5 Learning Models

The next two chapters review two categories of learning model. The first category, discussed in the body of the present chapter, consists of models that were made when behaviourism, of rather a strict and narrow variety, was rampant. As a result, they are coloured by this philosophy. For all that, it was necessary to introduce mechanisms responsible, non-trivially, for anticipation, expectation, and also for the type of cooperative interaction that Grey Walter (1953) so delightfully imaged with a gaggle of adaptive 'tortoises' and that the ethologists, like Tinbergen (1953) and Lorenz (1952), adopted as an essential part of the scientific stock in hand.

As a result, the models did not behave in a narrowly behaviouristic fashion. In fact, they exhibited many of the tricks (fuzzy computation, for example) that are introduced more deliberately in our later work. But an unfortunate consequence of the tradition in which they were spawned was that they did not contain the hierarchical structures necessary to expose these tricks very clearly to view.

The category of learning models described in the next chapter do have an hierarchical structure. Although they represent rather simple operations, compared to more recent models made by our group, they are cognitive systems. However, they were built in the spirit of symbol processing, a more modern dogma beset by quite different constraints; for example, list-search paradigms and serial processing. Because of that and also their smallish size, they are unable to image many of the processes which, today, seem most important.

Neither category of model embodies my current view of mentation, though both contributed to it. The reader is asked to accept the fact that it is possible to model very different and more interesting facets of cognition; either in the metier of artificial intelligence (Winston, 1970; Winograd, 1972) or in the more experimental (man-machine interaction) terms pursued in my own laboratory. He is also asked to leave this aspect of modelling as pending; it receives due attention, analysis, and discussion in the next volume.

1 The EUCRATES Machines

Inspired by Wittgenstein's insistence upon the primacy of relations, Bailey, McKinnon Wood and I built a machine to learn the relationships between events in its environment. In parallel with this project, we devoted some effort to teaching the machine a subset of useful or desirable relationships and concluded that a further mechanism was needed for that purpose. This mechanism was embodied in another piece of hardware and C. E. G. Bailey christened the learning system 'EUCRATES' (the original sorcerer's apprentice).

There were several EUCRATES systems. The first of these was commissioned by the Solartron Electronic Group and was exhibited in the Physical Society Exhibition of 1955 in London. The finally engineered version is an analogue computer for use in designing adaptive teaching machines by examining their interaction with a simulated student, and was completed by Solartron in 1960. These systems were mentioned in an earlier book (Pask, 1961a) where both of them are illustrated (Plate 1(i) is the production machine; Plate 1(iii) the original machine catalysing the growth of a system of metallic dendrites), but the account of them is terse and incomplete.

1.1 Philosophy Basically, EUCRATES is a reflex 'learner' which is sensitive to reward and brings about rewarded relations between the stimuli it receives and the responses it produces. In this respect, its design owes a great deal to the pioneering work of Uttley (1959) with conditional probability machines. In addition, however, EUCRATES contains a mechanism interpretable as an expectancy mechanism and another which simulates response anticipation (or, at any rate, something akin to Thorpe's (1956) specific action potential). Further, EUCRATES is designed around concepts to do with 'curiosity' and 'attention' (it has a primitive curiosity system and there is a sense in which its attention must be occupied). Briefly, EUCRATES is a machine which looks for problems and, having found one, is impelled to learn how to solve it. This property is so basic to the design of the system that if the machine is placed in a situation that prevents the exercise of 'curiosity', then it becomes functionless.

The action of the machine can be analysed in several ways. We might, for example, adopt the elegant techniques developed by Steinbuch (1961) in connection with his learning matrices (which are special types of conditional probability machine). For the present purpose, however, the activity can be analysed quite satisfactorily by saying what the machine does and how it does it, without mathematics.

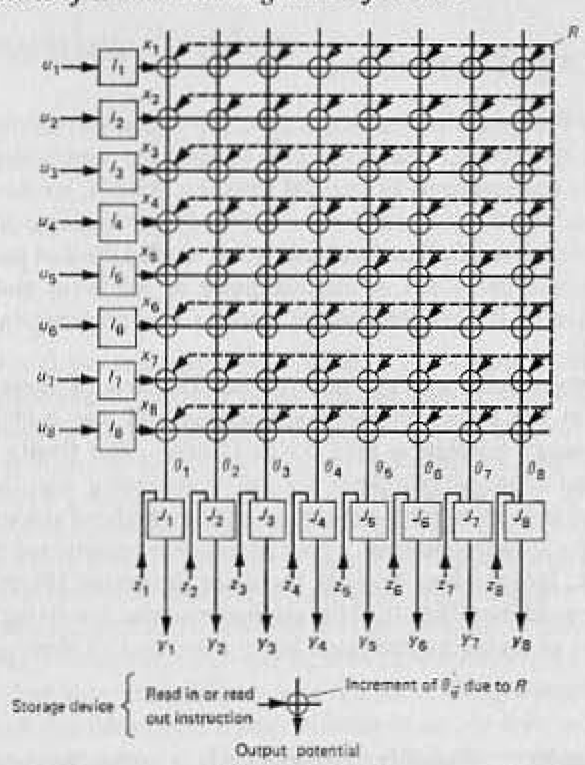

Figure 37 The plan of the machine. Each storage device retains the value of a biassing variable θ_{ij}.

1.2 Overall Plan The layout of the machine is shown in Fig. 37. It consists of a stimulus unit with components $I_1, \ldots, I_8$, an output or response unit with components $J_1, \ldots, J_8$ and an array of 8 × 8 analogue storage devices (retaining the values of variables θ_{ij}, $i = 1, 2, \ldots, 8$ and $j = 1, 2, \ldots, 8$). In the first machine to be constructed, this 'array' consisted of sixty-four capacitors with read-in, read-out and lock facilities provided by clamp diodes. (Read-out, even into a high impedance, appreciably modifies the state of an individual store.) In later machines, this rather crude arrangement was replaced by sixty-four capacitors, each of which was associated with a Miller Integrator, amplifier and impulse circuits for read-in, so that the state of a store is substantially unchanged by an interrogation.

The circuitry of the stimulus unit is shown in Fig. 38. The output of the unit is a vector, X, of eight binary variables x_i, each of which is associated with one of the components I_i. Any of the I_i consists of a local storage element, an averaging circuit, a current limiter and a lock circuit. The current limiters are cross coupled so that an increase in the current passing

through the ith circuit inhibits the action of all the remaining circuits. If a stimulus 'search' instruction is delivered (as a positive potential) to the common cathode valve, then a decision process is instituted as a result of which the current passing through one or more of the I_i exceeds the predetermined value built into the limiter. When the limit is exceeded by the ith circuit, $x_i = 1$. The lock circuit holds this condition (irrespective of the state of the machine) until a responsive event $y_j = 1$ takes place (the process whereby $y_j = 1$ is discussed below) and it delivers a 'response search' instruction to the response unit. Next, the event $x_i = 1$ delivers a negative, inhibitory charge to the ith local storage capacitor and a small positive charge to each of the remaining local storage capacitors. Then, after a short delay of δt, this event cancels the stimulus search instruction so that the decision process terminates. Finally, the event $x_i = 1$ 'opens' a row of the 8 × 8 storage array so that a vector θ_j of values θ_{ij} is delivered as part of the input to the response unit.

The outcome of the decision process depends upon the values assumed by the variables u_i in a vector U (the external input to the I) and the values assumed by eight variables μ_i (derived, as an internal input, from the storage circuits). If all of the $u_i = 0$, the action of the stimulus unit is fairly straightforward. The inhibitory cross coupling between the I_i fosters (though it does not strictly enforce) the condition that no more than one $x_i = 1$ at once. Suppose that the stimulus search instruction, cancelled by the event $x_i = 1$ is again instituted (by a mechanism to be described) and that x_i is set equal to 0. The decision process now searches for and produces the least 'likely' or the most 'unusual' event.

These conditions no longer hold true if the external input is energised; for example, it is possible to make several of the u_i positive valued and to secure several x_i conjointly equal to 1. Even so, the external input, U, still acts upon an underlying search and memory process of the sort described. In particular, because of the local storage feedback signal, the machine is prone to reject repeated and familiar inputs. Conversely, the value of u_i required to elicit the event $x_i = 1$, increases with repetition. In other words, the stimulus unit exhibits habituation.

