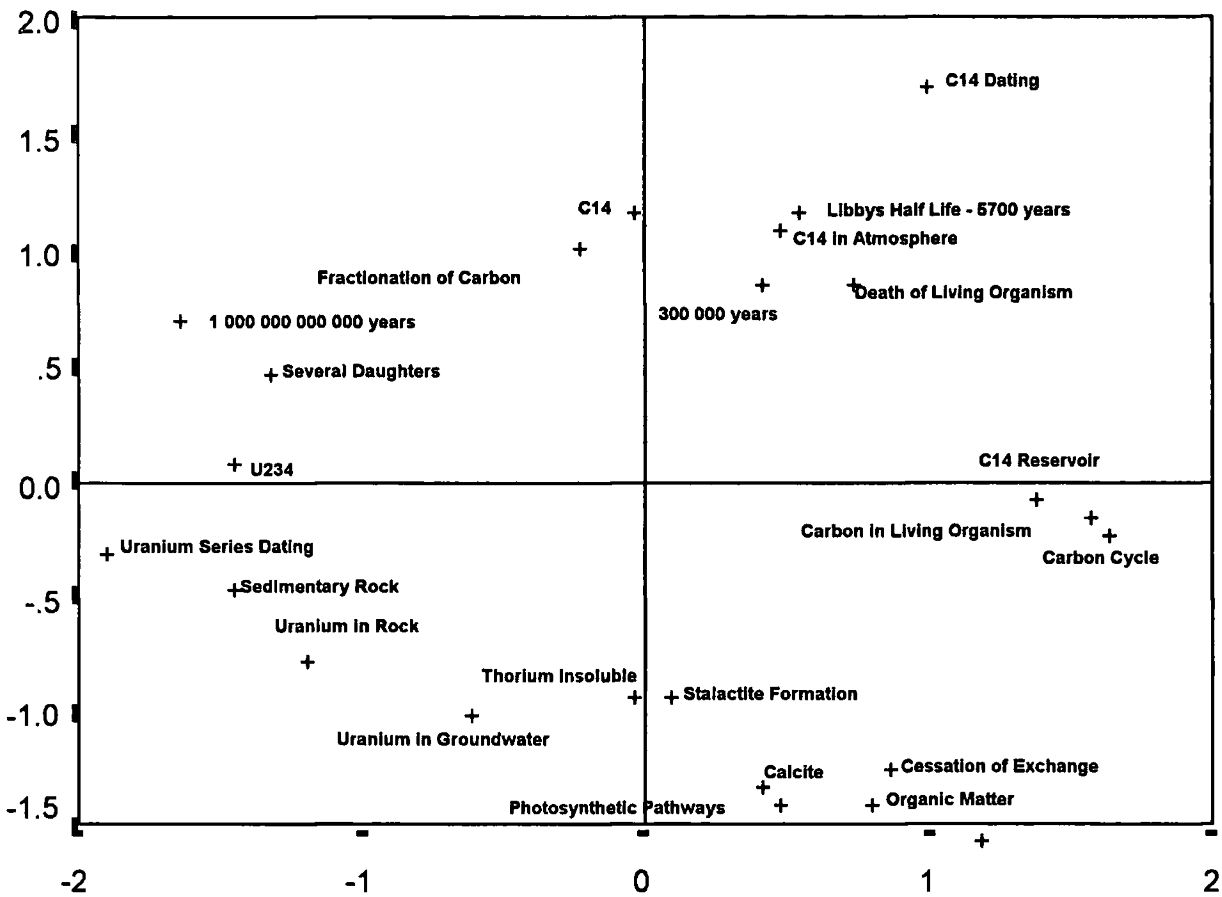
Figure 10 - MDS Plot for ARC & USD Expert Response.



This plot can be roughly divided with USD on the left half and ARC on the right half. Stalactite Formation (USD) is grouped with Isolate from Modern Carbon though (ARC) right on the right hand side, and Calibration (ARC) is grouped with 1 000 000 000 years (USD) at the bottom middle. Otherwise, groupings in the techniques are as follows. USD: Uranium Series Dating, U234, 300 000 years & Several Daughters (left) and Uranium in Groundwater & Thorium Insoluble (top). ARC: High Energy Mass Spectroscopy & 30 000 - 40 000 years (top right), Accelerator Mass Spectroscopy & Accelerator Dating (bottom right).

