International Journal of Modern Physics A, Vol. 15, No. 26 (2000) 4069–4082

MEMORIES AND MEMORY: A PHYSICIST'S APPROACH TO THE BRAIN*

LEON N. COOPER
Thomas J. Watson, Sr. Professor of Science
Director, Brain Science Program and
The Institute for Brain and Neural Systems,
Brown University, Providence, Rhode Island 02912, USA
E-mail: *Leon_Cooper@brown.edu*

Received 26 June 2000

The good materialist fervently believes that everything in our world, from vacuum polarization contributions to the electron's magnetic moment, to chemical reactions, mental states and social behavior — can be constructed from such objects as protons, electrons, quarks, strings or branes: objects that obey those few remarkably concise rules known as the "laws of physics."[a] Why then do not these "laws" explain everything? We might as well ask why they do not explain Hamlet. No small number of stageworks (good, bad or indifferent — as well as a great deal of gibberish) can and have been constructed using sequences of letters, spaces and punctuation made of quarks, branes etc. that happily obey the "laws of physics." Whatever it is that distinguishes Hamlet from, let us say, the Importance of Being Earnest is not easily gleaned from these "laws" since they seem perfectly content with both.

Such an argument is, perhaps, appropriate for theatre; but does it apply to science? I believe it does. Consider Darwin's theory: Evolution dominated by competition for limited resources and natural selection is considered by most (excluding, perhaps, the State of Kansas) to be the major principle governing the development of those species that are here on earth. But though Darwin's evolution is consistent with the "laws of physics," it is not an inevitable consequence. (Consider, for example, an environment with relatively unlimited resources dominated by recurring catastrophic events that randomly destroy most living creatures.)

*This article is based in part on lectures given at the Robert Serber Memorial April 30, 1998 and the CN Yang Symposium May 21, 1999.

[a]Whether or not this is so remains unanswered; but surely it is the working hypothesis of most scientists.

Explanation for complex systems, from stage works, human behavior, the evolution of the species, to the properties of systems of neurons, requires choosing among those many objects or rules consistent with the laws of the physics, to distinguish what is from what might be.[b]

In this article, I would like to describe the path I took in attempting to understand important aspects of one complex biological system — the human brain.[c]

The first question one asks one's self before embarking on such a hazardous intellectual journey, surely, is: Can it be done? The second, perhaps, is: Can it be done in our lifetime (being too far ahead of one's time is not wise either in science or in business) a third, reasonably, is: Can I contribute?

In the early 70's it was generally believed that the special properties of nervous systems — in particular the brain — are the results of interactions between neurons — specialized cells that transmit information chemically and electrically[d] and are linked to one another in networks of incredible complexity. Although the properties of individual neurons as well as the connection, for example, between neuron and muscle activity were more or less understood, higher level "mental" activity such as the means and place of memory storage seemed completely mysterious.

Here then was a possible opportunity. Although the brain is certainly a very complex piece of biological machinery, the functions it performs seemed to require organizing principles whose elucidation might require the talents of, and be amenable to analysis by, a theorist.

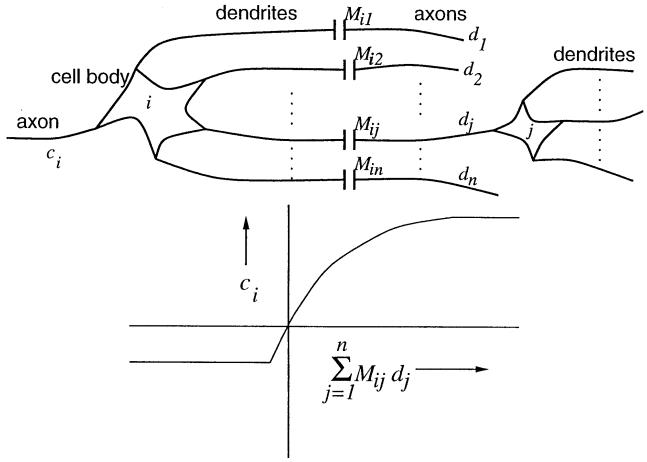
My first effort in this area was to attempt to construct a network of neurons that would display some of the qualitative features associated with what I called animal memory.[e] Initial success was the seduction that lured me further and further into this exotic domain. It is this journey I will describe.

[b]One could take the point of view, at least in classical theory, that the "laws of physics" with a complete specification of initial conditions have as a consequence everything that is (shades of Laplace). But then, of course, we have to distinguish the initial conditions that lead to what is — opposed to others that lead to what is not. Someday, I suppose, a theory of everything may give us a unique set of laws and initial conditions so (ignoring such problems as delicate dependence on initial conditions or the seeming inherent lack of absolute predictability of the future in the quantum theory) in a sense — explain everything. But even in situations totally within the domain of physics, we want something more illuminating than "Superconductivity is a consequence of the Schrodinger equation with 10^{23} electrons and ions subject to Coulomb's law."