The response unit, also shown in Fig. 38, resembles the stimulus unit. It differs in the following two respects: (a) in addition to the external input Z (analogous to U), and the internal inputs η_j (analogous to the μ_i), the J_j averager receives an input θ_{ij}, and (b) the lock circuit of a component I_i is replaced, in a component J_j, by a circuit that holds the condition $y_j = 1$, (analogous to $x_i = 1$) for an interval Δt. With these comments it will be evident that on receipt of a response search instruction (analogous to the stimulus search instruction) a decision process is set up in the response unit which leads to one or more events of the form $y_j = 1$. However, even if

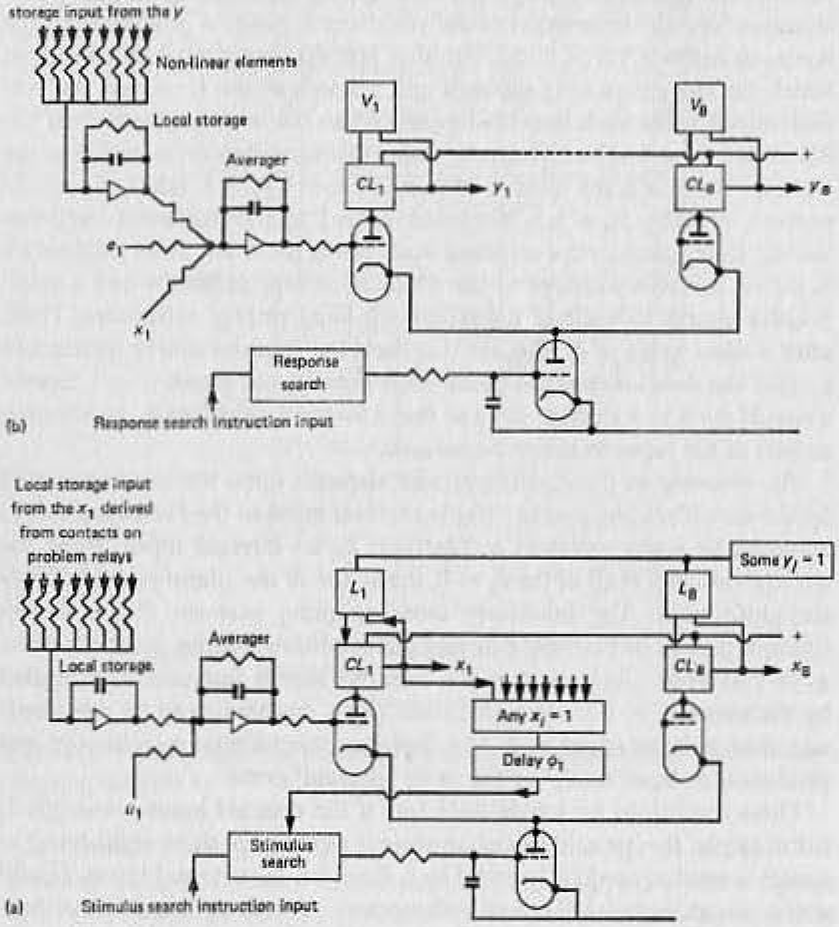

Figure 38 Circuits of the stimulus and response units: (*a*) stimulus unit. Component 1 shown in full. *CL*, current limiter; *L*, lock relays. (*b*) Response unit. Component 1 shown in full. *CL*, current limiter; *V*, relay circuit for retaining event, $y_j = 1$, over an interval of Δt.

the variables z_j in Z are zero valued, this process is biassed by an input from the 8×8 array of Fig. 37.

In considering the form of the bias, notice that the feedback to the local storage elements in the response unit is derived directly from the current limiting circuits through non-linear components. The time constants of these circuits are short compared with the time constant of the stimulus unit storage elements, and the feedback acts within the compass of a single

decision process. The effect of a high valued input θ_j at the jth component is to drive this circuit into the state $y_j = 1$. However, if the input does not do so, within a given interval, its action is inhibited by an increasingly negative η_j from the local storage circuit output. As a result of this, the θ_{ij} biassing effect is not unique, for example, the event $y_j = 1$ may occur either because θ_{ij} is higher than the other entries in the biassing vector θ_j, or because there is an ordering of the form $\theta_{il} = \theta_{im} > \theta_{ij}, \ldots$ when the tendencies to produce $y_l = 1$ and $y_m = 1$ compete and are inhibited by increasing η_l and η_m so that ultimately $y_j = 1$. A large number of competitive impasses are resolved in this fashion.

The response search instruction is delivered whenever $x_i = 1$; further, the particular row of the 8×8 array presented as a biassing input to the response unit depends upon the particular event x_i (this event 'opens' the row of storage devices and presents a vector θ_i to the response unit); hence, the array serves to couple the stimulus and the response units. To close the cycle, note that the stimulus search instruction is delivered at the termination of any event $y_j = 1$ (that is, Δt seconds after such an event has occurred).[1]

The cyclic action gives rise to sequences of events $x_i = 1, y_j = 1$, which overlap by Δt seconds to form conjoint events $\langle x_i, y_j \rangle = \langle 1,1 \rangle$. The values θ_{ij} retained by the 8×8 storage device in the 8×8 array depend upon the value assumed by a variable, R, at the moment when a relevant conjoint event occurs. Thus, at trial n in a sequence, if a single event $\langle x_i, y_j \rangle = \langle 1,1 \rangle$ occurs, the value of θ_{ij} is incremented or decremented by an amount proportional to Δt multiplied by the difference between R and θ_{ij}.

Conversely, the values of all of the entries in the vector θ_i are decremented towards 0 by an amount proportional to the values of these entries and the interval between the event $x_i = 1$ and event $y_j = 1$. This interval is later called the 'anticipation latency'. The variable R (which is later interpreted as a 'reinforcement variable') can assume values in the interval $-1, 0, +1$ and a null value $\emptyset$, if $R(n) = \emptyset$, then $\Delta\theta(n) = 0$.

The incrementation process is more complicated if a pair of events $x_i = 1$, $x_k = 1$, occur simultaneously or if they overlap before some $y_j = 1$. Suppose, to begin with, that $x_i = 1$ and $x_k = 1$ and $y_j = 1$ at the nth trial and that $R(n) = \emptyset$. In this case, if $\theta_{ij} > \theta_{kj}$ then the entry θ_{ij} receives an increment $\Delta_1\theta(n) \approx c(\theta_{ij} - \theta_{kj})\Delta t$. If $R(n)$ is not equal to $\emptyset$, then the total incrementation is complex and consists of a component $\Delta_2\theta(n)$, the 'extrinsic reinforcement' and a component $\Delta_1\theta(n)$, the 'intrinsic reinforcement'. So far as decrementation is concerned, if $x_i = 1$ and

1. If the θ_{ij} are constant valued, this coupling is such that the machine acts as a stochastic matrix multiplier. When, as below, the θ_{ij} are variables, the entries in the stochastic matrix are functions of the performance of the machine.

$x_k = 1$ all of the entries in θ_i and all of the entries in θ_k, including θ_{ij} and θ_{kj}, are reduced in the direction of zero in proportion to their several values and the length of the anticipation latency. Because of this, and because of the active characteristics of the stimulus and response units, the maintenance of a pattern of θ_{ij} values entails some sort of dynamic organisation.

1.3 The System Let us talk about the system anthropomorphically. On receipt of a stimulus search instruction, the stimulus unit 'expects' to receive a 'problem' (some non-zero value of X). The problem it does receive will be biassed by 'evidence' U, from its environment but, even if there is no evidence, some problem will ultimately be posed (as a result of the stimulus unit decision process). This state of 'expectancy' gives rise (when the problem is posed) to a state of 'anticipation' regarding a solution to this problem. A solution is *denoted by* some non-zero value of Y, and it is achieved by the response unit decision process, instituted through the response search instruction by some $x_i = 1$. The solution depends, even in the absence of the external input Z, upon the θ_{ij} values in those rows of the 8×8 array that are selected by the particular problem. The state of 'anticipation' terminates after a solution is *designated* and its termination gives rise to a further state of problem 'expectancy'.

In all this, the θ_{ij} values play a crucial part by determining what solutions will be given to various problems. An assumption that values of R represent values of some reward or desirable commodity, and that the machine would like to solve problems in such a way as to maximise the average value of R, is effectively built into the design. For, as a result of the process just described, the machine 'learns' to solve problems in a manner that is compatible with maximising the average value of R (the θ_{ij} assume a pattern that leads to this result, providing that specific and consistent values of R are associated with specific problem solution pairs).

1.4 Modification of Habituation Mechanism The habituation of the machine requires further comment since it is, in some respects, unrealistic. If a stimulus is repeated, the strength of stimulus required to elicit the problem state $x_i = 1$ increases. This is more or less as it should be, provided that $x_i = 1$ does not lead to a reinforced response (a significant event). If the result of a stimulus is significant (if $\langle x_i, y_j \rangle = \langle 1,1 \rangle$ is reinforced) then the present machine performs as it would do if the stimulus were not significant (at least, it does so with respect to the expectancy process; the anticipation latency, the interval occupied by the response unit decision process, is reduced). This is not altogether satisfactory from a biological point of view because in a real organism habituation with reference to significant stimuli is suppressed or nullified. The model works the right way

round for the anticipation latency; but it is perverse in respect of the expectancy latency for significant events.

The right characteristic can be secured in several ways. It was in fact, obtained by modifying the local storage feedback to the stimulus unit so that, instead of a simple inhibitory signal, this feedback became, at the nth trial and for the event $\langle x_i, y_j \rangle = \langle 1,1 \rangle$ a quantity $1 - \theta_{ij}(n)$ where $\theta_{ij}(n)$ is the quantity delivered as the internal input to the jth component of the response unit. If the problem event induced by a stimulus is not significant, $1 - \theta_{ij}(n)$ is high (and habituation takes place as before). If the problem event is significant, then $1 - \theta_{ij}(n)$ is low valued and habituation is suppressed. This arrangement gives the machine the required characteristics, but it does so 'unfairly'. Additions of this sort, constitute an attempt to represent, in a primarily single level model, the organisation of a many level, and hierarchical structure.

1.5 Experimental Facilities For experimental purposes the EUCRATES 'learner' is connected to a console from which the experimenter can act upon and observe the main variables as well as receiving an 'illicit' (from one stance) view of the interior via a θ_{ij} display. As shown in Fig. 39 the learner may either be manipulated directly or coupled to a EUCRATES teacher (not yet described, but essentially the same as the learner apart from a full representation of the task relation to be learned and circuits for computing criteria of performance, i.e. of the extent to which behaviours satisfy the task relation).

In addition, the teacher can be coupled to a real-life student so that comparative observation is possible (student/EUCRATES teaching machine in contrast to EUCRATES learner/EUCRATES teaching machine). Though important for the main application (design of smaller, degenerate, teaching machines) this facility is not discussed.