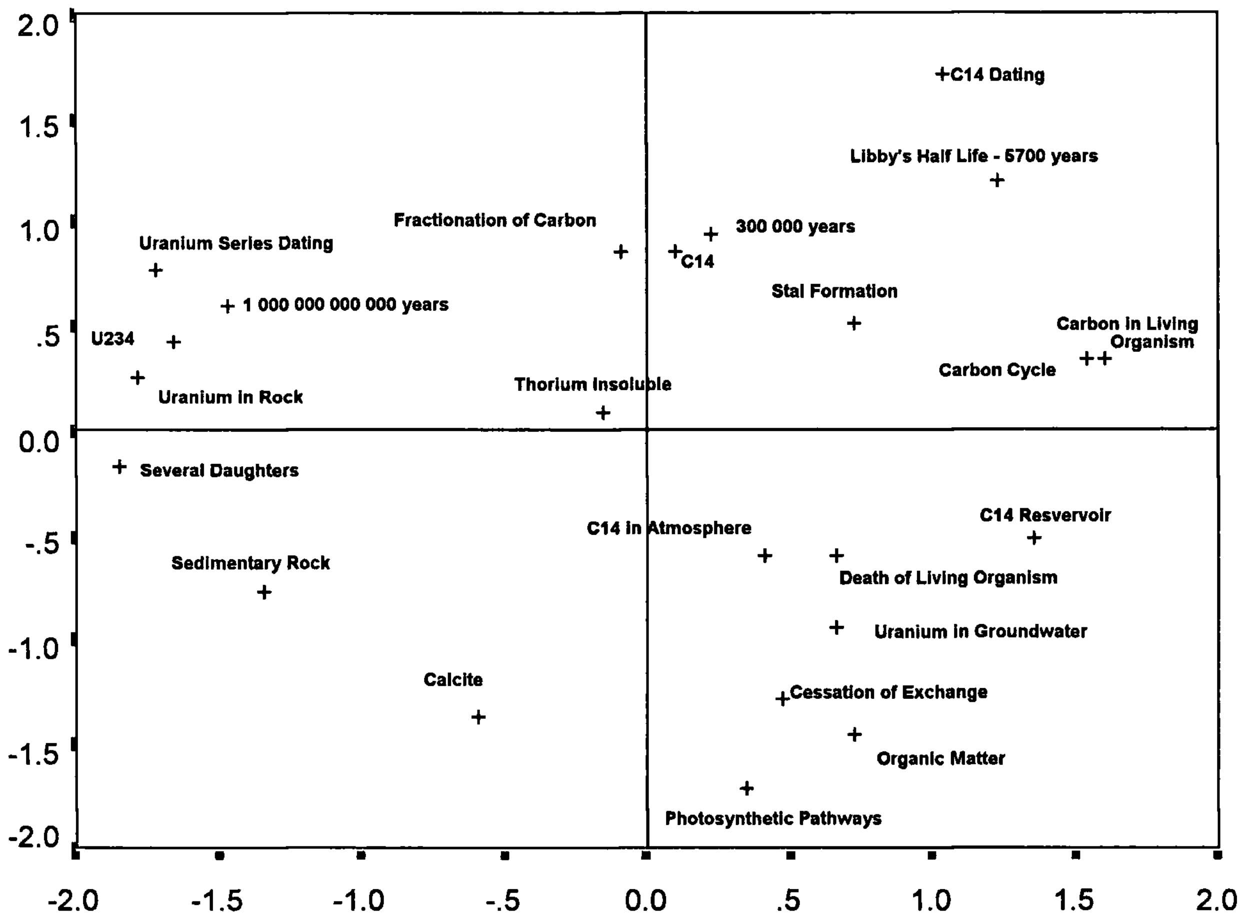
USD & NRC

Figure 11 - MDS Group Plot for NRC & USD Novice Responses.



This Novice group plot shows a roughly left right split, with USD concepts to the left and NRC concepts to the right. The USD concepts group with 1 000 000 000 000 years & Several Daughters top left; Uranium Series Dating, Sedimentary Rock & Uranium in Rock bottom left and with Thorium Insoluble, Stalactite Formation & Uranium in Groundwater bottom. Libby's Half Life - 5700 years & 300 000 years also group, but they are with the NRC concepts of C14 in Atmosphere & Death of Living Organism. NRC also groups with C14 Reservoir, Carbon in Living Organism & Carbon Cycle (right), C14 & Fractionation of Carbon (top) and Photosynthetic Pathways, Cessation of Exchange and Organic Matter (bottom right) which also group with the USD concept Calcite.

