1

Our Cognitive Architecture

Consider a student learning a new area such as elementary algebra, a new word processing program, a computer programming language, how to wire or repair an electric appliance or how to solve problems using equations of motion. On the surface, assuming appropriate motivation, this student's activity is likely to be universal. Time will be spent listening to a teacher, studying a text or manual, working through some worked examples and then gingerly attempting to accomplish a task or solve a problem. Performance is likely to be hesitant, slow and error-laden. Each step is likely to be accompanied by long pauses, considerable thought and frequent consultation of a tutor, if one is available. Failing an available tutor, then text- or manual-based information will be sought with, depending on the quality of the resources, greater or lesser degrees of frustration. The entire performance is likely to be seen by both the performer and any (probably unwelcome) audience, as difficult, time-consuming and clumsy.

Let us leave our universal student to study and practise for a few weeks, months or perhaps years, depending on the complexity of the material. When we return, a transformation has been wrought. Our awkward, ungainly student has metamorphosed into a confident, fluent virtuoso. Algebraic transformations can be completed easily and accurately, word processing commands can be used to alter text almost instantaneously, elegant computer programs can be written to accomplish complex tasks efficiently, word problems can be solved rapidly and easily. While on the surface, our novice student has been transformed into an expert, what has changed at a deeper level? Our student has learned but what does that mean? What are the cognitive processes involved in acquiring a complex intellectual task? Most important of all, what can educators do to facilitate the transformation of a poorly performing novice into an effective, proficient expert?

My research team and I have been concerned with these questions, especially the facilitation of skill acquisition, for about the last 20 years. Our task has been to find new teaching techniques and new instructional designs that can assist students' learning in complex areas. We have been motivated by two major factors. The first factor is that traditional instructional designs are just that: traditional. They tend to be governed by nothing more than tradition and intuition. There is no discernible theoretical or empirical research base determining the manner in which material is presented to students or the activities that students are encouraged to engage in during the learning process. Tradition and intuition are fine if no better alternative is available.

Our second motivating factor leads directly from the first. A better alternative is now available. In the past 20 years cognitive science has progressed impressively, giving us a substantial corpus of knowledge concerning our mental processes. Considerable information is available about our cognitive architecture — how we remember, learn, think and solve problems. Cognitive science has provided us with a plethora of information, based on theory and data, that potentially might be used to generate new instructional procedures. Thus, knowledge concerning our cognitive processes can be addressed to questions of instructional design. We no longer have to rely purely on tradition and intuition.

This book describes a variety of instructional designs that my research team has devised and the rationale behind them. For reasons of convenience, the research program has been concerned primarily with technical areas: mathematics, science, and technology education. In these areas, knowledge acquisition can be easily tested using well-defined problems that can be analysed in a relatively objective fashion. The nature of errors and misconceptions can be clearer than in more discursive areas. Nevertheless, the findings can easily be generalised to language-based, non-technical areas. I will begin by discussing those aspects of our cognitive architecture relevant to instructional design.

Working memory

Consider the following problem:

> Suppose five days after the day before yesterday is Friday. What day of the week is tomorrow?

Most of us find this a difficult problem to solve. Why? Are we using an inappropriate problem-solving strategy? Are we failing to use thinking skills that we should have learned? The answer to the problem is Wednesday and for those who feel that really was not a hard problem at all, here is another, adapted from Kotovsky, Hayes and Simon (1985).

> Three monsters, one small, one medium and one large, were each holding a globe. The globes came in three sizes only — small, medium and large — and each globe could be expanded or shrunk repeatedly to any one of these sizes but to no other size. The small monster was holding the medium globe, the medium monster was holding the large globe and the large monster was holding the small globe. They could change the size of the globes according to the following rules:
>
> 1 Only one globe could be changed at a time.
> 2 When two globes are the same size, only the globe held by the larger monster may be changed.
> 3 A globe must not be changed to the same size as the globe of a larger monster.
>
> What sequence of changes would allow the monsters to hold globes proportional to their size?