2 Ways of Looking at the System

The EUCRATES learner can be viewed in several ways. During the late 1950s it was fashionable to take a strongly behaviouristic point of view and it will be instructive to start with this image in mind and to modify it as needs be.

2.1 The Machine Regarded as a 'Black Box' Consider the machine as a 'black box', in Ashby's sense. It has an input U, which may be interpreted as a stimulus, perhaps complex in form, which gives rise, internally (and beyond the behaviourist image) to a problem vector X. It has an output, Y, which may be interpreted as a response vector designating an hypothetical

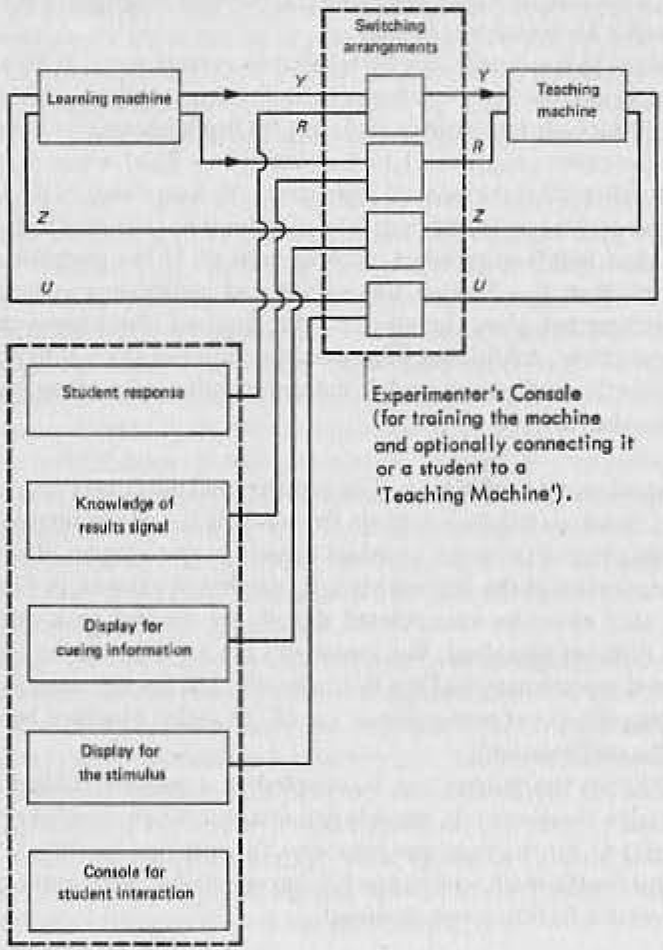

Figure 39 Switching arrangements for connecting either the learning machine or a real-life student into the teaching machine.

solution naming vector, if all of the y_j are brought out to external connections. But, in many arrangements of the system, only some of the y_j are brought out. In any case, there is a great deal of difference between a solution (the response *process*) and a response which may select in context amongst possible solution *names*.

The machine has an input R, which, given a correspondence between stimuli U and responses Y, acts as a reinforcement variable. The machine

has an input Z which manifestly guides it in selecting its response (we can prevent the machines from making response j by holding the jth component of Z negative-valued and force it to make response j by holding the jth component of Z positive-valued). At least this is true in *most* conditions; to be *certain* of doing so, it would be necessary to delve inside the black box. Hence, Z is identified with a cueing vector.

If this identification is adopted, it is convenient to relabel a vector U, in which one component is positive and the rest are zero valued by a *letter* (calling such a vector 'stimulus a' $= U = \langle 1,0,0,0,0,0,0, \rangle$ or 'stimulus b' $= U = \langle 0,1,0,0,0,0,0 \rangle$; there are eight of them in all).

Similarly, the response events $y_i = 1$ are named 'response A' $= Y = \langle 1,0,0,0,0,0,0,0 \rangle$, or 'response B' $= Y = \langle 0,1,0,0,0,0,0,0 \rangle$. We may, of course, apply more than one stimulus at once, in which case the vector U has more than one non-zero component. (Thus, the conjunction of 'stimulus a' and 'stimulus b' is the vector $U = \langle 1,1,0,0,0,0,0,0 \rangle$.) It is also possible for the machine to make more than one response at once.

At this point, it is apt to comment upon an ambiguity in the use of words like 'stimulus' and 'response'.

'Stimulus' can mean.

1. The state of the environment that is appreciable by the machine at a given trial.

2. An appreciable and separable event (the state of some part of the total environment).

If 'stimulus' is interpreted according to (1), then the several components, u_i in U, are properties of the stimulus, designating appreciable attributes. In this case, it is possible to talk about stimulus discrimination (that the machine comes to select certain properties as important and to disregard others as unimportant). However, it would not be reasonable to say that a pair of stimuli occurred simultaneously or that the machine rejected one of this pair and respected the other. (This situation might be described by saying that the machine attends to one property of the total stimulus and disregards the others.) Conversely, if definition (2) is adopted it is entirely possible to have coincident stimuli. But stimulus discrimination effects will be suppressed and will only be manifest in connection with phenomena such as habituation that are predicates of sequences of trials.

The 'stimulus a' and 'stimulus b' nomenclature opts in favour of definition (2) rather than (1). It is important to realise that this decision is not forced upon us by the machine; it is made to suit the experimental context in which the machine's behaviour is examined. The machine's behaviour could be equally well described within the framework provided by (1) or by (2). Many phenomena are the consequences of the *nomenclature* that has been selected.

Similar comments apply, more obviously perhaps, to 'response'. According to a definition that parallels (1), the machine, at any trial, makes a response which is a vector Y (to be more specific, adjoin latency information, stating when the various components y_i of Y undergo a transition $0 \rightarrow 1$). If, on the other hand, a definition akin to (2) is chosen, then the machine may make several 'responses' at a given trial and specificity is achieved by saying how and when these several responses are made. The word 'response' is itself slightly misleading (it would be more usual to say 'the machine's response has several parts—these parts of the total response are made in such and such a fashion'). As a result we shall take the liberty (commonly taken in the literature) of using the word 'response' to denote either the total configuration or the parts of this configuration, providing that the practice does not give rise to confusion.

2.2 Interpretation of the Learning Model Is this model a finite automaton, albeit a probabilistic one, i.e. is the artifact a finite-state machine? It certainly is, if regarded as a rather simple and straightforward physical object. It is, under interpretation, too; provided we subscribe to one or other of the definitions of 'stimulus' and of 'response' explained in the last section. But, however useful these constructs may be as approximations, there are good reasons for rejecting all of them.

The trouble is that the machine has several asynchronous loci of control or several asynchronous automata that act as parts; for example, the 'horse race' of the response resolution unit or even one component of it (each 'horse' in the race). Hence, strictly speaking, we cannot, in the capacity of users or model builders, regard aggregates of these asynchronous components as *having* states to begin with, just because the units *are* asynchronous. As conceded in the last paragraph, the physicist who regards the machine as a chunk of material is in a somewhat different position.

The crucial issue is independent of degree of resolution. Clearly the input and output 'states' posited earlier in section 2 are of low resolution; so are the internal states that might go with them. If the machine *could* be regarded as a finite-state machine the state specification could be refined; either by partitioning or by using behavioural histories to delineate irredundant (algebraic) internal states. But, as it is, the notion of state is undermined (deliberately so, in order to secure the anticipation and expectation properties) by the existence of asynchronous parts.

It should be emphasised that the machine is composed of asynchronous or autonomously clocked parts and is not a parallel computer such as a perceptron (where the ongoing computations, though occurring simultaneously, are centrally clocked). The crucial feature of the machine is that otherwise asynchronous computations *become* synchronous (or partially

so) by virtue of information transfer due to the global computation that the machine is currently executing; very similar entrainment phenomena occur in Kilmer and McCulloch's model for the mammalian attention directing system, SRETIC (Kilmer, McCulloch and Blum, 1969). Further the same comments apply if the model is realised as an actual machine (it has been, using the PSV modules of Barron (1968) and Gilstrap as the units for computation) rather than simulated in a digital computer with the housekeeping needed to make a serial processor mimic a concurrent process. Similar remarks apply to various mesh-like collections of components, derived from Beurle's original model (Beurle, 1954, 1959), noted in Pask (1961a), provided there is no degeneracy due to simulating concurrency as a randomisation of serial events. Here, the activity consists in the propagation of wave-like modes of activity. There is an interesting correspondence between initially asynchronous operation and incoherent (or 'random-phase' wave propagation) and partial synchronicity and coherent (or 'in-phase' wave propagation) which is tacitly employed, for example, in Longuet-Higgins and Michie's (1970) holographic model for memory.

So far as EUCRATES is concerned, the preferred interpretation is as follows. The machine was probably the first device to realise a fuzzy algorithm in the sense of Zadeh (1973) and for this reason is rescued from present-day triviality. In any fuzzy computation (Chapter 2, section 2) it is possible to insist that the (generally) fuzzy output is made to select one element from amongst a set of alternatives (using any convenient rule such as choosing a maximal or a minimal element) and, as a matter of fact, this kind of expedient is used in coupling the machine to its environment. But, as will be shown, it is neither necessary nor usual to convert fuzzy outputs into 'alternative selections' *within* the machine; one part can, and often does, accept (as a fuzzy input) the fuzzy output of another part.

