COMMENTARY

LEARNING FROM MISTAKES

D. R. CHIALVO*†‡§ and P. BAK*‡

*Niels Bohr Institute, Blegdamsvej 17, Copenhagen, Denmark

†Division of Neural Systems, Memory and Aging, University of Arizona, Tucson, AZ 85724, U.S.A.

‡Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, U.S.A.

Abstract—We re-examine the commonly held view that learning and memory necessarily require potentiation of synapses. A simple neuronal model of self-organized learning with no positive reinforcement is presented. The strongest synapses are selected for propagation of activity. Active synaptic connections are temporarily "tagged" and subsequently depressed if the resulting output turns out to be unsuccessful. Thus, all learning occurs by mistakes. The model operates at a highly adaptive state with low activity. Previously stored patterns may be swiftly retrieved when the environment and the demands of the brain change. The combined process of: (i) activity selection by extremal "winner-take-all" dynamics; and (ii) the subsequent weeding out of synapses may be viewed as synaptic Darwinism.

We argue that all the features of the model are biologically plausible and discuss our results in light of recent experiments by Fitzsimonds *et al.* on back-propagation of long-term depression, by Xu *et al.* on facilitation of long-term depression in the hippocampus by behavioural stress, and by Frey and Morris on synaptic tagging. © 1999 IBRO. Published by Elsevier Science Ltd.

Key words: self-organization, learning, memory systems, brain, mathematical models, criticality.

CONTENTS

1. INTRODUCTION

Despite the abundant knowledge of a myriad of cellular and molecular aspects of neuronal development, memory and learning, there is no clear understanding of which fundamental principles are at work in organizing the brain in order to allow it to act, remember, forget and adapt. We do not know the answer to the question: what is the essence of the biological mechanism that allows the brain to work at its incredible speed and precision?

One obvious strategy is to obtain detailed information of reduced parts of the problem in the hope that the complicated mechanism could be resolved from the bottom-up. However, even the most detailed mapping of the individual parts of the brain does not constitute understanding of brain function, just as complete knowledge of the function of transistors and wires does not in itself add up to an understanding of how a computer works. How can the enormous experimental information that has been gathered be synthesized to create even a rudimentary picture of how one may explain behavior which can be identified as "intelligent".

The point of view has often been expressed that the brain is so complicated that it cannot be

---

understood by the brain. Indeed, it consists of trillions of neurons each connected to tens of thousands of other neurons. We do not subscribe to this view. Although there is an enormous difference between a lobster brain and a human brain, our working hypothesis is that they are built and organized by the same few fundamental principles. By this, we do not imply, for instance, that lobsters must have a thalamus as mammals do, but that there might be some "universal" and "simple" mechanism which allows a large number of neurons to connect autonomously in a way that helps the organism to survive.

Our goal is to identify this mechanism. Our strategy is to construct a simple model that demonstrates the principle without specific reference to the underlying complicated physical and chemical properties of the neurons, etc., which are responsible for the implementation in a real brain.

Our model is based on two simple and biologically plausible principles, where one concerns the activity (i.e. neuronal firing), and the other the plasticity (i.e. the changes in synaptic strength): (1) Activity only propagates through the strongest connections. Thus the overall activity is kept low by this "winner-take-all", extremal dynamics. (2) There is only negative feedback from the environment to the active synapses, i.e. we are talking about learning with a stick and no carrot. As a result, the model operates at a highly adaptive, and possibly critical state. "Efficient, self-organized, and adaptive behavior can take place only when the two principles, negative feedback and extremal dynamics act together. One is useless without the other". We show that our simple mistake-driven neuronal learning model can out-perform other more complicated and sophisticated models that use positive reinforcement.

The paper is organized as follows. In order to motivate our views, we would like in the first sections to dwell on the train of thought that eventually led us to our particular model. Section 2 contains a succinct account of how mainstream ideas, in our opinion, have overemphasized the role of enhancing or reinforcing synapses in response to examples. The strong belief in Hebbian reinforcement has definitely biased both the experimental and theoretical research to "search under the light", preventing the objective examination of alternative scenarios. We argue that negative reinforcement is essential, and that any positive reinforcement reduces the ability to adapt. Section 3 presents a chronological description of theories supporting a Darwinian view of the problem at hand. Section 4 establishes the restrictions we impose in attempting to solve the problem in a biologically reasonable way. The reader who is not interested in all of this may proceed directly to section 5 in which our basic model is defined and analysed. This section includes our main results. The unique features of the model are discussed, emphasizing its

ability to fast adaptation. Its robustness against noise and the case of arbitrary geometry are discussed. The performance is illustrated with a simple simulated task. Section 6 closes the paper with a discussion of recent experimental findings of synaptic depression in light of our model.

2. POSITIVE VERSUS NEGATIVE REINFORCEMENT

There is a good deal of hard-wiring defined by the genetic code as it has evolved through biological evolution, but most of the connectivity has to be self-organized through the life of the individual. There is simply not enough information in the human (3×10^9 base pairs) DNA to provide a blueprint for all (10^{18}) the connections. Nowadays no one argues the fact that the connectivity of neurons could not be specified in the genoma.[30] DNA can encode for a variety of neuronal morphologies, different neurotransmitters, diverse adhesion molecules, etc., but certainly cannot specify with which other a given neuron should be connected.

Thus any brain, from the simplest to the most highly evolved, must be able to wire itself starting from structureless connections without consulting another brain. Furthermore, it must remain adaptable, i.e. capable to alter the connection matrix when the environment changes. The correct response to a given event tomorrow may not be the same as today.

It is widely believed and accepted that learning in the brain resides in the plasticity associated with alterations in synaptic efficacy. Without exception, contemporary theoretical formulations of such learning follow Hebb's ideas[26] of reinforcement, where synaptic connections between neurons excited during a given firing pattern are strengthened. This mimics the process known as long-term potentiation (LTP), which has been widely investigated in mainstream experimental neuroscience.

The corresponding mathematical models are inclined towards similar ideas. To be specific, let us mention the most relevant artificial neural network models that have been suggested as mechanistic explanations of learning and memory. These are the "attractor" neural networks (ANN) and various forms of feed-forward networks trained by back-propagation algorithms.

The ANN were introduced 20 years ago by John Hopfield.[27] Memory was represented by patterns of up and down "spins" representing firing and non-firing neurons, respectively. In "back-propagation" neural networks the environment provides an error signal used to correct synaptic connections by a Newton iteration method.