Most people find this a very difficult problem. Kotovsky et al. report that university students require about 30 minutes to solve it. What makes this problem and the day of the week problem difficult to solve while others are relatively easy? After all, there are no complex conceptual issues involved, with most people readily understanding both problems and their requirements. Furthermore, the solutions to the problems do not require a long, drawn out series of moves. The monster problem can, in fact, be solved in five moves. The solution is shown in Figure 1.

Figure 1 Solution to the monster problem

Monster size	Initial globe size	New globe sizes				
S	M	M	S	S	S	S
M	L	L	L	L	M	M
L	S	L	L	S	S	L

Now that you have seen the solution to the monster problem or solved it yourself, try solving the problem again. It should be easier, of course, but it is still likely to be quite difficult. Why should it be difficult to re-discover five moves that you have just been shown? The answer lies in one of the most basic characteristics of our cognitive architecture, brought to our attention many years ago.

Miller (1956) pointed out and codified a peculiarity of our memory system of which most of us are vaguely aware. Under some but not all circumstances, our memory seems to be exceptionally poor. If, for example, we are presented with a series of digits, we only are able to remember about seven of them. Miller called it the magical number 7, plus or minus 2, because the number of elements we can remember varies from about five to seven. An element can be anything: a digit, a letter, a word, etc. Irrespective of the nature of the elements, if there are more than about seven, we cannot hold them in memory.

It gets worse. Not only are we unable to hold many elements in memory, under some conditions, the few elements that we can hold are held in a very transient fashion. Peterson and Peterson (1959) found that if they gave people three-letter syllables to remember and prevented them from rehearsing by requiring them to count backwards in threes, they could not remember the letters for more than a few seconds.

The structure we use to initially hold information presented to us is, these days, called working memory. (Until fairly recently, it was called short-term memory.) The easiest way to think of working memory is in terms of consciousness. We are conscious of what is in working memory and not conscious of anything else. Indeed, my explanation of consciousness is in terms of working memory. The characteristics of working memory are the characteristics of consciousness and the most obvious characteristic of working memory is its extremely limited nature. We can hold few elements in working memory and those we can hold will remain for only a few seconds unless we are able to refresh them by rehearsal.

In fact, working memory is even more limited than the above description would suggest. The early studies on working memory investigated the extent to which we could hold or memorise a series of discrete elements. Normally, we are not in the habit of memorising lists of digits, letters or other such symbols. We use working memory to process information, where processing information involves contrasting, comparing, combining, relating or working on elements in some way. While we may be able to hold no more than about seven elements at a time, the number of elements that we can relate in some meaningful way is much smaller. We cannot measure the number of elements that can be worked on simultaneously in working memory because that number will depend on the nature of the processing that we are engaged in and as will be discussed below, what constitutes an element is much more complex than has been assumed to this point. Nevertheless, it is fair to guess that we can process no more than about two to four elements at any given time with the actual number probably being at the lower rather than the higher end of this scale.

We now are in a position to apply our knowledge of working memory to our attempts to solve problems such as the ones above. Consider the procedure you followed in your early attempts to solve the monster problem. You probably read the problem once and having assimilated the general context, found you could not even begin to proceed. In order to be able to solve a problem, you must be able to hold the start state of the problem in working memory. In this case, the start state is the initial size of the three monsters and the size of each of their globes. Most people are unable to hold so many elements and their relations in working memory. Pen and paper must be used to assist. A representation of each monster and its globe at the start of the problem must be drawn. Now for the first move. But what moves are legal? We cannot just change the size of any monster. Rather, we must carefully follow the rules for allowable moves. Unfortunately, we cannot hold those rules in working memory either. Each time we make a move, we have to check that move against each of the rules separately — our working memory is able to handle only that amount of information. To make matters worse, there is no point in simply making a legal move. We want to make a move that assists in achieving the goal of the problem. That means keeping in mind (i.e. working memory) that goal, the start state, differences between them that we wish to reduce, and the legal moves. Of course, with the limitations of our working memory where all of this has to happen, we cannot possibly manage such a task and so many, perhaps most, of our moves are essentially random. Many may be illegal because of the difficulty of holding in working memory the three rules governing legal moves. It is not surprising that it may take 30 minutes to solve the problem.

