11 Learning Strategies, Teaching Strategies, Matching and Mismatching

Learning depends upon the strategies used by a student in order to direct his attention and to partition the educational goal into separate subgoals. At one extreme, these strategies may stem from the student himself; they are *learning strategies* which he brings to bear upon an otherwise unstructured situation. At the other extreme, the strategies may be imposed upon the student as *teaching strategies* by a programme, a teaching device, or a training routine. Many real situations lie between these extremes. One of them is a tutorial conversation in which methods of learning are open to discussion and in which the strategy is selected as a result of a compromise between the student and the teacher.

Though strategies are important, it is also necessary to consider the student's competence in executing strategies of a given class. For example, an individual student may be good at 'seeing things as parts of a whole', or conversely, he may have a special aptitude for 'stringing subproblems into sequences', which (on resolution) lead to the solution of a large problem. There are, of course, many other types of competence; for example, competence demarcated along the visual/verbal dimension or determined by the relative efficiency of 'short-term' and 'working' memory.

Certain types of strategy call for certain types of competence. For instance, some strategies can best be (or ultimately can only be) executed by the 'holist' student, whereas the execution of other strategies rests upon a 'serialist' competence. By hypothesis, effective learning takes place if, and only if, the individually-selected (learning or teaching) strategy is matched to the student's existing competence.

Students are generally unaware of their competence, and a free learning student is unlikely to learn well. The combinatorial problem of mating the strategy choice to a (largely unknown) competence profile is considerable on its own, but there are also specific factors which militate against effective learning by positively encouraging mismatch. One of the most important is 'cognitive fixity', a tendency to adhere to an originally selected strategy even in the face of evidence showing it to be inappropriate.

The position is somewhat different if a teacher or a teaching programme

is introduced into the situation, because, so far as he (it) is obeyed, the teacher (programme or machine) usurps the role of the student's internal attention directing mechanism and guides his learning according to a teaching strategy. In the simplest case, the teaching system is designed on the basis of a single strategy. This is a provident design provided there are grounds for holding that a single strategy is, in some sense, apposite; for example, 'there is only one possible strategy' (rarely, if ever, true), 'one strategy is better than the others, i.e. there is an optimal strategy' (true in a few cases, note, for instance, Matheson (1964) on list learning), or 'one strategy is far more economically administered than the others, and individual differences have little influence upon the rate or effectiveness of learning' (quite often true and generally used to justify an instructional program that is neither branching nor adaptive).

Generally, however, such grounds for uniqueness do not exist, and in this case, the design of an effective teaching system must be founded upon a class of strategies. Sometimes, one member of this class can be chosen for use on a given occasion with a given individual by an adaptive procedure that employs feedback from performance measures, and the like. If so, the most important process in reaching such a decision is a matching process whereby the student's competence is weighed up and a strategy is selected to suit it. More often, the set of strategies is not given *a priori* and its members must be determined by a (human or mechanised) tutorial conversation of the sort alluded to earlier. In practice, also, there are other factors, to do with the regulation of uncertainty and the student's motivation, which often dictate the use of a conversational teaching system. But, within the conversational format, there must still be a tendency and even an authority to maintain matching. The compromise solutions accepted by the teaching system must be those that engender matching even if they do not completely secure it.

1 General Learning Situations

As a preliminary, it is useful to distinguish between performance strategies concerned with the execution of a skill and the learning (or teaching) strategies that build up performance strategies (Pask, 1969b). The arbitrary character of this distinction is stressed in Chapter 6, where it appears as a distinction between levels of control; namely *Lev 0* and *Lev 1*.

1.1 Performance Strategies Performance strategies have been studied with respect to a great many tasks. In the context of a perceptual motor skill, such as typewriting or vehicle control, the performance strategy is an operational or imperative interpretation of the hierarchical organisation in which subskills, associated with the achievement of subgoals, are integrated into the skill as a whole. Skill organisation, in particular the strategies involved in even the simplest of tasks such as compensatory tracking, is discussed in detail by various authors, for example, Pew (1966), Gaines (1968), Angel and Bekey (1968), and Pask *et al.* (1969b). Skill organisations differ appreciably between individuals, although detailed scrutiny reveals that the idiosyncratic performance strategies cluster around no more than four or five basic types.

Over a wide range of more intellectual tasks, performance strategies are manifest in the hierarchical organisation of problem solving procedures. This is clear, for example, in the protocols produced by the subjects of Newell, Shaw and Simon (1962) or Reitman (1965) (subjects who addressed themselves to logical demonstration; in a weak sense 'theorem proving').

Another instance is provided by the work of Bruner, Goodenow and Austin (1956) on concept acquisition. Their subjects adopted several strategies which were externalised, in the conditions of the experiment, as stretches of behaviour. Any strategy is a procedure for setting up and testing hypotheses about an unknown conjunctive category of the graphical exemplars in a finite universe of discourse, which is described in terms of four attributes (of the form 'shape of figure' or 'colour of figure'). Two main conditions were employed, receptive and selective. In the receptive condition, the experimenter has a conceptual category in mind (unknown to the subject) and he presents the subject with a sequence of exemplars (figures characterised by certain values of the descriptive attributes) furnishing, for each one, the information that it does or does not belong to the unknown class. From time to time, the subject submits an hypothesis about the unknown class and, if necessary, is corrected. In the selective condition, the subject (rather than the experimenter) has control over the evidence in so far as he is able to select exemplars which are submitted for tests of class membership. Subjects run in the receptive condition generally adopted one of two strategies, 'holist' and 'partist'; those run in the selective condition adopted one of four; 'successive scanning', 'part scanning', 'focusing' and 'focus gambling'. The nature of these strategies is detailed in the original work; the immediately important point is that though different individuals set about solving the problem in different ways, the variation is, in practice, quite limited. Although repeated performance of the task engenders a few novel strategies, most, if not all, of these are mixes of the basic ingredients noted above (Lewis and Pask, 1964).

As a final example, strategies occur at a level of cognitive activity which is generally deemed innovative or, at least, insightful. These strategies have been studied by Elshout and Elshout (1969), using the apparatus test as the experimental task (the subject is asked to provide two improved versions of

a given piece of 'apparatus'; to cite one of the author's examples, a 'chair'). Two strategies were distinguished: 'successive transformation' and 'locating problems'. Using the first, the subject reduces the problem of finding an improvement into (a) the subproblem of finding some property of the apparatus that needs improvement (usually culled from his own experience), and (b) the subproblem of transforming this property to yield an improved apparatus. Using the locating problems strategy, the subject (a) considers the set of properties of the given apparatus and selects one (e.g. the chair 'stands on the floor'); (b) replaces this by an imagined related property (another value of the same attribute, e.g. the chair 'lies on the floor'); (c) finds a problem encountered in connection with the original apparatus (a chair that stands on the floor) that would be solved if the apparatus were modified (i.e. if chairs in fact, lay on the floor), and (d) solves this problem by finding an innovation (some sort of collapsible chair, e.g. a deckchair) that can both lie on and stand on the floor.

Broadly, therefore, performance strategies characterise a very wide spectrum of the mental processes, which can be captured and objectified in laboratory studies. There is, of course, abundant discursive evidence that strategies are involved in all sorts of cognition and recollection (for example, in answering questions about a certain period of history or a certain field of mathematics).

1.2 Learning Strategies Given a realistically sized task (and assuming that he cannot already perform it), a student is unable to generate the required performance strategy all at once. Instead, he directs his attention to various facets or subtasks and musters subroutines that build up a performance strategy bit by bit. The process is carried out according to a learning strategy which, in the free learning subject, may be innate or acquired and which, for the student, is imposed externally by a teacher or a teaching system.

A learning strategy is comparable in kind with a performance strategy. Each sort of strategy entails decomposing goals into subgoals and applying mental subroutines to achieve the subgoals concerned. The necessary difference between learning strategies and performance is in the domain upon which they operate. Whereas the performance strategy solves problems posed by states of the (usually symbolic) environment, the learning strategy solves the problems posed (in the context of a goal like 'learn to solve apparatus test problems') by deficiencies in the current repertoire of relevant performance strategies; the solutions produced by a learning strategy are performance strategies.

In unmechanised and relatively uncontrolled situations there is naturally some confusion between learning and performance strategies, but the ambiguity disappears if the task is well defined and the extent to which learning is externalised in behaviour is fully stipulated. For example, are the concept acquisition strategies of Bruner, Goodenow and Austin (1956) one sort or the other? Well, it depends. If, as pictured in the last section, the subject is solving a problem, posed by the experimenter ('describe the unknown class') then the strategy is a performance strategy (as maintained). But the concept acquired in the process is not just a class description; it may be regarded as a procedure (in fact, just another name for 'performance strategy') designed to recognise members of this class. If so, then the Bruner, Goodenow and Austin strategies are learning strategies and the conduct of the experiment externalises a learning process. For instance, in some of the small group experiments carried out in my own laboratory (Lewis and Pask, 1964) subjects were required to use the concepts they attained; in this context the Bruner, Goodenow and Austin strategies are unequivocally learning strategies. Likewise, in the simulation of this situation by Hunt, Marin and Stone (1966), the 'artificial intelligence' program represents the student using (extended versions of) the Bruner, Goodenow and Austin strategies, as the learning strategy whereby it constructs a concept of the form 'performance strategy'.

So far as the other examples from the last section are concerned, there is little room for doubt about a reasonable interpretation. Without diffidence, the action of a typist in tackling subsets of the keyboard separately is assigned to the class of a learning strategy which builds up the motor programs (performance strategies) required in order to perform the skill. Nor is it too difficult to trace the modification of these programs as the skills develop (Pask *et al.*, 1969a). Again we have no hesitation in distinguishing the organised acquisition or inculcation of the 'successive transformation' or the 'locating problems' strategy from the performance itself.

In conclusion, the first aspect of a learning strategy is a contingent plan (i.e. the plan may depend upon indices of success, or the like) for selecting a field of attention. Thereby, the student directs his attention to different parts of the task, to different subgoals or subproblems. In this definition, the field of attention is the domain of some performance strategy and it may either be part of the external environment or an internal representation of it; the latter possibility is pertinent when learning involves rumination, covert rehearsal, and other processes which have no direct *behavioural* correlates. My early work is primarily concerned with situations in which precautions have been taken to ensure that the domain in question is related to the external environment, so that attention directing can be objectively scrutinised. But this condition is not required by the basic statement. Generally, the performance strategies, whose domains are selected by a learning

strategy, are incomplete or even embryonic entities (figuratively, 'boxes waiting to be filled in'). If so, then the learning strategy musters operations that act upon the form of the performance strategy; operations that remedy its defects or, in the limit, that construct it. This is the second aspect of a learning strategy. So, by way of a summary, a learning strategy is first of all a contingent plan for selecting performance strategy domains (fields of attention) and secondly a plan for building these strategies or for repairing them.

1.3 Individual Competence to Execute Strategies Learning strategies call for the execution of mental subroutines which are relatively permanent features of the mind; for example, the subroutines involved in abstraction, in concatenation or 'stringing together', in substitution (of one operation by another), and so on. By the same token, performance strategies make use of several relatively permanent subroutines, notably those that organise the short-term and working storage of the brain into coherent systems (as shown, for example, by Atkinson and Shiffrin, 1967). There is ample evidence that the efficiency of different subroutines, at either level, varies a great deal from person to person, perhaps also from day to day. The distribution of efficiency evaluations over the subroutines is the subject's competence (or competence profile).

A subject's competence is determined by presenting him with paradigm problems of a given type and finding how efficiently he solves them. There is nothing new in this. Multiple aptitude and ability tests provide the requisite data so far as performance strategies are concerned, and since it is maintained that there is no necessary distinction in kind (only in domain) between performance and learning strategies, this data is just as good with respect to the 'higher level' operations. For example, the styles or dispositions observed by Kagan (impulsive/reflective) or Witkin (field dependent/field independent); by Bruner, Luria and the Piaget school, who distinguish certain dominant or preferred modes of problem solving.

