# On the computational architecture of the neocortex

## I. The role of the thalamo-cortical loop

D. Mumford

Mathematics Department, Harvard University, 1 Oxford Street, Cambridge, MA 02138, USA

**Abstract.** This paper proposes that each area of the cortex carries on its calculations with the active participation of a nucleus in the thalamus with which it is reciprocally and topographically connected. Each cortical area is responsible for maintaining and updating the organism's knowledge of a specific aspect of the world, ranging from low level raw data to high level abstract representations, and involving interpreting stimuli and generating actions. In doing this, it will draw on multiple sources of expertise, learned from experience, creating multiple, often conflicting, hypotheses which are integrated by the action of the thalamic neurons and then sent back to the standard input layer of the cortex. Thus this nucleus plays the role of an 'active blackboard' on which the current best reconstruction of some aspect of the world is always displayed. Evidence for this theory is reviewed and experimental tests are proposed. A sequel to this paper will discuss the cortico-cortical loops and propose quite different computational roles for them.

## 1 Introduction

When one attempts to theorize about the functioning of the human brain, the task seems nearly impossible because of the extraordinary complexity of the brain. Anatomically, the brain is divided into hundreds, even thousands of areas and nuclei and structures, connected in a totally bewildering way, while physiologically more and more chemicals, exchanged by neurons, are being discovered which cause more and more types of effects of one neuron on another. Likewise, if you start by analyzing the computational task that the brain faces, using the tools of either pattern recognition, artificial intelligence or control theory, only the most simplified versions of these tasks seem tractable. In the setting of the real world, sensory data and motor requirements are so complex and need to be robust in the fact of so much noise, so many unforeseen disturbances, as to be totally beyond present techniques.

The idea of the present two-part paper is that, at the appropriate level of analysis, there are certain uniformities in the structure of the brain that suggest that some simple general principles of organization must be at work. If this is the case, looking at the tasks performed by the brain from a computational perspective, it may be possible to link the structures observed in the brain with elements in a theoretical analysis of what is needed to perform these tasks. The first part of this paper will put forward a proposal for the role of the thalamus based specifically on the existence of pathways between the cortex and the thalamus which are at least roughly topographic, i.e. preserving the two-dimensional layout of the cortical sheet, and inverse to each other. The second part of the paper will make proposals for the computational significance of the reciprocal cortico-cortical pathways and its relation to the pyramidal cell populations in different layers of the cortex and the fast oscillations recently observed in the cortex.

The uniformity which I have in mind is the uniformity of the neocortex of mammals. In essentially all species of mammal, including the very primitive opossum, the neocortex has an extremely similar structure throughout: it has six layers, a small number of cell types, one of which, the pyramidal cell, accounts for over half of all cells and a standard pattern of connectivity, locally, globally within the cortex and subcortically. Phylogenetically, this structure has not been changed in the evolution of mammals (except for being simplified in some orders). The major expansion of 'association' cortex in the primate order has not involved revision of this basic plan, but, apparently, simply replicating it over ever larger areas. This suggests very strongly that this structure embodies a basic computational module so versatile that it can be hooked together in ever larger configurations and still function, with ever increasing subtlety, to both analyze sensory input and organize motor actions. Even in producing the most remarkable achievement of the brain – language – the areas of the brain involved have used the identical structure. The fact that this structure

is present in much simpler animals, moreover, suggests that it may not be that hard to understand and that its mode of operation may be fully revealed in quite simple tasks. On the other hand, these structures are much more specific, with their own characteristic architecture, than those normally studied from a theoretical perspective under the name 'neural nets'. This suggests the possibility of studying the architecture embodied in these specific structures, looking for their computational significance.

Since this paper deals with a proposal for applying computational ideas to biological structures, we have tried to present the ideas so as to be clear both to computer scientists and biologists. In order to do this, it has been necessary to include a considerable amount of basic anatomy for the sake of the former and of basic computer science for the sake of the latter. I found with preliminary versions of the paper that when this background was left out, there were frequently misunderstandings and confusions and for this reason, I feel it is essential to develop my ideas at this length. I want to thank Francis Crick, Terry Deacon, Stephen Kosslyn, Adam Mamelak, Ken Nakayama and Steve Zucker for critical readings of various drafts of this paper that helped me immensely in refining and clarifying my ideas. In particular, while preparing this paper, I learned of the work of Erich Harth (1983), who made proposals in a similar direction for the role of the thalamus.

