Göteborg Psychological Reports

University of Göteborg Sweden

Helge Malmgren & Olof Östensson

Selective Attention is Selective Learning

Number 3 Volume 19 1989

Selective Attention is Selective Learning

Helge Malmgren and Olof Östensson

University of Göteborg

Malmgren, H. & Östensson, O. Selective attention is selective learning. *Göteborg Psychological Reports,* 1989, 19, No. 3. - After a discussion of some theoretical problems which arise in connection with learning algorithms based on the Hebb rule and neurophysiological theories concerning the function of known Hebb-type synapses, the first author's previous work on classical and operant learning in randomly constructed deterministic automata is reviewed. In the model of classical conditioning in collections of such automata, learning progresses because of a process of auto-selection: with time, automata which happen to be in states which are stable under the relevant input pattern come to dominate the output pattern of the collection. Some new results concerning classical conditioning in this model are presented. An integrated theory combining elements from the earlier theories of classical and operant conditioning is sketched. The theory of operant conditioning, according to which negative reinforcement was conceptualized as a temporary rise in input variability, is developed further using neuron-like elements. In the new model, reinforcement takes the form of a generalized lowering of neuronal thresholds, tentatively mediated by the arousal system. In a hierarchically arranged nervous system with pseudo-random connections guided by a vicinity principle and reciprocity, some trace activity will usually persist in the vicinity of the controlling neurons for some time after the release of an S-R pattern. Because of this trace activity, and because the neurons are non-linear, the neurons which affect the S-R connection in question will at that particular time be more sensitive to input than are neurons which control other responses. This differential sensitivity is tentatively identified with selective attention. Because of it, the general process of negative reinforcement will affect the controlling neurons more than it affects neurons which are less relevant to the control of the S-R connection in question. This proposed mechanism may provide a way - without introducing ad hoc "teacher" mechanisms - of accounting for at least part of the selectivity shown by actual learning processes, i.e. of partly solving the so-called "problem of credit assignment" for operant learning. A computer model designed to test our hypothesis is presented.

Keywords: Arousal, automata, learning algorithms, neural networks, random composition, reinforcement, selectional processes, selective attention.

In three previous publications (Malmgren 1980, 1984, 1985), it was shown that randomly constructed deterministic finite automata, or collections of such automata, tend to show phenomena similar to adaptation, habituation, classical (Pavlovian) conditioning and operant (Skinnerian) conditioning. The demonstration of these phenomena only presupposes certain simple and physiologically plausible ways of organizing the input-output relations of the automata, and - in the case of operant conditioning - a certain kind of information barrier within the system.

One main point in the previous work is that elementary forms of learning in random systems can be modeled without the use of certain very strong (but admittedly natural) assumptions concerning basic mechanisms. Thus, no "pre-arranged" ways of changing the transition functions of the automata, or the connections between them, are needed in order to make the systems conform to some simple learning rules. In particular, there is no need for any analogue of the Hebb rule (see below).

This work was supported by a grant from the Swedish Council for Research in the Humanities and Social Sciences (no. 451/87). We want to thank Alvar Ellegård, Anders Lundberg and Holger Wigström for constructive criticism.

Instead of using such specific mechanisms, the models are strongly "selectional". In the case of habituation and classical conditioning, learning progresses because with time, those sub-automata which happen to be in states which are stable under the relevant input pattern come to dominate the output pattern of the whole system. In the theory of operant conditioning, the corresponding selection (reinforcement) of stable states is accomplished by means of a temporary "arousal", and hence intrusion of environmental noise, in case of "incorrect" behaviour on the part of the system as a whole.

In this paper, after a discussion of some general problems in contemporary abstract learning theory and in neurophysiological research on memory mechanisms, the earlier results on habituation and classical conditioning in randomly composed automata are reviewed, and some additions to these results are presented. The theory of operant conditioning in random automata is reviewed. It is then integrated with the theory of classical conditioning and developed further in terms of neuron-like elements. In the last-mentioned model, selective attention - conceptualized as a selective rise in reactivity - is the factor which solves the so-called problem of "credit assignment" for operant learning.

Some notes on associative mechanisms in recent learning models

An algorithm for learning in a model system is a rule for changing the state of a system, given a certain performance under a certain input, so that the performance of the system eventually comes to match a certain standard. Given that the system can in principle perform in the desired way, it is often possible to devise algorithms which solve this problem in a fast and efficient way - if, that is, one does not pay attention to the constraint that the model has to have some chance to be implementable in real nervous systems. One of the main goals for a theory of memory mechanisms in biological systems must certainly be that it minimizes the number and complexity of the specific information exchange and storage processes that are postulated in order to explain the change in performance; else there is a real risk that the nervous system lacks the necessary resources to realize the theory.

It is often thought that Hebbian algorithms (Hebb 1949), which assume gradual increments in synaptic strengths in proportion to pre- and postsynaptic activity, place rather small such requirements on the nervous system while at the same time being sufficiently efficient to be able to explain the behavioural phenomena of associative learning. We wish to point to some problems connected with this view, in order to give a rationale for our own research strategy.

First, however, we want to emphasize that models based on Hebbian mechanisms certainly have several advantages. They provide a straightforward and economical explanation of how a specific association between a stimulus S and a response R can be "stored", namely, in the synaptic connection(s) between the neurons encoding S and R, respectively. Also, empirical evidence indicates that certain phylogenetically older parts of mammalian nervous systems have indeed developed specific Hebbian mechanisms (cf. below). However, we do not think that such mechanisms form the sole basis of associative learning. If simpler (albeit less specific) solutions are at hand, why shouldn't the nervous system use such solutions too?

We start with the problem of the so-called "teaching input", using an example from the recent Parallel Distributed Processing ("PDP") tradition. In the models constructed within this tradition, some version of Hebb's rule is usually exploited. This rule says, in essence, that a synapse from a cell A to another cell B is strengthened in proportion to the simultaneous or near-simultaneous activity in A and B. (The term "activity" can denote different processes; let us assume here - with Hebb - that on a real neuronal level, activity is the occurrence of action potentials.) The connection between Hebb's rule and the problem of associative learning becomes clear if one imagines that A transmits the conditioned stimulus (CS), that B produces the unconditioned response (UCR) and also receives the unconditioned stimulus (UCS), and that UCS alone, but not CS, is able to fire B at the beginning of learning:

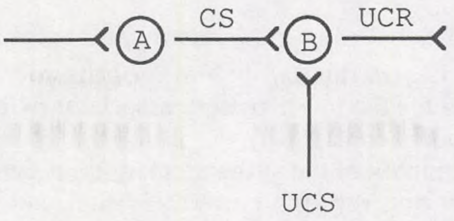

Figure 1. Example of how Hebb's rule might work.

If CS is fired in conjunction with UCS a number of times and some suitable version of Hebb's rule is followed, the strength of the synapse from A to B will with time be enough to make CS fire B.

Let us take a look at how this presumed process is modeled in the linear "pattern associator" using the simple Hebbian rule. In such a device, synaptic weights are changed in proportion to the product of presynaptic activity and a so-called "teaching input". Cf. also Rumelhart, Hinton & McClelland (1986), p. 62; in order to understand what the teaching input is, one should think of it as the desired output of the system. We choose the present example only because it is so simple. It is well-known that linear networks and the simple Hebb rule have certain limitations. However, our argument can be reformulated for non-linear networks and modified Hebb rules too (see below).

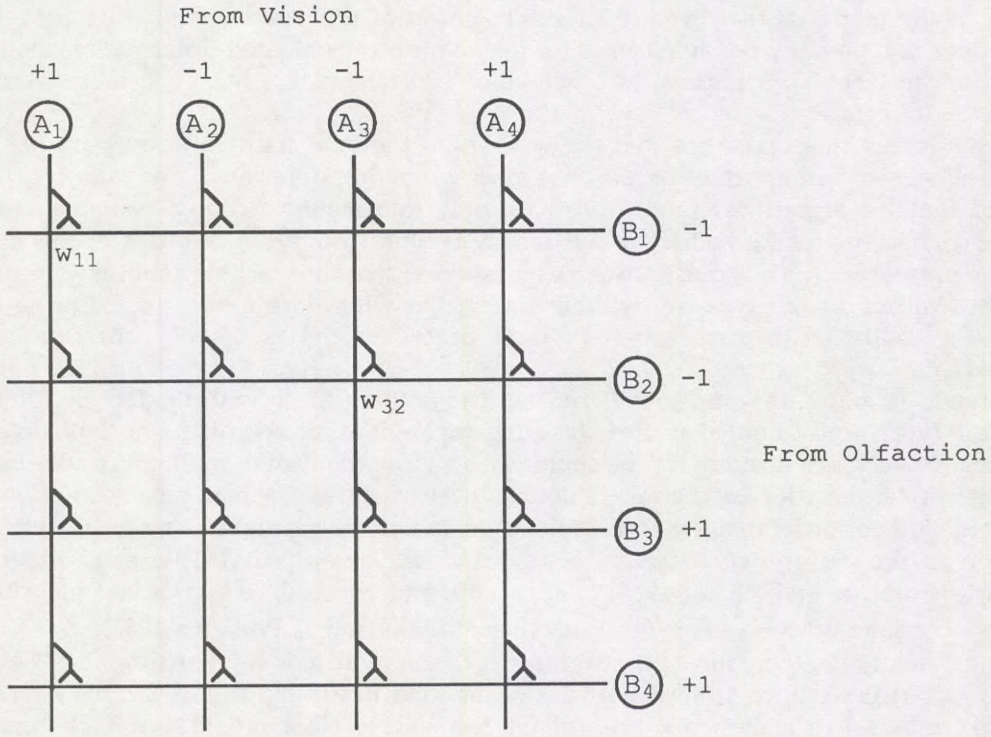

Figure 2. A linear network (pattern associator) using the Hebb rule.