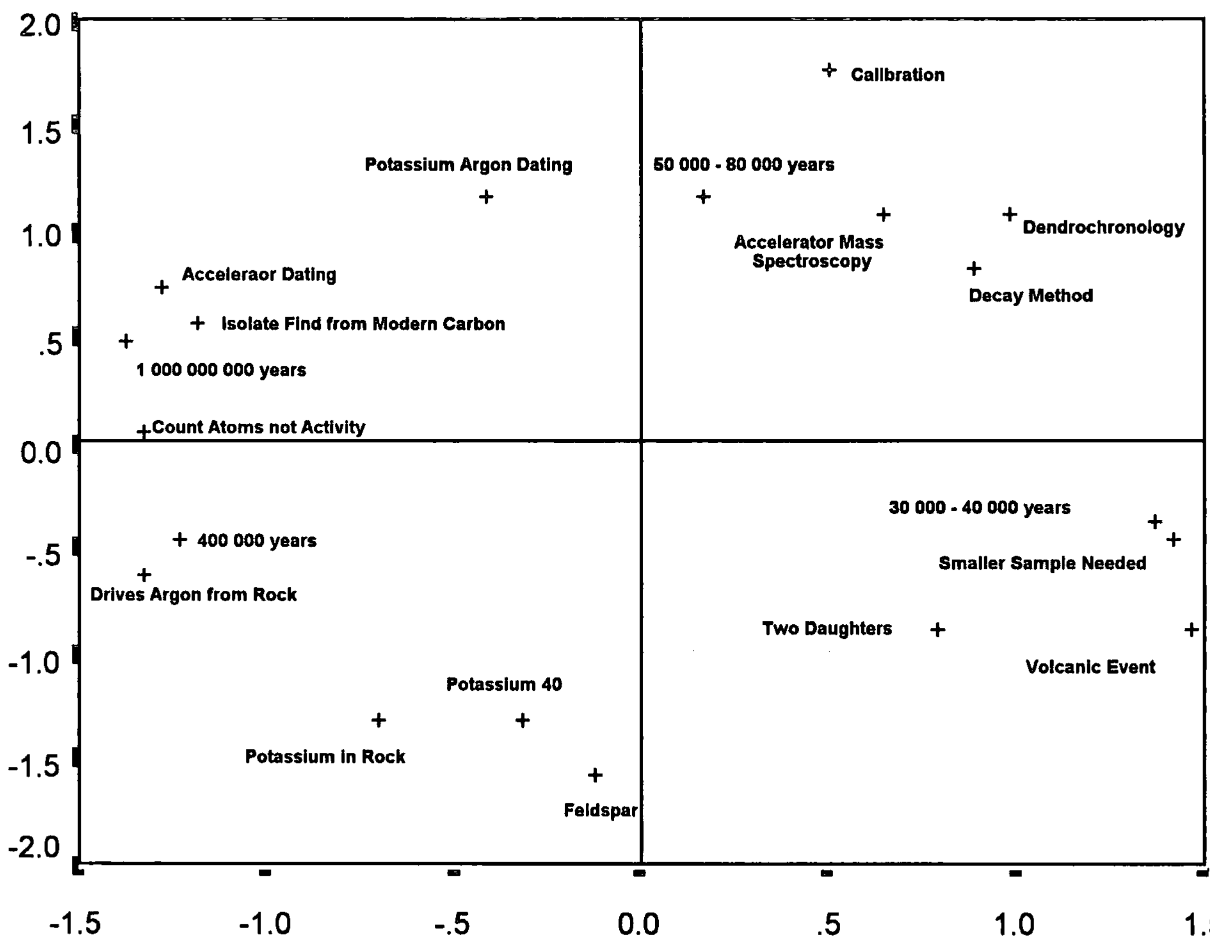
Figure 12 - MDS Plot for NRC & USD Expert Response.



The Expert's plot is mainly divided with USD concepts in the left hand side and NRC concepts the right. The main USD grouping is Uranium Series Dating, U234, Uranium in Rock, 1 000 000 000 000 years and Several Daughters (left). Two fingers of USD go into the right hand side, and around the edge of those fingers are Stalactite Formation, 300 000 years, Thorium Insoluble & Uranium in Groundwater. The NRC grouping between these fingers is Fractionation of Carbon, C14 & C14 in Atmosphere. Death of Living Organism & Cessation of Exchange are split by Uranium in Groundwater. Carbon Cycle and Carbon in Living Organism also group closely.