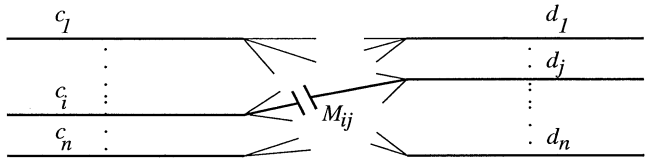
[c]There might, as some believe, be laws governing all complex systems; but it is my view that the particularity of each system — what distinguishes one system from another dominates the discourse. We might ask if rules for complex systems always exist? Surely for some systems whatever rules there are result in a situation that approaches randomness. What should surprise us is not that some systems seem beyond analysis but rather that so many are remarkably regular and that analysis is possible.

[d]Argument over the electrical properties of neurons goes back to Galvani and Volta.

[e]A more detailed account of the events and people involved is given in "How We Learn, How We Remember: Toward an Understanding of Brain and Neural Systems" (World Scientific, Singapore, 1995).



(a)



(b)

Fig. 1. (a) Neurons are very complex cells. Their information processing properties, in a much simplified form, can be summarized by the following input-output relations: $c_i = \sigma\left(\sum_{j=1}^{n} M_{ij} d_j\right) \approx \sum_{j=1}^{n} M_{ij} d_j$ (in the linear region) where d_j is the input from the jth incoming cell, M is the matrix of "ideal" synaptic junctions and c_i is the output of the ith cell. (b) A simple network of neurons. M_{ij} is the "ideal" synapse connecting the jth input cell with the ith output cell.

Neurons are very complex cells. The result of large number of simplifications[f] leads to a relation between neuron inputs and output as in Fig. 1(a); a simple network of such neurons might look as shown in Fig. 1(b).

Suppose that an "external auditory event," such as a particular tone, is eventually mapped into a n-tuple of signals (presumably in auditory cortex).

$$d = (d_1 \cdots d_n) \longleftarrow \text{tone}\,.$$

Suppose, further, that in the experience of the animal, this tone is accompanied by or followed by a visual event — the sight of food (presumably in some region of visual cortex) that triggers salivation.

$$\text{salivation} \longleftarrow c = (c_1 \cdots c_n) \longleftarrow \text{sight of food}\,.$$

[f]It should be understood that every symbol represents a tremendous simplification. For example, c is related to an integrated potential at the axon hillock of the outgoing cell that determines the average firing rate. M_{ij} is an "ideal" synapse — a logical summary of many actual synapses.

It is known that with a sufficient number of such experiences the animal comes to "associate" the tone with the sight of food and eventually begins to salivate upon hearing the tone

$$\text{salivation} \longleftarrow c \leftarrow \text{-----} \; d \longleftarrow \text{tone}\,.$$

We might have n such associations summarized by the mapping

$$c^1 \leftarrow \text{-----} \; d^1\,,$$

$$c^k \leftarrow \text{-----} \; d^k\,,$$

$$c^n \leftarrow \text{-----} \; d^n\,.$$

If we assume for extreme simplicity, that the mapped input signals $d^1 \cdots d^n$ are orthogonal to one another (presumably meaning that the events they represent are easily separated from one another) we can write the mapping in a form transparent to all physicists

$$M = \sum_{k=1}^{n} |c^k\rangle\langle d^k|$$

so that

$$|c^l\rangle = M|d^l\rangle\,.$$

Could such "associative" and "content addressable" mappings be constructed in simple neural networks? If we look more closely at the above projection operator and ask what is happening biologically, what is happening at a particular synaptic junction, we see that the required strength of the ijth synapse, M_{ij}, is

$$M_{ij} = \sum_{k=1}^{n} c_i^k d_j^k$$

which is a sum over all associations $(1 \cdots k \cdots n)$ of the product of the output of the ith neuron and the input from the jth neuron (See Fig. 2(a)).

Such a synaptic strength could result if synaptic modification (or learning) follows the famous Hebbian rule:

$$\Delta M_{ij}^k \sim c_i^k d_j^k\,.$$

This requires that information of the summed post-synaptic potential be propagated back from the cell body to individual synapses. (See Fig. 2(b)). (It is immediately clear that Hebbian learning can be only part of the story since synapses would grow in strength without bound. Thus one early question was: How could such learning be stabilized?)