Figure 2 is adapted from McClelland, Rumelhart & Hinton (1986), p. 34. The activity levels of all units can have any positive or negative real values. The A units are supposed to influence the B units via modifiable synapses with weights w_{ij}. The contribution of the activity a_i of an A unit to that of a B unit, b_j, is the product of a_i and w_{ij}. The activity of a B unit is the sum of these contributions from the A units. At each step of time, the weights of the synapses change with an amount proportional to pre- and postsynaptic activity.

Suppose that the sight of a rose produces the visual pattern shown above in the A units, and that the smell of a beefsteak would produce the olfactory pattern shown in the B units. It is now theoretically possible to make the visual pattern produce the desired olfactory one by making a certain one-step adjustment of the weights, namely by setting w_{ij} proportional to a_i

times b_j with a suitable proportional constant. Evidently, the same result can also be achieved in a more stepwise manner: i.e., starting with zero weights and then following the above-mentioned Hebbian algorithm for the linear pattern associator, with teaching input = desired output, and a smaller proportional constant.

Now it seems to be the opinion of the authors - cf. ibid., p. 36 - that this algorithm can be realized by repeatedly presenting the two patterns (visual and olfactory) together to the network of Fig. 2, and letting the modifiable synapses do their work. However, this is not the case, which we will now show.

If we are to present the desired output pattern to the B units, these must of course have another actual input source - the olfactory one - beside the A units. The need for such an extra input line is, indeed, acknowledged by denoting the desired output, "the teaching (or teacher) input". But if there is such an input, and the general rules of the model are to be followed, this input will interact linearly with the input from the A units. Therefore the activity of the B units will be the desired output plus the contributions from the A units. Hence, the intended application of the Hebbian algorithm, in which it was supposed that an amount proportional to the product of input and desired output was added to w_{ij}, will not be realized by the actual synaptic changes.

To be sure, in the present kind of system the desired final behaviour can be obtained by a proper choice of the proportional constant even if the activity of a B unit is supposed to be the sum of actual output and desired input. But that is quite another matter. We only wish to point out that the system cannot realize the very algorithm which it is intended to embody. The crux of the matter is that even if a teaching input of the desired kind is present, the network does not offer any possible means for making "uncontaminated" information about it available to the Hebbian synapses. In other kinds of systems this has even more serious consequences (cf. below).

Certainly, one may presume that the workers in the PDP tradition are aware of this problem. However, its importance seems not to be properly appreciated. For example, it is often said that the algorithms can work using only information "locally available at the [synaptic] connection" (cf. Rumelhart, McClelland & Hinton 1986, p. 37). In order for this to be true in the present example it must however be assumed that the teaching input is somehow physically realized "at the synapse" without interacting with other inputs according to the rules of the model. This necessary extension of the model is usually not explicitly acknowledged.

It should be noted that the above criticism has nothing to do with the fact that linear systems are intrinsically limited in their learning capabilities, or with the fact that there is no learning convergence theorem for the simple Hebbian mechanism in multi-layer non-linear systems. A similar objection can be raised against those modified Hebbian schemes for which such learning theorems exist, e.g. the "delta rule" where the weights change in inverse proportion to the difference between desired and actual output (in the symbolism of Rumelhart, Hinton & McClelland 1986, p. 62: $t_i - o_i$), and which does work as an algorithm for multi-layer semi-linear systems (cf. Rumelhart, McClelland & Williams 1986).

For the delta rule, the point of the argument is even more evident. For simplicity's sake, suppose again that we have a linear network of the kind illustrated in Fig. 2. Also suppose that the synapses in the network are still of the simple Hebbian kind and change in proportion to pre- and postsynaptic activity. If the reader feels that this contradicts the delta rule, note that we now "only" have to suppose that the postsynaptic activity in the B neurons can be set equal to $t_i - o_i$ (and cf. below). Of course, "o_i" cannot then designate the "new" postsynaptic activity but must refer to the output which was (or would be) produced by the A neurons alone in the absence of the teaching input. This in turn means that the subtractive operation which feeds the relevant differences into the system must have access to some independent representation of this hypothetical output.

Now our main objection is not the rather common one, that the teaching input may be difficult to realize as such. Therefore, let us suppose that in the present case, this can be done by means of the following extension of the model. There are connections from the A units to a parallel set of neurons, B'_i, which do not receive the teaching input. Therefore their actual b'_i = the sum of the contributions from the A neurons. If the differences $t_i - b'_i$ are fed into the B neurons as inputs, the actual b_i will be $= t_i - b'_i + b'_i = t_i$ (since the A neurons will still contribute to b_i), and the system as a whole will follow the simple Hebb rule. It seems that if, somehow, one instead feeds the amount t_i - 2 b'_i into the B units, the mechanism will work in the intended way.

By now, the structure required is evidently rather complicated. But that is not the end of the matter. In order for the B' neurons to be able to "upgrade" their information about the progress of learning, so that they do not simply keep feeding the same magnitude into the B neurons, they too have to be plastic; i.e., the A-B' synapses must also be modified - in the same way as the A-B synapses. But what possible mechanism will manage to do this? Yet another parallel set of neurons?

Maybe the reader now objects that we have misconstrued the delta rule: all the time we have been supposing that on the synaptic level, the simple Hebb rule is still followed! But of course - so the objection goes - the delta rule says not that the synapses change in proportion to the simple product of pre- and postsynaptic activity, but instead that they change in proportion to the product of (i) presynaptic activity and (ii) the difference between t_i and the postsynaptic activity! Hence, it is not at all a question of making the postsynaptic activity itself equal to the difference in question. - However, with a little afterthought one realizes that this alternative would require both that the value of t_i is somehow independently presented directly to the synapses, and that the synapses are sensitive to three simultaneous activity levels. This certainly does not remind us of any known neuronal mechanism. We conclude that it is completely unclear how the delta rule is to be implemented.

Of course we do not want to argue that no integrated and plausible Hebbian models can be constructed in which the neural mechanism of the teacher is specified, and which do work in the intended way. (See for example below, p. 6.) Our objection is instead that if one wants to argue that one single basic mechanism can solve the problem of associative learning, one must first take care to ascertain whether the chosen process really can do the work required without relying on implicit "supporting mechanisms" of a very different character.

Neither is our point that there is no knowledge of real "associative" neural processes which could serve as a basis for memory on an elementary level. Let us have a look at some known processes and discuss them in relation to the argument presented above.

Associative vs. non-associative basic mechanisms

Some of the most important findings concerning associative and non-associative neural plasticity (with a long enough time-scale to possibly serve as a basis for medium or long-time memory storage) are the following (cf. also Kazmarek & Levitan 1987):

1) Homosynaptic depression: in some neurons, high activity is followed by a period of lowered transmittor output (thought to underlie habituation in certain invertebrate nervous systems).

2) Post-tetanic potentiation (PTP): certain neurons have the property that if they are fired at high frequency for a while, they will for a short period be more effective in activating the post-synaptic neuron (because of an increase in transmittor release). This mechanism is not dependent on post-synaptic activity.