While both produce interesting results, it is widely recognized that they lack biological realism. In ANN the necessary assumption of fixed ("clamped") states is not realistic. In regard to back-propagation it is clear that the environment cannot provide the required information about the

specific error of the output signal. Furthermore, elaborate external computations are needed to set up the synaptic connections. A "miracle" must take place at some stage in those models through the interaction with another brain, thus none of these networks is "self-organized".

There has been little success relating these models to real brain function, although they have turned out to be very useful in an engineering context for various problems. In fact, we believe, in accordance with a recent review (see Table 1 in Reeke)[35] that essentially all contemporary artificial neural network models are biologically implausible.

The work coming closest to being biologically consistent is probably Barto's and colleagues[9,10,33] on "reinforcement learning". Here, the environment simply provides a feedback signal informing the brain on the success (or lack thereof) of its efforts. The environment acts as a critic rather than a teacher. However, the learning is extremely slow and the system has to be completely re-programmed from scratch when presented with a new learning task, even one which it has experienced before. The system has little ability to adapt. We shall argue that this is due to the presence of positive reinforcement rules.

All of the above models emphasize the role of enhancement of synaptic strength as the crucial mechanism for learning. However, long-term synaptic depression (LTD) in the mammalian brain is almost as prevalent as potentiation, but there appears to be little or no understanding of its functional role. Working hypotheses cover a wide range, but depression is always given an auxiliary function to potentiation. Early ideas by Cooper[15] motivated a detailed description of the dual depression/potentiation character of the plastic changes, first in the hippocampus and then in the neocortex.[20,31]

A recent review,[7] reflecting on the current variety of ideas regarding the functional role of LTD, speculates: "Although it is conceivable that LTP is the critical phenomenon used for storing information, and that LTD may exist simply to reset LTP, it must be noted that it is also conceivable for the converse to be true."

Turning things upside down, we suggest that "depression" of synaptic efficacy is the fundamental dynamic mechanism in learning and adaptation, with LTP playing a secondary role.

There are two fundamental differences between the classical view of learning by reinforcement and the view proposed here.

1. Learning by reinforcing good responses is a process that by definition never stops. No explicit rule ends the reinforcement when the goal has been reached. Without such an extra rule, learning by reinforcement could be regarded as addictive. However, learning which proceeds only by correcting mistakes implies a process that stops as soon as the goal has been achieved. This prevents the formation of "deep holes", or highly stable states from which adaptation to new rules is difficult and slow, requiring, perhaps, a significant amount of random noise.

2. If an adaptive system is placed in a new environment, or otherwise subjected to learn something new, the likelihood of making mistakes is generally larger than the chance to be initially right. Therefore, the opportunity to shape synapses is larger for the adaptive mechanism that only relies on mistakes, leading to faster convergence.

Negative feedback is the first of our two principles.

3. THE DARWINIAN VIEW

There is some important prior work suggesting some sort of negative feedback in the process of learning.

Young[42–44] was probably the first to suggest that brain changes involved in adult learning could be the result of the elimination of neuronal connections. He phrased learning in the octopus as "closing of unneeded channels":

"In discussions of possible synaptic changes during learning it is usually assumed that the change is an increase in the ease of passage along a particular pathway, and that this increase occurs gradually, that is, partially on each occasion of learning. The opposite hypotheses have been put forward here: that learning is by the inhibition of an unwanted pathway, and that this occurs suddenly and completely in each unit or module, the mnemon that is concerned."

Young's mindset about octopus learning and memory systems was footed on the fact that negative feedback was already implemented in simple monosynaptic or polysynaptic reflex systems even in brainless animals. Thus, one solution that evolution could explore and further modify would be a kind of multisynaptic reflex system in which the essential mechanism would be temporary or permanent inhibition of unneeded channels until the action fits the demand, following an initial random action. This simple strategy of trial and error can be efficient enough for certain organisms, and if supplemented with a way to amplify the bias towards the right solution while preserving previous successful trials, it could implement the more demanding requirement in higher organisms. Our own ideas endorse this early clever thinking.

Richard Dawkins[18] speculated that selective death of neurons could underlie storage of memories. Subsequently, in the 1980s Albus[1] hypothesized that, "pattern storage must be accomplished principally by weakening the synaptic weights rather than strengthening them."

Albus' ideas later inspired some of the work on "reinforcement learning", mostly articulated by

Barto.[9,10,33] Over the last decade, Changeux[11–14,19,39] has been a leading proponent of similar ideas, as delineated in his book "Neuronal Man":[11]

"The 10,000 or so synapses per cortical neuron are not established immediately. On the contrary, they proliferate in successive waves from birth to puberty in man... One has the impression that the system becomes more and more ordered as it receives instructions from the environment. If the theory proposed here is correct, spontaneous or evoked activity is effective only if neurons and their connections already exist before interaction with the outside world takes place. Epigenetic selection acts on preformed synaptic substrates. To learn is to stabilize pre-established synaptic combinations and to eliminate the surplus."

Edelman[21,22] and colleagues[35–37] have articulated a selectionist scheme known as neuronal group selection, a Darwinian mechanism able to explain the formation of higher functions in the working brain. These ideas emphasize: (i) generation of variability, both structural and dynamical, within and between neuronal populations; (ii) interaction of the neural circuitry with the organism's environment; and (iii) differential amplification or attenuation of neuronal connections (depending upon its functional contribution).

It is illuminating to read the accounts of these early supporters of a Darwinian view of the brain (the American psychologist W. James[28] was the first we know of to talk in these terms as early as 1890). However, to support any sensible claim about the advantage of this perspective, there is a need for mathematically tractable models such as the one discussed in the following sections.

4. ESTABLISHING THE GROUND RULES

How do we visualize the problem that the brain has to solve, and what tools are biologically acceptable?