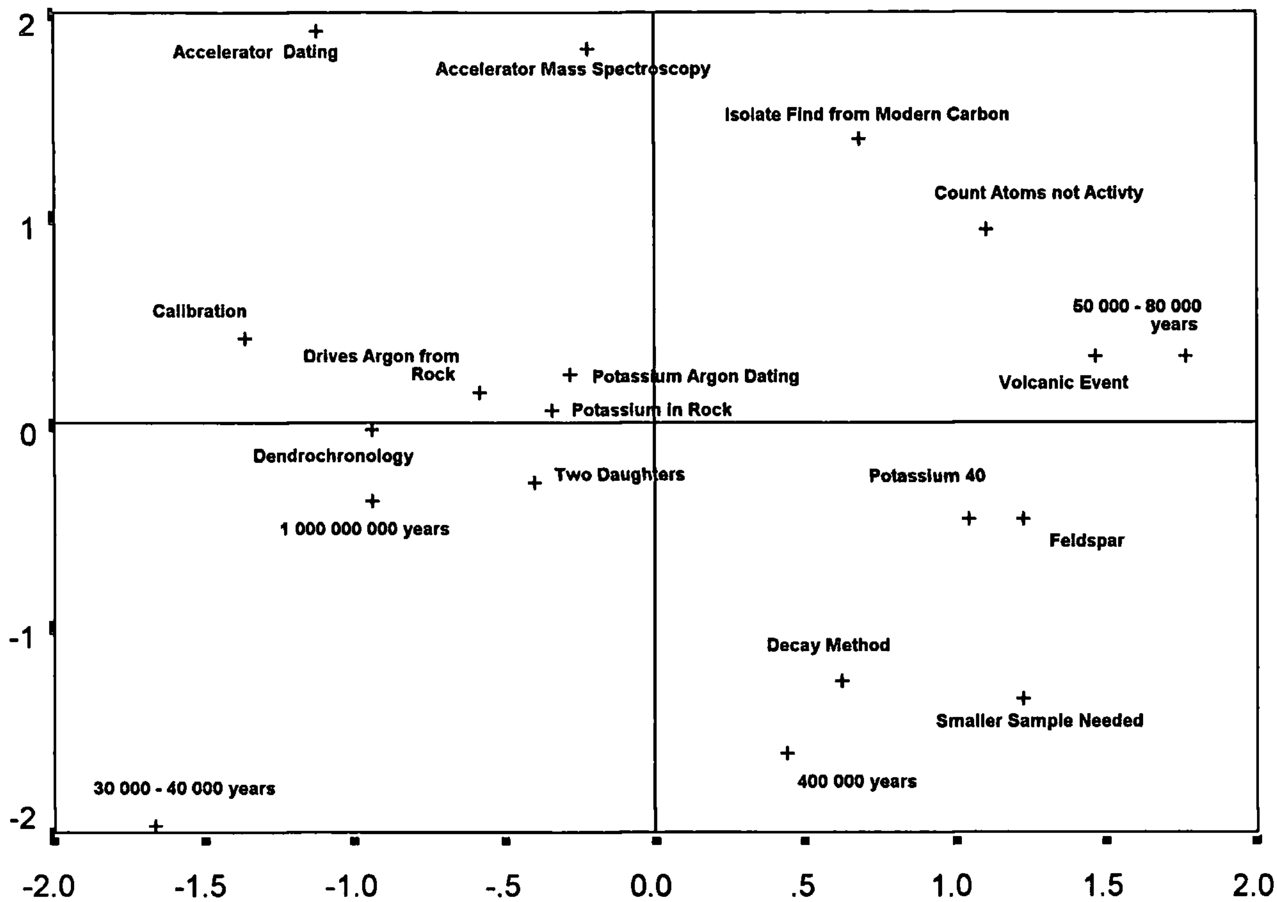
ARC & KAr

Figure 13 - MDS Group Plot for KAr & ARC Novice Responses.



The prominent groupings of concepts in this plot are ARC in the top and right and KAr in the left and bottom. To the top, the ARC concepts of 50 000 - 80 000 years, Calibration, Accelerator Mass Spectroscopy, Dendrochronology & Decay Method group; to the right it is Smaller Sample Needed & 30 000 - 40 000 years, with an interloper of Volcanic Event from KAr in the grouping. In the KAr groupings we have Potassium in Rock, Potassium 40 & Feldspar at the bottom with High Energy Mass Spectroscopy from ARC. To the left we have Accelerator Dating, Isolate Find from Modern Carbon, 1 000 000 000 years, Count Atoms not Activity, 400 000 years and Drives Argon from Rock.