3) Diffuse (non-specific) heterosynaptic facilitation: the activity in certain neurons causes other neurons to increase their responses to different inputs for a long period (known to occur in certain neurons of molluscs, and proposed as an explanation of behavioural sensitization and certain kinds of classical conditioning in these organisms; cf. Abrams & Kandel 1988); and

4) Long-term potentiation (LTP): in some neurons, the efficiency of a synapse can be raised by simultaneous or near-simultaneous pre- and postsynaptic activity, and the change may last for a long time (days or weeks). It has been demonstrated that post-synaptic activity which is not released by the presynaptic stimulus in question is also effective in controlling this plasticity. I.e., if a certain weak input (which does not by itself activate the post-synaptic neuron much) immediately precedes a strong input via other synapses to the same neuron, then the neuron will for a long period be markedly sensitized to the first input, but not to others not active at the time of "learning" (cf. Wigström & Gustafsson 1988).

5) Long-term depression (LTD): this is the "inverse" of LTP, i.e. pre- and postsynaptic activity causes a longstanding depression of the efficiency of a synapse (thought to be important in cerebellar motor control).

Especially mechanism 4 (LTP) is prima facie well suited to explain associative memory. The augmentation of the activation (now in the sense of membrane depolarization) of the postsynaptic neuron which is produced by "associative training" of the just-mentioned kind is rather small (at most of the same order of magnitude as the baseline). Hence the depolarization produced by the neuron giving the weak input (tentatively identified with the CS) is always only a small fraction of that produced by the strong input (the UCS, or teaching input). Since this means that at the level of the basic membrane process, the CS never produces a response of the same order of magnitude as the UCS does, there need not arise any disturbing interaction between the two inputs (of the kind discussed in the previous section). Hence the simple Hebbian algorithm with teaching input = desired output may be at least approximatively implemented. Given a suitably low treshold for the release of action potentials it seems that with time, the presynaptic neuron transmitting the CS may come to produce the same action potential response by itself, as does the UCS.

Also, the fact that neurons showing LTP are found in the hippocampus is a strong argument for the hypothesis that LTP is involved in associative learning (since it is well-known from clinical studies that the hippocampal region is important for memory; cf. Lindqvist & Malmgren 1986, Ch. III:7).

We want to discuss another issue in relation to this scheme (and to Hebbian schemes in general). In order for the specific "associative" mechanism just outlined to be the explanation of associative learning on the part of the whole organism, it ought to be a "dominant" mechanism, so that its effects is not cancelled by other, non-associative plastic mechanisms. Here, the fact that the LTP effects are small is of course no longer an advantage. Suppose, e. g., that we try to explain the conditioning of the response UCR to low-frequency stimulation with a stimulus CS by showing that CS and UCS are processed in a neuron N which shows LTP:

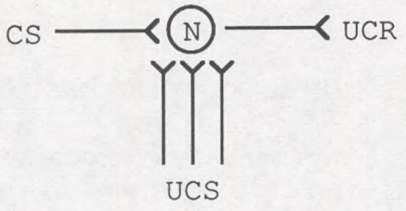

Figure 3. Conditioning in a neuron which shows long-term potentiation.

Suppose there is also strong homosynaptic post-tetanic potentiation of neuron N by stimulus CS. Then high-frequency stimulation with CS, without any coupling to UCS, will have the result that for a certain period of time, low-frequency CS:s can by themselves produce UCR:s.

Also, ordinary temporal summation will probably make N respond with UCR during PTP training! No effects analogous to this are of course seen in behavioural experiments.

In order for such confounding phenomena not to occur, the LTP kind of interaction between the different synapses of the neuron N must be so strong that these non-specific effects are overshadowed, and/or N must be supposed to always work under conditions that make the latter effects negligible. In the example, N might be coupled to a narrow frequency filter in order to hold temporal summation and PTP constant; several other arrangements would of course also be able to do the job required.

The hypothesis that the specific associative mechanisms on the neural level are sufficiently dominant to be able to manifest themselves as reliable, specific associative behaviour on the part of the whole organism has not yet received much experimental support - mainly, we think, because the problem is not often recognized. However that may be, our main point here remains, namely, that a number of non-trivial extra assumptions are needed also in the neurophysiological versions of Hebb's model for associative learning.

The main reason why we have entered into these detailed considerations about the use of neural associative mechanisms in explanations of associative learning is that they prepare the ground for the following turn of mind. Once it is admitted that a great many such extra assumptions about the nervous system must enter such an explanation, one should ask oneself if it is really an decisive advantage to start with the Hebbian synapse at all. If the similarity between a certain neural process and a certain piece of molar behaviour is not sufficient for the one to be the explanation of the other, why should it be necessary?

Let us suppose instead that the nervous system uses a great many different mechanisms as bases for memory - even for classical, associative memory - and that not all basic processes are similar to the molar ones. Any way of producing a long-term change from a short-term process could, theoretically, be the basis for an association. One has only (!) to supplement the basic mechanisms with an appropriate neural network; the burden of explanation will then be carried by "structural" assumptions. A very simple example: the strengthening of the behavioural connection between CS and UCR might be the result of the depression of a neuron N innervating a muscle which antagonizes UCR. (In a similar vein, LTD has been proposed as the explanation of some cases of associative learning.)

Now it must of course be asked whether it is possible to find plausible structural principles which could do the explanatory job. In our opinion, this problem can to a large part be solved by introducing selectional processes. The depression of N in the example above need not be the result of the working of any one single basic synaptic mechanism in any one specific kind of pre-wired "associative network"; it could instead be the product of a process which more or less randomly adjusts some parameters of the neurons in a large pool, until the desired response is produced by the collective output of these neurons. Of course it remains to describe a plausible implementation of such a process; but this goal is exactly what motivates our own research strategy.

This strategy is as follows: try to show that by means of certain plausible organizational principles and selectional mechanisms, the nervous system could in principle make use of almost any possible basic mechanism for accomplishing the over-all purposes of learning. The burden of explanation in a theory of the desired kind would rest not on the postulated basic mechanisms (which could vary within wide limits), but on the invoked macro-structural and macro-functional principles. Trying to find such a theory is of course not at all to devalue research on basic physiological mechanisms. But we really believe that for physiological learning theory to advance beyond the tentative stage, research efforts must be directed also to global organizational processes. In other words, what is said here does not amount to a recommendation to move away from the "implemental" level of description (Marr 1982), but to a shift of emphasis concerning which kinds of data on this level (i.e., the physiological one) are seen as relevant.

The most difficult challenge which a selectional strategy faces is, probably, the problem of "structural credit assignment". This phrase refers to the necessity of changing only, or at least mainly, the relevant parameters (cf. Barto 1987). If a learning algorithm tends to change not only those parameters which are relevant to the learning task, and/or tends to change them in random directions, learning will be very slow and interference between different tasks will make discriminatory learning almost impossible. In the last section of this essay, we will present an outline of a selectional theory which we think will be able to assign credits properly - to some extent at least. However, to be able to present that theory it is first necessary that we review two simpler selectional models. (For a selectional theory with some affinities to ours, cf. also Fogel et al. 1966).

The habituation and classical-conditioning models

First, a brief sketch of some results in Malmgren (1980) and (1984). We will here only give a qualitative treatment of the basic principles.

A finite deterministic automaton is an abstract machine with a finite number of states, moving in its state-space in discrete time from a starting point on. The state A of the automaton at any time t (after the start) is determined by its immediately preceding state and input I; symbolically:

$$A(t) = f(A(t-1), I(t-1)).$$

Its output O is taken to be a function of state, so that

$$O(t) = g(A(t)).$$

In the habituation and conditioning models that we are concerned with, g is the identity function.

The transition function for an n x m automaton (i. e. with n states and m inputs) can be represented by an n x m matrix F, where F_{ai} is f(a,i), i. e. the state to which the automaton goes from state a with input i. A uniformly randomly composed automaton is generated by filling in such an n x m matrix with uniformly random numbers from 1 to n. There are of course several other kinds of randomly composed automata (see below); the present one is chosen mainly for the sake of simplicity. If not otherwise stated, "random" in the following means uniformly random.

Note that each randomly constructed automaton is a deterministic device, defined by the transition matrix it happened to receive during its construction. Thus, there is no probability or uncertainty involved in its internal working. When in the following we talk about the "probability" of a certain behaviour, we are referring to the probability that a randomly chosen automaton, which has been constructed according to the stochastic rule, will show the behaviour in question.

The time course of the probabilities (in this sense) of any behaviour, such as that behaviour of an automaton which consists in its being (at time t) in its initial state, can now be regarded as a stochastic process. To repeat: this is not because the automaton, once constructed, is stochastic (it is not). But the rule for the construction of the automata determines that they will exhibit a certain statistical mixture of deterministic processes, and it turns out to be helpful to regard certain aspects of this mixture as stochastic processes.