Consider the response-resolution process (much the same comments apply to the stimulus-resolution process). If the machine is constrained to select one and only one response, which is counted as *the* output then the computation does select amongst alternatives (the response set). But, left on its own the machine generally selects *many* responses in various orders so that the entire output is fuzzy. This statement can be interpreted in several ways; for example, by noting that the entire output determines an element with grades of membership in several sets, which might be numericised by noting the order of the selections and the moments at which they occur. But the output can also be passed on to another part of the machine as a fuzzy output *simpliciter*, without numericisation. This occurs, for example, if the response resolution unit is back coupled to a stimulus unit when the machine interpretation of the fuzzy output (*as* a fuzzy input) depends upon the state and the clocking of the stimulus units. Concurrent computation occurs in so far as the otherwise asynchronous clocks in distinct parts of the machine are synchronised because of such an interaction.

In the experiments to be described, pains are taken to make the machine act *as though* it was a finite-state machine; for example, by external constraints that ensure that (or maximise the probability that) one and only one response will be selected. Similar precautions are used to secure a 'begin and end' ordering; that a response *follows* a stimulus. This part of the design is introduced *simply* so that an observer can synchronise events in the machine with his own stopwatch. It is important to bear in mind the fact that all such gambits involve constructs that are imposed upon the machine in order to make it behave *as though* it was a finite-state machine or (at most) a fuzzy-state automaton with numericised input and output. Nearly all of the interesting or emergent behaviours are due to the fact that the machine is not *actually* a finite-state machine and because the efforts to make it act as though it was such a thing are not and *cannot* be altogether successful.

Experimenters use similar tricks to make real organisms (human beings included) act as though they were finite-state machines, so that they can be coupled to a further finite-state machine representing the environment and described within the strict behavioural format. If these tricks were successful the experiment would be of small consequence (a point taken up in section 6). The interesting behaviours appear in so far as these tricks do not always work for real organisms any more than they do for the EUCRATES machine.

3 Learning Experiments

In a suitable environment, the machine is able to mimic many features of classical conditioning and instrumental conditioning. To make it play these tricks, we tacitly identify the device with a laboratory animal such as a pigeon or a rat.

A few preliminary remarks are needed. In the context of classical conditioning, a stimulus is partially causative (it is something that gives rise to, or helps to give rise to, an observable response). Responses would not occur without stimuli of some sort (though these may be internal stimuli or stimuli that are concealed from the observer). By way of contrast, the field of instrumental conditioning is built around the concept of an operant, that is, a response emitted by the organism autonomously. Stimuli are chiefly discriminating stimuli; events which, in Skinner's words, provide the occasion for a response but do not give rise to it. They influence and bias an autonomous response process.

The distinction, in experimental psychology, between a stimulus and a discriminating stimulus is tenuous (but the philosophical orientations of

these different fields are distinct enough to deserve respect). So far as the machine is concerned, the distinction between a stimulus and a discriminating stimulus is a matter of degree – some stimuli are more causative than others; there is an underlying autonomous activity; the hidden stimuli that produce it are not made explicit but could be invented.

3.1 The Classical Conditioning Paradigm If a stimulus is applied (say stimulus a) and, if this gives rise to the problem state $x_i = 1$, then the machine will emit one, or usually one, response (say response A). The sequence of events 'stimulus a, response A' can be interpreted as a reflex.

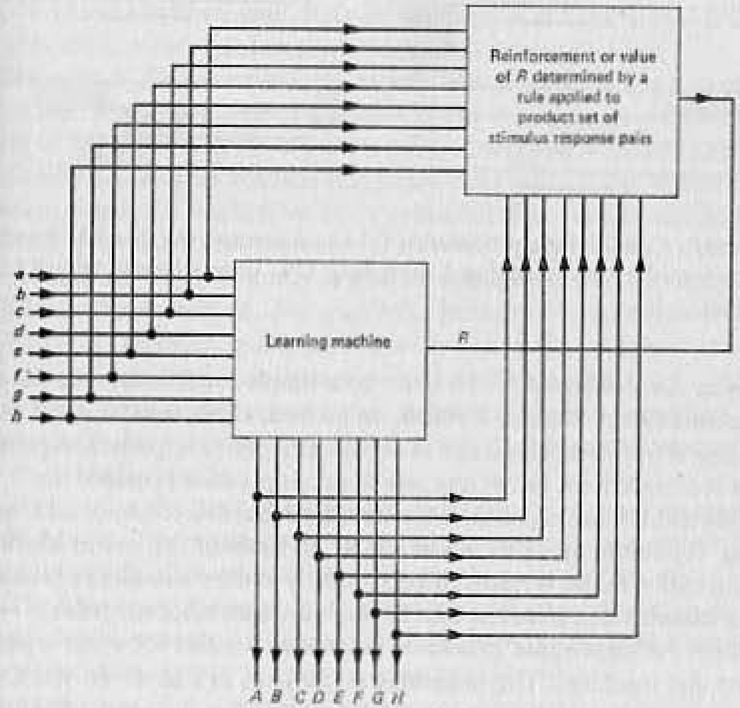

Figure 40 Reinforcement procedure.

To prepare the machine for experiments in classical conditioning, it is first necessary to embed in it certain unconditional reflexes – by hypothesis, innate and inherently reinforcing reflexes. The arrangement shown in Fig. 40 is used for this purpose. A subset r of outcomes or stimulus response pairs is associated with a high reinforcement value, the remaining outcomes being associated with a low value reinforcement. After a number of trials in the conditions of Fig. 40 the machine adapts so that it has a number of ingrained unconditional reflexes, the members of r. We shall assume that $a \to A$ is a member of r and will mention any other reflexes that belong to r, explicitly.

In its adapted condition, the machine demonstrates some of the simple features of classical conditioning. Although this point is not developed, the changes in response latency with conditioning also replicate the form of the latency changes that are observable in animal conditioning.

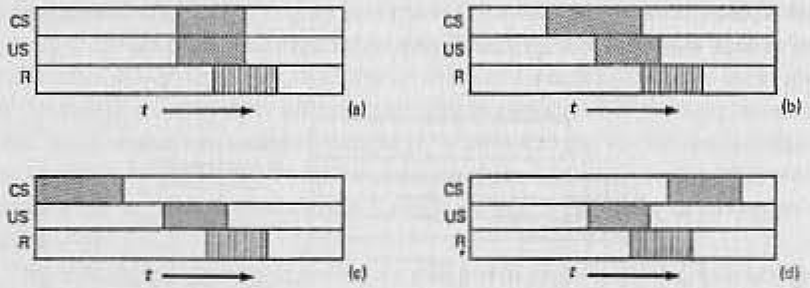

Figure 41 Conditioning procedures: (*a*) simultaneous; (*b*) delayed; (*c*) trace; (*d*) backwards. CS, conditional stimulus; US, unconditional stimulus; R, response.

Reflex Establishment To build up a simple conditional reflex choose an unconditional stimulus b which, on its own, elicits some unconditional response B (or which elicits one of several alternative responses, depending upon the conditions, but at any rate does not produce response A).

The conditioning procedures for simultaneous, delayed trace and 'backwards' conditioning are shown in Fig. 41, in terms of the way in which the conditional stimulus is related experimentally to the unconditional stimulus a (the stimulus that gives rise to A through the unconditional reflex $a \to A$). Of these, the 'backwards' procedure is ineffective (either for a real organism or for this machine). The remaining procedures in Fig. 41 do work. For organisms, both simultaneous and delayed conditioning are readily instrumented but delayed conditioning is more effective than simultaneous.

So far as the machine is concerned, trace conditioning is impossible. This occurs because there is no mechanism for retaining the 'trace' of the conditional stimulus which, in this procedure, is presented in the past and terminates before the unconditional stimulus. However, 'trace' conditioning is possible in certain more elaborate versions of the machine which contain loops capable of retaining a representation of the past stimulus. Both simultaneous and delayed conditioning work well, as they do for the organism, and delayed conditioning is the more effective. Thus, if the

unconditional stimulus b is paired (before and over-lapping with) the conditional stimulus a, and if this pairing is repeated, the following effects may be obtained: (1) response B either drops out or is modified; (2) stimulus b comes to elicit response A (or, occasionally a modified form of A) in the same way that $a \to A$; (3) the connection $b \to A$ is associated with a decreasing b, A, latency; (4) the connection $b \to A$ is eventually established, even in the absence of a. Hence, (5) $b \to A$ is a simple conditional reflex. Further, once $b \to A$ is well established, it may be used as though it were an unconditional reflex to produce another conditional reflex. Thus, if $c \to C$ is associated with $b \to A$, we can establish the reflex connection $c \to A$ and so on. Whereas the primary mechanism at work in establishing $b \to A$ is the 'extrinsic reinforcement', the only mechanism at work in establishing $c \to A$ is the 'intrinsic reinforcement' described in section 1.2.

Since each of these conditioning experiments is beset by the caveat 'or a modified form of response' a sequence of experiments can give rise to a variety of behaviours, some of which will be modified responses that are not manifest within the conditioning experiment itself. Thus, before conditioning, perhaps $d \to D$; after having established $b \to A$, it may be that $d \to E$. Bear in mind that the experimenter is grappling with a dynamic, restless machine that does something on its own accord if he fails to stimulate it in a suitable fashion. The machine's behaviour has a richness comparable to that reported by Pavlov and his school in their original writings (though its conditioned activity is more or less deterministic).

The chief distinctions between the behaviour of an animal and the behaviour of the machine occurs at a level of complexity stressed by Konorski (1962), namely:

1. Only certain features of a stimulus are relevant for conditioning (the animal has an innate structure, resembling a releaser organisation, in respect to aspects of its environment that may become significant).