The above description may strike a chord with any readers who teach in areas where presenting students with problems to solve is a common pedagogical tool. Students, just taught a new area in, for example, geometry or physics, are faced with exactly the same difficulty as we have in solving the monster problem. While students may understand the new rules, theorems or equations needed to solve a problem in the same way as we understand the rules of the monster problem, understanding those rules, while essential, does not get one very far. In order to use the rules to solve a problem efficiently, both the rules and the specific characteristics of the problem to which they are to be applied must be held and manipulated in working memory. That, of course, is not possible given the limited capacity of our working memory, and the result is either time-consuming, inefficient solutions or a failure to find a solution. Clearly, mechanisms to circumvent the limitations of working memory must be available.

Let us consider again the implications of working memory limitations. All of our conscious remembering, learning, thinking and problem-solving activity occurs in a structure that apparently is minute in its capacity to handle the elements that we need to process in order to engage in normal cognitive activities. Furthermore, at times some people engage in cognitive processes that considerably exceed anything that we would wish to call 'normal'. How could we possibly engage in the intellectual heights that we know humans are capable of? Fortunately, we have another cognitive structure that, as has become increasingly apparent over the past 20 years, is critical to our ability to process information and is probably a central factor in our intellectual prowess.

Long-term memory

We all know we have a long-term memory because we can remember things we learned some time ago, in some cases, decades ago. Nevertheless, in many respects, long-term memory is a mystery. Exactly what does it hold, in what form is information stored, what

can be said about the quantity of information held in long-term memory? We have few intuitions concerning these questions because we have no direct consciousness of long-term memory. Working memory, not long-term memory, is the seat of our consciousness and we are only aware that something was stored in long-term memory when it is brought down into working memory. Otherwise, we have no awareness of the contents of long-term memory.

Probably because of its concealed characteristics, until fairly recently it was assumed that long-term memory played little part in complex cognitive processes such as thinking and problem-solving. Learning and memorising obviously rely on long-term memory but there seemed every reason to assume that higher cognitive functions depended largely on different structures. It took some time for the field to realise fully that long-term memory is more than just a passive store. Initial evidence concerning some unexpected characteristics of this structure came from a surprising source: the game of chess.

What distinguishes a chess master from a weekend player? What cognitive activities allow a chess master invariably to defeat a weekend player? It was questions such as these that de Groot* (1966) asked.

There are several obvious hypotheses that can be tested. Firstly, one can plausibly suggest that chess masters look ahead more moves than less capable players. By looking ahead further, they are more likely to discover a better move. Consistently looking ahead further should result in victory over one's opponent. In fact, de Groot found no suggestion that masters consider more moves in advance than weekend players. The difference in skill appeared to lie elsewhere.

While players differing in skill may not differ in the extent to which they look ahead, there is another equally plausible hypothesis that requires testing. At each choice point, skilled players may consider a greater variety of moves than less skilled players. If done consistently, this too should result in victory by enhancing one's chances of finding better moves. De Groot's results were no more promising than when testing the hypothesis that masters look ahead to a great depth. Masters did not seem to consider more moves at each choice point than novices. An extensive search for alternative moves at each point in the game seemed to be no more a characteristic of masters' play than an extensive search ahead.

*His initial work was published in the 1940s but was not widely available to an international audience.

If masters do not consider more moves than less skilled players, what do they do to win consistently? Clearly, there must be some cognitive factor that can account for the differences between different grades of players. De Groot's next test provided the first evidence of a clear distinction in the cognitive processes of players who differed in skill. He showed his masters a board configuration taken from a real game, removed it after a few seconds and asked them to reproduce it. They were able to do so with a high degree of accuracy. In contrast, less skilled players indicated a far lower level of accuracy on the same task.