The latter processes can be shown to have some rather surprising properties - especially, they do not in general have Markov properties. For example, the process defined for different t's by the respective probabilities of being in the initial state is not a Markov process with time-independent transition probabilities (cf. also below). This means that, in a certain sense, the process has a memory. For some general considerations see Malmgren (1980, 1984). Here we will only illustrate the principles by means of certain behaviours which are especially relevant to our purpose.

To make the reasoning more concrete, imagine that by some uniformly random procedure, we generate a series of automata, until we have a large collection C of different automata. Then let all of them receive the same input (i. e., the input to the system is divergent to all subsystems). Of course the real collection C is as deterministic as its subsystems; there is no stochastic element in the internal working of such a collection, once constructed. However, in a sufficiently large collection the proportion of sub-systems showing a certain behaviour can of course be expected to approximate the corresponding probability of the behaviour in a randomly chosen automaton.

The latter fact, in turn, indicates that if the development over time of a behaviour of the system C is described in terms of the proportions of automata showing the behaviour in question at different times, this development will approximate the corresponding stochastic process. The basic reason why this is of interest for the study of the behaviour of organisms is that the organization of the output of real nervous systems often follows the principle of competitive convergence, which means some kind of "majority" vote; therefore it is often decisive what proportions of neurons show this or that pattern (cf. Arbib 1981).

For simplicity, we will talk of proportions and probabilities as if they were interchangeable. Let us look at the behaviour of a very large, randomly composed collection of automata under different sequences of input signals. Three basic ingredients in this behaviour will be the following:

(i) If a certain sequence of inputs (stimuli) $I_{j_1}...I_{j_k}$ occurs twice within a longer sequence, the subsystems will tend to be in the same state at the end of the second sequence as at the end of the first.

This is so, because if any subsystem happens to be such that it goes to the same state at step j in the second sequence as at step j in the first, it must (since it is deterministic) continue in the same path for the rest of the sequence; and if the system is not in this way "locked" at step j it may well become so at step j+1 - etc. (Cf. Ashby 1956, pp. 138f for an interesting practical application of this point.) For the first occurrence of an input sequence repetition in the life history of the large system (i. e. at a time when the system is "naive"), the exact probability of recurrence of state at step k of the second input sequence - i. e. the proportion of subsystems which at this step go to the state which they visited at step k of the first sequence - can be calculated. If t_1 and t_2 mark the onset of the two identical sequences, the probability (in question) is

$$(1) \quad P(A(t_1+k)=A(t_2+k)) = \frac{1}{n} \sum_{1 \le i \le k} \frac{n!}{(n-i)!n^i} = G(n,k)$$

(ii) If a long sequence of identical inputs is given, all subsystems will sooner or later be confined to periodic behaviours (basins) with period b = 1,2,... or n.

This follows immediately from the deterministic nature of the subsystems. The probability distribution of the length b of the basins for a sequence of identical inputs given to the naive collection can be calculated as

$$(2) \quad P(b=k) = \frac{1}{n} \sum_{k \le i \le n} \frac{n!}{(n-i)!n^i}$$

(iii) If a sequence of identical inputs is given, the subsystems will (for a time corresponding to the number of states of the system) more and more tend not to change their state between two successive stimuli.

Of course, (iii) is a special case of (ii). If, under constant input, a system has once gone from a certain state to the same state, it must continue to do so; and among the subsystems which have not yet entered any basin, the probability that they will in next step enter a basin with period 1 is = 1/n. From this, it can be calculated that in the freshly started system the probability of constancy of state after k identical stimuli is (cf. Eq. (1))

$$(3) \quad P(A(k)=A(k-1)) = G(n,k)$$

(For a different context, the same formula was derived by Kruskal 1954; cf. Griffith 1971).

With a more variable input, the corresponding probability is much lower. Especially, if after a sequence of k identical inputs the system receives an input which it has never before encountered, the choice of next state is independent of all previous choices and the probability of constant state at this step is of course only = 1/n.

In order to give an explanation of behavioural adaptation to constant stimuli, principle (iii) was applied (Malmgren 1984) in a straightforward way. It is supposed that there is a mechanism which sums all state changes in each moment (by means, inter alia, of convergent output), and the size of the gross behavioural reaction of the organism is taken to be proportional to this sum. Then it can immediately be seen that in a naive system, the size of this gross reaction will decline with time (t) according to the value of 1-G(n,t).

To explain habituation proper (which involves interpolated stimuli between the recurrent ones) and classical conditioning, some extra reasoning is needed.

(iv) If a sequence of (more than one) identical inputs is given, at the end of the sequence the subsystems will tend to be in the same state as they came from at the beginning of the sequence (the starting state, not to be confused with the initial state A(0)).

At each moment during the constant input sequence, a certain proportion of the subsystems enters into some pattern of stable behaviour. In order to enter such a pattern at exactly step k, an automaton must then (for the first time) go to one of the k previously visited states, the starting state included. Evidently, since the starting state is at each k in the set of visited states, and no other state is so at each k, the most probable "first return address" will be the starting state. For exact probabilities see below. From (iv) immediately follows:

(v) Under constant input the probability that the system will be in the initial state at time t is, for any t, > 1/n.

It can easily be seen, that the probability that the initial state will be contained in the basin B to which a freshly started system finally goes under constant input is (cf. Eqs. (1) and (3)):

$$(4)\quad P(A(0)\in B) = G(n,n)$$

However, since during a long sequence of identical inputs any one of the states in a basin of size b recurs only in one step out of b, the probability that under such an input, a randomly selected subsystem will go to the starting state at step k will depend on whether b is a proper divisor of k:

$$(5)\quad P(A(k)=A(0)) = \frac{1}{n}\sum_{1\leq b\leq n;\, b|k} \frac{n!}{(n-b)!n^b}$$

(where b|k means that b is a proper divisor of k). This means that there will be an uneven distribution of returns to the initial state, depending on the character of k; P will for example be only slightly larger than 1/n for large prime numbers k. For the long run, a mean distribution over different times k can be defined as

$$(6)\quad P(A=A(0)) = \frac{1}{n}\sum_{1\leq k\leq n} \frac{1}{k}\frac{n!}{(n-k)!n^k}$$

This is, then, the probability that given a random time k (>n) after the birth of a system which is given constant input, a randomly selected subsystem will be found to be in the starting state. (For some numerical calculations of this probability see Malmgren 1980).

With the proviso that no stronger, counteracting determinant of a specific state is at work, the same kind of reasoning is of course qualitatively valid also for the case where the constant input sequence in question starts at any other time in the automatons' history than at birth.

(vi) During a regularly alternating sequence of sequences of two inputs, the states to which a subsystem goes at the end of the sequences of the first type will more and more tend to be identical with the states arrived at after sequences of the second kind. This increase will be more stable and, depending on the length of the sequences, greater or smaller than in the case where there is alternation between a larger number of different stimulus sequences.

For simplicity we will in the following use digits to represent inputs. We can then write the relevant alternating sequence as

$$\text{(A)}\quad \underbrace{11\ldots1}_{k_1}\underbrace{22\ldots2}_{k_2}\underbrace{11\ldots1}_{k_1}\underbrace{22\ldots2}_{k_2}\ldots$$

As during constant input, every subsystem will of course eventually enter into a periodic behaviour under this constraint. More exactly, they will reach this stage at the point when, for the first time, they happen to be in the same state as they were at any previous "corresponding" time - e.g., if they happen to go to the same state at two following end-points of 2-sequences. Since the last response on a "1-sequence" is identical to the starting state for next "2-sequence", the reasoning under (iv) and (v) can be applied directly in order to show that the probability of same response to a "last 1" as to a "last 2" must grow during the presentation of sequence (A).

This is not the whole story, which is best seen from the fact that the argument in question is equally valid for the case of the input sequence

$$\text{(B)}\quad \underbrace{11\ldots1}_{k_1}\underbrace{22\ldots2}_{k_2}\underbrace{11\ldots1}_{k_1}\underbrace{33\ldots3}_{k_2}\underbrace{11\ldots1}_{k_1}\underbrace{44\ldots4}_{k_2}\ldots$$

where there is no global recurrent pattern imposed upon the subpatterns. In case (B), the system will start "afresh" on each sequence of new stimuli. Whatever has happened before the beginning of a certain such sequence - e g, a sequence of 4's - the probability of "same state" at the end of this sequence as at the end of the previous sequence of 1's will therefore be determined directly by Eq. 5, with $k = k_2$ = the length of the sequence of 4's.

However, in case (A) there is a global input pattern of a kind which is especially conducive to the formation of regular behaviours of the described kind. If, in either case (A) or case (B), a subsystem for the first time goes to the same state on a "last non-1" as it did on the preceding "last 1", then - again because of Principle (iv), but now applied "backwards" in accordance with Bayes' rule - the starting state of the 1-sequence was probably also the same. If this was the case, if the stimulus constraint is of the kind (A), and if the 2-sequence in question is not the first one presented, the behaviour of the subsystem in question is now "locked" - since the starting state of the 1-sequence is identical to the end-state of the previous 2-sequence.