Nor is the interpretation of ability test results in terms of a competence profile altogether original. The 'structure of intellect' model of Guilford (1956) is proposed in a similar spirit. But Guilford has an essentially descriptive approach; he maintains that mental factors, by the factor analysis of test batteries, can be categorised and refined in several ways to yield the familiar three-dimensional figure with operations (evaluation, convergent thinking, divergent thinking, memory, cognition) along one side, with contents (figural, symbolic, semantic) along the next, and with products (units, classes, relations, systems and implications) along the last. It is possible to obtain specific estimates of differential competence in respect to many of the ninety cells so distinguished. In contrast, our own theory is process orientated and, consequently, words such as 'operation' have a different meaning. Any strategy is made up of operations which are akin to the TOTE units[1] of Miller, Galanter and Pribram (1960), and which are assembled into hierarchical or interactive structures. Such a structure is a possible strategy and, when executed, it gives rise to a process. The most elementary operations are thus of the same type and the subroutines made up from them differ from one another in respect to context; i.e. their differences rest upon the operations or processes upon which they depend.

In general, ability tests may be expected to sample the efficiency of common subroutines. But any 'operation' (in the present sense) entails entries, in many of Guilford's cells; indeed, it involves most if not all of Guilford's operations. The multiple reference is neither surprising nor disturbing but it does highlight the question of whether Guilford's taxonomy is appropriate for a process model. It is also pertinent to remark that the currently available ability tests are far from optimal devices for sampling the basic competence data when a process model is in mind. For example, there is no test that distinguishes between people's ability to build up association-discrimination processes and their ability to construct rule-like procedures (which is an outstanding distinction from a process point of view).

1.4 Matching between Competence and Strategy The criterion of matching employed in experimental work is based upon specific tests for an individual subject's competence in executing particular operations or learning about a miniature universe of knowledge (under conditions that are designed to make his approach to the matter objectively recordable). Tests of either kind can be carried out in the course of an experiment but are usefully buttressed (they might be replaced) by an appropriate mix of mental test-data. The result, after analysis, is a competence profile which can be compared with each of the (rather few) strategy classes that actually exist. A matching index is obtained by correlating the measure on each proficiency index of the competence profile with the occurrence of each corresponding operation in the strategy or (in case the subject has learned about the miniature universe) by direct pattern superposition. In particular, if a definite strategy is actually adopted by a subject for a learning task, it can be assessed as more or less matched to the competence profile of the individual subject who adopts it.

1. As in the last chapter. However, I am giving a much broader connotation than usual to the familiar notion of a TOTE. It is crucial that the test and the operations are usually non-deterministic programmes or fuzzy algorithms (rather than serial programmes equivalent to finite automata).

1.5 Misperception of Competence and Fixity In free learning conditions most subjects are bad at selecting matched strategies. As a result, they learn slowly if at all (Pask and Scott, 1971; Pask, 1969b). To some extent this may be due to ignorance of the strategies available, and it is certainly true that the imposition of any well defined strategy (for example, via a teaching programme) generally enhances learning efficiency.

But, on the other side of the coin, students commonly bring preconceived notions to bear upon the process of learning and these notions are the germs of strategies, even if they cannot be verbalised as plans. Generally speaking, any imposed strategy is in competition with strategies that already exist. Hence, in a teaching system, it is essential to discover what the competing strategies are by allowing the subject to assert them. This is one reason why the system is one in which the subject is encouraged to talk about the learning process.

Moreover, even if the student is fully aware of the strategic possibilities (for example, as a result of a preliminary discussion), he is still prone to adopt the 'wrong' strategy. An important factor in the decision process is undoubtedly the student's misperception of his own characteristics. On the whole, students are unaware of their competence and the situation is improved if appropriate competence-evaluating information is provided. But misperception is not the only factor. When given the requisite information, individuals are still apt to disregard it, or (putting the matter differently) they need to accumulate much evidence before they will autonomously change their minds and adopt a more appropriate strategy than the one they chose originally. This tendency is commonly manifest; as before, it is called 'cognitive fixity' since in many ways it resembles the 'cognitive dissonance' of Festinger (1957), i.e. a strong tendency to reject or pervert as affirming, any information that disconfirms an hypothesis which the individual has decided to adopt and in which he has invested resources or effort. For example, the purchaser of an expensive motor car is liable to reject evidence that it is in fact unsuited to his needs. Cognitive dissonance is a special case of cognitive fixity. To minimise cognitive fixity, it is thus not only necessary to provide competence evaluations but also to actively dissuade the subject from pursuing misguided lines of action. This is the other reason why a conversational design is often mandatory.

There is often a lot of cognitive fixity over writing a paper, for instance. The author starts with a goal of sorts and realises that the paper must satisfy limits of length and comprehensibility and must put across certain salient ideas. The goal generates a strategy; the paper has a general form and its parts are tackled in a given order. The fixity occurs as follows:

At one stage in writing the paper a great deal of effort is invested in a section developing, for example, the theoretical underpinning of an argument. This, of course, constrains the form and content of the rest of the paper and, if the theoretical argument happens to be inappropriate (not necessarily false), makes it virtually impossible to present an understandable case. Nevertheless, the effort has been spent and the author likes the style of the section. In fact, the paper gets nowhere (and is probably unwriteable in principle) until the author eventually collects sufficient evidence to make him expurgate the offending section. Moreover, a sort of trapping state is involved. The longer that the inappropriate strategy persists (in this instance the strategy engendered by the theoretical section); the more it colours the rest of the structure and the more difficult it becomes to dislodge. As a final point, cognitive fixity is operative even though the author knows all about it, and the snares in which he may be caught. The fact is, human beings are not very good at self-observation or self-control and cognitive fixity (which is perhaps the major obstacle in the path of effective learning) can only be reduced by an outside influence.

2 Specific Investigations (Code-Rule Learning)

Due to the complexity of real-life learning situations, it is quite difficult to obtain individualised data of a type that will illuminate these issues. The learning strategies germane to geography, history, or chemistry may extend over days or even weeks and regularities are obscured by interruption and extraneous activity. More definitive results stem from laboratory studies. On the other hand, brief and small-scale learning experiments do not allow for strategic diversity and a special choice of task is called for.

2.1 Suitable Experimental Tasks

1. The task is learnable but learning is a lengthy business (5–7 hours is perfectly practicable).

2. The problems and subproblems belong to well defined classes so that the experimenter and the subject are in agreement regarding which class a problem belongs to.

3. Although the subject is given the goal of learning to solve full problems (of the main class) these are far too difficult for the novice to tackle *in toto*. Thus, any subject is forced to adopt a strategy of partitioning the task by learning to deal with subproblem classes one after another (eventually becoming proficient with respect to the full problems).

4. Problem solution is rapid so that the experimenter can present the subject with representative sequences of problems drawn from each class.

5. It is possible to evaluate the subject's performance and to offer him indices of success and goal approximation.

6. The learning strategy (which any subject is forced to adopt, as in (3) above) can be externalised as a stretch of behaviour. It is convenient to concentrate upon the attention-directing aspect of a learning strategy, and to regard the strategy as externalised in so far as the subject attends to a sequence of subproblem classes, which terminate in the class of full problems.

7. The subject can be given reasons for externalising a process which is, from his point of view, more conveniently carried out in his head (when, of course, it is hidden from the experimenter). For example, he may be asked to physically select a class of subproblems by pushing buttons on a console, as a result of which a sequence of problems drawn from this class is presented to him. If so, the subject *could* partition the full problems mentally and solve part of each full problem in his head. To induce him to make objectively-detectable selections in a veridical fashion (rather than the capricious manner he might adopt following an arbitary fiat), the following expedients are used.

(a) The subject agrees to maintain at least a certain average score, over and above agreeing to aim for the educational goal. A score is computed with respect to the selected class only (note, if the novice selected the full problems the score would be so low that the subject could not keep his agreement).

(b) Knowledge of results is provided in respect to the selected class (note, this gives the student detailed as against macroscopic information).

(c) Facilities, such as evaluative data, are provided in the framework of the classes so far selected.

Externalising gambits like (a), (b) and (c) constitute a primitive form of cooperative externalisation technique or CET (Pask, 1969b; Pask and Scott, 1970, 1972b).

8. It is possible to determine the subject's competence by tests that are relevant to the task. Usually behavioural tests are called for; paper and pencil ability tests have been used but they are insufficiently specific and are inconvenient for on-line administration.

9. It is possible to ascertain the subject's performance strategy. Once again, behavioural measurements are preferred.

10. The investigation is greatly simplified if one learning strategy is used to build up just one performance strategy.

11. The task is related to a useful category of real-life tasks.

2.2 Free Learning Experiments Experiments have been carried out in a situation that approximates conditions (1)–(11) and is a microcosm embodying certain aspects of real-life relation learning. The task employed is an elaboration of the code learning tasks of Chapters 6 (simulation), 7 and

8. The display and response facilities are shown in Fig. 96. A full account of the experiments is given in Pask and Lewis (1968), Pask (1969b), and Pask and Scott (1971).

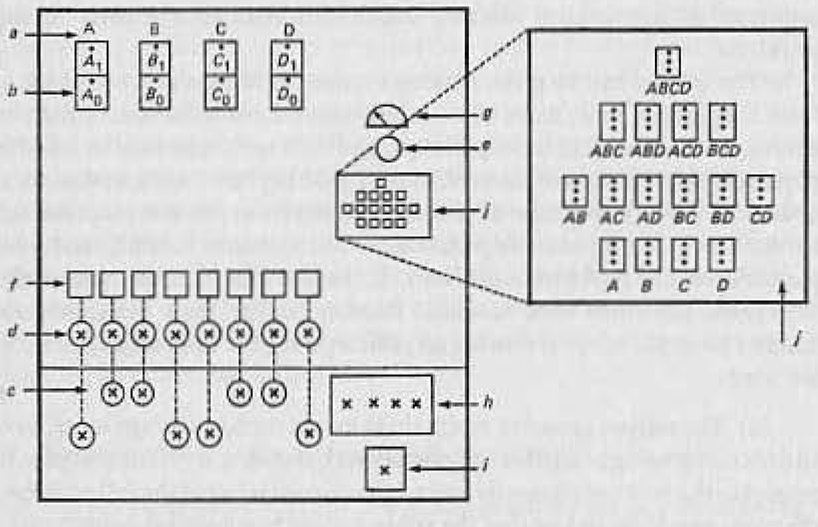


Figure 96 Display and response board: (*a*) signal variable names, *A, B, C, D*, one for each box; (*b*) problem lamps, $A_1, A_0; B_1, B_0; C_1, C_0; D_1, D_0$; in boxes. The lamp in the upper box indicates the value 1 of an attribute, and the lamp in the lower box, the value 0. Problems, *x*, are configurations of illuminated lamps (see text); (*c*) eight response keys, $y_1 \ldots y_8$; (*d*) external register, retaining key-pressed information until the end of any trial; (*e*) partial knowledge of results lamp (complete response is or is not correct, σ_0); (*f*) complete knowledge of results lamps indicating correct value for each component of the response; (*g*) score meter; (*h*) attribute selection buttons; (*i*), submit button (see text); (*j*) last value of score display. Three lamps for each subset: blue (upper), $\rho > \rho_2$; green (middle), $\rho_2 \geq \rho \geq \rho_0$; and red (lower), $\rho_0 > \rho$.

In the full problem situation (which must ultimately be mastered) the subject is presented with a sequence of four, variable, visual signals and must solve each problem by making an appropriately coded four-component response. Calling the signal variables *A, B, C* and *D*, the full problems belong to a class named (*ABCD*), i.e. its members present all of the variables at once. Under the temporal constraints imposed upon the task, a novice is quite unable to learn how to deal with problems in (*ABCD*) and is thus impelled to direct his attention to subproblem classes of the form (*A*), (*B*), . . ., (*AB*), (*BC*), . . ., (*ABC*), (*ABD*) . . . (there being fifteen classes in all). The structure of the problem environment is shown in Fig. 97. Clearly, the

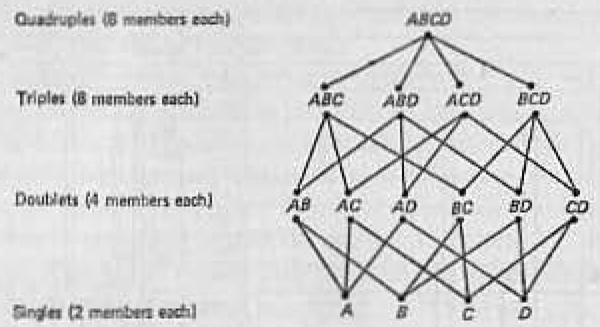


Figure 97 Structure of the problem environment.

terminal task is isomorphic with the terminal task for code learning described in Chapter 8, section 1, only, in this case, the environment has been partitioned so that there are many visually distinct ways of directing attention to different subproblems. The experiment is fully automated and the subject is asked to externalise his attention-directing strategy at the end of each block of trials (each sequence of problems) by keying into his console the name of the subproblem class which he will consider during the subsequent block. This selection determines the subproblem class from which stimuli (designating the problem) are actually drawn. A CET involving the provision of scores, and the evaluations and knowledge of results is used to secure reliable externalisation.[2] In these conditions a learning strategy is manifest; (a) As a sequence of subproblem class selections, and (b) Through information obtained, on subsequent interrogation, about how these selections depended upon the evaluation and scores.