While this memory test finally produced a difference between masters and less skilled players, in one sense it merely substituted new questions for the old one. Instead of asking what cognitive processes distinguish players at different levels, we are left with the questions: Why should differences in memory characteristics result in differences in playing performances? Do players with varying levels of skill differ in working memory characteristics? The answer to these questions becomes apparent when we consider some additional data. Chase and Simon (1973) provided evidence that the superior memory of masters is not general. It only applies to memory of realistic board configurations. Chase and Simon replicated de Groot's experiment and in addition, tested memory of random board configurations rather than configurations taken from real games. After viewing a random board configuration for a few seconds, chess players were required to attempt to replicate it. Unlike a real game configuration, there were no differences between more and less skilled players. Differences in working memory should show up whether the configurations are taken from real games or whether they are random. It seems chess masters only have a superior memory for board configurations taken from real games, not a superior working memory per se.

Why should chess masters be readily able to memorise real board configurations and only real board configurations, and more importantly, why should this enhance their level of skill? In fact, these results provide us with a possible, although surprising, explanation for the superiority of masters over less skilled players. It requires many years to become highly proficient at chess and we need to ask what is being learned during this period. It now appears probable that the major cognitive change is the accumulation of many general board configurations with appropriate moves associated with each configuration. The number of such configurations may be immense.

For example, Simon and Gilmartin (1973) have estimated that a grand master may have memorised as many as 100 000 board configurations.

The hypothesis that a major, possibly the major, difference between experts and novices lies in their knowledge of problem states (or configurations) and their associated moves can explain a very broad range of results. It can readily explain the fact that chess masters have an excellent memory of realistic board configurations but a no better than average memory of random configurations. It also explains why masters do not seem to need to engage in a more extensive search than less skilled players in order to find a good move. They know from previous experience which is the best move for many of the board configurations they are faced with. The hypothesis also explains another interesting characteristic of highly skilled chess players: their ability to play several games against several opponents simultaneously, acquitting themselves well against each opponent. The normal procedure in such exhibitions is to make a single move on each board and then move on to the next opponent, eventually returning to the initial opponent and repeating the exercise until the games are finished. How can a single person devise and remember each of a dozen successful game strategies against opponents who have far more time to think and who have before them continuously the materials about which they are thinking? In fact, grand masters do not need to engage in any such complex activity. As they arrive at each board, they merely have to recognise whatever configuration they are faced with, make the best move appropriate to that configuration and move on to the next board. Devising and remembering a unique strategy for each game is unnecessary. Knowledge of problem states and their associated moves is all that is needed.

While the initial results on memory of problem states were obtained using chess, similar results now have been obtained in a wide variety of intellectually complex tasks. For example, Egan and Schwartz (1979) briefly showed novice and expert electronics technicians electronic wiring diagrams which they were then required to reproduce from memory. The experts performed at a higher level. This difference disappeared when the experiment was repeated using random configurations. Chiesi, Spilich and Voss (1979) had learners read prose passages related to baseball. Learners who had a high level of knowledge of the game demonstrated a recall of event sequences that was superior to that of learners with less knowledge of baseball. Sweller and Cooper (1985) found that university students and senior high school students had better memory of simple algebraic equations than junior high school students. This difference was not apparent using random algebraic symbols.

These findings provide us with a startling conclusion. Long-term memory is not simply a repository of rote learned facts. Rather, it contains sophisticated structures that permit us to perceive, think and solve problems. Skilled intellectual performance comes from the ability to recognise appropriate circumstances and the actions required by those circumstances. This ability comes from long, deliberate practice (Ericsson & Charness, 1994) that permits the appropriate cognitive structures to be acquired and held in long-term memory. What are the characteristics of those structures? This question is addressed in the next section.