Although such conjectures seemed attractive, in the early 1970's there was little if any evidence for synaptic modification of any kind. The appearance of vaporware was difficult to avoid. Many ideas, some possibly interesting, but no real connection with the world in which we happen to live — a limited contribution to a field, if

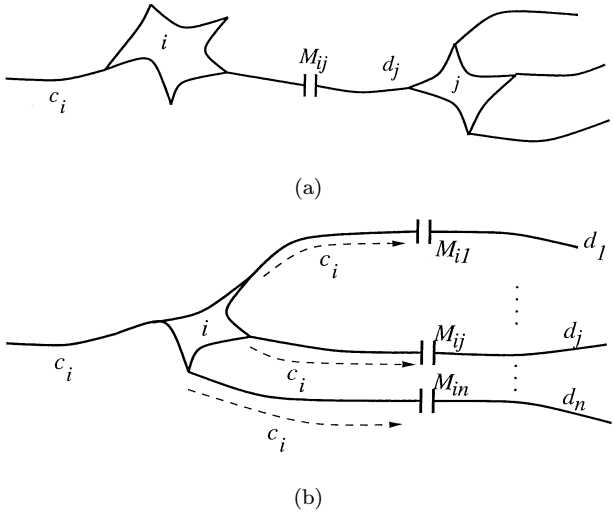


Fig. 2. (a) The axon of the jth incoming neuron is connected to the dendrites of the ith outgoing neuron by many actual synapses. Their net effect is given by the strength of the single "ideal" synapse, M_{ij}. (b) In order for the information required for Hebbian synaptic modification to be available locally at each cell's synapses, $M_{i1}, \ldots, M_{in}$, the integrated potential, c_i, must be propagated backwards (in a direction opposite to the usual information flow) from the cell body to each of the synapses. "Back spiking" that could carry this information has recently been observed, and associated with changes in synaptic strength.

one could call it a field, that was and is plagued with excessive mathematical and philosophical wheel spinning.

It seemed essential to me that theory be made sufficiently concrete so that it could be confronted by experimental results. At the time theory was a somewhat novel idea for biologists: plausibly so since, for the most part, that specialty, the spinning out of consequences of ideas in long and complex arguments, while accepted (although occasionally the subject of some ridicule) in the community of physicists, had not really been required in biology. There, the connection between idea and experiment was straightforward enough so that every self-respecting experimentalist insisted on doing it himself. This skepticism was also justified, in my opinion, since with a few striking exceptions, many previous so-called theoretical attempts were totally removed from reality. Physicists, in particular, displayed an arrogance in talking to biologists that was not designed to inspire friendly relations. I recall, a presentation at a conference in which an eminent physicist said, in effect, "*Here is the Schrodinger equation, here we have* 10^{23} *electrons and ions subject to electrical forces. One of the consequences is life.*"

We thus began a series of attempts, to make specific connections between fundamental ideas of synaptic modification and testable experiments in actual animals, to produce a theoretical structure concrete enough so that one would know precisely what the assumptions were, and so that one could see one's way through the arguments and know exactly which conclusions followed from which assumptions.

The primary object is not to be right, (although that certainly is one of the hopes) it is to be crystal clear so, to paraphrase Galileo, "*One knows what follows from what one has said before.*" His teachers of mathematics taught him this method.

We chose visual cortex because of the large number of experiments that had been done in that region of the brain. At the time, in addition to the work of Hubel and Wiesel, there was a large body of (very controversial) experimental work suggesting that the response properties of cells in visual cortex depended on the visual experience of the animal. This indicated to us that one might be observing experience-dependent cellular changes, analysis of which could reveal the systematics of synaptic modification. The collaboration between experimentalists and theorists that resulted has continued for more than twenty years.

The primary questions at that time seemed to be: Can we find any evidence for synaptic modification? If so, what is its form? Further, what is the cellular and molecular basis — thus the cellular and molecular basis for learning and memory storage?

One path toward making contact with experiment began with the observation that in most situations, the mapped inputs $d^1 \cdots d^n$ are not orthogonal so that there would be confusion among the outputs. To better separate incoming events (if they do not map into n-tuples orthogonal to one another) we can try to form selective neurons (those that respond more to the mapping of one external event than another).

$$M^s = \sum_{k=1}^{n} |c^k\rangle\langle d^{sk}| \,,$$

where

$$\langle d^{sk}|d^l\rangle = \delta_{kl}$$

so that

$$|c^l\rangle \sim M^s|d^l\rangle \,.$$

Such selectivity is relatively common in the nervous system. Hubel and Wiesel observed edge detectors in area 17 (V1) of visual cortex of kittens. Rolls and colleagues observed face detectors in inferotemporal cortex of monkeys. (Humans with lesions in inferotemporal cortex suffer from a rare disorder called prospagnosia — a severe disturbance in the ability to recognize familiar faces). Further the development of such selectivity in visual cortex had been shown to depend on the animal's visual experience.