Preliminary studies revealed the existence of four types of learning strategy (Fig. 98).

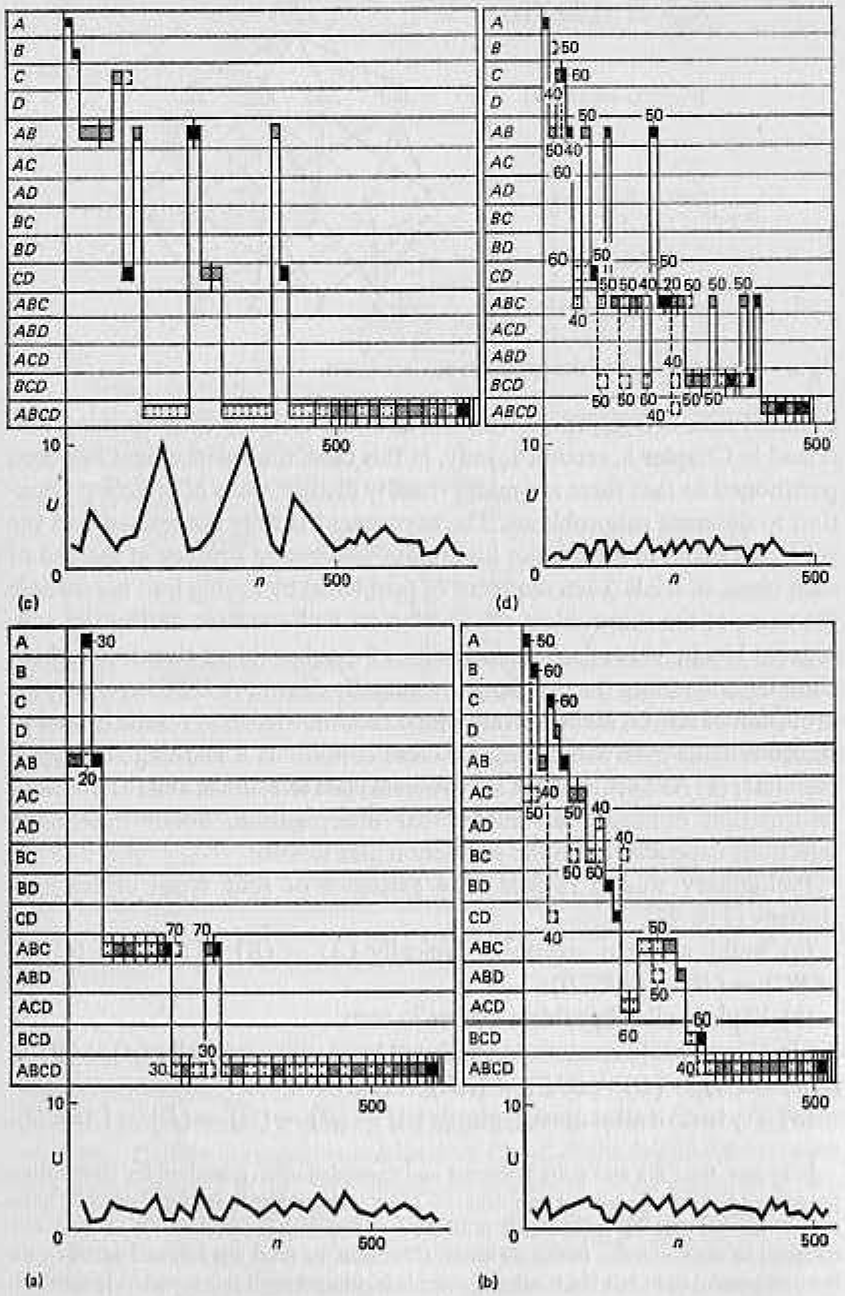
(*a*) Build up in a sequence, typically (*A*) → (*B*) → (*AB*) → (*C*) → (*ABC*) → (*D*) → (*ABCD*).

(*b*) Explore all subproblem classes in turn.

(*c*) Group problems into increasingly large units, typically (*A*) → (*B*) → (*AB*) → (*C*) → (*D*) → (*CD*) → (*AB*) → (*ABCD*).

(*d*) Try to do it all at once, typically (*A*) → (*B*) → (*C*) → (*D*) → (*ABCD*).

2. In fact, the CET was quite elaborate and depended upon a method for committing the subject to a certain level of performance. On entering a block, say (*AB*) or (*ABC*), the subject knows that he cannot exit until he has responded correctly on at least one occasion to each stimulus in the sequence. Of course, he need not respond correctly on the first presentation but the block sequence is continued until this criterion is satisfied. The score is the number of 'first presentation correct responses' (divided by the number of different stimuli). From the number of repetitions or guesses we derive a direct measure of the subject's uncertainty (i.e. Figures 98, 101, and 102).



One variant c^* of the type (c) strategy is noted in the source paper Pask and Scott (1971) but will not be discussed.

Of these possibilities, the performance strategy that goes with learning strategy type (d) leads to triviality in so far as the subject who uses it is failing to solve the entire problem. In contrast, he deals, at his leisure, with as many separate items as there are stimulus variables (it is doubtful whether, in this case, his behaviour is a problem solving behaviour). Hence, a time constraint was imposed to prohibit the offending performance strategy. In fact, the time allowed for solving each problem was separately adjusted for each individual (in a preliminary adaptively controlled session), so that no more than a pair of elementary operations could be carried out at any one trial. As a result, subjects were told and knew that learning strategy (d) was bound to be abortive. Moreover, they experienced this fact since before the main experiment (but after the adaptive period) each subject had practice in learning the skill with a different coding rule in force.

Two types of competence were observed as response processes, manifest in regular patterns amongst latencies of the several components of the complex response; (a) a serial competence which at the microscopic level is called stringing (of response producing operations) and (b) a holistic competence, interpreted as a grouping of operations.

The main findings of the free-learning experiments were that though a few subjects learned the skill successfully (in 400 trials or so) the great majority did not (some taking 750–1000 trials and a few hardly learning it at all). Those who failed did so for quite definite reasons.

1. The subject tries to use strategy (d) even though he knows that it cannot succeed. The lure of getting straight to the goal is remarkably seductive. Eventually, strategy (d) is abandoned in favour of some other plan and the student takes about as long to learn the skill as he would have done starting out this way.

2. Subjects who are good at stringing but bad at grouping try to use strategy (c) (which is fitted to grouping). Their learning is retarded and occasionally they are forced to change strategy.

Figure 98 Typical free-learning curves. Upper part of figure: problem class selections along the vertical coordinate; n (number of trials), along the horizontal. Proficiency, ρ, is shown by shading for each block of trials:

[black] $\rho \geqslant 0{\cdot}75$; [grey] $0{\cdot}75 > \rho > 0{\cdot}25$; [open] $0{\cdot}25 \geqslant \rho$.

Probabilistic problem class selections are shown as P numbers associated with bifurcating lines. Lower part of figure: U (behavioural uncertainty) along vertical coordinate, n along horizontal. (a) type A subject; (b) type B subject; (c) type C subject (also demonstrating unsuccessful attempts to use strategy D); (d) type C^* subject.

3. Subjects who are good at grouping but bad at stringing try to use strategy (*a*), which is adapted to stringing. Their learning is retarded and occasionally they are forced to change strategy.

4. Subjects adopt strategy (*b*). In most cases, it would be more accurate to say that these subjects have no global strategy. (Certain exceptions exist.) In later experiments where pre-experimental and post-experimental protocols were taken, one subject did express the intention of exploring and produced type (*b*) selections. But most of the apparent explorers turn out to be people who are in doubt about what to do. This contention is supported by subjective confidence estimates over the set of subproblem classes, which were obtained after each block.

In all, twenty-six subjects were run in the free-learning situation. Ten of these were run in conditions which are strictly comparable with those of subsequent experiments, and for this subgroup the mean number of trials to criterion was 574·5, σ (standard deviation) = 153.

2.3 Adaptive Machines The most obvious way to remedy some of these defects is to impose a teaching strategy upon the subject. Since there is great individual variation in how long it takes to achieve criterial proficiency, teaching must be a score sensitive procedure mediated by an adaptive machine, which like the controllers described in Chapter 8 was programmed on the special purpose computer shown in Plate 5. The machine is designed on the basis of just one strategy which is incorporated into the machine design in the same way as the skill structure of Chapter 8, and the machine (rather than the subject) makes all subproblem class selections.

Learning of the code learning and signal translation skills has been extensively computer simulated (see Chaper 6) and there are good grounds for supposing that strategy (*c*) is generally the best. It is not the best strategy for each individual, of course; strategy (*a*) is better for someone with a proclivity for stringing. But on average it is as good as any other single alternative and, in order to obtain a standard condition for comparison and reference, the experimental machine was based upon a type (*c*) strategy. The flow-chart of the adaptive system is shown in Fig. 99.

The subject is advanced to the next selection in the strategy if his score exceeds an upper threshold 'sent back' if his score is less than a lower threshold and 'left' where he is 'if his score lies between these limits'.

The results obtained from the single strategy adaptive experiments are quite unequivocal. This type of instruction is faster than free-learning. (Mean trials to criteria = 463, σ = 167, N = 10.)

Learning rate (adaptive) > learning rate (free-learning); the difference is significant at the 0·1 per cent level. Care was taken to check that the

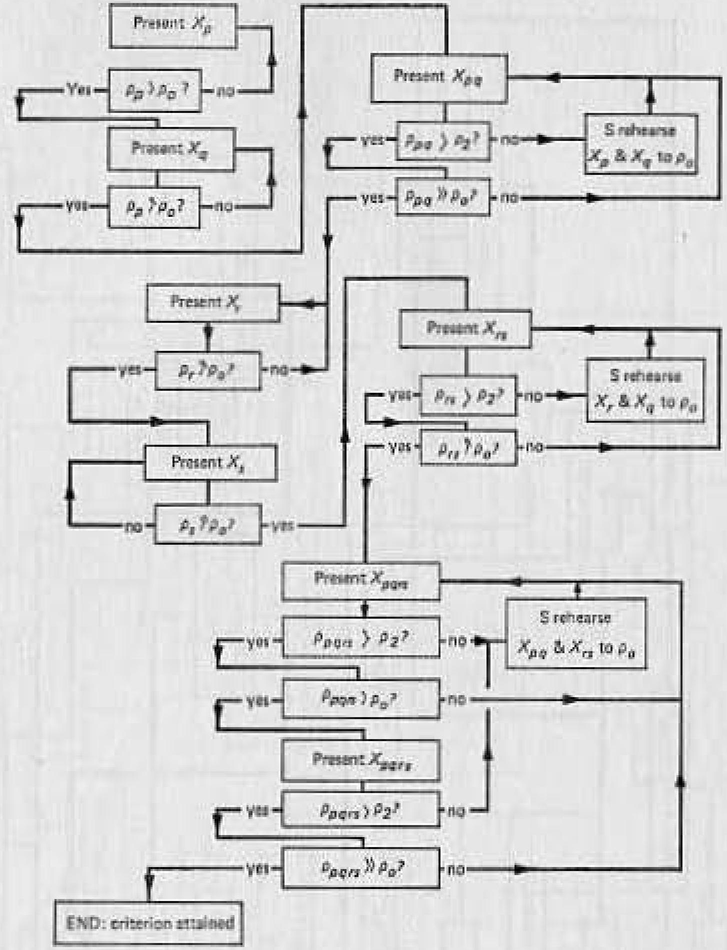


Figure 99 Flow-chart for the adaptive system based on type (*c*) strategy. *Explanation.* The X_i are the subproblem classes. The subscript values p, q, r, and s correspond to the signal variables, A, B, C, and D in the order that they are *first selected*. In the adaptive system, the machine makes all selections and is set so that $p = A, q = B, r = C$ and $s = D$. Thus, e.g. $X_{pqr} = (ABC)$.