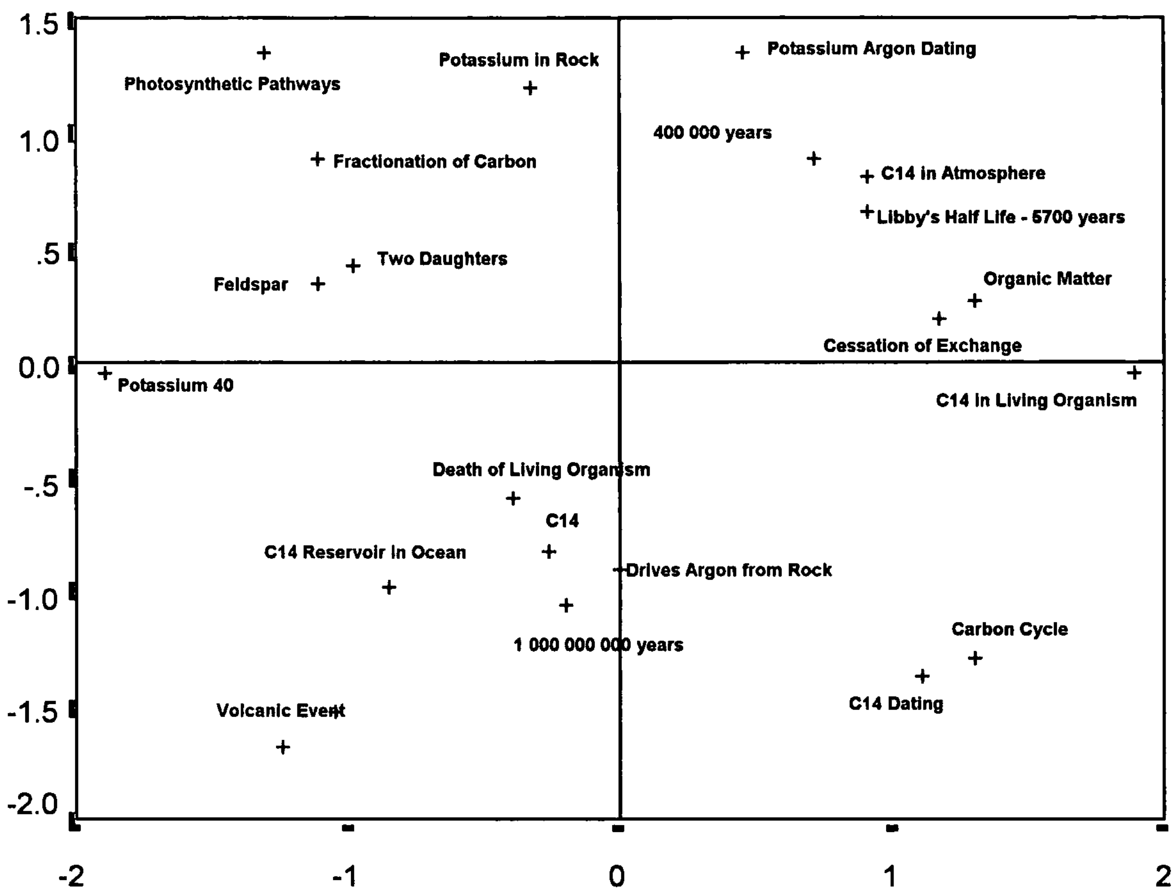
Figure 14 - MDS Plot for KAr & ARC Expert Response.



The Experts plot roughly divides with KAr running from centre right through to the left of centre, and with ARC above, below, and to the left of it. For KAr 400 000 years is isolated at the bottom below an ARC grouping of Decay Method & Smaller Sample Needed. Volcanic Event, Potassium 40 & Feldspar group to the right and Potassium Argon Dating, Potassium in Rock, Two Daughters, 1 000 000 000 years & Drives Argon from Rock group in the centre. For ARC 50 000 - 80 000 years, Count Atoms not Activity & Isolate Find from Modern Carbon are top right, Accelerator Dating & Accelerator Mass Spectroscopy are top left and Calibration & Dendrochronology are centre left.