## 2 The connections of the cortex and the thalamus: a review

In order to put our proposals for the role of the thalamus in perspective, I need to lay out the basic facts about the structure of the cortex and its connections with the thalamus. Everything in this section is standard neuroanatomy, but it is included here so that readers with other backgrounds can follow our ideas.

The neocortex has an area of about 200,000 sq.mm. in humans, a thickness of 2–3 mm., and a neuron density around 100,000/sq.mm.[1] Over half of these cells are so-called pyramidal cells, characterized by the fact that at least one branch of their axons projects to distant, e.g. a centimeter or more, targets. The neocortex has a uniform structure with 6 layers, characterized by their cell populations, which I will discuss in greater detail below. There are local variations in the thickness and prominence of the layers,[2] but in general the same structure is there. (Reproduced in Fig. 1 is a composite photo of the six layers in three different stains.) The
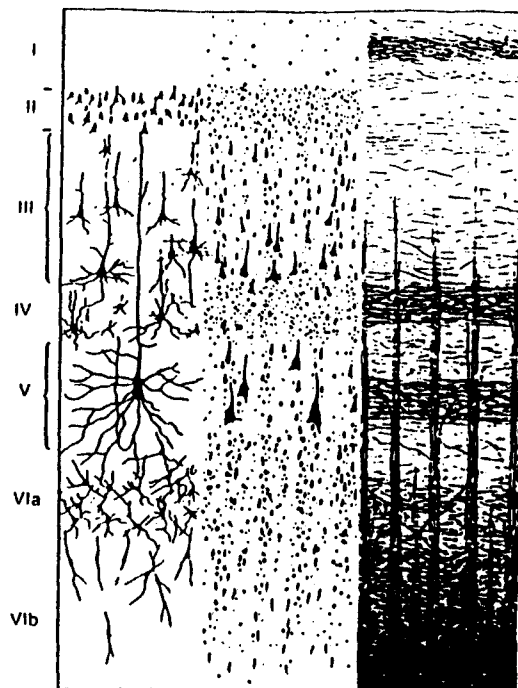


Fig. 1. The 6 cortical layers in Golgi, Nissl and myelin stains (from Brodal 1981)

cortex as a whole includes two more primitive parts, the paleocortex and the archicortex, with 4% of the total cortical area in man (Blinkov and Glazer 1968, p. 381). This will play essentially no role in this paper so that I usually refer to the neocortex simply as cortex. These more primitive parts have a similar but simplified pattern with a less elaborate pattern of layers.

The cortex of every mammal seems to be divided into areas, each with a specialized role. The original identification of these areas was based on tiny differences of cell types and cell distributions and led to maps of cortical areas due to Brodmann and others. There are, of course, species differences, so that the primitive mammals have relatively few areas and primates more, but the general map and often many of its details are closely homologous for all species studied. More recently, the possibility of tracing pathways in cortex very accurately using chemicals which move both forward and backward through axons has modified and frequently subdivided Brodmann's areas (e.g. the third visual area, Brodmann's area 19, turned out to be composed of more than one area), but the picture of a map-like division of the surface of the brain into independent computational modules with specific interconnections has been repeatedly confirmed. For a very recent and comprehensive review of the areas known in the Macaque monkey, see (Fellemann and Van Essen 1991).

All input to the cortex, except for the olfactory sense, comes to it via the thalamus, which sits at the top of the brain stem in two parts, one in the middle of each cerebral hemisphere. It is not easy to expose because it is totally surrounded by the white matter of

---

[1] There is a wide discrepancy in estimates ranging from 12,000/sq.mm. to 160,000/sq.mm., but 100,000/sq.mm., resulting in 20 billion cortical cells, seems to be a frequently cited figure e.g. Rockel et al. (1980), after allowing for 18% shrinkage in each dimension, or Cherniak (1990). Note that the primary visual area is an exception with at least twice as many cells per sq.mm