The ρ_i are scores of first presentation correct responses to the signals in the subproblem class X_i.

ρ_0 and ρ_2 are, respectively, the upper and lower thresholds, where ρ_0 = 75 per cent, and ρ_2 = 25 per cent first presentation correct responses.

All commands are to the adaptive machine except for those prefaced by '*S*' which are commands to the subject.

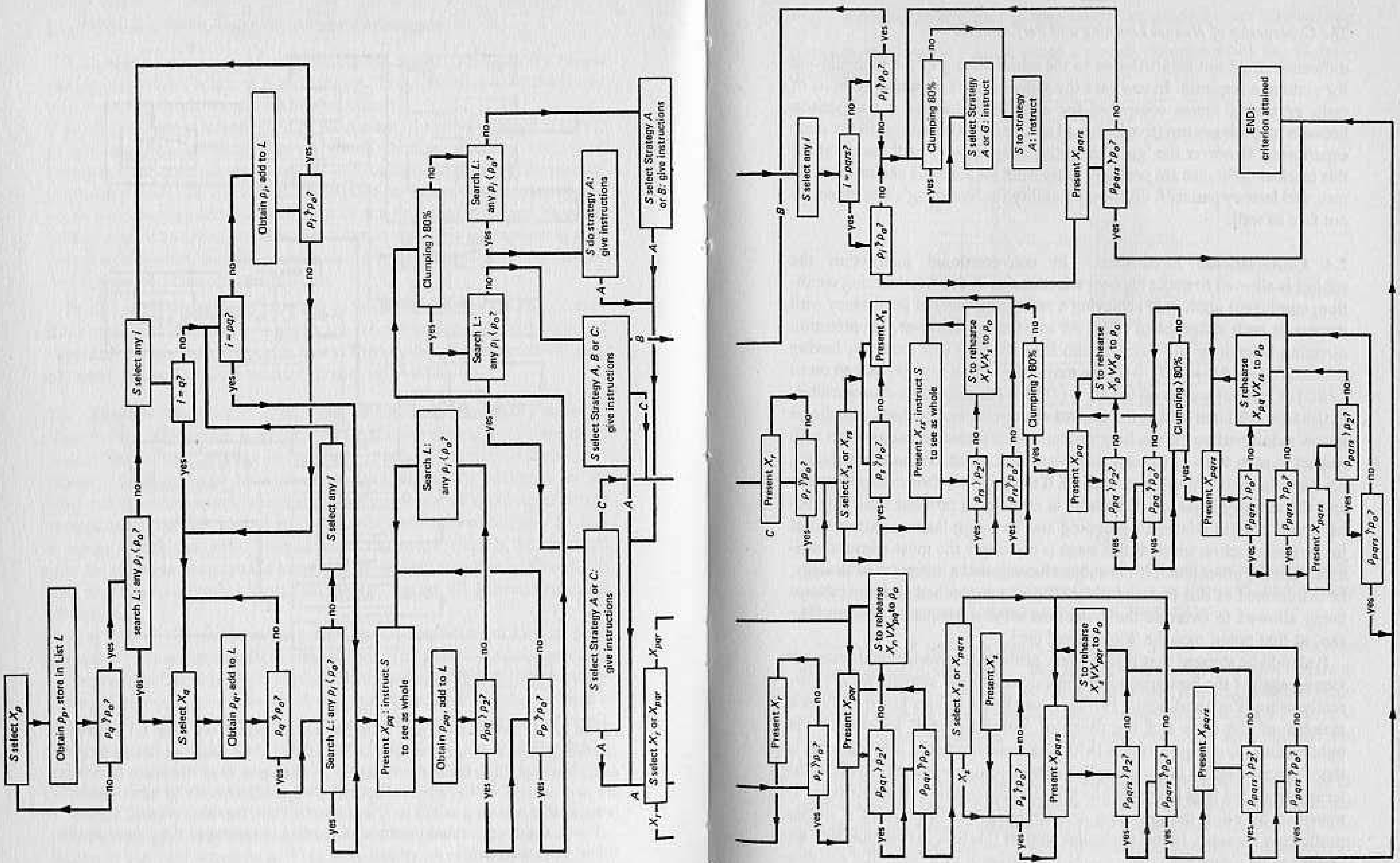


Figure 100 Flow-chart for the conversational system. *Explanation.* As for Figure 99 with one addition: the test of a subject's ability to make grouped responses ('clumps') is represented by the test box 'Clumps 80 per cent,' i.e. to pass the test at least 80 per cent of a subject's 'first presentation' responses must be correct *and* grouped.

difference could not be attributed to the initial setting of the time allowed for making a response. In any case the same result is obtained in terms of ratio or saving scores computed for each individual as a comparison between preliminary run (free learning) and main run (adaptive) in the same experiment. However the 'good' adaptive subjects who really account for this marked difference are precisely those who, on the basis of continuously recorded latency patterns, do show an ability for 'grouping'. The others do not fare so well.

2.4 Conversational Instruction In conversational instruction the subject is allowed to make his own selections as in the free-learning condition, contingent upon only achieving a reasonable level of proficiency with respect to each subproblem class. At some stage, however, his attention directing behaviour will commit him to a strategy (for example, having selected $(A) \rightarrow (B) \rightarrow (AB) \rightarrow (C)$ he may either adopt type (*a*) and go on to (ABC) or type (*c*) and go on to $(D) \rightarrow (DC)$. At this point the strategic alternatives are exhibited so that the student can opt for one of them. His choice is now monitored by the teaching system, which assess his competence with respect to each strategy (by carrying out a 'look ahead' test and determining his disposition to 'group' or to 'string'). If the subject's choice is appropriate and the strategy is matched, then he is allowed to proceed satisfying behavioural criteria related to grouping and stringing latency patterns (not just to overall score, which at this stage is no longer the most relevant variable). On the other hand, if the subject has selected a mismatched strategy, he is informed of this fact and diverted on to a matched strategy (always being allowed to override the system on some subsequent occasion if he can, at that point, pass the 'look ahead' test).

It should be stressed that this is a very primitive conversational system if viewed against the background of the conversational systems we are currently using (Pask and Scott, 1972b), nevertheless it does just qualify as a member of the class and the usage of 'conversational' appears in the publications for our group up to 1970. The main deficiency, a serious one, is that the conversation lacks an explicit criteria for *understanding* (a technical term introduced later in the chapter). To an appreciable extent this defect is remedied by using the grouping and stringing test in place of a simple proficiency measure. In this particular context this test, in common with an understanding condition, guarantees the existence of a 'memory' (section 2.6).

The design of a conversational system is based on a class of strategies (Type *a*, *b*, *c*) and not just one strategy. Its flow-chart is shown in Fig. 100 and allows for all sorts of remission and recycling which may complicate the picture sketched out above. Typical subject records for this type of

learning are shown in Fig. 101 which is usefully compared with Fig. 98 (free-learning).

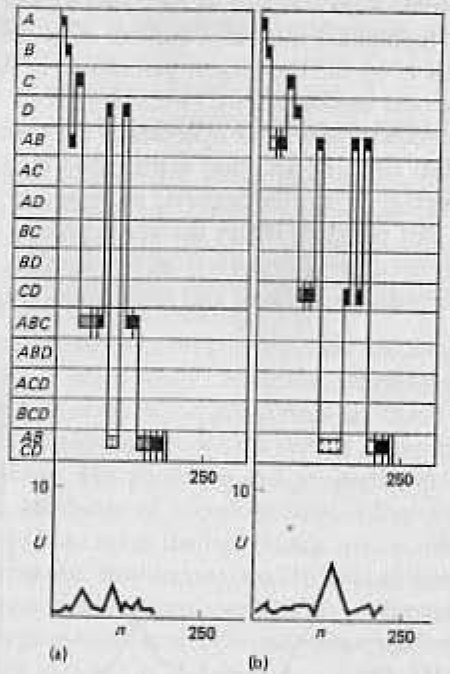


Figure 101 Typical learning curves for conversationally controlled subjects: (*a*) type-*a* strategy; (*b*) type-*c* strategy (in notation as Figure 98, each strategy adopted as a compromise between subject and the machine).

The results of the experiments demonstrate a clear advantage in favour of the conversational system. The results were as follows. Mean trials to criterion = 220, $\sigma = 40{\cdot}6$, $N = 10$; the difference, learning rate (conversational) > learning rate (adaptive system) is significant at the 0·1 per cent level (the same precautions were taken with respect to time allowed and ratio scores).

Although the type (*b*) (exploratory) strategy was offered as an option, no conversational subject chose it and the confidence-estimation data indicates that this was so because no subject had any doubt about what he ought to do (though all subjects had the liberty to choose selections). By definition all strategies were matched to the subjects competence so, not surprisingly, 'grouping' subjects followed a type (*c*) plan and 'stringing' subjects a type (*a*).

The most dramatic difference between the free-learning and the conversational subjects is in the elimination of cognitive fixity, especially with respect to the type (*d*) strategy. On average the free-learning subjects spent 39·8 per cent of their subproblem-selections rehearsing problems in class (*ABCD*), and most of the trials were abortive as indicated by the high uncertainty over these blocks (computed from the 'number of guesses' data provided by the CET). This is by far the largest percentage of selections for any problem class; the next largest is (*ABC*) with 15·6 per cent. In contrast, the conversational subjects spent only 25·2 per cent of their selections rehearsing problems from class (*ABCD*) and they experienced relatively little uncertainty. This percentage is not the highest; more selections were spent rehearsing (*AB*) at 26·1 per cent. Hence the whole pattern of learning in the conversational system differs markedly from free-learning. In particular the students uncertainty decreases faster and more smoothly (Fig. 102).

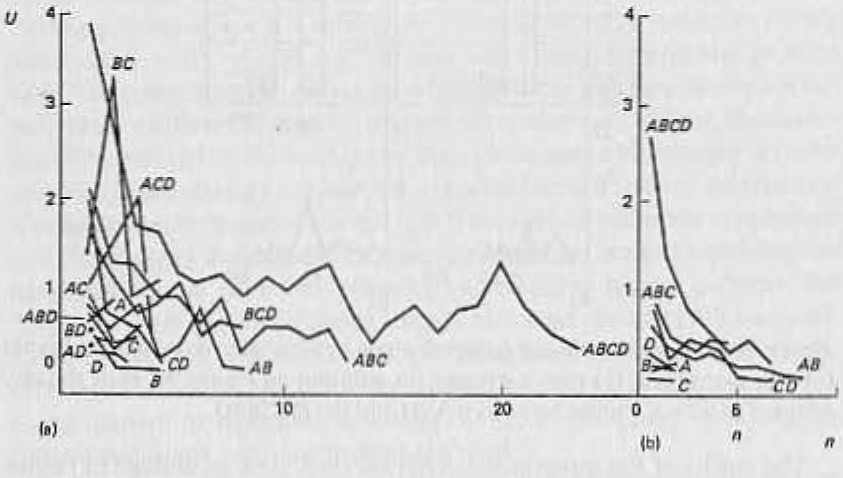


Figure 102 Plot of *U* (behavioural uncertainty), averaged over: (*a*) free-learning subjects; (*b*) conversational subjects with respect to separate problem classes plotted against number of trial blocks (*n*) spent dealing with each class.

2.5 Other Tasks, Isomorphism Between the Perceptual Motor and the Cognitive Domain The experimental desiderata of section 2.1 are satisfied by certain intellectual tasks; notably learning maps and layout schemes, moderately sized taxonomies, the operation of mechanical and biological systems (including cyclic processes such as the mammalian oestrous cycle and the operon/repressor control system). If the proper conditions are satisfied, then the distinctions manifest at the perceptual motor level of activity are recapitulated in the cognitive or intellectual domain; especially.

(a) Cognitive fixity.

(b) The existence of strategy types that are exclusive with respect to *one* experiment.

(c) The existence of a differential competence to execute learning strategies of different types.

(d) The effect of matching and mismatching.

(e) Interference and incompatibility of certain mental operations.

In fact, as hinted in section 1, it has been shown that these and other processes are characteristic of learning and cognition of a much more general kind; for example, in learning academic subjects; in design, in creative and innovative activity, and in the learning that accompanies and forms part of real-life decision making.