Hebbian modification can be written

$$\Delta M_{ij} \sim \phi_i^{\text{Hebb}} d_j \,,$$

where

$$\phi_i^{\text{Hebb}} = c_i \,.$$

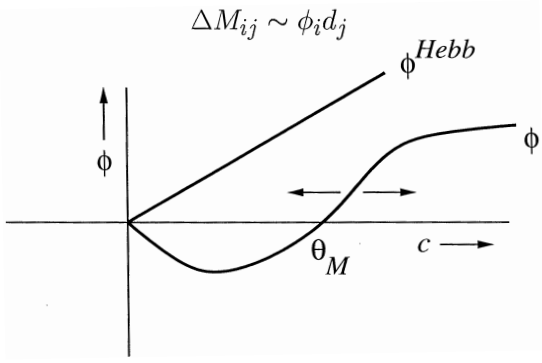


Fig. 3. BCM Synaptic Modification. $\Delta M_{ij} \sim \phi_i d_j$. The BCM synaptic modification function, ϕ, is contrasted with Hebbian modification function, ϕ^{Hebb}. (The index referring to the cell i is suppressed for clarity). For active synapses ($d_j > 0$) when $c < \theta_M$, M_{ij} is decreased, when $c > \theta_M$, M_{ij} is increased. Allowing the crossover, θ_M, to vary with cell activity as, for example, $\theta_M \sim E[c^2] = \frac{1}{\tau}\int_{-\infty}^{t} c^2(t')e^{-(t-t')/\tau}dt'$ stabilizes the system and gives agreement with experiment. Among the important results are: Fixed points depend on the environment. Only selective fixed points are stable.

But such synaptic modification yields no selectivity and needs stabilization. To construct M^s by a learning process that proceeds without instruction on the presentation of nonorthogonal inputs, we introduced a form of synaptic modification that combined Hebbian with anti-Hebbian modification. Then, allowing the crossover point between these to vary rapidly enough with cell activity led to stabilization. The resulting form has become known as BCM synaptic modification (Bienenstock *et al.* 1982). See Fig. 3.

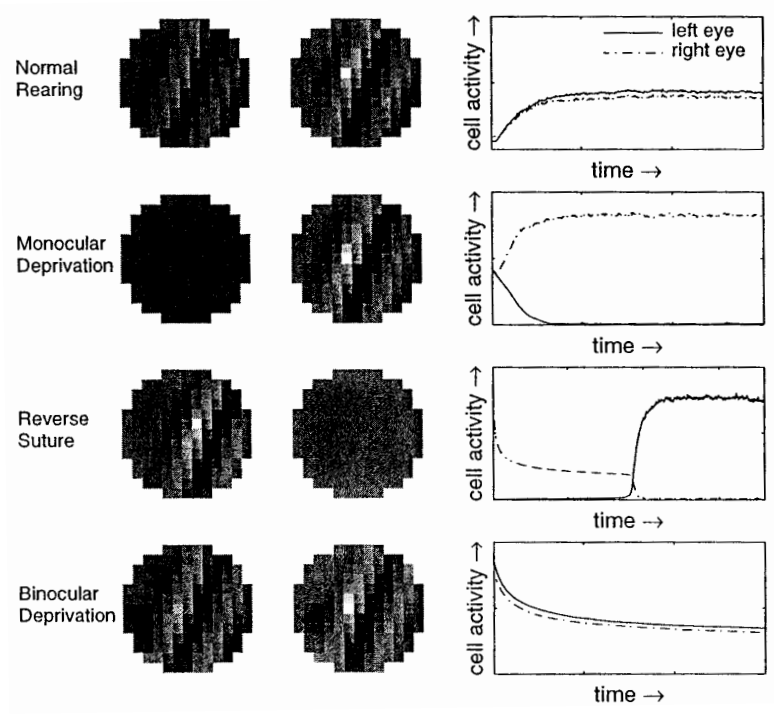
In general one can test and/or distinguish between theories by comparing predicted consequences of theory with experiment, or by attempting, more or less directly, to experimentally verify the underlying assumptions.[g]

Consequences of BCM synaptic modification have been shown by analysis and in many simulations to be in agreement with experimental observations on the receptive field properties of visual cortical cells for animals reared in normal and various deprived environments. (See Fig. 4(a)).

These might be characterized as post-dictions since the deprived rearing experiments were known for the most part (with the exception of the reverse suture result) before the BCM theory was constructed. Although there is no logical distinction between post-diction and prediction (all theorems must be in agreement with observation) it is psychologically satisfying to have a clear prediction that is confirmed by experiment. One such prediction, that has recently been experimentally tested, is described below.

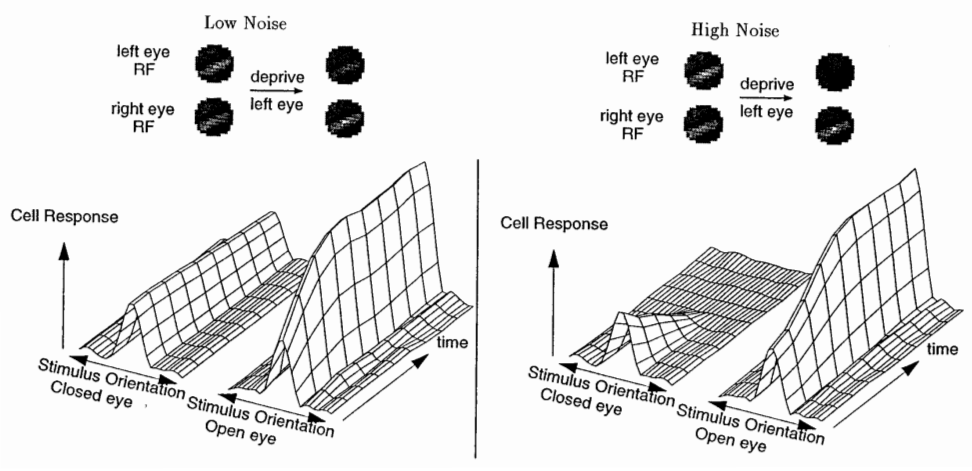
[g]These are connected, of course, by the logical structure. For Newtonian theory, consequences would include the planetary elliptical orbits; an underlying postulate is the inverse square gravitational force.