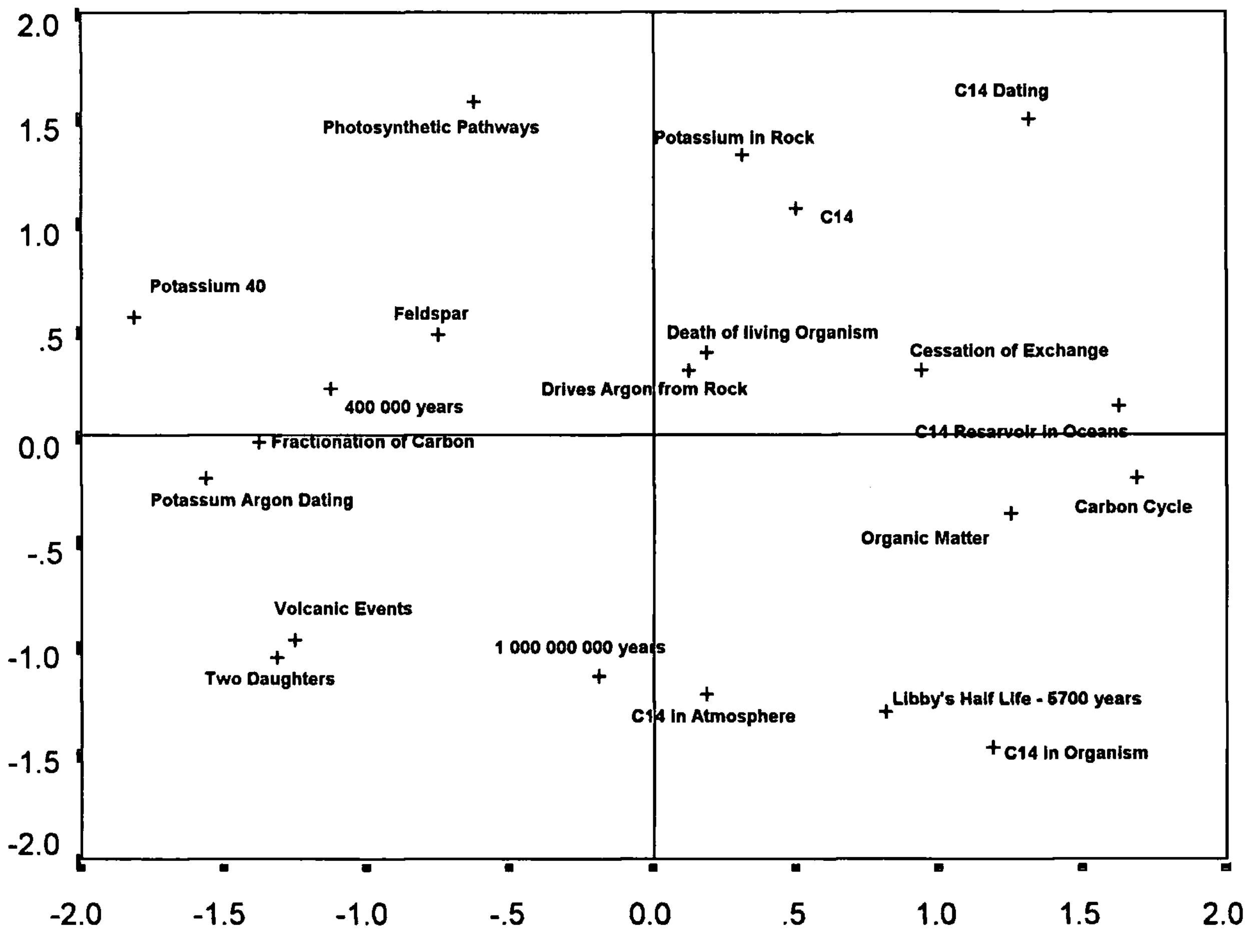
NRC & KAr

Figure 15 - MDS Group Plot for KAr & NRC Novice Responses.



The Novice representation appears to group the NRC concepts to the right and KAr to the left, but with an incursion of NRC into the left at centre, and an isolated pocket top left. Photosynthetic Pathways & Fractionation group top left, and C14 Reservoir in Oceans, C14 & Death of Living Organism group just under centre. To the right C14 Dating & Carbon Cycle group at the bottom, Cessation of Exchange & Organic Matter group in the middle and Libby's Half Life - 5700 years & C14 in Atmosphere group to the top. For KAr 400 000 years & Potassium Argon Dating group top left of centre, Two Daughters & Feldspar groups middle top left and Volcanic Event, 1 000 000 000 years & Drives Argon from Rock group bottom left.