First of all, the purpose of the brain has to be defined in advance. If there is no goal of the game, how can one even start talking about solving the problem? Specifically, we assume that any brain is able to judge what is good and what is bad by a mechanism which is not self-organized on the time-scale of brain function, but has been established through Darwinian evolution. There has to be a "Darwinian good-selector". Roughly speaking, one might think of the sugar level in the blood resulting from getting food as providing such a signal. In more elaborated organisms one can reasonably assume that basal ganglia projections to all areas of the brain will also be able to broadcast elementary outcome signals such as good/bad, pain/comfort. Specifically, we assume that the elementary brain (or the brain region being modeled) receives a global signal which simply states whether the result of the action caused by the output of the brain led to a favorable result or not. This signal is all that we are allowed to act upon. The environment does not tell us which specific neurons should have fired or which muscles should have been activated. The environment acts like a critic, not a teacher. If we think of a brain playing a game, the only information which is provided is a signal that tells the system whether it won or lost at the end of the game. The Darwinian good-selector interprets winning as good, losing as bad. Of course, one might think of the signal as being highly preprocessed by hardwired mechanisms before it reaches the brain, or one might even think of the signal as coming from other areas of the brain, which play the role of environment for a given area.

In contrast, everything else has to be self-organized. We assume that initially all inputs from the environment (visual signals, etc.) are fed in randomly, the connections between neurons are random, and the outputs are randomly connected with muscles. The system is then on its own, acting upon the feedback signal only in a pre-established way. We assume that the feedback signal for plasticity is broadcast to all neurons, which are then modified using only the information which is locally available. One might suspect that the limitations evoked by the demand of self-organization and biological feasibility may limit our choices to the extent that the resulting learning process would be slow and inefficient. To our surprise, the resulting learning process turned out to be extremely fast, as described later.

4.1. On self-organized criticality

Some of the intuition behind our modeling approach comes from the study of self-organized critical phenomena.[3,5,6] A good metaphoric picture is that of a pile of sand on to which sand is poured at a very slow rate. In the beginning, the pile is flat and stable. As the process continues, the pile becomes steeper and small avalanches will occur. Eventually there will be avalanches of all sizes, up to the full size of the system. The pile has reached the self-organized critical state, where most of the surface is very steep, with grains very close to the instability threshold. Consequently, very small modifications of the configurations of the sand-pile can lead to large changes in the flow of sand; the susceptibility is high.

The mathematical modeling of sand-piles bears some striking resemblance to models of neuronal function. In sand models, each site receives sand grains from neighbor sites. When the height exceeds a critical threshold, there is a toppling event where grains of sand are sent to the neighboring sites. In neurons, each neuron receives a signal, the "action potential", from other neurons. When the sum of the inputs from other cells exceeds a threshold, the neuron fires and a signal is sent to other neurons. Could it be that the brain operates at a critical
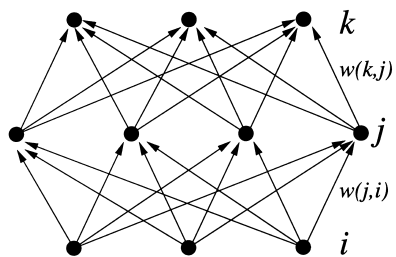
Fig. 1. Neurons in layer *J* have synaptic connections with all inputs in *I* and connect to all neurons in layer *K*.

state, just as the sand models? The sand-pile does not perform any useful task so at the very least a mechanism for plastic change of the toppling rule has to be added. Already, Turing[40] has speculated that perhaps the working brain needs to operate at a barely critical level, to stay away from the two extremes, namely the too correlated subcritical level and the explosive supercritical dynamics.

Instead of adding the sand, one could simply tilt the pile slowly. The avalanche will then start at the steepest site. This type of dynamics is called "extremal dynamics"[34] and has been utilized in recent models of evolution,[4] where the least fit species was always weeded out.

Extremal dynamics (winner-take-all) is our

second principle. It serves to take the system to the adaptive critical state. Also, it plays a major role for the credit assignment by keeping the number of neurons eligible for punishment small.

5. THE MODEL

Just about any arbitrary architecture, for instance a completely random one (we return to that case later), can be chosen, but for pedagogical reasons we begin with a rather simple geometry. We consider a two-layer network where *K* represents the outputs, *I* the inputs and *J* the middle layer (Fig. 1). The inputs are very simple: they consist of a single input neuron firing. This could represent a visual signal. The output cell could represent a motor response. Indeed, the usefulness and thus the evolution of the brain started when biological species achieved the ability to escape from predators, and search for prey, by moving, so it makes sense to describe brain function in connection with the sensory-motor system integration. No immobile species has a brain. The argument had been made[1] that higher brain functions evolved also as similar input–output transformations but for more sophisticated environmental demands.
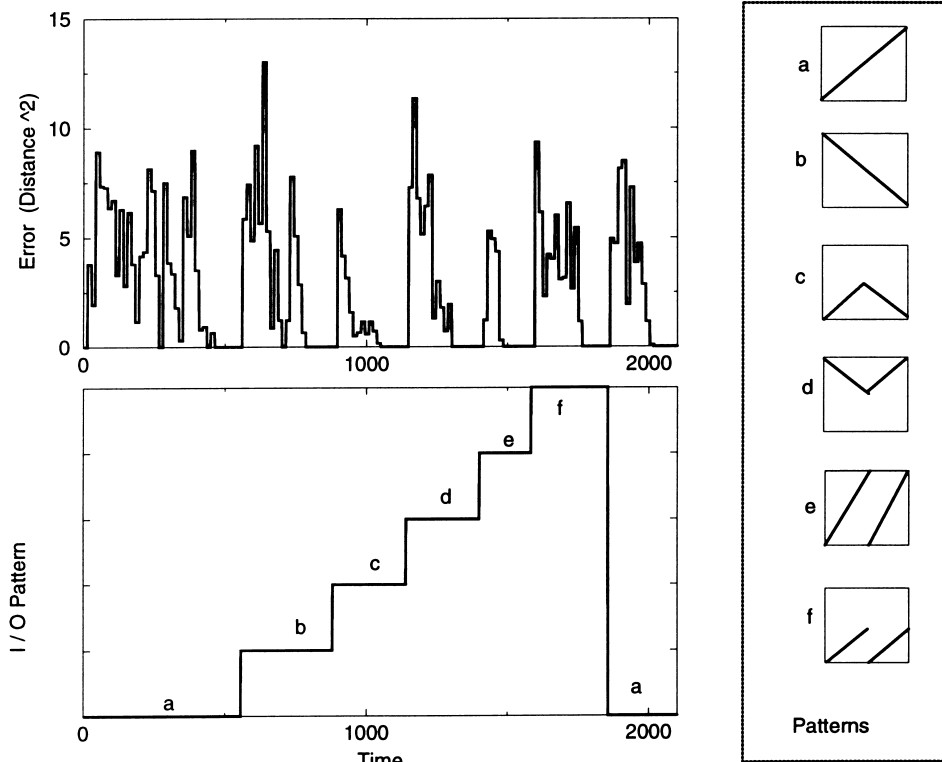
Each input is connected with each neuron in the