Hence it is argued, firstly, that certain processes are common to perceptual motor skill and to a small scale intellectual learning provided that the 'proper conditions' of observation are preserved. These conditions are, in fact, a conversational situation which may either approximate free learning or a type of teaching. The processes and properties in question are thus, strictly speaking, attributes of a conversation rather than an individual though clearly they bear upon the individuals who converse. In respect to these qualities (processes, distinctions, etc.) there is an *isomorphism* between the perceptual motor field and the cognitive or intellectual field.

Secondly, it is argued that the same isomorphism has a much wider scope; spanning education for certain, and perhaps extending further.

The second contention (like the first) has been empirically substantiated but it is necessary to look at the nature of the 'proper conditions' with some care, i.e. to spell out in detail how a 'conversation' is interpreted in this context and to surmount certain difficulties that become clear on examining the intermediary and laboratory sized learning situations.

2.6 Explanation Prior to considering the intermediate (small, intellectual learning) experiments it is essential to give a better account of understanding. In a perceptual motor skill, taught conversationally, the criterion for understanding is an ability to *reconstruct* whatever problem solving or goal directed procedure is employed to *construct* a complex response. Within the intellectual domain (for the moment, also, verbal conversations) the goal directed procedure is equated, as in the last chapter, with a *concept* which brings about or satisfies a relation *R*. The evidence for the existence of a *concept* of *R* is a verbal *explanation* of the relation in question.

Again, following the last chapter, suppose that the concept is assigned to a level of control, *Lev 0*. The entity which reconstructs a *concept* of *R* is a special case of an entity which constructs it in the first place; namely a *Lev 1* goal directed procedure. It might also be called a concept (for the distinction of levels is a matter of convenience in description) but to maintain this convenience and to follow ordinary usage[3] it will be called a *memory* of R; (that is, a constructive operation that recapitulates, reconstructs, or relearns a *concept* of R). The evidence for its existence is a verbal explanation of how to construct the companion *concept*. Hence, the evidence that a concept of something exists and is reconstructible, permanent, or memorable is an explanation *of* an explanation; which (if properly attested) is called an *understanding*.

This definition tallies with the perceptual motor paradigm in so far as the *Lev 0* explanation can be elicited non-verbally if there is an appropriate environment in which the relation it brings about can be modelled (literally as a physical model, a computer program for example, which, on execution, brings about this relation R). Any verbal explanation is a description of a model, either one that has been built in an environmental facility or one that does not, in fact, exist. For example, as Loefgren (1968, 1972) points out, a proof chain or a derivation of R (for *Lev 0* terms α, β; with α as an axiom) is the sequence

$$\pi_0(R) = \langle \alpha, \beta, \ldots, R \rangle$$

which is either the model of R or a description of it (the explanation of R) which is just a verbalisation of the *concept*, π_0, of R. Further, an explanation of this explanation (the verbal evidence for a *memory*) is a sequence (in *Lev 1* terms $a, b, \ldots$) like,

$$\pi_1(\pi_0) = \langle a, b, \ldots, \pi_0 \rangle$$

or, equally, this is a description of an 'internal-to-the-subject modelling operation' that reconstructs the concept, π_0, in the subject's *Lev 0* repertoire (and π_0, on execution, brings about R).

This imagery is stilted in so far as it represents serial operations (π_1 and π_0 are Turing machines, and π_1 is a reproductive automaton that replicates π_0). In general (hence our continual insistence upon the non-serial character of most mental procedures), there are many explanations and ways of modelling. This contingency is quite easily encompassed in the theory (Pask, 1973). However, the 'proof-chain' image makes it clear that the perceptual-motor paradigm and the intellectual paradigm, are in register at *Lev 1*. A *memory* is the same in either case; similarly, in either case, the constructive operations that build concepts to be remembered are described as the *learning strategy*. In contrast, at *Lev 0* all of the concepts (alias

3. Not in the run of psychological literature where 'memory' is often regarded as a storage.

strings or *groups*) consist in models *composed in the subject's brain* and generally not described by the subject (though they are described by the experimenter when he talks of 'strings' or 'groups' or contrives a 'learning model' of the type discussed in Chapter 6; *not* to be confused with the *subject's* model). This internal-to-the-subject-model, though not described by the subject, is executed autonomously to produce a complex response and, if it were described, that description would be a *Lev 0* explanation (in the experiments under discussion an explanation of Ω or of a part of Ω).

Hence, the general criterion for *understanding* is to elicit in respect of each relation R (each part of Ω in the perceptual motor task and each topic in the intellectual task) an explanation and an explanation of this explanation. Several methods may be used to do so, all of them constituting means for externalising one or (generally) many chains, like,

$$\pi_0(R) = \langle \alpha, \beta, \ldots, R \rangle; \; \pi_1(\pi_0) = \langle \alpha, \beta, \ldots, \pi_0 \rangle$$

in the conversational dialogue, using a CET for this purpose. One method is called 'teachback'.

As a matter of fact teachback is commonly used in education for ensuring that a topic is understood, though its function is usually appreciated intuitively.

Teachback goes as follows: the teacher says of the student (or 'subject') that the student understands a topic to the extent that he can teach it back to the teacher. That is, understanding is inferred if the student can furnish an explanation of the previously discussed topic and can also explain why he gave that explanation or how he constructed it. The crucial point is that the student's explanation and the teacher's explanation need not be, and usually are not, identical. The student invents an explanation of his own and justifies it by an explanation of how he arrived at it (in fact an identical explanation is generally rejected unless the student can give a reason *why* the teacher's explanation was particularly good). It is required, of course, that the explanations (the *Lev 0* explanations) due to the student and the teacher explain the same topic R; moreover, that the *Lev 1* explanations produce a concept of R. But, in general, an indefinite number of explanations have this property (though there is also an indefinite number of psuedo 'explanations' that do not).

2.7 The Validity of Teachback Operations This technique (teachback) has been partially formalised using a restricted form of natural language dialogue, checked out by content analysis. The technique is crucial to the argument and is validated by experiments of the following kind.

In asserting that teachback (explanation of explanation of what has been explained in teaching) is an indication of understanding, we are positing that if a topic R is taught back then R will be retained at any rate whilst the student is learning the same material. Conversely, if he does not have a *memory* and a *concept* of R, the student has not (in the present sense) learned R. But the existence of a concept of R and its memory (though it may be inferred from an externalised explanation) does not imply that an explanation will be given. That depends upon the quality of the CET (if it were perfect, all concepts/memories would be externalised as explanations; otherwise there is no guarantee to this effect). Hence, we may expect to show, experimentally, that *any* topic R that is 'Taught Back' *is* retained; of those that are not taught back, some *may* be retained.

2.8 A Study This hypothesis was tested in studies involving over fifty students learning about thirty-five *related* topics (Pask and Scott, 1972a) using the following design.

Session 1		*Session 2*	
Learning with teachback Group 1	→ Test Group 1	→ Fresh teachback Group 1 + 2	→ Test Group 1 + 2
Learning with ineffective teachback Group 2	→ Test Group 2		

'Dummy' or 'ineffective' teachback (session 1) is a procedure in which the requirement to explain is replaced by a requirement to give *correct* responses; up to a criterion of perfectly correct response, which is obtained by appropriate cueing and remedial action. The test in session 1 contains questions that demand explanatory replies. The test in session 2 contains some multiple-choice questions but mostly questions that demand explanations in reply. The latter are of primary interest (the multiple-choice questions act as a control). The session 2 'fresh teachback' is introduced to provide additional data about how explanations are changed and perverted by interfering events the student has experienced in the week or more between sessions.

First, every taught-back topic is recalled, with explanation, in the session 1 test; though many correctly answered though unexplored topics are not. The score difference between group 1 and 2 at session 1 is significant at the 0·1 per cent level ($0{\cdot}001 > p$). Next, the taught-back topics are nearly all retained until session 2; rather few of the unexplained topics are retained. Using the explanation-demanding question score, the difference between group 1 and group 2 at session 2 is enhanced, though the same level of

statistical significance is achieved ($0{\cdot}001 > p$). The difference is also significant in respect of the multiple-choice test component at session 2 but to a lesser extent (1 per cent significance, $0{\cdot}01 > p$). The fresh teachback data from session 2 is chiefly of interest in connection with a variation upon the experiment. It can be argued that although retention of taught-back items will be perfect *within* one session (it is), the resilience of a memory will depend upon the number of *explanations* produced in teachback; for example, that a student impelled to give *many* explanatons will fare better at session 2 than a student required to give only one. He has many ways of reconstructing a *concept* and this redundancy will combat the effect of interfering and incompatible learning experiences during the intervening week. This result is valid, and, in experiments designed to exhibit the effect by deliberately interpolating destructive learning experiences, it is highly significant (Pask and Scott, 1970). But in the study described, the effect is best exhibited by comparing and contrasting the detailed teachback records from session 1 and session 2 as the differences reveal both distortions and the operation of the reconstructive procedures that remedy them.

3 Intermediary Scale Experiments (Taxomnoy Learning)

The intermediary scale experiments were all concerned with modes or styles of learning and were designed to determine the free-learning strategies adopted by individual subjects and to assess the efficiency, for individuals with differing mental makeup, of different teaching strategies. In order to externalise the free-learning process (in which a student is at liberty to ask questions of his choice and direct his attention to different subproblems), the design embodied a conversational technique. For some of the experiments, the dialogue took place in a formalised language, via an interface (as in section 2) with a partly mechanised tutorial system. In others, the subject conversed directly with the experimenter who was, however, preprogrammed by fairly rigid rules.

All of the experiments can be conceived as experiments in relation learning and the relations in question are complex and redundantly specified. For example, one task in the series involved the relations inherent in a taxonomy; another concerned the relations describing a cyclic system. Embedded in each redundantly specified set of relations there is a kernel of essential relations which a subject has to learn and recapitulate. Though the subjects were aware of the essential relations, they generally learned and used redundant relations in order to access and manipulate essential relations while performing the test task which succeeded each experiment.

These are clearly combined experiments in learning and teaching. The design was partly engendered by educational objectives, but also by the general contention that theories of learning and theories of teaching do not exist as separate entities. There are theories of learning and teaching; that is all.

3.1 Relations and Types of Strategy If a relation R is described in the context of its domain, then it has an apparent dimension, n, equal to the number of properties in (descriptors of) its domain and an order or adicity m equal to the number of properties that R necessarily unites. The terms apparent dimension and order are easily interpretable (in the case that R is stated extensionally as a subset of the product of variables (U_i) indexing the related properties), i.e. as $R \sqsubseteq U_1 \times \ldots, \times U_n = U^0$ when an n-tuple $\langle u1, \ldots, un \rangle \in R \sqsubseteq U^0$ satisfies R. Ashby (1964a) provides an algorithm for determining the order m of an arbitary subset of U^0 (his 'cylindrance'), and generally, m is less than n. In particular, following Ashby's line of argument:

(a) if $n > m$ (as usual), then R is the intersection of subrelations (cylinder subsets) of generally lower order and dimension;

(b) there will often be several families of 'cylinder subsets' that intersect on R; any one family specifies R; the set of families provides a redundant specification of R.

Relations may also be specified in a language where they are named by words or described by phrases, for example, 'father of' or '$>$' or 'the T.C.A. cycle' or '*A* allows *B* to do *C*' or 'happiness' or 'antelope' (a class of animals defined by relations between behavioural or physical properties; the taxonomies we actually employed classified mythical, Martian, creatures since these were necessarily unfamiliar to the subjects). Sometimes the order of the relation and the dimension of its domain are explicit; for instance '$>$' has order 2 and is irredundantly specified (for numerical domains of dimension 2; similarly, '*A* allows *B* to do *C*' has order 3, though the dimension is not explicit). But, linguistically stated relations can be quite respectably under-specified. It is possible to list situations that engender happiness or relations supporting the antelope character *ad infinitum*.

There is nothing mysterious about the idea of an underspecified relation. On the other hand it is an important and interesting idea. An individual who understands the language in which the relation is stated can extensionally specify the relation, at a given instant, by ostending the n-tuples which belong to R and separating them from n-tuples that do not. But this specification is tied to the individual and to the instant; if R is underspecified, then n (and possibly the value of m) is variable in any general statement that is independent of instants and individuals. Hence, it may be deemed prudent, and it is certainly quite harmless, to regard underspecified relations as the open

union of fully specified but redundant subrelations of possibly variable dimension. For example, Banerji (1970) distinguishes between input properties (in a fine-structure family) and non-input properties. Phrased in these terms an individual need not specify all non-input properties at the outset. But, if his goal is underspecified, he will use them and could specify them at some later stage.