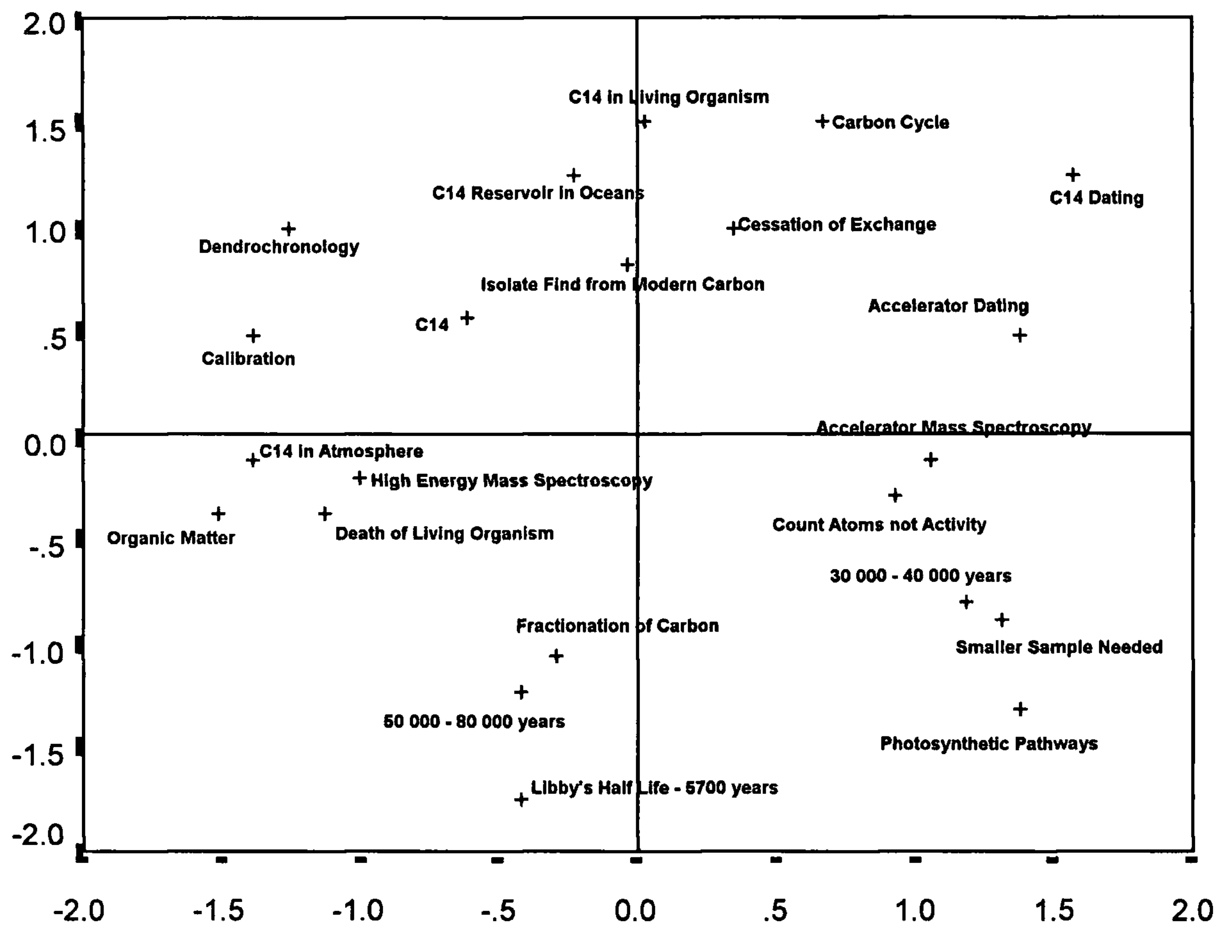
Figure 16 - MDS Plot for KAr & NRC Expert Response.



The Expert plot divides KAr on the left hand side and NRC on the right, with one concept grouping, Photosynthetic Pathways & Fractionation of Carbon (NRC) incurring on the top left hand side. There is also an obvious boundary grouping between techniques of Death of Living Organism (NRC) and Drives Argon from Rock (KAr) at the centre. In the KAr half, Volcanic Events & Two Daughters group together bottom left, whilst Potassium 40 & Potassium Argon Dating group left of the NRC incursion and Feldspar & 400 000 years group to its right. On the NRC half of the plot C14 & C14 Dating are at the top; Cessation of Exchange is close to Death of Living Organism; C14 Reservoir in Oceans, Carbon Cycle & Organic Matter group centre right and C14 in Organism, Libby's Half Life - 5700 years & C14 in Atmosphere are at the bottom.