2. The features of the stimulus that are relevant often depend upon the form of conditional response (certain features are significant in respect to some responses and other features in respect to others).

The deficiency of the machine as a model compatible with (1) is due entirely to the chosen identification of the stimuli as unitary entities (without this constraint it will mimic an organism). To achieve a performance compatible with (2), it is necessary to introduce a hierarchical structure.

Reflex Extinction If a conditional reflex such as $b \to A$ is elicited repeatedly in the absence of a it becomes extinguished. The b, A latency increases and ultimately b fails to produce A. The phenomenon of extinction is not a forgetting process (which may also occur due to the decay of the θ_{ij} values). Unlike forgetting, it depends upon the repetition of stimulus b.

Similarly, the phenomenon of 'spontaneous recovery' (which used to be pointed out as a demonstration that extinction is not merely forgetting) can occasionally be observed as part of the machines' behaviour. If $b \rightarrow A$ is extinguished to a long latency L_2, and if (after a rest interval) $b \rightarrow A$ is tested by applying b, then, if the rest interval is long enough, $b \rightarrow A$ will again be manifest at a shorter latency L_1 where $L_2 > L_1$. Spontaneous recovery in the machine is due to the fact that the response latency is the sum of the expectancy latency and the anticipation latency. Extinction increases the length of each latency period. After a rest interval the expectancy latency is likely to be reduced by the normal feedback to the stimulus unit local storage circuits so that the total response latency is L_1 rather than L_2.

3.2 The Instrumental Conditioning Paradigm For experiments in instrumental conditioning, the reinforcement variable is under the experimenter's control and is continually manipulated. (This is in direct contrast with classical conditioning, where reinforcement is a function of the organism's state or of an innately determined relationship between the organism and its environment.) The reinforcement may be interpreted as primary, examples of which are the delivery of food, events closely related to the reduction of a basic drive or, even better, the electrical stimulation, through an embedded electrode, of a positive centre in the brain. (Flood uses this technique in adaptation studies and the results are dramatically consistent. As a philosophical premium, this is what 'reinforcement' really is.) Alternatively, the reinforcement may be interpreted as *symbolic*; really, a misnomer, for the condition would be better stated as confirmation of an internal hypothesis.

Primary reinforcement will be considered first, in particular, reinforcement of an operant response. An operant is a response produced autonomously by an organism or by the machine, and the instrumental mode of conditioning depends upon the fact that autonomous responses are emitted to meet with reward or punishment. (The machine contains no analogue for punishment, i.e. a negative centre in the organism's brain. Thus it is only possible to achieve positive reinforcement or negative reinforcement of a response.) Discriminating stimuli are introduced to modify the autonomous activity or are provided, by the experimenter, in response to the activity of the system. For instrumental conditioning experiments the initial condition of the machine is a *tabula rasa* (apart from any residual charges that assign values to the θ_{ij}). In this condition, some response, say A, will be fortuitously rare and the experimenter selects A as the 'operant response' with which he is going to deal. In particular, he concentrates on response A and disregards other responses.

Interval Reinforcement Interval reinforcement consists of assigning a positive value (say of $R = +1$) to the reinforcement variable on the first occasion that A occurs after a fixed interval from the last occurrence of A. With respect to this procedure, the machine simulates an organism such as a pigeon in so far as the rate of response A is inversely proportional to the interval chosen. The reinforced response A is, of course, gradually extinguished if the reinforcement is withheld. Extinction is reflected in the response A latency and consequently the graph of the number of responses A per minute shows the short-term slowing down fluctuation that is characteristic of the animal subject. The fluctuations may be reduced by using randomly chosen reinforcement intervals around a fixed mean value (in place of a fixed interval). Although this procedure leads to higher response rates, it does not lead to a greater resistance to extinction of the operant conditioning (as it does in the pigeon).

Ratio Reinforcement In ratio reinforcement, the response A, is reinforced after a fixed number of occurrences, regardless of when they occur; the 'ratio' concerned being the ratio of unreinforced to reinforced responses. The experimental finding for animal subjects is that the larger the ratio, the more rapidly are responses emitted.

The machine parallels this behaviour within limits, and this is, at first sight, surprising. On closer inspection (or an illegitimate peep inside the black box) the reason is fairly straightforward. With larger ratios, the response A is reinforced in connection with a greater diversity of problem events $x_i = 1$. The latency is compounded from a stimulus unit latency and a response unit latency and the mean rate of response A depends inversely upon the average latency. But the average latency is less, as the problem events in connection with which the response A has been reinforced are made more diverse.

As with interval reinforcement, the behaviour shows the characteristic slowing down fluctuation which, once again, may be reduced by choosing a ratio at random (around some fixed average) in place of the fixed ratio. This procedure leads (as might be expected in view of the glimpse at the internal mechanism) to rather high response rates.

3.3 Discriminating Modes In the psychological laboratory, an animal may be trained according to several alternative reinforcement schedules, each of which produces a characteristic pattern of behaviour. If each reinforcement schedule is associated with a discriminating stimulus, such as different lights α and β, the appearance of a particular discriminating stimulus elicits the characteristic behaviour in the fully trained animal. The machine can be used to simulate this sort of experimental result if we choose a subset of the 'stimuli' as representative of one light and another subset of the 'stimuli' as representative of the other. Thus, light α is simulated by the vector $U = \langle +1,+1,+1,+1,-1,-1,-1,-1 \rangle$, and light β is simulated by the vector $U = \langle -1,-1,-1,-1,+1,+1,+1,+1 \rangle$. In this case, the machine is segregated, functionally, into a pair of subsystems α^* and β^*, one or the other of which is trained in connection with a particular stimulus α, β. Subsequently, the appearance of α will elicit one behaviour (α^*) and the appearance of β will elicit the other behaviour β^*). So far, the machine is fairly lifelike. Unfortunately, the machine cannot react as a real organism and test the reinforcement conditions of its environment to determine how it should perform. To simulate this 'choice' or 'test', it is necessary to use an hierarchically structured machine of a sort described later.

Next consider an extremely important characteristic of stimuli that act as discriminating stimuli; namely, that they become secondary reinforcers. The paradigm experiment consists of reinforcing a response, say response A, and associating the reinforced occurrence of response A with a stimulus b which would not, on its own, elicit response A. After the conditioning is established, the response is extinguished by withholding stimulus b and allowing the organism to respond with A for trials that are not reinforced. If the reintroduction of stimulus b without any reinforcement leads to the production of response A (or to a decrease in the response A latency) then stimulus b is a secondary reinforcer (in the initial pairing process it acquired the property of acting like a reinforcement).

Obviously, if this experiment is performed with the machine it will yield a positive result. Stimulus b, having been reinforced in connection with response A, is part of a conditional reflex $b \to A$. Hence, by previous definition, stimulus b does act like a positive value of R. Further, as in the classical conditioning experiments, it is possible to build up conditional reflexes such as $c \to A$ by association with $b \to A$. Finally, because a response is not determined uniquely by the problem state evoked in the machine, it is possible to manipulate and modify the mode of response by using secondary reinforcing stimuli.

Unfortunately, in the machine as described, the secondary reinforcing stimulus will not play the crucial trick of reinforcing the emission of some response other than A, say response B. Secondary reinforcers are not generalised as they are in animal learning. It is a comparatively simple matter to modify the machine (by introducing a further mode of 'internal' reinforcement) so that 'generalisation' of this kind does take place. The modification works. But, as it probably works for the wrong functional reason, we shall not dwell upon it. (To make it work for the right functional reason entails building a machine that can accept several sorts of reward, rather than the single reward of value R.)

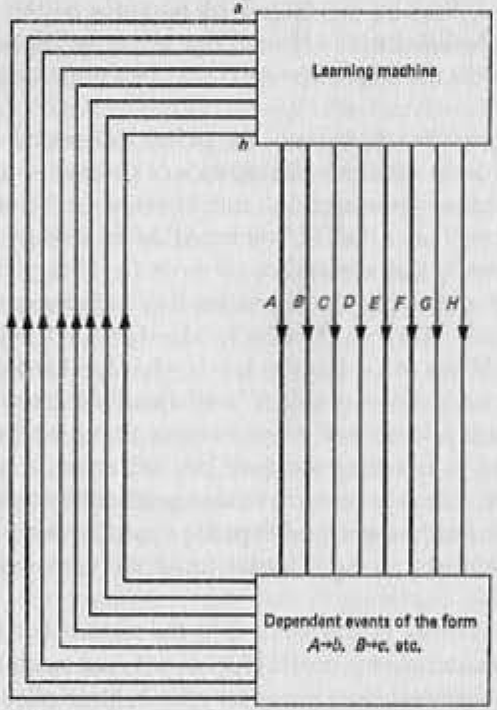

Figure 42 Contingent environment.

3.4 Chain Responses It is easy enough to establish chains of reflex actions. If the production of response A gives rise (in the environment of Fig. 42) to stimulus b and if $b \to C$ is a conditional reflex, then the chain $A \to b \to C$ is built up. The process may be repeated up to a chain length of eight as a maximum; for example, to produce chains such as $A \to b \to C \to c \to D \to d \to E \to R \to F \to f$.

As in the case of animal conditioning, the production of chain reflexes relies upon the initial establishment of the component reflexes and a later assembly of these components into the composite reflex entity.