It is profitable to distinguish between two major kinds of problem-solving computation. At one extreme, problems are solved by the serial application or basic routine, produced by a type (*a*) strategy (section 2). At the other extreme, problems are solved by mustering these routines concurrently (an effectively parallel computation produced by type (*c*) of section 2). Amarel (1969), taking the position of a formalist, conceives the first type of problem solver as a macro-assembler and the second as a director and compiler combined. If we make the same distinction in terms of artificial intelligence and psychology, we can say that the first sort of problem solving resembles the operation of a programme such as Newell, Shaw and Simon's 'logic theorist', whereas the second sort of problem solving resembles the operation of an associative network like Quillian's (1969) 'TLC'. When R is redundantly specified (as it is) then the second sort of problem solver aims for several subgoals at once and is likely to achieve R as the intersection of various families of subrelations.

The crucial point and the reason for making this distinction, rather than many others, is that (in common with some other pairs of processes) these two are incompatible. Data used by one process cannot be used without complete recoding and reconstruction, by the other.

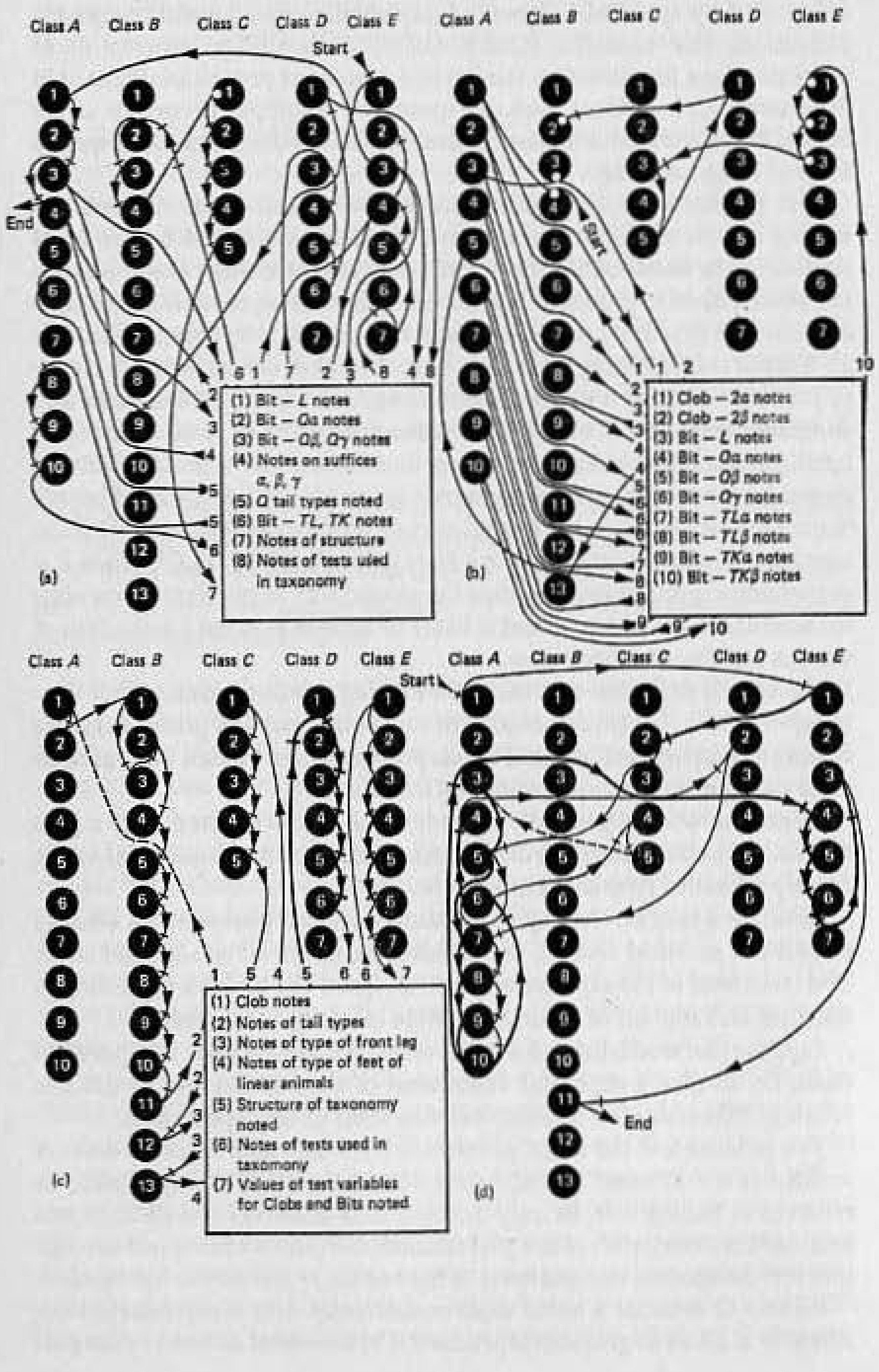
If, as is nowadays generally agreed, a concept is equated with a goal directed procedure, then the distinction demarcates two conceptual types, namely a serialist type and a holist type.

Consider a task involving a redundantly-specified goal relation. (Nearly all tasks of practical concern are of this type, as are all educational tasks. The tasks used in the experiments were designed to be of this type, though the form and amount of redundancy were carefully controlled.)

Suppose that a serialist and a holist solver both know what is required of them, i.e. to give a canonical description of the relation and to use this relation. This condition was also satisfied in all of the experiments.

The serialist and the holist problem solvers use quite different data. A serialist is embarrassed by redundant data, unless it is clearly marked as redundant. Failing that, he only succeeds if he discovers and sticks to one irredundant description of the goal relation, and, since a canonical description is to be elicited, this one must be the one the experimenter has in mind.

By way of contrast a holist deals concurrently with many descriptions, and even if asked to give the (experimenter's) canonical account of the goal



relation he generally constructs it by cross-reference to a set of different and redundant descriptions.

The same comments apply to the way in which both serialist and holist problem solvers use the goal relation in solving problems.

An individual's strategic type may be determined in various ways. If the subject is allowed to free-learn a taxonomy of animals (for example by gaining access to indexed cards on which are inscribed facts about the classification and the creatures classified), he exhibits a characteristic pattern of exploration (see Fig. 103).

Content analysis of the questions asked about the study material during free-learning (based on the cylindrance measures of section 3.1) also provides an independent discrimination; the serialist either strings through the card packs according to a linear plan or poses hypotheses which test *one*, generally *unary*, predicate value at once, and then questions the experimenter about the truth or falsity of the hypotheses. In contrast, the holist poses relational hypotheses; his questions are concerned with the validity of predicate functions of many variables.

These discriminations are more readily elicited during *teachback* where the hypotheses are fully externalised for scrutiny.

Further, there are characteristic differences in uncertainty and correct belief (as determined by confidence estimates; Chapter 1, section 7). As a slight oversimplification, the holist entertains a measurable uncertainty and correct belief for topics 'ahead of' the topics he is dealing with, and this uncertainty is reduced (with an indrease in correct belief) as he studies a cluster of topics from which he garners data and relates it to the topic 'ahead'. A serialist, on the other hand, is unable to give a confidence estimate in respect of a topic 'ahead' of his current focus (at any rate, a topic 'far ahead') and he typically does not explore further until the one (or occasionally the few) topic relation currently occupying his attention is

Figure 103 Patterns of exploration for the learning of a zoological taxonomy: (*a*), (*b*) serialists; (*c*) redundant holist; (*d*) irredundant holist. *Key*: Class *A*, pictures of typical members of subspecies; Class *B*, contextual information (behaviour, habitat, and so forth); Class *C*, statements about structure of taxonomy; Class *D*, physical characteristics used for making distinctions between subspecies; Class *E*, meaning of subspecies names and codes.

Card number of a particular column of the array.
General search.
Testing many predicate hypothesis.
Testing single predicate hypothesis.
Failure to find information sought after.
Transition between questions.

fully established (he is completely *certain*; ultimately he is certain of the *correct* topic relation).

All this begs the question of what a topic 'ahead of' or 'far ahead' of may be; this question is answered in due course and the reader is asked to accept a dictum that the goal relation is ahead of all others, and that topic relations elsewhere in the study material can be sampled (by posing questions and eliciting confidence estimates).

Finally, there are some paper and pencil tests which (less reliably) discriminate the holist and serialist.

A holist strategy and a serialist strategy are cognitive processes. They are also incompatible in the *same* context (though a student might change between contexts). In the context of a given chunk of study material, for example, the description of a species to be classified according to a taxonomy, the action of cognitive fixity should ingrain one type of strategy or the other.

This effect occurs. It is very pronounced, to the extent that, during a particular piece of learning, an individual (once started) can be typed dichotomously as 'holist' or 'serialist' (as tacitly assumed in the discussion up to this point). That is, in *one* conversation the strategic types are 'all or none' distinct. It *is* possible to shift type by external instructions. But the amount of disconfirming data needed to effect this transformation increases with time and, as a result of the transformation, the previously erected cognitive equipment has to be discarded; the student who is impelled to change 'starts afresh' and relearns.

In contrast, it is possible to characterise an individual as having a statistical *disposition* to use one kind of strategy or the other by examining his free-learning performance over different tasks, preferably several of them. If his successful performance is scrutinised, the disposition is *one* index of competence to execute the same type of learning strategy (there are several other methods of determining competence, to be introduced in the next volume).

Since the holist/serialist distinction rests upon a process distinction it is possible to design teaching strategies that are *matched* to a holist and *mismatched* to a serialist as well as strategies *matched* to a serialist and *mismatched* to a holist.

The design process is systematic (again, it is detailed in the next volume) and it entails finding paths through the topics, i.e. the part relations of a goal relation, that lead to the goal relations and guide the student's attention in a way that maximises (holist) or minimises (serialist) the degree of concurrent processing required.

With these preliminaries, it is possible to make sense of the intermediary-scale learning experiments (Pask and Scott, 1972). The design is fairly

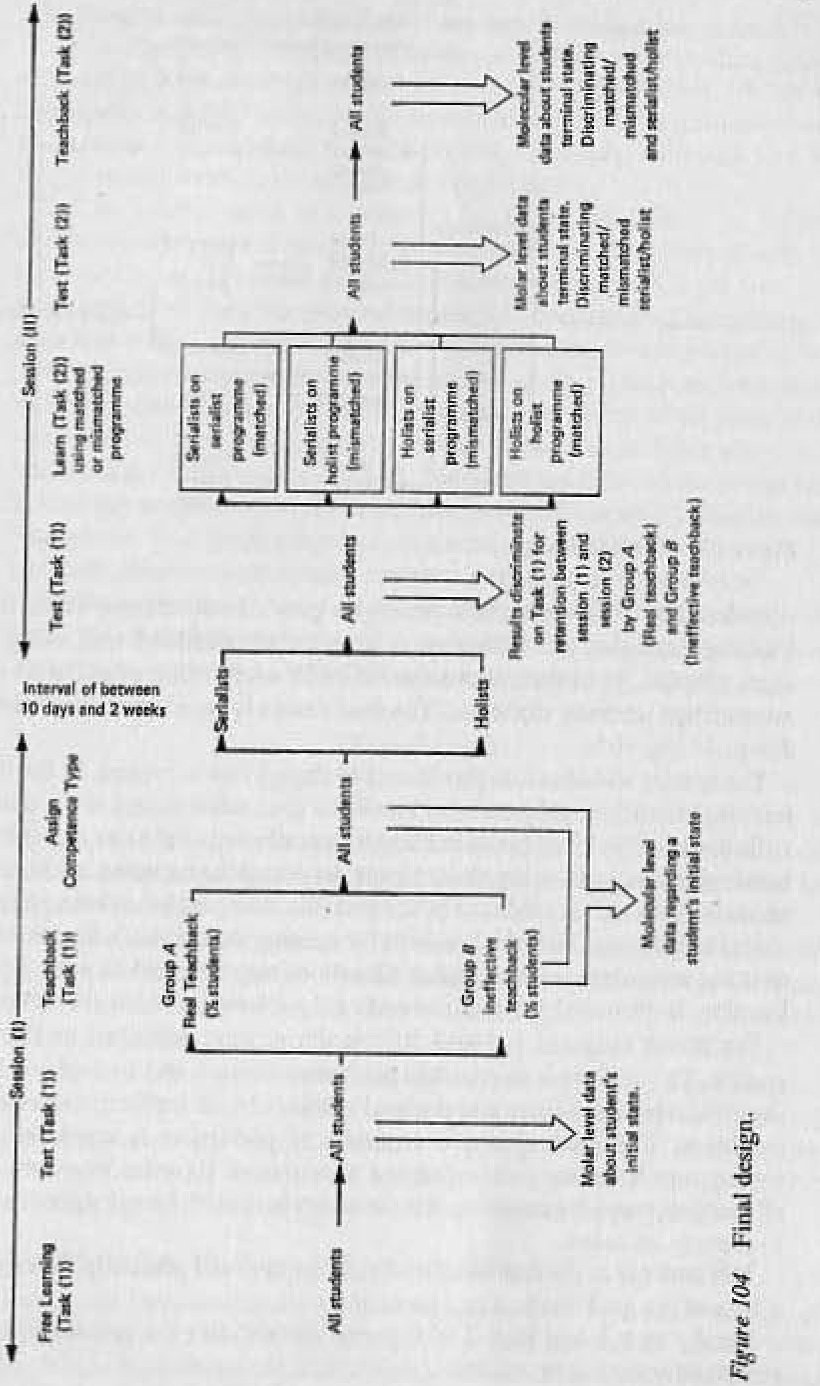


Figure 104 Final design.