Fig. 2. Learning six associative mapping (patterns illustrated in the right box), showing the outputs (*y*-axis) to be connected with the inputs (*x*-axis). Seven inputs and outputs and 300 neurons in the middle layer were used with δ = 0.01. The bottom diagram shows which of the six patterns was presented at a given time. The top diagram shows the relative distance between the desired output and the net's current output. A distance of zero meant that perfect learning had been achieved; after 50 additional iterations (for visualization purposes), a new pattern was presented.
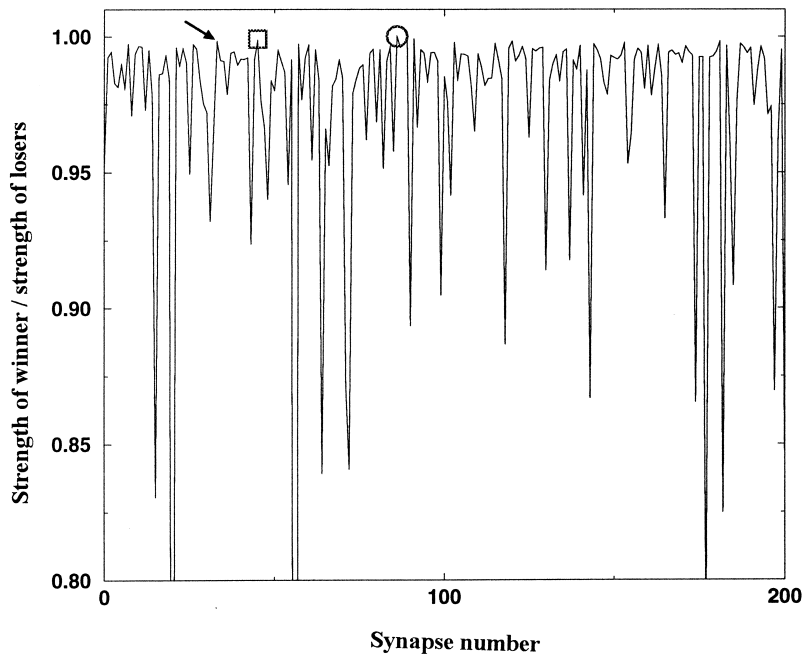
Fig. 3. Landscape of synaptic strengths between one input and 200 neurons in the middle layer. The encircled value corresponds to the strongest connection between the input neuron and the output layer. The arrow and the box indicate other connections likely to be used in the future.

middle layer which, in turn, is connected with each output neuron, with weights W representing the synaptic strengths. The network must learn to connect each input with the proper output for any arbitrary associative mapping (see Fig. 2). Let us arbitrarily choose a map (labeled "a") where the goal is to connect each of seven inputs 1–7 to the corresponding seven outputs, 1–7. The weights are initially randomized, $0 < W < 1$.

In order to achieve efficient self-organized learning, it is essential to keep the activity low. In previous work on self-organized learning,[2,38] this was done by a global threshold control. Here, we invoke winner-take-all dynamics: only one neuron k, namely the one that has the largest $w(k,j)$, fires at each time step.[45]

The entire dynamic process is as follows:

An input neuron i is chosen to be active. The neuron j_m in the middle layer with the largest $w(j,i)$ is firing. Next, the output neuron k_m with the maximum $w(k,j_m)$ fires. If the output k happens to be the desired one, nothing is done, otherwise the two synapses which were involved in the bad decision, $w(k_m,j_m)$ and $w(j_m,i)$, are both depressed by a fixed amount δ. This is the only parameter in the model. The value of δ is not important—it is the natural unit in our model, like the grain of sand in the sandpile model. In order to prevent the synaptic strength from decaying forever, the amount δ is redistributed equally among other synapses. This redistribution does not have to be related to any particular event, but can proceed at a uniform compensatory rate. Then another random input neuron is chosen and the process is repeated.

The physiological process that we have in mind is one where the synapses $w(k_m,j_m)$ and $w(j_m,i)$ connecting all firing neurons are "tagged" by the activity. The tagging lasts long enough to ensure that the result of the process is available—the time-scale must match typical processes of the environment rather than firing rates in the brain. If a negative feedback signal is received all the synapses which were involved and therefore tagged are punished, whether or not they were "responsible" for the bad result.

The iterative application of this rule leads to a quick convergence to any arbitrary input–output mapping. Figure 2 shows this for the map "a" where the seven inputs are mapped to the corresponding seven correct outputs in a few hundred time steps. After this, there are no further modifications since all actions are correct.

It is now clear how the use of winner-take-all extremal dynamics solves the crucial "credit assignment problem"[10] (which has been the stumbling block in previous attempts to model self-organized learning), in a simple and elegant way. The dynamics assures that only very few synapses are active in the process and therefore eligible for later punishment. This allows patterns associated with other inputs to remain largely intact. In traditional reinforcement schemes this is not the case, and the learning becomes extremely slow.[10]

It is worth reflecting on the basic principles behind this simple model. There is a process of selection followed by a process of possible depression. The former assures that activity only propagates through the strongest available connections.
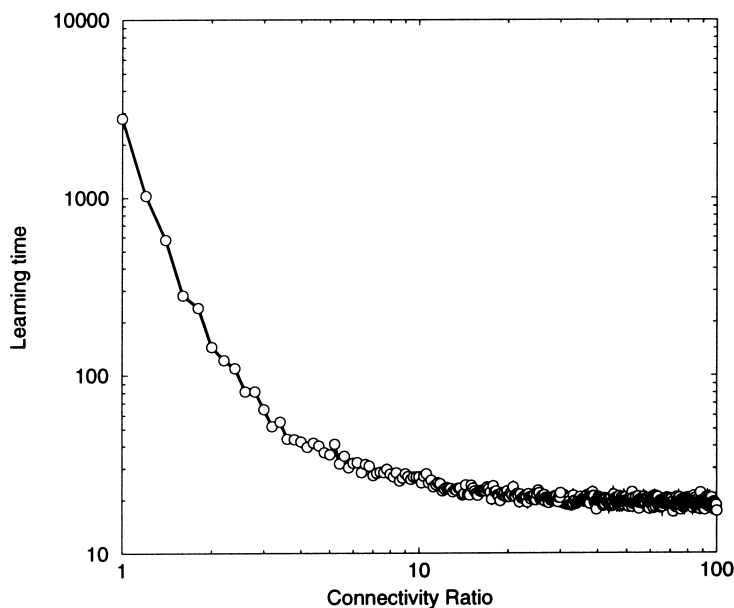
Fig. 4. Scaling of the average learning time (per input) as a function of the middle layer/input layer ratio. Values represent means of 128 realizations. Learning time was defined as the number of steps required to learn all input–output associations.