If the inputs are of the kind (B) this "locking mechanism" is not at hand. It does not matter for the future whether, in the case described, the starting state of the 1-sequence was the same as its end-state, since the previous non-1 sequence was not of the same kind as the present one. Therefore the system has, so to say, a certain freedom to "choose" another response at the end of next non-1 sequence.

Whether constraint (A) or constraint (B) will be superior in producing similarity of state at the end of the different sequences depends on the choice of lengths of these sequences. If, e g, this length is a large prime number, there is an appreciable risk that the behaviour of the system under (A) becomes "locked" in another pattern than that in which the end-states are similar (cf. below, and the details of Principle (v) above). However, even in this case the global behaviour of the system will be more stable over time under constraint (A).

Principle (vi) can also be shown to be true in another way, which better illustrates what is meant by the "selectional" character of the present theories. For a given sub-automaton, sequences of 1's often have more than one possible stable response pattern (basin); which of these possible basins is actually entered during a certain sequences of 1:s of course depends on the subsystem's previous history. Now, under input condition (A) there will be a certain kind of dependency between the final "choices" of stable pattern on 1-sequences (an "1-basin") and the corresponding "choices" under 2-sequences, in that the choice of a 2-basin ending with state a is more probable to be the final choice if the next choice of 1-basin ends with state a, than if it does not. In this way, there is an auto-selection, or auto-reinforcement, of similar stable response patterns on associated stimuli.

(This way of looking at the effect of associated input sequences also generalizes in a self-evident way to other multi-stable systems than the uniformly randomly composed automaton.

For example, it is easily seen that a randomly composed nervous net with semi-linear activation functions and symmetrical connections - cf. Rumelhart et al. 1986 - has the potentiality of similar behaviours.)

The present result holds also for the case where the length of the 1-sequences is = 1. In this case - i.e. with a global stimulus pattern of the form

$$(C)\quad \underbrace{22...2}_{k}1\underbrace{22...2}_{k}1\underbrace{22...2}_{k}1\underbrace{22...}_{k}$$

one can still apply Principle (iv) to the 2-sequences. Therefore, if the 1-stimulus has as its result that a subsystem does not change its state, this increases the probability that the system's behaviour is from now on closed. From this, in turn, it follows that in the whole collection the state changes on consecutive 1:s tend to be fewer and fewer.

This is, in essence, the proposed explanation of habituation proper. In this explanation, the same auxiliary assumptions as in the case of adaptation to a constant stimulus are used (cf. above and Malmgren 1984). Because the effects are small some amplifying mechanism must also be presupposed (cf. below, "Outlines of an integrated model").

However, if the lengths of both kinds of stimuli is set = 1, the argument fails because one cannot use principle (iv) at all. (For inert random systems - see below - the argument still works, however.)

Let us now study the case of classical conditioning. The basic idea is here to exploit the above result concerning habituation, while adding the condition that the responses of the subsystems to one of the inputs are in some way pre-wired. The latter condition is, of course, the counterpart of the assumption of a fixed unconditioned reflex (UCR).

If such an assumption is to be incorporated in the present model, one must not assume that all subsystems of the organism have a completely fixed response, say state (=output) 1, to a certain stimulus, say stimulus 1. (For simplicity, we will from now on use digits to denote both inputs, states and outputs; context will make clear which is meant.) Since any sub-automaton is deterministic, such a fixed response would mean that the behaviour of the automaton after any 1-input is completely independent of its previous input. This in turn means that the system as a whole cannot modify its responses to any other input as a result of experience. In short, if the UCR is fixed at the level of the subsystems, so is the CR, and no learning will occur.

Instead, the UCR was usually modeled in the following way. Let us suppose that all subsystems are stable under input 1, if they are already in state 1 - in other words, that they have a possible 1-basin of length 1, consisting of state 1 only. In all other respects they are uniformly random. If this is the case, repeated 1-stimuli will of course lead to a growing number of 1-responses in the large collection of systems.

We further suppose that the nature of the gross output of any collection of automata is determined by a majority rule, applied to the outputs (=states) of the subsystems. For simplicity, we call the gross output corresponding to a majority of state k, "gross output k". The strength of the output can be taken to be proportional to the difference in frequency (or the quotient) between the dominant and second-dominant states. (Cf. above on competitive convergence.) It is clear that with the proposed arrangement, the response of the collection to UCS = 1 will ordinarily be gross output 1, with increasing strength for increasing stimulus lengths.

The main point of the present modeling of the UCR is, however, that it allows the subsystems' responses to the UCS (input 1) to be modified by other stimuli, and therefore also allows modification of their responses to the CS (input 2). In the simplest case, namely, with a stimulus sequence of type (C) (cf. above), this works as follows. If a subsystem ends up in state 1 given a 2-sequence, it must start in 1 when next 2-sequence begins. At the same time, the fact that it ended in 1 on the 2-sequence makes it probable that it also started in 1 (Principle (iv), backwards). Therefore, it is more probable that its behaviour will from now on be locked, than it would have been if the response on the last 2 had not been a 1. Repeated application of this argument leads to the conclusion that the frequency of 1-responses on the last 2 will rise gradually.

With a stimulus sequence of type (B) (and with 1-sequence length = 1), i.e. with different stimulus sequences interpolated between the 1:s, the rise in 1-responses on the last input in these sequences will, under certain assumptions concerning the length L of non-1-sequences, be lower than with recurrent 2-sequences (cf. also above, p. 11). It turns out that if L has many small proper divisors, condition (C) has a definite advantage, while with a reasonably large prime number L, condition (B) may actually perform "better".

The above reasoning can be extended to the case where the length of the 2-sequences is not constant, but is for each sequence a randomly chosen number (between n and a fixed larger number N). In this case there will of course be another stochastic element in the process which makes the argument more complicated. For more details see Malmgren (1984); some simulations are also presented there which show that - especially for larger n - the effects with recurring random-length 2-sequences are clearly greater than with the corresponding type (B) constraints. Under the former constraint, the system shows a distinct, progressive modification of its behaviour when presented with the repeated stimulus association - i.e., it learns associatively, albeit in a primitive way.

Some further notes on classical conditioning

In Malmgren 1980, 1984 the argument was not explicitly extended to conditioning proper - i.e. to stimulus sequences in which there is also an interstimulus interval. However, it is easily seen that the logic holds also for this case. Consider, e.g., the constraint

$$\text{(D)}\quad \underbrace{33\ldots3}_{\text{rnd}}\underbrace{22\ldots2}_{\text{k}}1\underbrace{33\ldots3}_{\text{rnd}}\underbrace{22\ldots2}_{\text{k}}1\underbrace{33\ldots3}_{\text{rnd}}\underbrace{22\ldots2}_{\text{k}}1\underbrace{33\ldots}_{\text{rnd}}$$

Repeated applications of Principle (iv) again show that there will be a successive rise in the frequency of 1-responses to the last 2:s.

However, the preliminary simulation results with this model are not promising ((B) type constraints are usually superior), so some modifications have been tried.

Here is one of them. By an inert random automaton is meant a randomly constructed automaton, which is uniformly random except for the fact that there is a bias toward stable states: the probability, for all states a and any input i, that the matrix for the automaton will contain a as the value of f(a,i) is some fixed amount greater than 1/n. The inert automaton is also used in the operant conditioning model; see below. For classical conditioning, inert automata have been used with the modification that the response to input 1 is always state 1 from state 1, but uniformly random elsewhere. The latter condition actually means that the automata are not inert for input 1, except when they are in state 1, and there they are extremely inert. Inert automata were chosen partly because they eliminate the need for using 2-sequences of greater length than one stimulus. Also, the just-mentioned construction of the UCR lowers the risk that the behaviour of the automata becomes "locked" by a CR different from 1, and so it should tend to favour constraint (D) in comparison to (B).

As an illustration, the result of a simulation with 10 000 subsystems and 8 trials for n=20 and k=1 is presented in Table 1.

Table 1
Conditioning in an Inert Random System with 10 000 Subsystems, with n=20 and k=1.