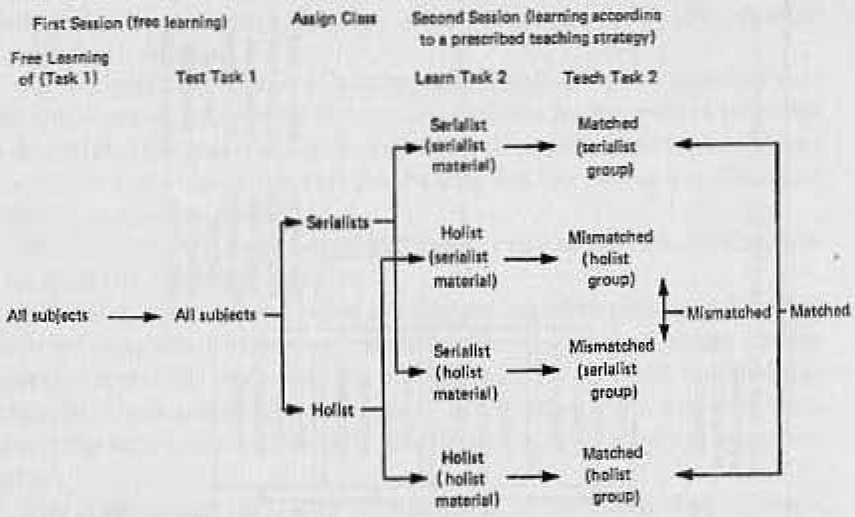


Figure 105 Outline design.

complex (Fig. 104), since the experiments provide information about free-learning strategies, their effect upon retention of teachback and, using the same subjects, yield data about the influence on learning of matched and mismatched teaching strategies. The final design is based upon the skeleton design of Fig. 105.

The species classification (taxonomy learning) task is typical. In the free-learning condition, subjects who know the goal relation and are provided with charts of the kind shown in Plate 10, are allowed to explore an environment of cards containing data about an imaginary species of Martian animals. The cards are indexed by content (for example: behaviour, physical characteristics and historical reasons for naming the animals). Some data is written; some data is pictographic. Questions may be asked as part of free-learning. Both verbal transactions and card pack explorations are recorded.

Teachback (Figures 104 and 105) is the process described in the last section. The student is required to achieve perfection and to explain all of the topic relations he uses and the goal relation. In the ineffective teachback condition, the same objective criterion of perfection is employed but explanation is neither demanded nor encouraged. In order to ensure that all students reach the criterion, it is necessary to furnish knowledge of results to remedy mistakes.

The tests (as in the last section of this chapter) call primarily for explanations of the goal relation and *parts* of it.

Finally task 1 and task 2 of Figures 104 and 105 are two distinct but structurally similar taxonomies.

3.2 Results 1. As expected, explanation chains do occur in teachback.

2. The explanation chains of serialists and holists are (as predicted) quite different in kind, especially with respect to the type and referent of the hypothesis tested during the production. (Subjects have been independently characterised as serialist or holist in task 1; simple statistical tests are inappropriate but pattern matches are significant.)

3. The establishment of a memory for a concept X may be inferred unequivocally from the existence in teachback of a chain reproducing the concept. Groups of twenty show differences significant at 0·1 per cent. This memory persists for a few weeks at least and for months in all subjects tested after that interval. There is reason to believe that the memory is indelible.

4. After learning to solve problems, one group of subjects were tested until they could achieve full proficiency. Another group of subjects, at the same level of proficiency, were submitted to teachback. After a couple of weeks, all the subjects were retested. The difference between the two groups is clear-cut to the extent that no statistical method is needed to discriminate their scores. Teachback subjects have a memory, test subjects did not do so, i.e. under these circumstances a memory is dependent on teachback.

Students previously assigned to the competence-type serialist (S) and holist (H) were instructed in task 2 via serialist training routines (SR) and holist training routines (HR). Of the four combinations; namely,

$$(A) = S \rightarrow SR$$
$$(B) = S \rightarrow HR$$
$$(C) = H \rightarrow HR$$
$$(D) = H \rightarrow SR$$

both (A) and (C) are matched, whereas both (B) and (D) are mismatched. Subsequently, all subjects were tested for retention and regeneration of the learned material.

Serialist subjects fare just as well, on the average, as holist subjects (the problems were chosen to obtain this result). However, the matched subjects have a much higher test score than the mismatched subjects. Groups of twenty give differences significant at the 0·1 per cent level ($0{\cdot}001 > p$).

It looks as though effective learning depends upon securing a matched condition. If the size of the effect is as large as we suspect, classroom education must be grossly inefficient and an improvement of several orders of magnitude could be achieved by the (entirely practicable) expedient of matching the tutorial strategy to the competence of an individual.

4 Outstanding Problems of Representation and Understanding

Two main requirements are outstanding; first, a reasonable means of representing what may be known (the goal relation and its various components or topic relations) and secondly, a reasonable method of detecting the occurrence of an *understanding*. It is clear enough at this point, that a conversation open to observation, measurement, and so on is (whatever else it may be) a sequence of *understandings*. It will later be possible to establish that the condition of understanding observed in teachback is fundamental.

The first requirement (a representation of knowledge) is obtrusive both on practical and theoretical grounds. From a pragmatic point of view it is necessary to deal with educationally realistic stretches of learning and topic or subject matters as large as a course (in statistics, biology, or sociology). Learning the background data for a taxonomic scheme with some dozen distinct tests in it (the Martian animals, for example) gives rise to explanations or learning strategies of the kind shown in Fig. 103. It is not difficult to appreciate that a study of chemistry learning, for example, would be altogether intractable if *this* subject matter was represented in the same way. Moreover, even if this representation were practicable and manageable, it would (for reasons to be given) prove inadequate; there is no way, in general, to express a teaching strategy in terms of such a figure though it happens to be possible in the specific cases under discussion. As an equally significant objection, there is no systematic method for combining chunks of knowledge to form curricula or semantic depositories of one kind or another. We seek, and have obtained, a representation scheme that does meet these and various other criteria quite satisfactorily.

From a theoretical point of view the representation in Fig. 103 (with categories of data cards corresponding to dimensions) is a description scheme or indexing scheme for learnable and knowable relations; it is not a representation of these relations, *per se*, together with the relations that unify them into a coherent body (the all-adumbrating 'goal relation'). In fact this goal relation (the species, membership of which is tested by the taxonomy) can be expressed as a subset of the values of these dimensions. Further, since the description is deliberately redundant, the goal relation can be expressed as a subset of the products of several distinct clusters of dimensions and as a subset of the products of *many* different *value* sets. However, the description is a description and it is not unique; many different descriptors and indexing schemes can be chosen to describe the same goal relation. Further, it is essential to admit this possibility because some subjects learn very successfully (the redundant holist of Fig. 103c is one example) in terms of additional properties and descriptors which they bring to bear on the task.

Of course, it is possible to represent the topic relations that are learned and to represent the goal relation as a relation between them. To do so, in general, yields a relational network similar to, but not identical with, the kind of network quite widely employed in artificial intelligence (for example, Winston (1970), or Winograd (1972); the latter stressing the descriptive and explanatory aspects of search in machine-interpretable terms). Such

structures are very large, even for modestly sized chunks of educational material; one is shown in Fig. 106 to give the reader an idea of the size

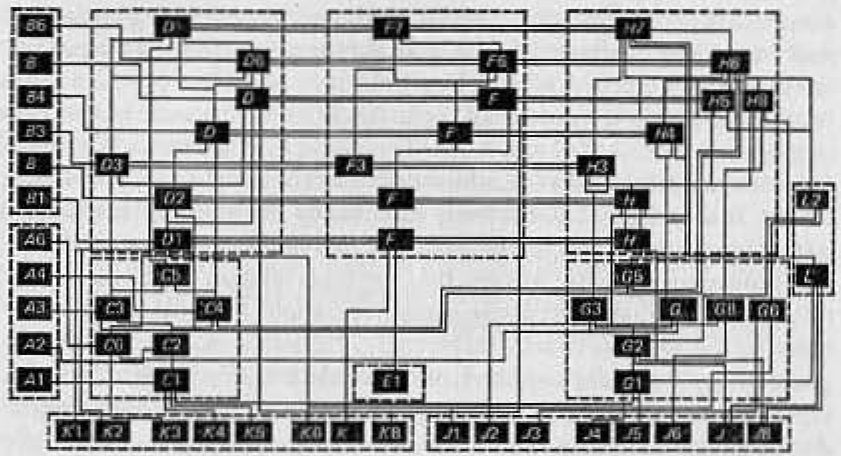


Figure 106 The 'Probability Theory' entailment structure: *A*1 to *A*6, peripherals for *C* (e.g. 'example from games of chance'); *B*1 to *B*6, peripherals for *D* (e.g. 'example from behavioural science'); *C*, deterministic experiments; *D*, random experiments; *E*, theory of simple results; *F*, probability theory; *G*, structural model; *H*, probability number model; *J*, *K*, primitives (e.g. 'order'); *L*, peripherals stating relations between parts of *G* and *H*. *Key for nodes of entailment structure*: *C*1, simple results; *C*2, composite results; *C*3, exclusive results; *C*4, inclusive results; *C*5, deterministic experiments; *C*6, random results; *D*1, frequency of simple results; *D*2, frequency of composite results; *D*3, frequency of exclusive results; *D*4, frequency of inclusive results; *D*5, conditional frequency of simple results; *D*6, conditional frequency of composite results; *D*7, random experiments; *E*1, theory of simple results; *F*1, probability of simple results; *F*2, probability of composite results; *F*3, probability of exclusive results; *F*4, probability of inclusive results; *F*5, conditional probability of simple results; *F*6, conditional probability of composite results; *F*7, probability theory; *G*1, event set; *G*2, composite events; *G*3, exclusive events; *G*4, inclusive events; *G*5, structural model; *G*6, probability numbers; *G*8, complement of composite event; *H*1, probability numbers (p. no.) of simple events; *H*2, p. no. of composite events; *H*3, p. nos. of exclusive events; *H*4, p. nos. of inclusive events; *H*5, conditional p. nos. of simple events; *H*6, conditional p. nos. of composite events; *H*7, probability theory; *H*8, p. no. of complement of composite event; *J*1, model/real world distinction; *J*2, inclusive/exclusive distinction; *J*3, arithmetic operations; *J*4, universe; *J*5, set and subset; *J*6, complementation; *J*7, intersection; *J*8, union; *K*1, long-run stability; *K*2, counting; *K*3, order; *K*4, qualities; *K*5, 'one at once' definition; *K*6, fractional numbers; *K*7, definition of probability relation; *K*8, definition of given knowledge; *L*1, relation between logic of structure and arithmetic of p. nos.; *L*2, relation between p. nos. and conditional p. nos.

involved and not (until the next volume) to reveal its structure. Figure 106 represents one thesis (of one subject matter expert) about elementary probability theory. Under execution, if a search or exploration is carried out, these networks are verb networks. Moreover, by performing operations that image the relations said to hold between the topic relations (and represented by directed arcs in Fig. 106), it is possible to construct the relation in question from other topic relations in the same mesh. This requirement is demanded and is called *cyclicity*; in other words Fig. 106 is not just a representation of an arbitrary thesis about elementary probability theory. It is certainly a thesis freely invented by the subject matter expert *apart from constraints upon what properties the relational network shall have.* One property is cyclicity; namely, that any topic relation can be derived by relational operations, that *image* mental operations from other topic relations in the same network. Incidentally, relational operations may be grouped together under a common rubric as relations of *entailment* (between topic relations) and the corresponding collation of mental operations is *discovery*.