A dramatic example of experience-dependent receptive field plasticity is the shift in ocular dominance of visual cortical cells that results from briefly depriving one eye of vision. Such deprivation leads to a very rapid disconnection of the deprived eye from cells in visual cortex so that stimulation through the deprived eye no



(a)

Fig. 4. (a) Simulations of the Development of Cortical Receptive Fields using BCM Synaptic Modification. Left: Final receptive fields and synaptic weight configurations. Each pixel represents a point in space over the retina, where white and black correspond to strong and weak synaptic strengths, respectively, from that retinal input. Right: Maximum response to oriented stimuli, as a function of time. Simulations from top to bottom are as follows. Normal Rearing: both eyes presented with natural scenes. Monocular Deprivation: following normal rearing, the left eye is presented with noise and the right with natural scenes. Reverse Suture: following monocular deprivation, the eye presented with noise is now presented with natural scenes, and the other eye with noise. Binocular Deprivation: following normal rearing, both eyes are presented with noise. It is important to note that if the binocular deprivation simulation is run long enough, selectivity will be lost. These simulation results are in agreement with experimental observations. (b) The Effect of Deprived Eye activity on the Disconnection of the Closed Eye in Monocular Deprivation with BCM synaptic modification. Shown are the results of monocular deprivation starting from the binocular state. Left and right receptive fields (above), before and after depriving the left eye. Each pixel represents a point in space over the retina, where white and black correspond to strong and weak synaptic strengths, respectively, from that retinal input. The responses of the cell to oriented sine gratings (lower plots), as a function of time during deprivation in a low noise environment (lower left) and a high noise environment (lower right). Note that with higher noise the disconnection is more rapid (see equation on page 9). This is in agreement with the experimental results of Rittenhouse *et al.* (Nature, 1999).



(b)

Fig. 4 (*continued*)

longer drives the cortical cells. Analysis of the BCM modification equations yields a time dependence of the strength of synapses from the closed eye:

$$\ln\left[\frac{M^{\text{closed}}(t)}{M^{\text{closed}}(0)}\right] \sim -\overline{n^2}t\,,$$

where $\overline{n^2}$ is the variance of the activity or noise from the closed eye. As the noise increases, so will the disconnection rate of the synapses from the closed eye. Thus, one has the nonintuitive result that higher levels of noise lead to a faster disconnection of the closed eye. This an experimentally testable prediction of the BCM theory, and one that distinguishes this theory from others.[h] (See Fig. 4(b)). A recently completed experiment, inspired by the above analysis, (Rittenhous *et al.*) is consistent with the BCM prediction that deprived eye connections in the visual cortex are more rapidly weakened with increasing noise level.

Experiments to test the underlying postulates of the BCM theory have been performed at Brown in the laboratory of Mark Bear as well as elsewhere. The central assumptions of BCM synaptic modification are that (1) active synapses are bidirectionally modifiable, (2) the sign and magnitude of the modification depends on the integrated level of postsynaptic response and (3) the synaptic depression-potentiation crossover point (θ_m) varies as a function of the history of postsynaptic cellular activity. (See Fig. 3). Much experimental work over the past decade has been devoted to determining if these assumptions are valid at excitatory glutamatergic synapses in the cerebral cortex.

[h]Typically, competitive models of cortical plasticity, such as variants of the Hebb rule with weight decay, predict that the effects of monocular deprivation occur at a rate that is inversely proportional to the level of activity from the deprived eye. Accordingly, as activity from the deprived eye decreases, the rate and severity of the synaptic disconnection in the visual cortex would increase.