In the case of unwanted results, the latter makes sure that some other neurons have a chance next time to offer a better solution. This is exactly the essence of natural selection in the context of evolution. The fittest organism is the fittest not because it has proliferated by enhancing its intrinsic abilities, but simply because others have been weeded out by being relatively weak. As learning (evolution) proceeds there is no further advantage for the successful (strongest) connection. In traditional Hebbian networks the differences between "good" and "bad" connections are always large.

The feature just commented on provides another interesting property to the model. Remapping to new cases is straightforward for this network, which demonstrates the ability to adapt quickly to arbitrary new situations. After learning the identity map the network was exposed successively to five other associative maps, labeled "b" to "f", respectively, where the definition of correct outputs has been modified. The small inserted diagrams in the figure illustrate which output is satisfactory for each input. The inverse map "b" can be thought of as turning the visual field upside down when compared with "a". It can be seen that the error (plotted in the top diagram) quickly returns to zero, indicating that the new pattern has been completely learned. This dynamic process has no analogue in traditional neural networks where the synaptic strengths have to be recalculated from scratch for each new input–output mapping, although this single input–output mapping can be much more complicated than the one studied here.

The reason for quick relearning (adaptation) is simple. As was already discussed, the rule for changing synaptic strengths ensures that synaptic changes occur only at neurons involved in wrong outputs. Therefore, the landscape of weights is only re-shaped to the point where the new winners barely support the new correct output, with the old pattern only slightly suppressed. Thus, only a slight suppression of a currently active pattern is needed in order to generate new patterns when need be.

The most spectacular feature of our model is its ability to remember quickly "old" patterns which have been correct once in the past. This is illustrated in Fig. 2, where the pattern "a" is learned much more quickly the second time around. The learning time decreases even further if the system is subjected to more learning cycles.

How can that be? Figure 3 shows the strengths of the synapses connecting one input with all of the middle layer neurons. These values are the ones from which one winner will be selected. It is seen that the landscape include very many values which are very close to that of the active one, a manifestation of the critical nature of the state. Only a small perturbation of the strength is needed to suppress the currently active one (the winner, encircled in the figure), and previously active synapses can easily be activated again. This contrasts with the classical reinforcement scenario where at the end of some pre-established learning period the correct synapses are dominating the incorrect ones. Then adaptation is slow since new learning has to start essentially from scratch, and there is no memory kept of old patterns. One could speculate that LTP appears to be useful only for permanent learning patterns.

Suppose, for instance, that the strongest synapse, indicated with a circle in Fig. 3, is the one associated

with the current correct input–output mapping, say "a". If the environment changes so that "b" is now the correct map, the current responses to the seven inputs will be incorrect, and punished by an amount δ. Since they were only barely larger than the other strengths, they will become inactive in a few steps, and synapses related with patterns which have been correctly learned in the past may re-emerge. Suppose that a pattern similar to "c" emerges utilizing the synapse indicated with a square. This will also be deemed wrong by the Darwinian good-detector, and suppressed in another few steps. Eventually the synapse (indicated with an arrow) providing the correct pattern will be selected. Thus, reading through past experiences, the correct mapping may become established quickly without having to consider the entire space of possible synaptic connections. Sometimes the recovery fails because of overlap between the patterns formed in constructing the correct connections. The fast retrieval of old, good patterns is further enhanced if synapses which have never been involved in correct outputs are punished much more severely than the synapses whose firing at least once has produced a good result.

Any amount of positive reinforcement will tend to interfere with this ability, by enhancing the currently correct connection above the level of other synapses, and prolong the time required to suppress it when need be.

5.1. Robustness against noise

The system is robust with respect to noise. Suppose that when selecting the winner, a small amount of noise is added to all weights. This might cause an incorrect synapse to be selected instead of the winner. As a result, the synapse will be depressed. This might happen for a while, but eventually the synaptic strengths will become suppressed below the level where the noise can activate them, and the input–output mapping becomes perfect. The price which had to be paid was the suppression of incorrect synapses below the critical level or, equivalently, the enhancement of the proper connections relative to the background. Thus, adaptation will take longer in order to depress the excess strength of the synapses, but the learning remains perfect. The network compensates for noise completely automatically without further modifications! We believe that this feature is important for the model to be biologically reasonable.

5.2. More neurons means better brains?

How does the performance change as the size of the brain increases? We know that generally larger brains with more neurons perform more sophisticated, much faster tasks. (Without entering into a very deep debate, it is a fact that the ratio between brains is determining which species become prey and which become predators, making we humans the predator by excellence which our environment has to endure.)

In contrast, the common result for traditional artificial neural networks, often hidden under the rug, is that as the networks get larger they perform worse! In our model, the scaling of the learning time with the size of the middle layer is interesting and different. A network with a large middle layer offers many more options for the system to select among when an incorrect pattern is suppressed. This speeds up the learning of correct associations, as shown in Fig. 4, where the average learning time is plotted for a network with a constant number (seven used here) of units in the I and K layers and an increasing number of neurons in the middle layer. It can be seen that the performance improves for increasing ratios, an increase of one order of magnitude in size speeds up learning time by two orders of magnitude. Results for re-learning scale in the same qualitative way. Bigger is better, in contrast with all other schemes that we are aware of where learning is slower in bigger systems. For constant middle layer size, the scaling of learning time t versus input (and output) layer size N is $t \sim N^2$, also in contrast with other models where the scaling is usually exponential in N.

Biologically, this is of great importance. Increasing the quantity of neurons is cheap, since it does not require further DNA, it only requires a few genes responsible for the turn-off of neurogenesis. In contrast, hardwiring is expensive since it requires new information for each connection. For our system, quantity is quality.