Trial no	Constraint	1 (IS/s)	2 (IS/e)	3 (CS)	4(UCS)
1	D	99	98	102	201
	B	99	94	99	201
2	D	131	123	117	220
	B	134	123	101	200
3	D	137	131	127	227
	B	132	128	107	201
4	D	147	142	123	223
	B	128	125	105	198
5	D	147	134	126	223
	B	134	127	104	205
6	D	152	137	125	221
	B	139	133	111	209
7	D	147	134	129	224
	B	146	140	117	217
8	D	149	136	126	225
	B	135	133	108	207

Note. The upper figures in each double-row refers to condition (D), the lower to a B-type constraint where the CS changes from one sequence to the next. The figures in sub-column 3 in each double-row in Table 1 shows the frequency (in percent of 1/n) of 1-responses on the successive last CS:s. Also shown are the frequencies with which the system goes to 1 on the start and end components of the interstimulus (IS/s and IS/e, columns 1 and 2), and the corresponding frequency on UCS (=input 1; in column 4).

From the table two things can be seen, in addition to the fact that the 1-frequency on CS rises in the desired manner (i. e., rises more, and to a more stable value, in condition (D) than in condition (B)). First, the frequency of 1-responses to 1 rises very much. This is a simple expression of the fact that in order for learning - i.e. modification of the response to CS - to occur, the response to UCS must also be capable of modification (cf. above). What is worse is that the frequency of 1:s on the last 2 is lower than the corresponding 1-frequency on the first and the last interstimulus. In fact - although it cannot be seen from Table 1 - the CS "generalizes" to the whole interval between the UCS:s. (A similar remark holds for the simpler model using constraint (C)).

This is of course a serious drawback of the model, since in real associative learning the conditioned response occurs mainly (although not exclusively) as a response to the stimulus immediately preceding the UCS. If this drawback cannot be eliminated, there are essential features of classical conditioning which the model cannot explain.

Although the results illustrated in Table 1 have one more promising feature which could be exploited - namely, the advantage for constraint D over constraint B seems to be larger for the CS than for the IS's - we are not sure that the just-mentioned problem can be solved within the model. This is mainly because of the way in which it makes use of Principle (iv) above. For in condition (D), the rise of 1's on stimulus 2 is directly dependent on the formation of basins under input 3, starting and ending in state 1. In other words, although the collection of automata does learn what to do when the CS arrives, it "codes" this knowledge by increasing the proportion of the same response even in the absence of the CS. What we would like the machine to do is of course instead to increase the proportion of automata which, from any starting state, will produce the desired response on the CS. For this a radically different principle is perhaps needed; for example, one could use elements of the model sketched in the next section. A combined model is outlined in the section following that.

Reinforcement as noise reduction

Considering that the behaviour of a randomly composed automaton is more varied the more varied its input (cf. (iii) above), manipulating the subsystems' sensitivity to input might be one way of enhancing the process of selecting and de-selecting possible basins. In the theory of operant conditioning (Malmgren 1985), the selection (reinforcement) of stable states of the subsystems is accordingly realized by means of a temporary "arousal", and hence intrusion of environmental noise, in case of "incorrect" behaviour on part of the system as a whole. The following illustration is adapted from Malmgren (1985).

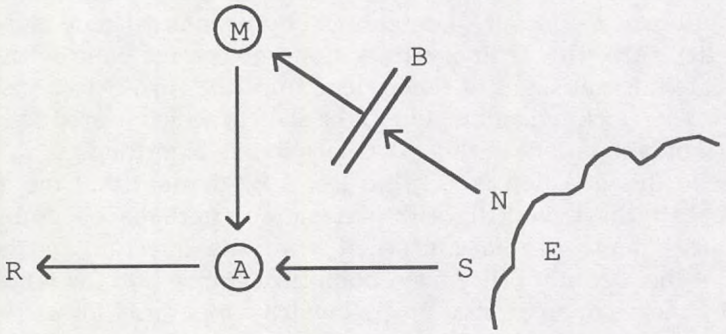

Figure 4. Selection of stable states through temporary intrusion of environmental noise.

Here M (for memory) is a controlling system which provides one of the inputs to the "reflex apparatus" A, which also receives input S from the environment E. The output of A is a uniformly randomly composed (deterministic) function of the input S and the controlling input (the state of M). M itself is an inert random automaton, i.e. for a given input/state combination, it is more probable that the state remains constant than that the automaton goes to another state (the amount of inertia is a parameter which can be varied in the model). Note that in this model, collective input and output functions of large systems are not exploited, but the process of learning takes place in single sub-systems of the kind illustrated in Fig. 4.

Ordinarily, M is protected from environmental input (in the sense that its input is constant = 1) by the information barrier B, which means that it usually tends heavily toward stable states. However, under certain circumstances - namely, in the process of arousal - B is lifted, the tresholds of M are lowered, and M becomes sensitive to a greater range of environmental input. For example, it can be imagined to become receptive to a random noise which is actually there all the time; in the model, this means that a pseudo-random number (within a certain range) is now given to M as an input. (Other kinds of environments have also been used in the simulations, but the larger the random element, the "better" the effects.)

If, now, a contingency is arranged between on the one hand the system's not having response R as its output on a certain stimulus S, and on the other hand this arousal process, M will tend to change its state as long as the system does not perform in the desired way, but to remain stable if the system performs well. The process is a kind of kinesis in the internal state space of M, hopefully ending in a basin consisting of a single state which controls the response of A in the desired way.

With properly chosen parameters, learning is quite impressive even for small state spaces (in the simulations with N=16 and "random noise" reported in Malmgren (1985), about 80% of the systems learned a four-way choice problem).

As in the earlier work, the choice of certain kinds of finite automata as elements in these models (inert randomly composed automata) was made for illustrative purposes only. The qualitative results of course hold for a large class of discrete and continuous systems with multiple equilibria. Certainly systems could be designed for which the results would be much "better" from a behavioural experimentalist's point of view, and/or which would be more realistic from a neurophysiologist's point of view.

It should be noted that in the model of Fig. 4 the inputs of M and A may be disjoint, so that the memory unit never receives S as an input. Indeed, M may be supposed only to "hear" a background signal and (sometimes) the random noise. In a certain sense, such an M does not know what it is learning but only when it has learnt. It should therefore be of interest to compare this model with real motor learning, which is often in a certain sense "unconscious". Not only is learning to ride a bicycle a trial-and-error process - one cannot even always tell what exact movement one is trying. This is consistent with the idea that real motor learning, like learning in the simple model presented here, essentially involves adjustments of parameters via pseudo-random descending commands to lower integrative centres (see also below).

It might now be objected that the model presented here is trivial. Certainly, there is a universally valid algorithm for getting a system to learn any behaviour that it can possibly perform. Let an automaton's output be determined by its internal state and its present input, i.e. $O(t)=f(A(t), I(t))$. Mix the ordinary transitions of the machine with a random walk through all possible internal states of the system, until the correct response to the signal in question is found. Then lock the automaton to the state in which it produces this output. Isn't the present theory only a variant of this extremely simple algorithm?

The answer to this question is Yes; but lest it be thought that the main point of the model is to present this trivial algorithm it should perhaps be emphasized that the substantial part of the theory is concerned with the question of the (approximate) implementation of the algorithm. The main point is of course that the random walk element is implemented by means of an arousal process which temporarily lowers the treshold of the "memory unit" (when performance is "bad"). Further, output as a function of state and input is modelled by means of a hierarchy of automata with state-dependent output. This kind of implementation, we think, is not quite trivial (although it is based on a well-known equivalence between different kinds of finite automata) and may also contribute to the application of the random walk model to real nervous systems.

The relevant objections against the "arousal" model of (non-)reinforcement presented so far are, we think, of another kind. It has two main disadvantages, both of which are connected with the problem of "credit assignment". First, for large systems the process will be extremely slow and simply quite impracticable, which is partly due to the fact that there is no such thing in the model as a solution's being approximatively correct. Therefore, gradual approaches to correct solutions by means of graded reinforcement are not possible. Secondly, discriminative learning (i.e. simultaneous learning of different responses to different stimuli) will be very difficult, since negative reinforcement of one kind of response will change relevant parameters no more often than it changes irrelevant ones. Indeed, there is only one parameter - namely, the state of M. This of course means that discriminative learning will take an enormous time - if it is at all possible in the system in question: even if there is an appreciable probability that there exists a state of M which makes A respond with R_1 to S_1 and a state which makes it respond with R_2 to S_2 it is, in general, rather improbable that there is a state which fulfills both conditions.

We will here only discuss the second problem. To get a better performance in this respect we must, we think, move to yet another class of models; see the last section below.

Outlines of an integrated model

The present section will be devoted to an outline of a model for classical conditioning, using an element from the the present theory of operant learning.