Schemas

Assume you are outside and see a tree. A simple enough event, you may think, but consider what has actually happened. Each tree is unique in the sense that every tree differs from every other tree. It is unique in the sense that the number, colour and shape of its leaves, the contour of its branches and its general profile are unlike that of any other tree. What you have perceived is an entity unlike any other entity and yet you recognise the object as a tree virtually instantaneously. The cognitive processes used to identify trees are equivalent to those used to identify a variety of entities including concrete objects, symbols such as those used in writing or mathematics, problem types and their solutions, and even abstract concepts.

The structures that permit us to recognise objects and other entities are called schemas. A schema is defined as a cognitive construct that permits us to treat multiple elements of information as a single element categorised according to the manner in which it will be used. Thus, in the above example, a tree schema, held in long-term memory, allows us to ignore most of the enormous perceptual detail of any tree that we see and instead, consider one tree in the same manner as we might consider any other tree.

Schemas are essential to cognitive functioning and the educative process. We are able to read because of the huge number of schemas acquired over many years. All text that we see differs from all other text, with the differences being particularly marked in the case of handwritten text. We can recognise a literally infinite variety of shapes as the letter *a* despite not having seen a particular, handwritten example of the letter previously because we have a schema for *a* that permits us to ignore irrelevant details. Furthermore, we not only have schemas for the letters of the alphabet, we have higher-order schemas for combinations of letters that form words and combinations of words that form phrases and sentences. That is why we are able to read so rapidly. We can ignore most of the detail of the squiggles that appear on a page and concentrate on meaning because we have appropriate, pre-acquired schemas, held in long-term memory.

Schemas are equally useful in problem-solving (see Chi, Glaser & Rees, 1982; Hinsley, Hayes & Simon, 1977). The chess masters discussed in the previous section are able to reproduce chess configurations taken from real games because they have acquired schemas for tens of thousands of configurations. These schemas tell them which moves are the best moves when faced with particular circumstances. Similarly, electronic technicians have schemas for wiring diagrams and computer programmers for computer code that can be used to solve problems. In effect, a schema allows us to see a problem state as a single entity rather than a collection of elements. It is far easier to recall material organised by a schema because of the reduction in the amount of material that needs to be held in working memory. If we have a schema for the equation $(a + b)/c = d$ then we may require little more than a single element in working memory in order to reproduce it. If we have not acquired an appropriate schema because we have not learned any algebra, reproducing the equation from memory will require memorising the nine elements that constitute the equation. Not only will that task be difficult, it also will be inordinately difficult to solve problems based on the algebra of this equation.

What are the functions of schemas? Firstly, and most obviously, they provide the means of storing huge amounts of information in long-term memory. On current assumptions, our intellectual prowess is derived largely from the schemas we have acquired over long periods of time. We are able to read easily and fluently because we have spent many years developing reading schemas. We are able to understand concepts in our areas of expertise because we have schemas associated with those concepts. The concepts of a pronumeral in algebra, a right-angle triangle in geometry, a force in physics or a computer's RAM are all intelligible to us because they are represented by a small number of the myriad schemas held in long-term memory. Some of these schemas can be combined with other schemas to form higher-order schemas that allow us to solve problems readily. Few people reading this book, faced with the problem $(a + b)/c = d$, solve for a, will attempt to solve this problem by attempting to remove the addend in the numerator on the left-hand side as their first move. It is appropriate to ask why. There is no basic mathematical reason why, for example, subtracting b from both sides should not be attempted as a first move and indeed, students who have just learned the relevant rules of algebra are highly likely to attempt to solve the problem in exactly this fashion. Most students will read from left to right and top down. Following this route, b will be the first pro-numeral encountered that needs to be dealt with to solve the problem. In contrast, we know that attempting to subtract b from both sides will simply result in an impasse and that the appropriate first move is to multiply out the denominator. How do we know? According to schema theory, we have a schema for this particular problem and other schemas for different problems such as $(a + b)c = d$, solve for a. These schemas are very numerous and were acquired over a long period of time, and their function is to allow us to recognise problems and their solutions. Fluent problem solutions are generated by schemas.