5.3. Random geometry

In order to illustrate the robustness of the model, which is important for our mechanism to have any biological relevance, we have studied an architecture where each of n neurons is arbitrarily connected to a number n_c of other neurons. A number of neurons (n_i and n_o) is arbitrarily selected as input and output neurons, respectively. As before, an input neuron is chosen. This neuron is connected with several others, and the one which is connected with the largest synaptic strength fires. That neuron is also connected with several others, and the one with the strongest connection is selected next. If, after a number of firing events, the correct output has not been reached, each synapse in the chain of firings is depressed as before. If the correct output is achieved, no modification occurs. A system with $n = 25$, $n_i = n_c = n_o = 5$ behaves like the layered structure presented above. This illustrates the development of structure even in the case where all initial connections are absolutely uncorrelated. We have systematically studied the performance of a system with 200 neurons, out of which five were chosen as
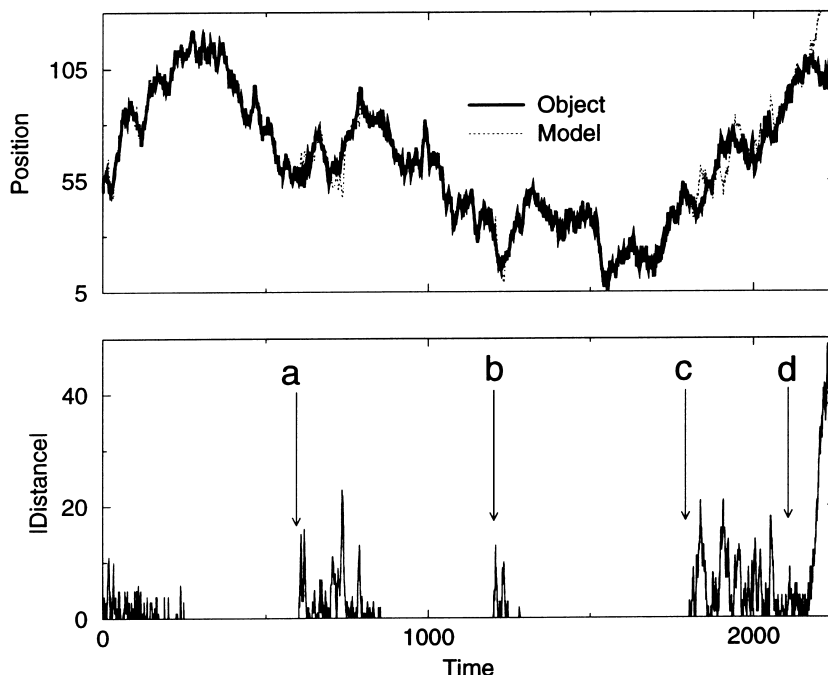
Fig. 5. An autonomous agent learning to track a randomly moving object, subjected to various (some drastic) perturbations. (Top) Position of the object (continuous line) and of agent (dashed line) as a function of time. (Bottom) Tracking error as a function of time expressed as the absolute value of the distance between the object and the model. At about 250 time-steps the model had learned to perfectly track the object. (a) At the time indicated by the arrow, the inputs are inverted. The network re-maps in a few hundred steps and the error goes back to zero. (b) Seventy-five per cent of the synapses were randomized at the time indicated by the arrow. The network quickly reorganized and decreased the error. (c) Seventy-five per cent of the synapses are removed from the middle layer. The error grew and although it returned to relatively small values, it was no longer able to perfectly track the object. (d) Seventy-five per cent of the input synapses were removed. The network was unable to cope with the damage and the agent took off in the wrong direction.

outputs, and five as inputs. After learning to correctly connect the inputs with the proper outputs, a single input–output assignment was changed arbitrarily, and the learning process continued. After several such reassignments, the typical learning time for a new connection was as little as five time steps, compared with perhaps 1000 time steps for the initial learning. However, sometimes the process failed, certainly because of destruction of old patterns due to interference with new patterns, and the process might take several hundred time steps. The mechanism is not perfect, and it should not necessarily be so. It would be interesting to study systematically the probability of breakdown versus the number of neurons and their connectivity.

5.4. *Simulated organism*

We now show how the model works on a "closed loop" situation similar to the way sensory-motor integration works in an infant, or in an analogous developmental stage of any organism. An "agent" has been constructed using a two-layer architecture similar to the one discussed above. The agent is

supposed to track a moving target, which makes one or more steps randomly to the left or to the right at each time step. The "sensory" input layer gives the position of the agent relative to the target. The output cells can be thought of as muscles moving the agent to the left or to the right by various amounts. A correct output is one which takes the agent closer to the target, a wrong output, triggering modifications of synaptic strengths, is one that does not. Seven inputs and seven outputs, and 300 units in the middle layer were used. Figure 5 illustrates the initial process of learning and the response to various drastic perturbations to the network.[1] The network is able to discover all by itself the proper map of connections between sensory and motor neurons that ensures a perfect tracking. The samples a–f can be thought of as six different bugs with six different behavioral patterns. The agent can quickly switch behavior when a new bug, or an old bug with a previously identified pattern, arrives at the arena.

6. DISCUSSION

The biological plausibility of our scheme depends on the realization at the neuronal level of two crucial features. (1) Activity propagates through the strongest connections, that is a winner-take-all action. This

can be fulfilled by a local circuit organization, known to exist in all cortices, where feedforward and recurrent excitations compete with lateral and feedback inhibitory connections. The robustness of such operations has been extensively studied on detailed neuronal models.[32] The co-existence of LTD in some of these structures might constitute a verification of our theory.[8,16] Alternatively, a global threshold mechanism keeping the firing rate low would suffice. (2) Depression of synaptic efficacy involves the entire path of firing neurons. A process must exist such that punishment can be relayed long after the neuron has fired, when the response from the outer world to the action is known. We conjecture a mechanism of "tagging" synapses for subsequent LTD, analogous to (but mirroring) recent result for LTP.[24] Frey and Morris recently reported a process where temporary modifications at the synapses in a slice of the hippocampus of a rat *in vitro* resulted from external stimuli.[24] The results suggest that there is a "tagging" mechanism, remaining for a period of approximately 1 h. As a consequence, the synapse which is re-stimulated within this period is the one where LTP is consolidated. We simply suggest that a similar process exists where the synapse is temporarily eligible for LTD.