Let us leave, for a moment, the machine drawn in Fig. 4, and return to the collections of randomly composed automata with an UCS defined as above. Suppose that there is, as assumed in the habituation model, a short-time comparator mechanism which for every sub-automaton registers whether or not it changes its state, but also an information barrier of the kind described in the previous section. Further suppose that at any moment, the number of state changes determines the magnitude of the temporary overall arousal reaction (in the sense explained above) of the organism: i.e., if there are many state changes in the system between two successive moments, the input will be more varied during the next moment than if there are few state changes. Evidently, this mechanism will tend to amplify the habituation and dishabituation effects (cf. p. 11).

Also suppose - and this may solve the problems posed above (pp. 14f.) - that for some reason, differences in arousal are also much amplified by the presence of the UCS. For example, there could be a long period of complete non-arousal if sufficiently many CR's of the subsystems are identical to their UCR, and else (this latter assumption may not be necessary) an unusually high arousal for the same period of time. Then the primary determinant of stability in the sub-systems will be whether or not the same response is produced by CS as by UCS, while similarities between responses on consecutive interstimuli will not have a comparable effect. Hence the system will primarily tend to be locked in sequences of states which fit the rules of classical conditioning.

Such a model - which we have not yet simulated - implies that the UCS has some special "reinforcing" character, which does not belong to all inputs. Not all input contingencies could be learnt by the proposed mechanism, only those involving certain stimuli. However, we do not think that this is an unnatural assumption. The same could be said about the following possible extension of the model: suppose that certain stimuli - let us call them the "salient" ones - are especially efficient as conditioned stimuli in associative learning since they tend to temporarily raise the arousal level and, therefore, to enhance the importance of any lowering of this level which may follow. Indeed, is it not the case that classical conditioning works best if an "interesting", arousal-raising conditioned stimulus is followed by an "satisfying", arousal-lowering UCS?

After this excursion we will return to the problem of structural credit-assignment.

Selective attention is selective learning

In general terms, the natural solution of the problem is a system where the responses to different stimuli are controlled by physically different memory units, and where these units are selectively negatively reinforced by the arousal process in case that they "misbehave". Probably, when an animal is punished after having performed a certain response, something happens with the nervous structure controlling this very response, which does not happen (at least not to the same degree) with the structures controlling other responses - otherwise the animal's behaviour would be changed in all its aspects by the failure on any single test.

On the level of algorithms one can of course easily devise a system analogous to the construction in Fig. 4 above, but with several independent memory units reacting specifically to punishments for errors in just those kinds of behaviour which they control. However, it is more difficult to design a way of implementing such an algorithm in real nervous systems, since how is the arousal process to know which unit was responsible for the organism's behaviour? Since we do not feel that the mentioned ad-hoc extension of the above model adds anything to the understanding of this problem (although it does offer a way of posing the problem), we turn directly to the problem of implementation.

Now, as the title of this essay says, we want to identify selective attention and selective learning. Of course, it is very well known that selective attention profoundly affects learning (see for example Mackintosh 1975). To our knowledge, this influence of attention on learning has to date no explanation in terms of basic mechanisms.

However, if (as we postulate in the model of operant conditioning) arousal means intrusion of noise in the system, and therefore a general tendency toward change of state in its sub-systems, then selective attention may simply be the process which distributes the noise unevenly so that certain states are more affected than others by the general tendency toward change. The state changes "on error" may still be supposed to be of a radically non-Hebbian kind: namely, "undirected", pseudo-random jumps, specific only in the sense that the process of selective attention determines where these jumps are to be largest and/or most frequent.

The postulation of a process with this property is not made ad hoc; below we will try to show that it is a natural consequence of the non-linearity of neurons and of certain very general structural traits of the central nervous system.

Of course it must also be shown that this process is equivalent to selective attention in some ordinary sense. We will not go into that question, since it is not at all essential for the explanation of credit assignment in operant learning that the postulated process should also be an explanation of attentional phenomena in general. However, we think that a good case could be made for the hypothesis that selective attention is a "localized" counterpart of global arousal. If this hypothesis is coupled to the above theory of arousal and operant learning it becomes very natural to make the proposed identification.

Before going into the details of the theory, we have to say some words about different concepts of activity. One should not identify (pre- and/or postsynaptic) neural activity (depolarization, spiking etc.) with information processing (or information transmission). Information transmission is essentially a relation of dependency: if changes in structure A cause, or are caused by, corresponding changes in structure B, then information (in a technical sense, not yet semantic information!) is passed between A and B. The nature of the changes do not matter: as well as sending information from a neuron by means of a burst of spikes, you can send information by making a pause in a background activity. This is analogous to the fact that written words can be in white on black, as well as in black on white.

Although the conceptual distinction between neural activation and information transmission is generally recognized, one rather often meets a tendency to confuse the two. The explanation of this confusion goes, we think, as follows.

Since the response curve of a representative neuron (i. e., the probability or frequency of spiking as a function of net input) is non-linear, the first input makes less difference than the fifth on whether an action potential is released. This means that, in general, a neuron which does receive a number of inputs, and which is maybe even already firing at a low frequency, is more sensitive to further inputs than is a silent neuron, and a fortiori more sensitive, or "reactive", than a neuron which does not receive any input at all. One can therefore say that the more a neuron is activated - within a certain range - the better it works as an information channel; pre-synaptic changes (in any direction) have their largest effects when the post-synaptic neuron is in this range.

To be sure, the response curve is usually approximatively sigmoid which means that the reactivity is greatest at medium activity. This is the reason for the phrase "within a certain range" above. But because the nervous system rather seldom works with pauses against a background of maximum excitation - probably for energy reasons - this rather low range is the most common one. Hence neural activation is, in general, positively correlated with information transmitting capacity, and hence the tendency to confuse these two concepts.

A neuron, which is in a state which makes it more reactive to further inputs (more reactive than it is in some reference state, or more reactive than an average neuron at the time), will here be called "sensitized". An important case of sensitization is bringing a neuron near firing threshold, and for simplicity we will mostly discuss this case.

Now, according to our theory, what basically explains credit assignment in learning is not that the structures which control a certain response to a certain stimulus are specifically activated at (or immediately after) the time of the behaviour in question, but that a major part of this activation occurs in a range where it results in a specific sensitization. Because of this specific sensitization at the crucial time of reinforcement, the general process of arousal will tend to change the relevant structures more than others.

We will now try to outline one way in which such a differential sensitization at the time of reinforcement could come about, starting with some well-known facts about neural organization, among which the principles of hierarchical organization and reciprocal innervation (Arbib & Szentagothai 1975) are the most important. We will mainly discuss the second one. It says that if neuron A has an outgoing connection to neuron B, the probability is greater than otherwise that B has a connection back to A. (Probably, this is at least partly a matter of proximity at the time of ontogeny.)

Suppose, for simplicity, that neurons are linear threshold devices, and that their long-term properties are dependent - in some way or other, cf. below - on their firing. Think of a set N of neurons which, at a certain time, fire to make a certain stimulus S release a response R. They probably also will cause a number of neurons in their "vicinity" (defined by connections) to fire. There are also many neurons in the vicinity of N, which are brought near threshold - i. e., they are sensitized - by these firings and by the input that they share with the neurons in N, but which do not fire themselves. From the principle of reciprocity it follows that among these latter sensitized neurons there are probably several which differentially control the reflex arc S-R in the sense, that their outputs have an influence on that behaviour which is greater than the mean influence from a randomly selected neuron in the nervous system.

Now imagine that the behaviour S-R is negatively reinforced by means of a general arousal process which lowers the thresholds of all neurons of the central nervous system by a certain constant amount (or, equivalently, adds the same excitatory input to each). How will this signal interact with the activity and sensitivity of the neurons in the pool controlling S-R?

Since the reinforcement signal always comes after the release of the behaviour to be punished, the neural activity in most of the neurons in N might well have ceased when the arousal occurs. There will however probably be some "traces" of the primary activity, in the form of the sensitization of some neurons a number of synapses away - in an hierarchical system, a number of levels higher up, but also in numerous places defined partly by reciprocal connections on the same level as the primary activity. Because of the reciprocity, the neurons which do show some such trace sensitization are probably among the controlling neurons for the same behaviour. Also because of reciprocity, it is not unreasonable to suppose that after the primary activity in the original set N of neurons has ceased, the same neurons are for a while more sensitized than an "average" neuron.

Note that the "traces" which we are talking about need not be constructed according to any specific algorithm. We just postulate a general (but not absolutely universal) reciprocity of connections, and therefore reciprocity of control.

Now, what happens when the arousal occurs? Because of the physical localization of the trace sensitization, a representative controlling neuron for S-R is at the moment more reactive than an average controlling neuron; i.e., it will have a greater tendency of firing than the latter. This we take to be equivalent to saying that the organism's attention still tends to be on the specific S-R connection. The long-term properties of a controlling neuron for this S-R will therefore tend to be more changed - in some direction or other - by the general arousal than what is the case for the properties of a randomly chosen neuron. This of course means that some structural credit-assignment has occurred.