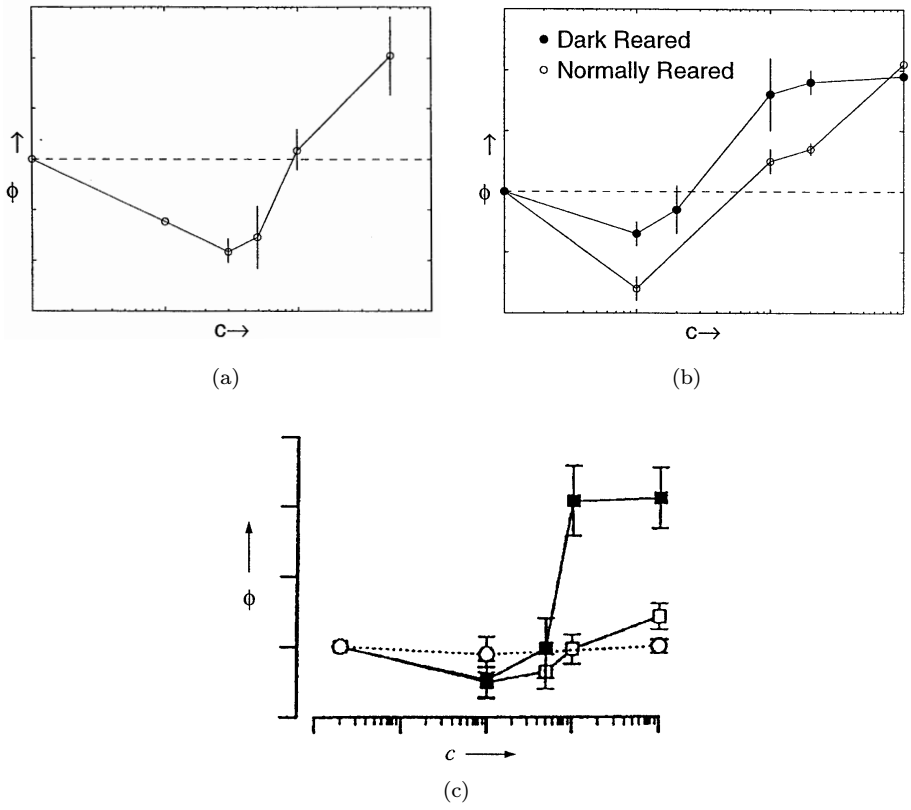


Fig. 5. (a) The synaptic modification function, ϕ, as constructed from the experimental results of Dudek and Bear (1992) in rat hippocampus. These results have been replicated in many parts of the brain, in young and old animals, and in many species — including humans. (b) The movement of the modification threshold, θ_M, as constructed from the experimental results of Kirkwood *et al.* (1996) on dark reared rats (filled circles) and those with normal visual experience (open circles). This activity dependent shift is consistent with the BCM postulate of the moving modification threshold, θ_M. (c) The movement of the modification threshold, θ_M, as constructed from the experimental results of Tang *et al.* (1999) using genetically engineered mice and those of Quinlan *et al.* (1999) linking cellular activity with the ratio of two distinct subunits of the NMDA receptor.

The basic shape of the BCM synaptic modification function was first confirmed by Dudek and Bear (1992) [See Fig. (5a)] in a region of the brain called hippocampus. Kirkwood *et al.* (1993) showed that the result was the same in visual cortex. Since then these findings have been confirmed in many different regions of neocortex in many species in both young and old animals. Of particular interest are recent data showing that the same principles of synaptic plasticity apply in the human inferotemporal cortex, a region believed to be a repository of visual memories (Chen *et al.*, 1995). Together, the data support the idea that very similar principles guide synaptic plasticity in many species in widely different regions of the brain.

To stabilize the system, the modification threshold, θ_m, must vary according to the history of post-synaptic cortical activity. An experimental test of this hypothesis

was first reported by Kirkwood, Rioult and Bear (1996). They compared the synaptic modification function in the visual cortex of normal animals with that in the visual cortex of animals reared in complete darkness and found a shift of this function in accordance with the theoretical postulate. (Fig. 5(b)). A very different experiment (Mayford *et al.*) using genetically altered mice has been interpreted as confirming this result. Also, recently published results of Tang *et al.* (1999) again using genetically altered mice (a different alteration) taken together with results of Quinlan *et al.* (to be published) that link cellular activity with the ratio of two distinct sub-units of the cortical NMDA receptor support the idea that θ_m is set according to the activation history of the cell. (See Fig. 5(c)).

On the basis of results, such as those sketched above, we conclude, with a certain optimism, that comparison of the consequences of BCM synaptic modification with experiment is satisfactory and further that the postulates that underly the theory are consistent with experiment. It has become accepted that the synapses modify in a manner described, at least in a first approximation, in Fig. 3.

Much effort is now devoted to investigations of the cellular and molecular mechanisms that underly these synaptic changes. Extensive experimental work has revealed the dependence of synaptic modification on the influx of calcium into cells when accompanied by activation of the synapses. (This is presumed to be the cellular event that correlates active synapses with the integrated post-synaptic response.) It is generally believed that various receptors in the post-synaptic membrane play critical roles and that the modification of some of them through such processes as phosphorylation or changes of subunit ratios are responsible for the learning curve and synaptic changes. Some of these, therefore, are thought to provide part of the molecular subtrate for memory storage. (See Fig. 6). The detailed molecular and genetic processes that are responsible for these changes as well as those that control the transfer of short or intermediate memories into long-term storage is now a subject of the most intense interest.