Propagation of activity-dependent long-distance LTD has been directly demonstrated by Fitzsimonds *et al*.[23] in cultured hippocampal neurons from rat embryos. Fitzsimonds *et al*. suggest that signals travel backward extracellularly for long distances on a much slower time-scale than the forward-propagating signals through the neurons. This is precisely what we need for our model to work! In contrast, traditional back-propagation models do not invoke such long-range LTD.

It is worth commenting on recent experiments on rats subjected to behavioral stress. Xu *et al*.[41] found that when rats were placed in an unfamiliar recording box, stable homosynaptic LTD was induced in the CA1 area, even long after the behavior was experienced. The same stress blocked the induction of LTP—again in precise accordance with our scenario. The biological plausibility of mainstream models of learning and memory are seriously compromised by very recent results[25] indicating regeneration of neurons in the adult primate hippocampus. Our ideas clearly welcome this novel experimental result, since learning in our model is not associated with a rigid permanent structure. New neurons will be automatically incorporated when need be.

In addition to the idea of negative feedback only (learning from mistakes), yet another concept turned out to be essential for the function of the brain. The activity must be kept low, i.e. only a relatively small number of neurons can be associated with a given task. This is important in order to solve the "credit assignment" problem, i.e. to identify the synapses to be punished for bad behavior in a self-organized

way. This is accomplished in the current model by a "winner-take-all" or "extremal" dynamics where only one neuron (or a small proportion) fires at a time. This is also in line with the fact that, in relevant brain areas, many pyramidal neurons only exceptionally fire at high rates. This type of dynamics (which has also been applied to models of evolution) leads to a critical state where very many neurons are near to the threshold, in contrast to all other neural network models where most synaptic connections are deeply suppressed, i.e. the activity is subcritical. At a critical point in physics, the susceptibility is high, which means that it is easy to change the pattern of activity. It is precisely this sensitivity that translates into high adaptability, and also fast convergence.

Although our model is only a caricature of a real brain, all the main ingredients are biologically reasonable and correspond to plausible physiological processes. The model is completely self-organized with no need for external computation of synaptic strengths. Nevertheless, it can outperform other more complex and sophisticated models based on the current paradigm of positive reinforcement. All control mechanisms are local at the postsynaptic site of the active neurons, and information for plasticity is fed back globally to all neurons. This is the first concrete model in which depressing synaptic efficacy leads to learning.

6.1. *A historical perspective*

Historically, processes that were thought of as directed instructional learning have been shown to be caused by selection. The Lamarckean view of evolution as a learning process, where useful acquired features are strengthened, was replaced by the Darwinian view of evolution as a selection process, where the unfit species are weeded out. A similar paradigm shift occurred in immunology, where the clonal selection theory has replaced the traditional instructionist learning template-type models. Ironically, if our thinking turns out to be correct, learning is not a (directed) learning process either, but also an evolutionary selection process in somatic time-scale where incorrect connections are weakened. A similar "selectionist" view for learning was put forward by Jerne[29] and more recently explicitly proposed, at the level of neuronal groups, by Edelman (reviewed in Ref. 22). A superb account of the evolution of these ideas is given by Cziko:[17] "Current research is under way to determine whether unambiguous physical evidence can be found for the overproduction and elimination of newly formed synapses in the adult brain in response to environmental changes. Such a finding would place the brain alongside the immune system as another striking example of how cumulative variation and selection processes during the lifetime of an organism

make it possible to adapt to complex, changing environments."

However, it seems that the current focus in this area is limited to collecting and arguing the evidence of pruning of neuronal synapses during development. As discussed recently,[36] a selectionism view has wider implications for brain function. With the exception of Edelman's work[22] and Dehaene[19] no well-defined mathematical models have been proposed to test the value of these ideas. We go much further by explicitly demonstrating that synaptic depression in conjunction with extremal dynamics is a simple, reliable, fast and robust mechanism of synaptic selection by which an initially structureless group of neurons can self-organize into a plastic structure of connections performing a relevant function in response to changing environmental demands.

Of course, our model is only a toy model, and as such, it leaves many things unspecified and much more work must be done. We leave open the possibility that the two crucial mechanisms proposed; (i) the selection of the strongest synaptic connection and (ii) the synaptic depression of the unfit, could manifest themselves in a variety of forms in biological reality. For instance, it might well be that some sort of long-lasting inhibition and not LTD can be the specific mechanism for weeding out the "unwanted" synapses in some cases. We are also comfortable with the idea that the winner-take-all scenario could be implemented in many different ways.

We feel that we have only scratched the surface of a vast new area of neuronal modeling and understanding. Among the immediate questions which arise one might mention more quantitative measures for the capacity and efficiency of our set-up. Although comparison with traditional artificial neural network model is like comparing apples with oranges (since these latter models cannot even formally be thought of as a part of an autonomous learning system), such comparison might nevertheless be useful for some purposes. Also, perhaps the brain unit constructed here should be seen as a part of a larger hierarchical system, where some parts of the brain act like the environment of other parts, broadcasting negative feedback signals.

So, coming to the end of the millennium, we might argue that selectionism rules essentially all biological processes, ranging from evolution to immunology and brain function, with the new twist that weeding out is the essential point rather than enhancement and unbounded proliferation of the fit.