A similar case could be made for our thesis on the presupposition that neurons are semi-linear elements with sigmoid activation curves; but we will not carry out that part of the argument in detail here.

It is perhaps objected at this point that what we say has no relevance for learning in the proper sense, since the non-linearity which we have referred to is just a property of processes on a very short time scale. So, the results of the interaction between arousal and localized trace sensitization do not, by themselves, entail any permanent changes in behaviour!

Here our reply is: Of course there has to be some kind of permanent changes; but it does not matter in this context how the nervous system translates the changes in short-time activity to more permanent traces. Maybe the "controlling" neurons use post-tetanic potentiation which in turn enhances transmitter production and/or synaptic growth, but they could as well utilize homosynaptic depression or, for that matter, have the character of self-reverberating circuits which turn on when the amount of innervation temporarily rises above threshold. Of course LTP and/or LTD mechanisms might come into play too.

As remarked above, any of these memory processes could have the desired result (in this case the inhibition of reflex R on next S), given a suitable nervous structure. For a mechanism built on random walk and selection it could even be an advantage to be able to make use of a number of different mechanisms, so that not all neurons change in the same way. This helps in giving the process enough variation.

We have designed a simple computer program by means of which different versions of our theory can be modeled. It is yet on a preliminary stage and we have no simulation results to report here. In it, an organism with input neurons, output neurons, and hidden units is constructed (cf. Fig. 5 below). Non-linearity is achieved by the use of discrete threshold units or semi-linear elements. There are four specific input-output monosynaptic "reflex arcs", corresponding to movements in four directions in the plane. Movement is determined by the balance between opposing outputs. Signals from the input units (I1-I4) also go - according to a mixture of a locality principle and a random function - to a number of "hidden units" in a first layer (HF); these in turn excite or inhibit (according to their randomly determined nature) output neurons (O1-O4), usually following some principle of reciprocity, or again according to locality mixed with a random element. The hidden units in the first layer connect in a more or less random manner with a second layer of hidden neurons (HS) which also connect randomly with each other, and which direct their output to the first layer only, usually in a reciprocal manner. Several such layers of hidden units can be added (and will probably be needed).

In Fig. 5, a small organism with two layers of hidden neurons is illustrated. The connections are not shown.

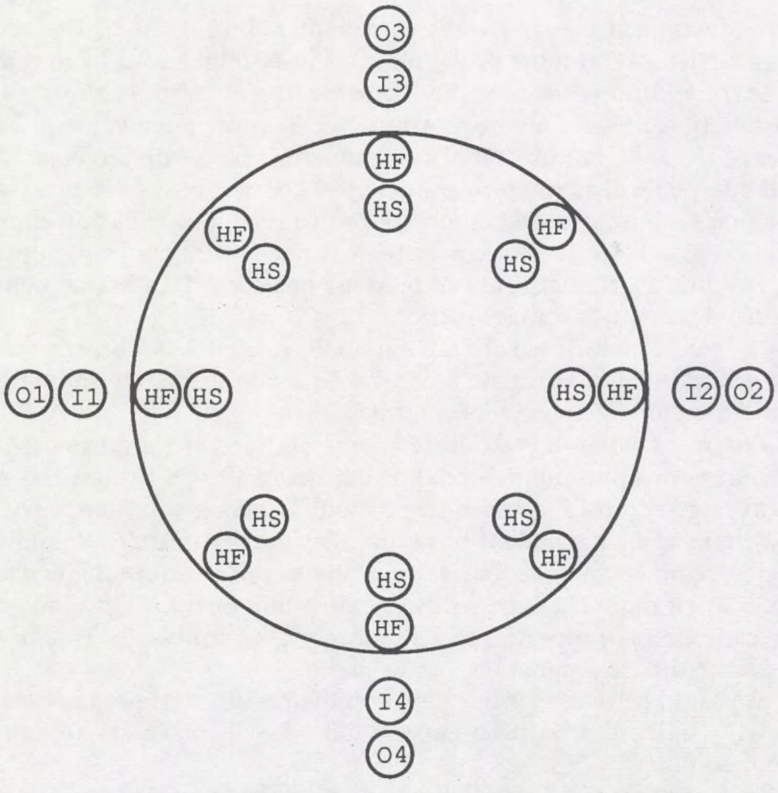

Figure 5. A small organism with two layers of hidden neurons.
I1-I4 = Input units
O1-O4 = Output units
HF = Hidden unit, first layer
HS = Hidden unit, second layer

The hidden units can be made to have different long-term plastic properties. Hebbian learning is not primarily used but instead some simpler kinds as for instance the PTP-like one which consists in a more or less permanent strengthening or weakening of all forward synapses of a cell if it is fired. In the linear threshold version, the reinforcement scheme works by temporarily lowering all thresholds of the hidden units by the same amount in case of inappropriate behaviour (e.g. no movement, or movement in the wrong direction, at a certain input) on the part of the whole system. Hopefully, the model will react selectively in that certain units, which are functionally connected with the reflex in question but at the time only near threshold, will fire for the first time on this "arousal". Since this will change their "membrane properties" for a long time, they may react differently at next presentation of S, and so may the "organism" as a whole. When the organism performs in the proper way, there will be no arousal, the neurons will not change their input-output relations, and the system will react in the same way to next stimulus of the same kind.

We are at present running some preliminary experiments using this model.

References

Abrams, T.W. & Kandel, E.R. (1988). Is contiguity detection a system or a cellular property? Learning in Aplysia suggests a possible molecular site. *Trends in Neurosciences, 11,* 128-35.

Arbib, M. (1981). Perceptual structures and distributed motor control. In V.B. Brooks (Ed.), *Handbook of Physiology. The Nervous System. II.* Bethesda, Maryland: American Physiological Society.

Arbib, M. & Szentagothai, J. (1975). *Conceptual models of neural organization.* Cambridge, Mass. : MIT Press.

Ashby, W.R. (1956). *Introduction to Cybernetics.* London: Chapman & Hall.

Barto, A.G. (1987). An approach to learning control surfaces by connectionist systems. In M. Arbib & A. Hanson (Eds.), *Vision, Brain and Cooperative Computation.* Cambridge, Mass.: MIT Press.

Fogel, L.J., Owens, A.J., & Walsh, M.J. (1966). *Artificial Intelligence through Simulated Evolution.* New York: Wiley.

Griffith, J.S. (1971). *Mathematical Neurobiology.* London: Academic Press.

Hebb, D. (1949). *Organization of Behaviour.* New York: Wiley.

Kaczmarek, L. & Levitan, I. (1987). *Neuromodulation. The biochemical control of neuronal excitability.* New York: Oxford University Press.

Kruskal, M.D. (1954). The expected number of components under a random mapping function. *American Mathematical Monthly 61,* 392-7.

Lindqvist, G. & Malmgren, H. (1986). *Organisk Psykiatri. Vetenskapsteoretiska och kliniska aspekter. Philosophical Communications, Red Series, 26.* Göteborg: University of Göteborg.

Mackintosh, N.J. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. *Psychological Review 82,* 276-98.

Malmgren, H. (1980). *Om sannolikheten för inlärning i slumpvis sammansatta, deterministiska system. Philosophical Communications, Green Series, 6.* Göteborg: University of Göteborg.

Malmgren, H. (1984). Habituation and associative learning in random mixtures of deterministic automata. *Göteborg Psychological Reports 14:2.* Göteborg: University of Göteborg.

Malmgren, H. (1985). On the nature of reinforcement. *Göteborg Psychological Reports 15:3.* Göteborg: University of Göteborg.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

McClelland, J.L., Rumelhart, D.E., & Hinton, G.E. (1986). The appeal of parallel distributed processing. In D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel Distributed Processing.* Vol 1. Cambridge, Mass.: MIT Press. Pp. 3-44.

Rumelhart, D.E., Hinton, G.E. & McClelland, J.L. (1986). A general framework for parallel distributed processing. In D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel Distributed Processing.* Vol 1. Cambridge, Mass.: MIT Press. Pp. 45-76.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). Learning internal representations by error propagation. In D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel Distributed Processing.* Vol 1. Cambridge, Mass.: MIT Press. Pp. 318-62.

Wigström, H. & Gustafsson, B. (1988). Presynaptic and postsynaptic interactions in the control of hippocampal long-term potentiation. In P.W. Landfield & S. Deadwyler (Eds.), *Long-Term Potentiation: From Biophysics to Behavior.* New York: Alan Liss. Pp. 73-107.