Given the level of skepticism displayed when ideas such as synaptic modification were discussed twenty-five years ago, I think that it is reasonable to say that we have made more progress than is generally appreciated. Our initial aim to build a theoretical structure relevant to a fundamental brain process that was sufficiently concrete so that it could be tested by experiment has been accomplished. It is particularly gratifying that theory has inspired experiments that, in addition to confirming the various postulates and predictions, have served to allow us to refine the hypothesis on which the theory was founded: to make sensible the introduction of complications to the originally very simplified assumptions that, hopefully, will lead more and more realistic descriptions of the processes of learning and memory storage.

Permit me then to project into the future with a certain optimism. Assume that the fundamental mechanisms are understood — that the neural (cellular and molecular) correlates of mental states (as Francis Crick is seeking) are known. Do

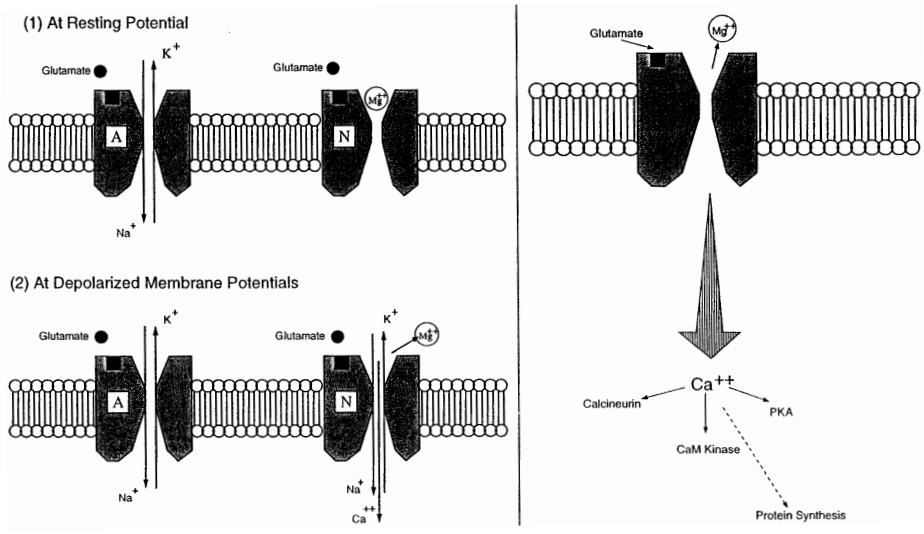


Fig. 6. (a) Postsynaptic Response to Glutamate. Among the receptors on the post-synaptic membrane that are activated by the neurotransmitter glutamate are two of particular importance: the AMPA receptor (A) and the NMDA receptor (N). The AMPA receptor responds to glutamate by opening its channel and allowing sodium and potassium ions to pass through. Even in the presence of glutamate, the NMDA receptor is blocked by Mg^{++}; its channel is opened only with a substantial depolarization of the postsynaptic membrane. Thus a coincidence of glutamate and sufficient depolarization are required to open the NMDA channel. When this channel is opened, it allows Ca^{++} to flow into the postsynaptic dendrite. (b) Modulation of the NMDA Receptor Effectiveness. The Ca^{++} that enters the cell is thought to initiate a sequence of molecular events that (depending on the amount of calcium that enters) result in an increase or decrease of synaptic strength. (This change in synaptic strength has been associated with the phosphorylation or dephosphorylation of sites on the AMPA receptor.) It has recently been shown that alterations in the ratios of subunits that make up the NMDA receptor change its response properties in a manner consistent with the BCM moving threshold postulate.

we then understand the human mind and its presumed origin in that biological organ, the brain?

Although much emphasis has been placed on algorithmic behavior or what are very loosely referred to as neural computational processes, these properties, in my opinion, will probably be easier to understand than mental states such as feeling, awareness or consciousness. It has been suggested by some that these latter properties will arise as a consequence of the execution of the proper algorithmic behavior, but my opinion is that algorithmic behavior and consciousness are relatively independent. After all, hand-held calculators execute algorithms and little dogs wagging their tails do not do much arithmetic.

I would phrase the question as follows: What are the steps required so that a machine could experience mental activity? Can we, in the extreme, construct a machine that is conscious? What we must understand is how mental states, such as consciousness, arise as a properties of a very complex physical system. This, in my opinion, is the profoundest mystery surrounding that biological entity: brain.

The question is sufficiently difficult so that we have been subjected (as is often the case) to various evasions — Cartesian dualisms: variations of homunculus proposals; Denial of the phenomena: consciousness and feeling are "epiphenomena" (whatever those are) and do not have to be explained or — in the extreme — do not exist; solutions of one mystery by invoking another: consciousness arises in the quantum measurement process or where gravity meets quantum theory; refusal to confront the issue: consciousness arises "somehow" when a machine executes the proper algorithmic processes; retreat under the cover of positivist philosophy: how would we know if a machine were conscious ... and so on.