The other constraint upon the person or group (or student, as a matter of fact) who specifies a thesis and thereby a network, is that each topic relation in the mesh shall be *Lev 0* explicable; either he can explain it or, given an appropriate physical device such as a programmable computing machine, a laboratory, or a game situation (for example, a role-playing game in management or economics), he can model it. It should be stressed that *Lev 0* explicability has no obvious connection with veridicial truth or even with formal rectitude. The expert (or the student) can explain scientific nonsense, if he wants to do so. This is quite intentional as the thesis is *his*. The explicability requirement is only that the model *works*; it is a criterion of *satisfaction* not a criterion of truth. Finally, a requirement of *consistency* is involved; namely, that if two or more topic relations are differently named and distinct in the mesh, they then do not have the same class of *Lev 0* explanations.

Between them, the canons of cyclicity and *Lev 0* explicability, both of which are imposed upon an acceptable mesh, guarantee that the mesh is learnable and memorable. They lend it the quality of a gestalt closely related to the interpretation of a gestalt given by McCulloch and Pitts (1943) though not quite identical with it.

To see this in outline, one should notice that cyclicity means that any subject equipped with the posited mental operations that carry one topic relation into another, is able to furnish an explanation of how he arrived at a *Lev 0* explanation for each topic relation in the mesh. This is a condition of *Lev 1* explicability. But the *understanding* of a topic relation is evidenced by the execution of a *Lev 0* explanation, together with its justification by a

Lev 1 explanation; hence, each topic relation in such a mesh *may be* understood. With reference to Chapter 10, the mesh is a permission giving structure for *learning*, i.e. as asserted before, it specifies what may be *known*. The class of models corresponding to the class of *Lev 0* explanations is no more nor less than a process representation, a permission giving structure for *doing*, which specifies what may be done. *Lev 0* explicability requires that one process representation is attached to each node in the relational network.

Such networks (generally called entailment meshes) have a couple of other apparently innocent but actually far-reaching properties.

(a) They are not really specified in extension but in terms of processes (goal directed or problem solving procedures); that is, intentionally.
(b) They are isomorphic to an indefinitely large number of other meshes.

Property (a) means that a topic relation could be explained in any context; the point was mooted in section 3.1 (the brief comment on verbally stated relations). A concept is a procedure that brings about a relation. If the procedure is used to build a model which, on execution, does bring about the desired relation then quite clearly this relation is defined in extension (by a listing). But the particular universe in which the model is built and executed may be chosen by the student (apart from limits of convenience) and in this sense the student is free to, and required to, engage in acts of predication (choosing the properties which *he* deems *relevant*), for his concept could be executed in many domains of interpretation.

Property (b) is of consequence when it comes to describing the relational mesh; in other words, imposing coordinates (descriptors) that act as the grid lines (in an often very curious topology) of a cognitive map. Because of property (b) it is possible to choose an indefinite number of descriptive schemes, all of which are compatible and satisfy the last requirements to be outlined.

(c) The description scheme uniquely indexes each node (representing a topic relation) in the entailment mesh, so that it can be ostended.

(d) By using the description scheme, it is possible to point at some 'blank' positions and place holders for nodes that do not currently exist, but might be brought into existence. By this means, it is possible to knit the ends of meshes together and it is possible for meshes to evolve, in a systematic manner, as a result of learning.

The hierarchical structure shown in Fig. 94 (page 252), an and/or tree is *not* a verb net; it is a *noun* tree. Its deficiencies, which were mildly criticised, can be summed up by the fact that it is not and, in general, is not derivable from, an entailment mesh though *this* one *can* be so derived. It does not represent knowledge in the ordinary sense; rather, it represents a peculiarly stilted description of knowledge. To show this point, consider any description of an entailment mesh. Amongst other *descriptors*, it is possible to choose (and to obtain the promised sense of a topic relation which is 'ahead of' another, it is necessary to choose) at least one descriptor with the broad connotation of subordinate/superordinate. In terms of *this* description the topic relations are partially ordered; that is, one aspect of their description is partially ordered and quite possibly this *ordering* looks like Fig. 94. Hence, within this ordering, some topic relations are 'ahead of' others and, as a rule, the goal relation is placed at the top as a 'head' in this description. The essential point is that there are many descriptions with different 'heads' for the same entailment mesh. In fact, Fig. 94 *is* such an ordering and it has one intriguing characteristic; namely, *any* topic could be chosen as head in some description of the underlying mesh. That is, it would be reasonable to learn about truth tables in terms of variables and connectives (as in Fig. 94) or, using a different ordering, to learn about connectives in terms of truth tables, or variables in terms of connectives and truth values.

The second requirement (a means of detecting the understandings that go on in a conversation) is best tackled in the context of an entailment mesh that forms the conversational domain; the topic or subject matter that is under discussion. In this context we need a mechanisable test for understanding. The 'mechanical' requirement upon the test or detection device is chiefly imposed to ensure that it is well defined. Of course, it is useful to have a mechanically executed test for other reasons (for example, in order to instrument a full-blooded system of computer-assisted instruction or computer-guided learning, or merely to relieve the experimenter, in a seriously large learning study, of an impracticable burden). However, there is nothing magical about this gadgetry; the test could be performed (as it is, in teachback) by a human experimenter. The crucial issue in either case is whether the test is definitive for it is to rank as an alternative to Turing's test for intelligence in conversation (Turing, 1963, 1969). We shall say that any conversation (whether between two men or a man and a machine) is characterised by a contiguous series of occasions upon which understanding is reached by a *process*, without too much commitment about *where* the process is executed (for example, in the teacher's brain, in the student's brain, or in some inanimate processor). This process might be deemed intelligent (i.e. an instance of intellect and not just 'able to pass intelligence tests', which is a trivial criterion, judged either by Turing's test or our own). As a matter of orientation, I happen to call the putatively intelligent process *learning* and to notice that it *is*, quite literally, symbolic *evolution*.

5 Taming Wild Understanding

The mechanical system which tests for and may also manipulate understanding in a CET externalised conversation is shown in Plates 11 and 12 (these pictures are chiefly intended to give the reader an overview of the equipment used in studies of learning; not, in this volume, to inform in any depth). The system is founded upon a theory of what understanding is (Pask, Scott, and Kallikourdis, 1973) and the following comments in the matter constitutes no more than an orienting outline; sufficient to indicate the flavour of this field of research.[4]

5.1 It is maintained that a conversation in a domain, R (satisfying the conditions of the last section), is a self-reproducing system. That is, using *Rep* to stand for 'reproduction' (in the symbolic and automaton theoretical sense as in Chapter 2) and choosing any topic relation (R_i) in R on which the conversation is momentarily focused at a given occasion, there is a stable entity

$$Rep\,(Rep\,(Rep\,(R_i)))$$

where $Rep\,(R_i)$ is a *concept* i of R_i

and $Rep\,(Concept\ i) = Rep\,(Rep\,(R_i))$ is a *memory* of R_i

If it were externalised in a conversation, by a CET technique, this entity would be symptomatised by a cycle of explanation (namely teachback) consisting of an explanation by one participant, agreed by the other, and a justifying explanation also agreed (the comment follows from our earlier specification of what an explanation or modelling operation *is*). Thus, if the conversational participants are A and B, the first part of the equation becomes (under these conditions) '*Lev 0* explanation by A of R_i to B and *Lev 0* explanation by B of R_i to A' (with agreement) and the next part becomes '*Lev 1* explanation by A of R_i to B and *Lev 1* explanation by B of R_i to A'; the last *Rep* term closing the system and stabilising its execution.

This 'stability' accompanies (is *due* to or gives *rise* to, as you prefer) the effect of cognitive fixity; i.e. it is a property of such processes that when executed in respect of a given domain they 'recognise' and reject classes of dissimilar processes executed in the same domain.

5.2 Such an entity is recursively specified with respect to R_i, or in general, R. But it may be executed in various processors, for example, in a teacher, a student, or a mechanical device. In order to *observe* understanding on the part of A and B, it is necessary (using a CET) to externalise the

4. Experiments using the system are described in Pask and Scott (1972b, 1973).

reproductive cycles that *might* be private to a student (say) and to represent them in terms of dialogue. For the facility shown in Plates 11 and 12, the dialogue takes place in a mechanised command and question language with various transactions, mediated through the interface, that point to topic relations and clusters of topic relations in the *described* entailment mesh (the large display). As a result of these transactions a conversation takes place between a regulatory heuristic (one participant, *B*) and one or more students (the other participant, *A*). In the course of it, at any occasion, agreement is reached regarding a topic or topics the student is currently aiming for (generally, topics he can describe by locating them in the cognitive map, but is not yet in a position to learn about) and topics that he takes as his goal and is learning about, as well as a backlog of topics that are already *understood* (the *understanding* condition has been satisfied). In order that goal topics shall be understood it is necessary for the student to explain them; in this facility we use non-verbal explanations that are models made on a computer-monitored, laboratory-like, modelling facility, which, unlike the rest of the equipment, is task specific and is labelled 'Statlab' in Plate 11.

All of the conditions of *any* node in the large display are marked by coded signal lamps; hence, the progress of learning, i.e. a learning *strategy*, is a wave of activity headed by *aim* nodes followed up by goal nodes, and leaving a train of *understood* nodes in its wake. These conditions are continually displayed to the student.

5.3 It can be shown that a system

$$Rep\,(Rep\,(Rep\,(R_i)))$$

can be dissected into two subsystems (Loefgren's (1972) reproductive automaton and its support in the domain *R*). The desired condition is to detect and externally observe an understanding so that the chains of explanation (alias reproduction) all penetrate the interface between a student and the heuristic governing the operation of the machine. If the student's mental organisation is *A* and if the machine heuristic is *B*, then the basic form of *B* is a procedure for securing this condition with respect to all topics in *R*.

5.4 The *Lev 0* component is the explanation or modelling operation. The *Lev 1* component need not be made explicit, since a description of the learning strategy as it is displayed *is* such a thing. In contrast to teachback, the *Lev 1* explanation or 'how a topic is to be learned' comes *before* the *Lev 0* explanation of the topic is furnished. If that is furnished, the topic *is* learned and understood.

5.5 There are several tricks in the system; amongst them is the trick that once an understood marker is assigned it cannot, under ordinary circumstances, be deleted. The justification for this particular trick relies upon the fact that the conversational domain is so constructed that a concept, once established, may be reproduced as a memory. Provided the conversation remains within this domain, the concept will not be forgotten.

5.6 It is possible to constrain the *B* heuristic in various ways. For example, it is possible to sample the student's uncertainty both regarding a *Lev 0* explanation and regarding any of the exploratory *Lev 1* explanations in the learning strategy (and so, for example, to give substance to the uncertainty distributions noted in section 3.1). The subjective uncertainties are computed by using BOSS as an input device (Chapter 1, section 7).

5.7 Further boundary conditions can be imposed to regulate uncertainty and to ensure rapid learning of concepts that are retained (at least whilst attention is concentrated on *R*). Other tutorial procedures (fostering learning to learn, for example) are possible and specific data gathering facilities can be instrumented.

5.8 The theoretical superstructure outlined here, to be developed in the next volume[5], allows us to predict rather than observe the isomorphism mooted in section 2.5. Both perceptual-motor, cognitive and, for that matter, social processes are identifiable and explicable as the components of an understanding (as it is interpreted within such a theory). The isomorphism is not merely an interesting correlation or a loose similitude.

5. Or, in different terms and using a more technical approach, the superstructure is discussed in Pask, Scott and Kallikourdis (1973).