Schemas have a second, equally important function. When presented with one of the above algebra problems, in order to solve it, we need to hold and algebraically manipulate the relevant parts of the problem in working memory. For an algebra student, it is a major task to consider all of the symbols, search for ones that require manipulation by comparing the initial state of the problem (e.g. $(a + b)/c = d$) with the final state (a isolated on the left side), then find an algebraic rule that will accomplish an appropriate transformation. In contrast, if a schema has been acquired, instead of processing all these elements in working memory, a single element — the schema — is all that needs to be held in working memory. The schema itself incorporates all of the elements including the solution. Most readers will not only be able to hold the expressions in working memory, but be able to mentally (i.e. using working memory) go through the solution. Furthermore, that can be done even if part of the expression, such as the denominator on the left side, is inordinately complex rather than merely the pro-numeral, c. A schema tells us that the complexity of that denominator is irrelevant to the solution. That schema, by reducing the burden on working memory, allows working memory to be used for other, more complex algebraic tasks.

The seemingly simple task of reading can be used to demonstrate the importance of schemas in freeing working memory space. Currently, in reading this page, you are faced with a hugely complex set of squiggles that in the normal course of events you could not hope to hold, let alone process, in working memory. A large number of schemas for letters, words, combinations of words and phrases, allows you to devote working memory to the more important task of considering meanings and ideas. Thus, schemas not only allow us to hold material in long-term memory, they massively reduce the burden on working memory, allowing us to accomplish intellectual tasks that otherwise would be impossible.

In summary, we can say that experts in an area differ from novices in that they have previously acquired schemas that can be used to encode the elements of each intellectual task encountered into a single entity or a limited number of entities. Novices, not having access to relevant schemas, must attempt to remember and process the individual elements that constitute the problem state. The difference in difficulty between these two tasks is reflected in the different memory performance characteristics of novices and experts and the differences in skill.

Instructional implications

The suggestion that skilled performance in complex domains is dependent on the acquisition of an enormous range of schemas has implications for the way in which we organise instruction. If processes similar to those used in chess are required for competence at solving, for example, mathematics or science problems, then this may change our conceptions of what needs to be learned in these areas. Commonly, the assumption of most educators has been that the major task of students is to learn the subject matter of the discipline. Effectively, this means learning the relevant rules (e.g. geometry theorems, kinematics equations, etc.). It may well be that learning the relevant rules, while clearly essential, is no more sufficient for competence in mathematics or science than is learning the relevant rules of chess for competence in chess. As in chess, the most important knowledge in subjects such as mathematics and science may be learned only after the rules have been learned. The major component of expertise that needs to be acquired may not be the rules to transform problem states but rather the schemas that allow one to recognise problem states and their associated moves. For example, learning the limited number of equations associated with kinematics may be simple. Learning which equations should be used with which problems may require far more time and effort.

Clearly, we must place a substantial emphasis on schema acquisition. Without schemas, there is no expertise. Students need to learn to recognise problem states and their associated moves. A student faced with the problem:

> A car uniformly accelerated from rest for 20 seconds has a final velocity of 60 kilometres per hour. What is its acceleration?

needs to be able to recognise this problem as one in which the solution is obtained by using the equation $V = at$ because V and t are given and a is the only unknown. A competent student will be able to scan this problem rapidly, note the two givens and note that the goal is the only unknown in the relevant equation. Furthermore, the student should be able to do this irrespective of the details of the wording. The relevant schema will allow the student, without undue reference to the surface structure, to categorise the problem correctly as one incorporating V and t as knowns and a as the unknown. Similarly, a competent algebra student must be able to immediately categorise a problem such as $(a + b)/c = d$, solve for a, as one requiring multiplying out the denominator as the first move. The ability to recognise these problems as ones belonging to certain categories requiring certain moves for solution appears to be no different from that of a chess expert who can recognise a board configuration and the best moves associated with it. These skills, based on schema acquisition, are an essential component of expertise. Procedures for facilitating the acquisition of schemas will be discussed in subsequent chapters.