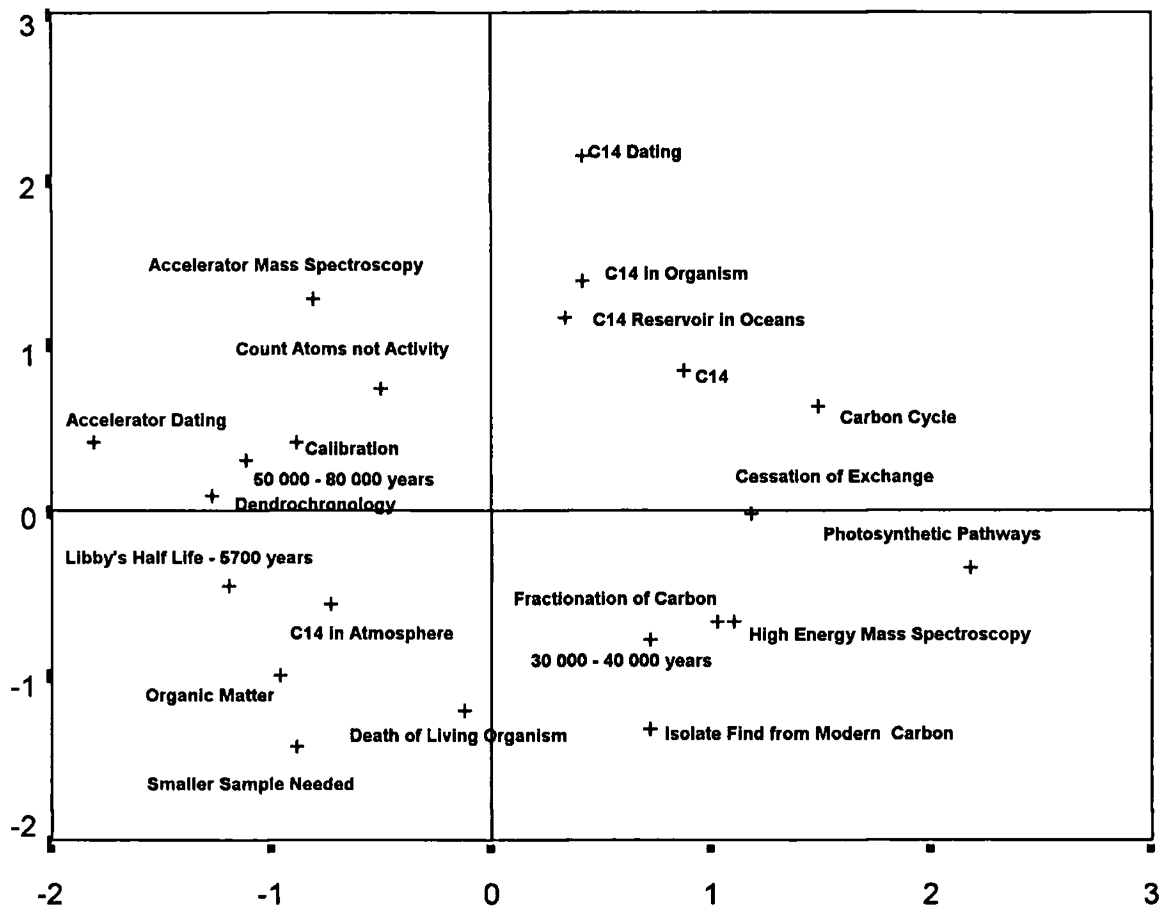
NRC & ARC

Figure 17 - MDS Group Plot for ARC & NRC Novice Responses.



The Novice Plot shows us NRC and ARC in all directions. For ARC concepts we have a grouping to the left of Dendrochronology & Calibration; to the right of Accelerator Dating, Accelerator Mass Spectroscopy, Count Atoms not Activity, Smaller Sample Needed & 30 000 - 40 000 years. For NRC concepts we have a grouping of C14, C14 Reservoir in Oceans, C14 in Living Organisms, Cessation of Exchange & Carbon Cycle to the top; to the bottom we have Fractionation of Carbon & Libby's Half Life - 5700 years together with the ARC concept 50 000 - 80 000 years. There is also a NRC grouping to the left of Death of Living Organism, Organic Matter & C14 in Atmosphere.

Figure 18 - MDS Plot for ARC & NRC Expert Response.



For the Expert this plot NRC is mainly in the top right with a finger extending centrally to the left. ARC is mainly left with an extension bottom right. In NRC the main grouping is C14 in Organism, C14, C14 Reservoir in Oceans & Carbon Cycle (right middle to top), Photosynthetic Pathways, Cessation of Exchange & Fractionation (right middle to bottom), and Libby's Half Life - 5700 years, Organic Matter, C14 in Atmosphere & Death of Living Organism (bottom left). For ARC the groupings are Accelerator Mass Spectroscopy, Accelerator Dating, Count Atoms not Activity, 50 000 - 80 000 years, Calibration and Dendrochronology (top left) and High Energy Mass Spectroscopy, 30 000 - 40 000 years, Isolate Find from Modern Carbon & Smaller Sample Needed strung out along the bottom of the NRC incursion.