REFERENCES

1. Albus J. S. (1981) *Brains, Behavior and Robotics.* BYTE Books, McGraw–Hill, Peterborough, NH.
2. Alstrom P. and Stassinopoulos D. (1995) Versatility and adaptive performance. *Phys. Rev. E.* **51**, 5028–5033.
3. Bak P. (1997) *How Nature Works: The Science of Self-Organized Criticality.* Oxford University Press, New York.
4. Bak P. and Sneppen K. (1993) Punctuated equilibrium and criticality in a simple model of evolution. *Phys. Rev. Lett.* **71**, 4083–4087.
5. Bak P., Tang C. and Wiesenfeld K. (1987) Self-organized criticality: an explanation of 1/f noise. *Phys. Rev. Lett.* **59**, 381–385.
6. Bak P., Tang C. and Wiesenfeld K. (1988) Self-organized criticality. *Phys. Rev. A.* **38**, 364–376.
7. Barnes C. A., Baranyi A., Bindman L. J., Dudal Y., Fregnac Y., Ito M., Knopfel T., Lisberger S. G., Morris R. G. M., Moulins M., Movshon J. A., Singer W. and Squirre L. R. (1994) Group report: Relating activity-dependent modifications of neuronal function to changes in neural systems and behavior. In *Cellular and Molecular Mechanisms Underlying Higher Neural Functions* (eds Selverston A. I. and Ascher P.). Wiley, New York.
8. Bartha G. T. and Thompson R. F. (1995) Cerebellum and conditioning. In *The Handbook of Brain Theory and Neural Networks* (ed. Arbib M. A.), pp. 169–172. MIT, Cambridge.
9. Barto A. G. and Jordan M. I. (1987) *Proc. IEEE Int. Conf. on Neural Networks* **2**, 629–636.
10. Barto A. G., Sutton R. S. and Anderson C. W. (1983) Neuron-like adaptive elements that can solve difficult learning control problems. *IEEE Trans. Syst. Man Cybern.* **15**, 835–846.
11. Changeux J. P. (1985) *Neuronal Man: The Biology of Mind.* Oxford University Press, New York.
12. Changeux J. P., Courrege P. and Danchin A. (1973) A theory of the epigenesis of neuronal networks by selective stabilization of synapses. *Proc. natn. Acad. Sci. U.S.A.* **70**, 2974–2978.
13. Changeux J. P. and Danchin A. (1976) Selective stabilization of developing synapses as a mechanism for the specification of neuronal networks. *Nature* **264**, 23–30.
14. Changeux J. P. and Dehaene S. (1989) Neuronal models of cognitive functions. *Cognition* **33**, 63–109.
15. Cooper L. N. (1973) A possible organization of animal memory and learning. *Proceeding of the Nobel Symposium on Collective Properties of Physical Systems* 1973 (eds Lundquist B. and Lundquist S.). Academic, London.
16. Crepel F., Hemart N., Jaillard D. and Daniel H. (1995) Long-term depression in the cerebellum. In *The Handbook of Brain Theory and Neural Networks* (ed. Arbib M. A.), pp. 560–563. MIT, Cambridge.
17. Cziko G. (1995) *Without Miracles. Universal Selection Theory and the Second Darwinian Revolution.* MIT, Cambridge.
18. Dawkins R. (1971) Selective neurone death as a possible memory mechanism. *Nature* **229**, 118–119.

19. Dehaene S., Changeux J. P. and Nadal J. P. (1987) Neural networks that learn temporal sequences by selection. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2727–2731.
20. Dudek S. M. and Bear M. F. (1992) Homosynpatic long-term depression in area CA1 of hippocampus and effects of *N*-methyl-D-aspartate receptor blockade. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4363–4367.
21. Edelman G. M. (1978) Group selection and phasic reentrant signaling: a theory of higher brain function. In *The Mindful Brain* (eds Edelman G. M. and Mountcastle V.), pp. 51–100. MIT, Cambridge.
22. Edelman G. M. (1987) *Neural Darwinisn: The Theory of Neuronal Group Selection*. Basic Books, New York.
23. Fitzsimonds R. M., Song H.-J. and Poo M.-M. (1997) Propagation of activity-dependent synaptic depression in simple neural networks. *Nature* **388**, 439–448.
24. Frey U. and Morris R. G. M. (1997) Synaptic tagging and long-term potentiation. *Nature* **385**, 533–536.
25. Gould E., Tanapat P., McEwen B. S., Flugge G. and Fuchs E. (1998) Proliferation of granule cell precursors in the dentate gyrus of adult monkeys is diminished by stress. *Proc. natn. Acad. Sci. U.S.A.* **95**, 3168–3171.
26. Hebb D. O. (1949) *The Organization of Behavior.* Wiley, New York.
27. Hopfield J. J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2554–2558.
28. James W. (1890) *The Principles of Psychology.* Holt, New York.
29. Jerne N. K. (1967) Antibodies and learning: selection versus instruction. In *The Neurosciences: A Study Program* (eds Quarton G. C., Melenechuk T. and Schmitt F. O.), pp. 200–205. Rockefeller University Press, New York.
30. Johnson M. H. (1997) *Developmental Cognitive Neuroscience: An Introduction*, p. 234. Blackwell, Cambridge, MA.
31. Kirkwood A., Dudek S. M., Gold J. T., Aizenman C. D. and Bear M. F. (1993) Common forms of synaptic plasticity in the hippocampus and neocortex *in vitro. Science* **260**, 1518–1521.
32. Kohonen T. (1996) *Self Organizing Maps,* pp. 132–136. Springer, New York.
33. Mazzoni P., Andersen R. A. and Jordan M. I. (1991) A more biological plausible learning rule for neural networks. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4433–4437.
34. Paczuski M., Maslov S. and Bak P. (1996) Avalanche dynamics in evolution, growth, and depinning models. *Phys. Rev. E.* **53**, 414–443.
35. Reeke G. N. (1994) Selection vs instruction. Use of computer models to compare brain theories. *Int. Rev. Neurobiol.* **37**, 211–242.
36. Sporns O. (1997) Variation and selection in neural function. *Trends Neurosci.* **20** (7), 291–293.
37. Sporns O. and Tononi G. (1994) *Selectionism and the Brain.* Academic, New York.
38. Stassinopoulos D. and Bak P. (1995) Democratic reinforcement. A principle for brain function. *Phys. Rev. E.* **51**, 5033–5037.
39. Toulose G., Dehaene S. and Changeux J. P. (1983) Spin glass model of learning by selection. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1695–1698.
40. Turing A. M. (1957) Computing machines and intelligence. *Mind* **59**, 236–241.
41. Xu L., Anwyl R. and Rowan M. (1997) Behavioural stress facilitates the induction of long-term depression in the hippocampus. *Nature* **387**, 497–500.
42. Young J. Z. (1964) *A Model of the Brain.* Clarendon, Oxford.
43. Young J. Z. (1965) The organization of a memory system. Croonian Lecture of the Royal Society. *Proc. R. Soc.* **163**, 285–320.
44. Young J. Z. (1966) *The Memory System of the Brain*. University of California Press, Berkeley.
45. Yuille A. L. and Geiger D. (1995) Winner-take-all mechanisms. In *The Handbook of Brain Theory and Neural Networks* (ed. Arbib M. A.), pp. 1056–1060. MIT, Cambridge.