3.5 'Symbolic' Experiments The experimental environment can be given a quasi-symbolic interpretation, as follows. In the case under discussion, some relationship, $\mathscr{R}$, is defined between the stimulus set and the response set. One $\mathscr{R}$ is a permutation of the numbering of the stimuli with reference to the numbering of the responses; but a 'many-to-one' relationship is equally valid. Given a stimulus the machine is required to produce the correct response, that is, the $\mathscr{R}$ related response to this stimulus. The variable, R, is identified with a 'knowledge of results' signal that assumes the value $+1$ if the machine response is a correct response and the value 0 if it is a mistake.

Within this framework, stimuli are to be interpreted as events that definitely occur or do not occur; the absence of a stimulus negates an event, i.e. complementation is presupposed, and, in a very weak sense, the machine entertains an hypothesis that is 'confirmed or denied' by 'knowledge of results'. To embody this constraint, we write '-1' in place of '0' in the stimulus vector and apply the corresponding negative potentials to the machine; thus 'stimulus a' is $U = \langle +1,-1,-1,-1,-1,-1,-1,-1 \rangle$, and 'stimulus b' is $U = \langle -1,+1,-1,-1,-1,-1,-1,-1 \rangle$. Responses designate solutions to a relational problem. 'Anticipation' becomes the characteristic of a selective process and it makes sense to regard the response unit decision process as resolving more or less 'uncertainty' regarding an $\mathscr{R}$ related response, given the problem state engendered by a particular stimulus. (The amount of 'uncertainty' depends upon the form of the decision process at a given trial; the greater the uncertainty, the longer the anticipation latency.)

The cueing variables z_j now come into the picture as constraints upon the response unit decision process. If $z_j = -1$, the machine is instructed that the decision process must not select $y_j = 1$. Generally, the assignment of negative values to some or all of the z_j for some or all of the anticipation interval at a given trial, will reduce the machine's selective uncertainty at that trial. (In conditioning terminology, this would be response differentiation.) Conversely, of course, if $z_j = +1$ the machine is instructed to select $y_j = 1$. Either mode of cueing is possible but it is sufficient to consider the case in which the z_j are assigned negative values.

The machine can be trained to use a skill entailing knowledge of $\mathscr{R}$. The training procedure amounts to repeatedly presenting each stimulus, constraining the machine to make a correct response by the use of the cueing variables, and positively reinforcing the correct response when it occurs. Further, the machine will continue to perform the $\mathscr{R}$ based skill as long as the environment provides it with a sufficiently varied and rapid sequence of stimuli to replace the variety of the stimulus search process. (Recall that the stimuli are associated with definite values of $u_j = 1$.) In particular, the machine performs adequately as a chain conditioned device where $\mathscr{R}$ is of the form $a \rightarrow A$, $b \rightarrow B$, and the environment completes the chain by consequential events of the form $A \rightarrow b$ and $B \rightarrow c$.

At first sight, these operations amount to a rather stupid renaming of entities; in particular, inputs have been constrained so that the operation that used to be called 'reinforcement' is plausibly relabelled 'knowledge of

results'. But, on closer examination, something more has been gained. The concurrent modes of operation and the synchronisation phenomenon that are latent in the machine's activity (and which give it non-trivial attention and expectancy properties) break through the framework of constraints imposed to *mimic* problem solving or symbolic transformations in such a way that certain conditions *cannot* hold and certain types of training are rejected. The machine, constrained to this extent, demonstrates a number of 'abhorrences' some of which are very lifelike.

The machine cannot learn to sit quietly until a stimulus occurs (perhaps for an indefinite interval). The condition in which all stimuli are negated is contrary to the tenets of the model. Hence, it cannot be trained to 'learn $\mathscr{R}$' in the commonsense meaning of the phrase, if $\mathscr{R}$ applies to the entire repertoire of stimuli and responses. The abhorred condition is 'all $u_i = \ldots 1$' (so that the stimulus search process is effectively inhibited and no problems are produced). The electronic consequence of this condition is that no 'response search' instruction is delivered, and, ultimately, all of the responses assume states $y_j = 1$. The philosophical consequence is that we have disobeyed the (tacitly stated) rules of a machine that should actively search for problems and, given problems, search for their solutions.

To avoid breaking the rules whilst giving a commonsense meaning to the phrase 'the machine learns $\mathscr{R}$', it is necessary to restrict the stimuli and responses related by $\mathscr{R}$ to a subset of the entire repertoire, for example, by saying that $\mathscr{R}$ relates stimuli, a, b, c, d, to responses A, B, C, D; that stimuli e, f, g, h, feature as 'internal stimuli' with $u_e = 0$, $u_f = 0$, $u_g = 0$, $u_h = 0$; and that responses E, F, G, H, are 'irrelevant responses'. Problems corresponding to e, f, g, and h, may thus occur without external stimulation, and responses, E, F, G, and H may be made without being counted as relevant responses. By this expedient the machine is given a field of attention on which it can concentrate even if the relevant (skill orientated) field of attention is held void by the experimenter.

4 Teaching and Training

A machine that has been partitioned in this fashion to mimic problem solving, cannot be trained or conditioned by the simple extrapolation of techniques that *would* work for any similarly dressed-up reflex making device. There are some broad, but from the foregoing comments, obvious principles of effective teaching. These will be stated. After that, the principles will be embodied in a suitable teaching machine. Although these comments apply to a machine in which the 8×8 array is a *tabula rasa*, it is much more interesting to consider the case of a machine with previous experience (so that the 8×8 array contains definite entries θ_{ij}). In particular, a part of this previous experience is likely to interfere with the acquisition of the $\mathcal{R}$ skill.

The principles are as follows:

1. The teaching procedure must retain the machine's attention. If it does not, the stimulus search process will give rise to irrelevant problems. Conversely, the rate of presentation of stimuli must not exceed the point at which the machine becomes overloaded.

An instrumentation of this principle involves a feedback loop, through which the instructor is informed whether or not the machine's attention is occupied. As a result of this information, he (or it) may adjust the rate and variety of the stimulating input.

2. Correct responses must be reinforced. If the knowledge of results were genuine it would be sufficient to say that a certain response is correct. Here, assigning a positive value to R is, of course, a trick standing in place of real confirmation. But another feedback loop involving the machine and the instructor is required for this purpose.

3. Effective teaching should allocate more effort to problems that give rise to individual difficulties (the instructor should present the machine more often with stimuli that produce mistaken responses, so that these aspects of the skill are well rehearsed). Once again, this principle involves a feedback loop. The instructor must determine which stimuli are associated with mistaken responses, and present a controlled sequence of stimuli to the machine.

4. The instructor should co-operate with the machine by introducing cues that reduce the response selection uncertainty. This will reduce the number of mistakes and avoid unduly long anticipation latencies.

5. There is, however, a requisite variety (in Ashby's (1964c) sense) built into the response search process. Hence, the response selection uncertainty must not be reduced beyond a point that would inhibit the response search process.

6. Again, cueing should be minimised on the grounds that if the anticipation latency is cut down (by excessively rigid cueing) the machine will be unable to get rid of mistaken tendencies (or θ_i entries) that interfere with the performance of the skill. The machine does, quite genuinely, learn best from 'near misses' (this is the argument used by Winston (1970) in a slightly different context). The required degree of cueing depends upon the form of the decrementation process. In general, mistakes can only be rectified if the machine is allowed to exhibit its mistaken tendencies and it is also generally

true that the way a skill is taught and learned should be adapted to the existing internal organisation of the machine.

7. Clauses (4), (5), and (6), can be satisfied by a delayed cueing procedure. Cues are delayed to a variable extent after the delivery of the stimulus. The delay in each cue, which may be so great that the cue never appears, is modulated as a function of the machine's activity to minimise the amount of cooperation needed to maintain learning.

One simple but effective instructional algorithm starts by delivering all the cue information with every stimulus. As the learning machine responds correctly, the cue information is delayed until a mistake is made. The cueing required to correct this mistake in respect of a particular stimulus is presented earlier after *that* stimulus until the mistake has been eliminated (with reference to *that* stimulus). This operation is repeated for all of the stimuli. Ultimately, when the machine responds correctly and with the required minimum latency to all of the stimuli, cooperation is withdrawn entirely since all the cues are so much delayed that none of them appear. In summary, this is a cueing algorithm that selectively increases the difficulty of the task (by selectively withholding cooperation) as the machine's proficiency increases. Its execution necessarily involves yet a further feedback loop.

8. The delay of a cue is a 'delay relative to the mean latency for a correct response' (that is, relative to the expected value of the sum of the expectation latency and the anticipation latency for a correct response) rather than an 'absolute delay'. A final feedback loop is needed to estimate the mean latency.

5. The EUCRATES Teaching Machine

It would be quite possible for a nimble-fingered, real-life instructor to use these principles to teach the learning machine a relational skill. But four separate feedback loops are involved, each of them requiring a certain amount of calculation. In practice, therefore, the instructors' task would be rather difficult. Further, his difficulties are unnecessary, for the entire teaching process can be automated using a machine that engages in partially cooperative interaction with the learning machine (or, alternatively, with a real-life student).

This automated system (outlined in Fig. 43) was realised as the EUCRATES teacher. The block outline of the teaching component is refined in Fig. 44, from which it will be clear that the EUCRATES teacher is an image of the EUCRATES learner constrained by the requirement that $\mathcal{R}$ is satisfied and by the rules listed below.