Automation

While schema acquisition is an essential component of problem-solving skill, automation is an equally important factor. Information can be processed in either conscious or automated form. An automated process is one that can be carried out with minimal conscious thought. If you were asked to read this text aloud to someone, you could do so effortlessly. You would not have to think carefully about the letters and words. They are processed automatically. A child, with

less reading experience, could read the text also but could not do so automatically. Considerable thought might need to be given to some of the less familiar words.

Kotovsky, Hayes and Simon (1985), working with the monster problem discussed above, provided a large impetus to the realisation that automation was an important variable in problem-solving performance. They found that if people memorised the rules of the problem to the point where they could repeat them effortlessly, problem solution was much easier. By not having to consider the rules consciously each time a move is made, working memory is freed to plan a problem solution. Thus, in one sense, automation has a similar effect to schema acquisition. Both processes have the effect of freeing working memory. We can conclude that automation may be one of the major keys to successful problem-solving performance. Without it, planning behaviour may be difficult or, in extreme cases, impossible because of insufficient cognitive capacity. Once automation has occurred, this capacity is released from monitoring the legality of problem moves to planning useful move sequences. The ability to plan should facilitate solution.

The concept of automation can be applied readily to schemas as well as rules. Once a schema has been acquired, it can undergo a process of automation in the same way as a rule. Initially, a student may need to engage in considerable mental effort to classify a problem state such $(a + b)/c = d$ and to remember which problem-solving operators are relevant to this state. With experience, this procedure may become more and more automated until, like the readers of this book, the student can accomplish it almost immediately with little cognitive effort. There is no reason to suppose that this process of schema automation is any different from the process of, for example, automating the rule which allows a denominator to be multiplied out from one side of an equation.

We are now in a position to provide a more complete profile of an expert problem solver. On being presented with a problem in his or her area of competence, there is a high probability that the expert will have an automated schema that can be used to classify the problem appropriately and generate a series of moves. The greater the degree of expertise, the greater the probability that a schema will be available to govern performance. Neverthless, most experts will on occasion be faced with novel problems for which they do not have pre-acquired schemas. Under these circumstances, they will need to revert to a general problem-solving strategy. In using this strategy, they are likely to demonstrate a high degree of competence because they are able to use automated problem-solving rules such as the rules of algebra, kinematics, computer programming, etc. Experts can devote most of their cognitive resources to searching for a suitable solution path rather than searching for an error-free use of an operator such as a mathematical or scientific rule.

In contrast to experts, novices will have few, if any, schemas available to categorise their problems and generate moves. Furthermore, any schemas they do have are not likely to be automated. Novices may have no option but to have recourse to a general strategy. They are unlikely to be able to use that strategy as efficiently as experts, not because they are unfamiliar with the strategy but, rather, because they must use it with non-automated rules on unfamiliar problem states. Simultaneously having to plan moves and consider whether the moves are legal may exceed available working memory capacity.

Instructional implications

The suggested influence of automation on problem-solving in mathematics and science has implications for the manner in which we view expertise and how we interpret problem-solving failures under some conditions. Assume, as we did above, that a student has been taught some new principle such as, for example, how to multiply out a denominator in order to move it from one side to the other side of an equation. Assume, furthermore, that the student has understood and learned this principle. Despite apparently understanding the principle, on being given a new problem to solve requiring the use of the newly learned principle, it is frequently the case that students are unable to find what appears to us to be an obvious solution. Since the students 'know' the relevant principle but are unable to apply it in an appropriate context, this common situation is frequently used to suggest that the deficiency is due to a lack of suitable general problem-solving skills. Taken in isolation, this explanation is plausible. Nevertheless, an alternative explanation using automation as the explanatory concept is also possible.