We have heard such arguments before. They seem to be typical responses to the frustration of failure in attacking really difficult scientific problems. First try and fail. Follow this by proving a solution is impossible or irrelevant. Toy with the notion that a new law of nature is involved. Then, when the solution is found, complain that it is really trivial or (even better) that it was suggested in some obscure comment once made in a paper you published a long time ago.

On a personal note, we experienced some of this in the course of developing a theory of superconductivity — also a complex and subtle consequence of an interacting many component system, in this case the quantum mechanics of electrons in a metal. After the fact one, rather well known, physicist expressed his disappointment that "such a striking phenomenon as superconductivity [was] ... nothing more exciting than a footling small interaction between electrons and lattice vibrations." — thus missing the point in operatic style.

The scientific problem, as I see it, is to construct from material components such as neurons and/or systems of neurons (thus from quarks, branes, etc.) the simplest entity that performs the most primitive conscious act. (To paraphrase Bourbaki, a beautiful problem — it can be stated very simply and will, no doubt, have a very complex solution.)

The simplicity of this statement has been obscured by the unwillingness on the part of some to accept the possibility that such a construction is possible (assigning a special status to the mental, while not being willing to divorce mental completely from physical) resulting in enormous complications that would be unnecessary if the construction were possible.

It has also been obscured by an excess of positivism — a fear of and/or aversion to mentalism or machine-like constructions of the internal workings of the mind that cannot be directly verified by experience. This, to my mind is, totally contrary to the nature and purpose of scientific thinking.

Successful science has given us just such machines (actual or conceptual) that work behind events in the world. The greatest include Newton's laws, Maxwell's equations and Schrodinger's equation. Such entities as molecules and/or atoms were assumed to exist (an assumption that was vigorously contested in the 19th century with positivistic-type arguments) long before they were "seen." It is not necessarily the case that every element of the "behind the scenes" machinery can be directly observed. In quantum mechanics, for example, the wave function is not

directly observable. The consequences of this sometimes invisible machinery can, however, be put into correspondence with experience. The essence of the positivist argument (as actually employed by Einstein and Heisenberg) is not that we may not introduce entities that are not directly observable but, rather, that if an entity is not observable (e.g. absolute time in special relativity or simultaneous position and momentum in the quantum theory) it *need* not appear in theory.

Thus a theory of mind not only is allowed, but, in my opinion, requires the introduction of mental entities. We will be satisfied only when we see before us constructs that can experience mental states such as consciousness, when we see how they work, how they come about from more primitive entities such as neurons.

It is possible, in my opinion, that there is no sharp demarcation between the categories conscious and nonconscious (Just as we would say today that no sharp distinction exists between the categories living and nonliving.) It could even turn out (shocking as this might appear) that we *must* invoke a new "law of nature:" follow Descartes and pour the conscious substance into the machine.

But the conservative scientific position is to attempt to construct this seemingly new and surely very subtle property from the materials available — those given to us by physicists, chemists and biologists (as has been done many times before: celestial from earthy material, organic from inorganic substances, living creatures, from lifeless atoms, the concept of temperature from the motion of molecules, or light from electricity and magnetism.) This unrepentant materialist believes that the construction can and will be made; further — it is important to emphasize this — that it will in no way diminish the value or significance of what has been constructed. To paraphrase Santayana,

"All our sorrow is real, but the atoms of which we are made are indifferent."

The great problem, therefore, is to construct real sorrow from hypothetical indifferent atoms. Success would no doubt be magnificent but failure might be more so. If we cannot make the construction, then we will genuinely have made one of the profoundest discoveries in the history of thought — consequences of which would shape our conception of ourselves in the deepest way.

References

1. M. F. Bear and A. Kirkwood, *Current Opinion in Neurobiology* **3**, 197 (1993).
2. E. L. Bienenstock, L. N. Cooper and P. W. Munro, *J. Neuroscience* **2**, 32 (1982).
3. S. M. Dudek and M. F. Bear, *Proc. Natl. Acad. Sci.* **89**, 4363 (1992).
4. A. Kirkwood, M. G. Rioult and M. F. Bear, *Nature* **381**, 526 (1996).
5. M. Mayford, J. Wang, E. Kandel and T. O'Dell, *Cell* **81**, 1 (1995).
6. E. M. Quinlan, D. H. Olstein and M. F. Bear, "Bidirectional, experience-dependent regulation of NMDA receptor subunit composition in rat visual cortex during postnatal development," *Proc. Natl. Acad. Sci.* **96**, 12876 (1999).
7. C. D. Rittenhouse, H. Z. Shouval, M. A. Paradiso and M. F. Bear, *Nature* **397**, 347 (1999).
8. Y.-P. Tang, E. Shimizu, G. R. Dube, C. Rampon, G. A. Kerchner, M. Zhuo, G. Liu and J. Z. Tsien, *Nature* **401**, 63 (1999).