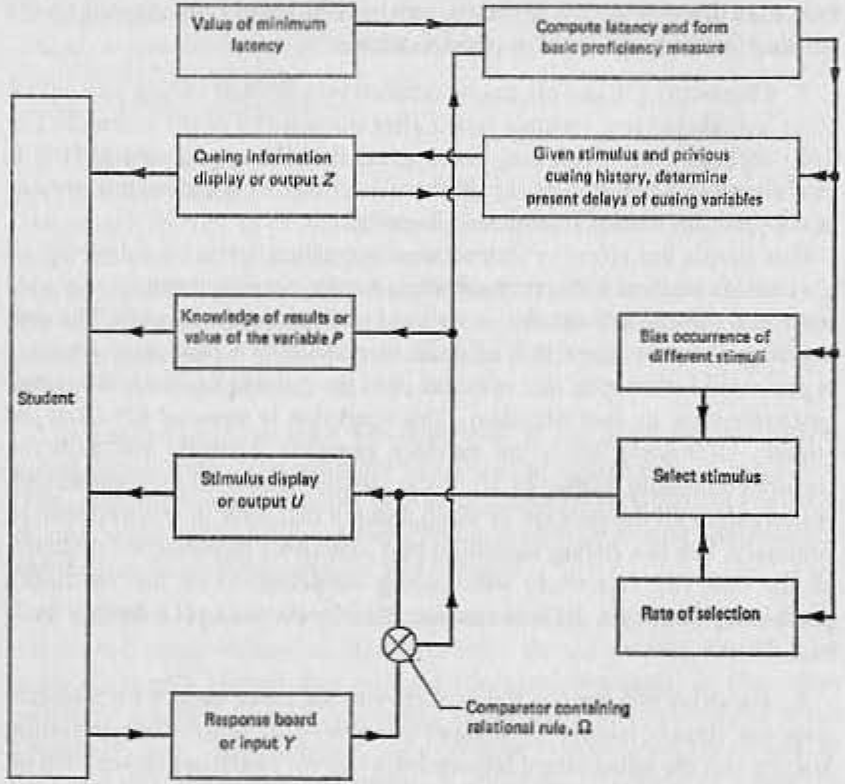

Figure 43 Plan of a teaching machine.

The student's or machine's response is first compared with the $\mathscr{R}$ transform of the stimulus delivered, to determine its rectitude and its latency. Both terms enter into the four feedback loops (Fig. 43) and the data are collated to give a basic measure of performance.

(a) The teaching machine specified in Fig. 43 forms an estimate of the sum of the expectation latency and the anticipation latency from an estimate of the mean correct response latency.

(b) Biasses a random selection of possible stimuli so that stimuli yielding correct responses are less often selected.

(c) Provides a knowledge of results signal which, in the case of the machine (rather than a human student), is returned as the value of variable R.

(d) Determines the degree of simplification of the problems posed under the selected stimuli by assigning a variable delay to the cueing information

(in the case of the machine by assigning a value $z_j = -1$ to each of the cueing variables after a variable delay interval).

The last process requires further comment.

(i) The delay introduced is relative to the estimate of latency as calculated in (a) above.

(ii) The delay associated with variable z_j must be computed separately for each stimulus, i. Hence, the basic measures from the comparator must be averaged with reference to an array of 8 × 4 contingencies; eight cueing variables combined with four stimuli.

The requirements of (i) and (ii) are instrumented (Fig. 44) by an inverse of the learning machine. The biassed random selector is replaced by a stimulus search process having a basic repetition rate determined by the mean expected latency. Values of 8 × 4 averages of the performance measure are compiled in the 8 × 4 analogue storage array of Fig. 44 (as indices θ_{ij}' analogous to the θ_{ij} in the learning machine), and the rows of this storage array (selected by stimuli) provide a slope determining input to a set of four sawtooth generating circuits. These circuits, which are analogous to the response circuits in the learning machine, contain voltage limiters. When the jth sawtooth voltage exceeds a predetermined value, the cueing variable z_j is set equal to -1. The sawtooth voltage is returned to 0 when the next stimulus is displayed.

6 Learner/Teacher Distinctions

Adopting the behaviouristic and causal stance, the learner/teacher system in Fig. 44 may be partitioned into two finite-state machines, namely a 'learner' and a 'teacher'. This picture is plausible if the actual mechanism of EUCRATES is constrained by the various tricks already described: for synchronising the separate machines, for ensuring that a stimulus *begins* a selective process which gives rise to a response that *ends* the process, and for delineating stimuli that are input states and responses that are output states of the machines in question. Although the tricks are not easy to play they can often, though not always, be played successfully and, under these circumstances, it is possible to construct contingency tables representing the frequency of events, joint events, etc.; for example, stimulus/response/cueing tables. If the frequencies in question are interpreted as probability estimates, it is also easy to plot statistical learning curves, for instance, in terms of correct response probability.

By adjusting the parameters of the learning machine and restricting the teaching machine to act as a simple environment (a degenerate finite-state machine) it is possible to replicate the results of statistical learning theory. By removing the restriction upon the teaching machine it is generally possible

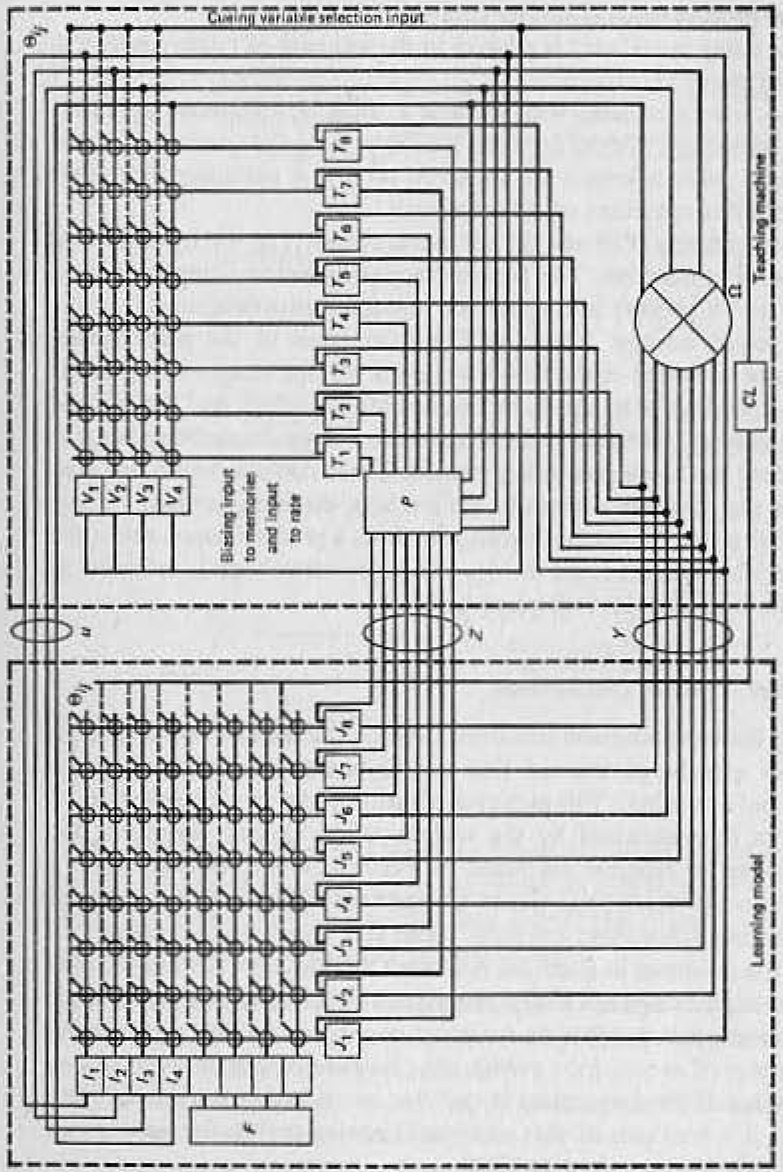

Figure 44 Learning and teaching system: *E*, inputs of $u = 0$ in learning model; *V*, stimulus selection generating units; *T*, sawtooth cue generating units; *P*, rule based on Ω determining which response should not be inhibited by the sawtooth cueing output, given a particular stimulus; *CL*, compute latency and form the basic proficiency measure.

to show that an adaptively modulated teaching process leads to more rapid learning than any of the commonly advocated stimulus, cueing and reinforcement schedules. Reams of these curves exist, but are of chiefly parochial interest in solving specific design problems. The general result (that the learning system converges in various ways and at various rates to an asymptotic condition) is better shown by the mathematical formulae of Bush, Estes, Mosteller and others in the field of statistical learning theory (*see* Luce, Bush, and Galanter, 1963) or by work on the statistical theory of optimal teaching operations due, for example, to Matheson (1964) and Smallwood (1962). It is easy to convert the response probabilities, etc., into indices of selective information and consequently to represent the data in information theoretic terms. There are some advantages in doing so, as pointed out by Wattanabe (1963).

7 Von Foerster's *Reductio ad Vacuum*

The formulae of statistical learning theory are unquestionably elegant but unavoidably correct. That is, any pair of finite-state machines coupled to one another (Fig. 45) is bound to form a large finite-state machine with a deterministically or probabilistically convergent behaviour. If the organism

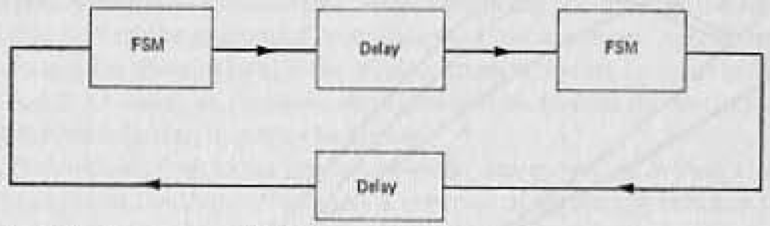

Figure 45 A pair of coupled finite-state machines.

or the learning machine is restricted to act in this fashion the result is inescapable; as pointed out by Von Forester (1971) in his discussion of finite-function machines. Of course, it is possible to obtain a large number of statistical behaviours within the compass of this framework, depending upon the structure and initial states of the finite-state machines. Some, mostly culled from Von Foerster (1971), are shown in Fig. 46 where the response frequencies are transformed and summarised as selective information measures. However, it is unnecessary to call any of these processes 'learning' and, in common with Von Foerster, I am inclined not to do so. These results show 'adaptation', no more and no less. Learning in a nontrivial sense, becomes possible in so far as the finite-state machines are augmented by an exploratory or attention directing mechanism able to modify their internal state sets, i.e. if, in Von Foerster's sense, at least one of them is a finite-function machine.