Full automation should occur well after a new rule has been learned and understood. For competent students, initial learning and understanding may be relatively rapid. In contrast, automation may occur only after substantial and lengthy exposure to the rule in question. Until such exposure, each time a student must use the rule, it must be either carefully considered, repeated or, in extreme cases,

even looked up in a text before use. Students may need to stop and explicitly ask themselves, 'How do you multiply out a denominator again?' Even after reminding themselves, application of the principle to a problem state may be slow and tedious. Attempting to use a non-automated rule in this way requires cognitive resources which consequently are unavailable for problem-solving search. It may not be surprising under these circumstances, that students who merely understand a new rule, as opposed to having it automated, have some difficulty applying it. Under similar conditions, more expert problem solvers may see the problem as being routine. The difficulty faced by novice students may be no different from our difficulty in solving the monster problem. We understand the rules of the monster problem just as novice students may understand how to multiply out a denominator or use the equation $V = at$. Nevertheless, we become competent at solving the problems only after we have automated the relevant rules. We may all need to progress through the same cognitive process before gaining competence.

There are several consequences for educational practice that flow from the work on automation. Teachers need to appreciate that a student may have understood a new procedure or concept despite being unable to use it in a new context. Competence in using the procedure may occur well after it has been understood and should not be confused with understanding. It is very easy to assume that because we are able to use a procedure easily and without mental effort, a student, once the procedure has been understood, will be similarly proficient. Understanding may occur substantially more rapidly than automation and failure to solve a problem may be due entirely to lack of automation.

In some areas such as mathematics, where knowledge is cumulative, lack of automation of prior concepts may prevent understanding of subsequent concepts. Attempting to assimilate a new concept or procedure that incorporates a previously learned but not yet automated process may be difficult for all of us. For example, learning how to solve kinematics problems using the equation $V = at$ may be very difficult for a student who understands how to transpose the subject of an equation but needs to revise the procedure mentally before actually doing so. If the student is not immediately aware that $a = V/t$, problems in which a is a goal or subgoal variable may be almost impossible to solve. Constructing an appropriate chain of equations is relatively simple when equation manipulation is automatic. Few of us would have the working memory capacity to change the subjects of equations and relate them to each other and the problem if we had to remind ourselves constantly of the procedures for changing the subjects of equations. It would be unreasonable to expect our students to be any better.

Differences in ability between students may be due in part to differences in speed of automation. Once a procedure has been understood, some students may require far less practice than others for a high degree of automation. From the above discussion, it is plausible that differences in understanding are caused by differences in speed of automation. If automation has failed, understanding of subsequent higher-order concepts may fail also because of reliance on automated lower-order concepts. Automation may be essential.

How should automation be encouraged? The process seems to be inevitably slow, requiring considerable practice. Nevertheless, there are techniques for presenting material in a manner that facilitates automation. These techniques will be discussed in subsequent chapters.

The argument so far

At the beginning of this chapter I described a student working tentatively and hesitantly in a new area. We now are in a position to relate relevant aspects of our cognitive architecture to that performance. The material is new to the student and so most or all of the entities and concepts associated with it must be manipulated in working memory. While the manipulation must be carried out consciously because working memory involves consciousness, very little can be done at any time because working memory is very limited in both duration and size. Our student has difficulty holding and manipulating sufficient elements for a sufficient period of time to complete the task at hand. A diffident performance is the inevitable result.

When the material has been learned and the procedures practised for a period of time, the transformation in our student's performance has occurred because of an increased reliance on a different cognitive structure: long-term memory. In contrast to working memory which is conscious and limited in duration and size, long-term memory can be characterised as being unconscious and effectively unlimited in duration and size. Long-term memory holds huge

numbers of automated schemas for indefinite periods and it is these schemas that govern performance once material has been properly learned. By moving schemas that incorporate many elements from long-term memory to working memory and allowing those schemas to determine performance, our student can function smoothly and accurately with little apparent effort.

The ease with which information can be acquired varies. Some material can be easily understood and learned while other material is much more difficult. What factors determine whether an area is easy or hard to learn and understand? The answer to this question is closely intertwined with the cognitive architecture discussed in this chapter and is provided in the next chapter.