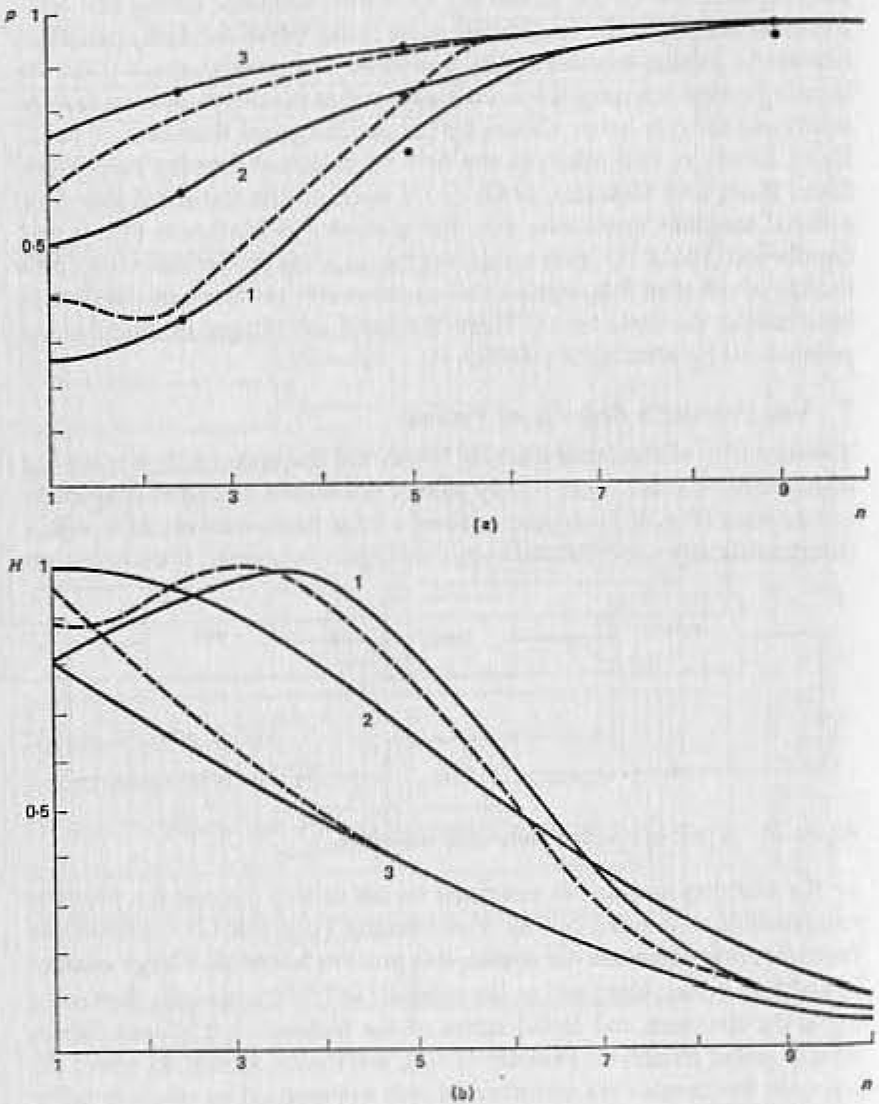

Figure 46 Typical correct response probability curves (*a*), and their transformed representation as indices of selective uncertainty/information with respect to an external observer (*b*). Three cases are from Von Foerster's simulation and two are from EUCRATES.

8 Attention Directing and Constructive Mechanisms

It is possible to obtain a (rather cackhanded) model for non-trivial learning by relinquishing the constraints imposed upon the system of Fig. 44 in order to secure two machines featuring as 'teacher' and 'learner'. The essential condition is that the learner's output can form a non-numericised fuzzy input to the teacher and, vice versa, that the teacher's output can form a non-numericised fuzzy input to the learner.

It will be evident that the resulting system is distributed; there is no functional distinction between the teacher and the learner. Or, to make a point also voiced by Von Foerster (and which is generally valid) *learning* entails a *teacher who learns*, just as teaching entails a learner who learns. Any theory of learning in the preferred sense is a theory of teaching, and vice versa.

Although the learner/teacher distinction is satisfactorily eliminated there remains an engine responsible for evaluating responses with respect to the relation $\mathscr{R}$ and determining whether or not the machine is reinforced. Crudely, but seriously enough, this engine is an arbiter of veridicial truth; if a response is taken to index a problem-contingent solution then only true solution-statements are rewarded. Since there is but one relation, the action of this part of the equipment trivialises the joint machines' operation by thwarting the attempt (which the joint machine is bound, by construction, to make) to direct its attention elsewhere and to explore its environment for further relations that may be learned.

The method of surmounting this obstacle, recommended in Pask (1960) is to augment the truth criterion by a criterion of agreement between sub-systems of activity (initially, the subsystems localised in the teacher and learner parts of the apparatus) and to enlarge the apparatus so that physical limitations, such as the number of storage locations, no longer play a dominant role. Under these circumstances, replicas of the teacher and learner *organisations* can coexist and cooperate in the enlarged network as dynamic but stable entities (the stable configurations in networks described by Aleksander (1973) are examples of such entities). It is still true that any aggregate of two or more cooperating subsystems must *satisfy* $\mathscr{R}$ and that they must also *agree* (in the sense that the joint system is stable in the domain of $\mathscr{R}$). But the several distinct pairs (or aggregates) of subsystems coexisting in the network correspond to different descriptions of $\mathscr{R}$; in particular, each one realises a distinct method of computing $\mathscr{R}$, i.e. they are distinct *intentions* of the same extensionally specified relations or, in psychological parlance, they correspond to different methods of solving the same problem or performing the same skill.

This proposal remains valid and it has been embodied in a physical artifact which operates in this manner. But the proposal is not unique and it is worth considering at least one alternative which can be usefully contrasted with the first system.

Instead of a single relation $\mathscr{R}$ let there be a plexus of related relations; a relational mesh of which $\mathscr{R}$ is a part. Further, let the constituent relations be named; for example, by an index i so that $\mathscr{R}$ becomes (for some i value) $\mathscr{R}_i$. Finally, let there be a description of this relational mesh so that the $\mathscr{R}_i$ can be represented in the context or neighbourhood of other constituent relations.

For the moment, the question of how a mesh of relations should be constructed, of what constitutes a learnable field of relations and of what constitutes a satisfactory description are all left open to be addressed and answered at a later stage. Regardless of the particular answer, it is certainly true that EUCRATES could direct its attention from $\mathscr{R}_i$ (having learned $\mathscr{R}_i$) to a differently named relation in some other part of the mesh.

It could do so haphazardly (a dubious argument because something, perhaps a random selection, tells it what to attend to next). More cogently, there might be an organisation in EUCRATES with the ability to comprehend not the *relations* but *a description of the relations*. This organisation is conveniently regarded as residing at a higher level in an hierarchy of control and acts as a specific orienting or attention-directing unit. By symmetry, it would also be responsible for constructing or monitoring the construction of those processes (alias cooperative aggregates or putative learner/teacher *organisations*) that solve problems or satisfy relations; that is, the attention directing unit is able to perform the other functions (demanded from time to time in the previous discussion) of an 'hierarchical' system design.

Where does the hierarchy come from? It might be built into the structure. On the other hand, it need not be. The processes that direct attention and/or construct problem solving processes are not necessarily distinct in kind from the problem solving processes themselves; they merely act upon a different domain. If that *is* so, then the required mechanism develops already, in the first system, even though the first system is not explicitly hierarchical. In fact, the hierarchy is a fiction, introduced for the convenience of an external observer, which may or may not be reified as an organisation or structure.

The difference between the first proposal and the second is simply that between internalised replication of methods for computing the same relation (first proposal) and an attention-directing exploration, in which fresh relations are discovered and learned (second proposal). These proposals are not in the least contradictory and it can be argued that both of the activities they presage are manifest in a competent learner, either man or machine.

In either case, the basic organisation consists in cooperative elements to be identified with teacher and learner (though unlike the original teacher and learner, they are seldom localised in one part of the physical equipment). Viewed from the point of view of cognition rather than learning, they appear as the 'proposer' and 'critic' subroutines that Minsky and others believe to be important constituents of any competent 'artificially intelligent' program with a goal of the kind under discussion. According to the present point of view, the distinction between teacher and learner, in common with the distinction between hierarchical levels, is chiefly a matter of observational convenience (perhaps, of necessity in so far as sharp valued measurement depends upon marking such a distinction). But it is not generally a matter of *fact*. These distinctions become matters of fact only if the machine designer, or nature in that capacity, happens to build up structures in which the teacher and learner or the hierarchical levels are specially localised. When such constraints are introduced, as they are in most of the models discussed in the sequel, they are acknowledged to be arbitrary and, however convenient they may be, they are quite inessential.