[2] e.g. the elaboration of layer IV in primary sensory areas and its sparseness in the primary motor area
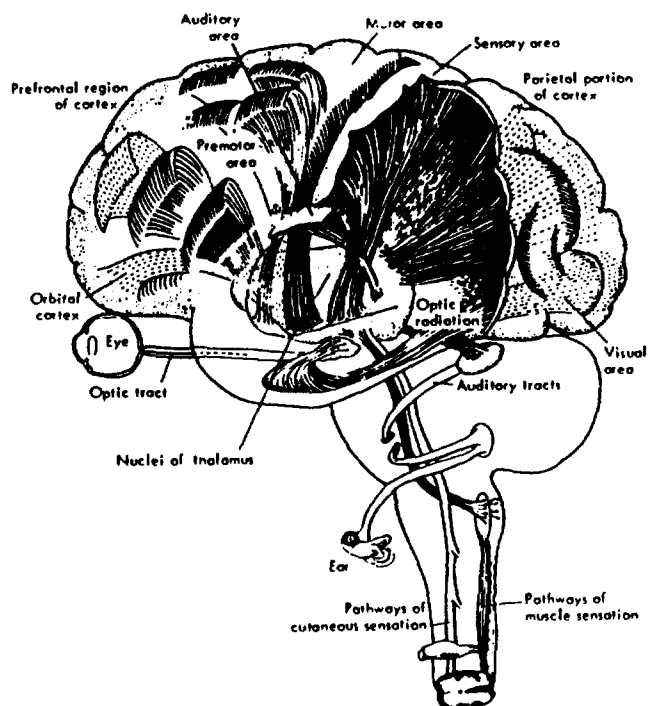
Fig. 2. The location of the thalamus within the cortex (from Luria 1969)

afferent and efferent axons, but it is shaped roughly like a pair of small eggs, side by side (see Fig. 2).[3] It is composed of a set of something like fifty nuclei (not all clearly marked). Each part of the cortex is reciprocally connected in a dense, continuous fashion with some nucleus in the thalamus. Two examples will be repeatedly referred to in this paper. The first is the primary visual area of the cortex V1 (Brodmann's area 17) which is reciprocally connected to the lateral geniculate nucleus, or LGN, in the thalamus. The second is the primary motor area, Brodmann's area 4, which is reciprocally connected to the posterior ventral lateral nucleus, or VLp, in the thalamus. It appears in the cases which have been closely studied that the connection is set-up by dividing up the thalamic nucleus into parallel columns (i.e. volumes extended in one dimension along some curve, but of small extent in the two dimensions perpendicular to this curve), each of which is connected to a column of cortical tissue cutting vertically across the 6 layers.[4] Geometrically, it is as if the thalamus were an elaborate 7th layer of the cortex, with a long (hence slower) neuronal loop tieing it to the cortex proper. The loop is made by the thalamus sending axons up to the cortex where they synapse mainly in layer IV or the deeper part of layer III, and receiving axons originating in pyramidal cells in layer VI or the deeper part of layer V of the cortex (Steriade and Llinas 1988). The thalamus has a small population of inhibitory local interneurons (about 25%, cf. Steriade and Llinas 1988, p. 659)

[3] When we talk of the thalamus, we shall always mean the dorsal thalamus, which is its largest part
[4] See Sect. 3.6.3 in (Jones 1985) and, especially Fig. 3.20

and the remaining neurons all project directly to the cortex with no collaterals (with one exception: see discussion of RE thalamus below). Thus, except for the RE nucleus, the nuclei in the thalamus are *not directly connected to each other*.

Where does the thalamus get its input? Some nuclei in the thalamus are the principal route for sensory signals and 'relay' these up to the primary sensory areas of the cortex, e.g. LGN to V1. And the nucleus VLp transmits the motor related signals from the cerebellum to area 4 in the cortex. Other nuclei get more elaborated sensory, motor or emotional signals from further subcortical structures – the superior colliculus, globus pallidus, amygdala, mammilary nuclei, etc. – but, by and large, their *largest* input is from the cortex itself, via the reciprocal cortico-thalamic pathways described above.[5] Roughly speaking, it seems as though each area of the mammalian cortex receives input, via the thalamus, from that sub-cortical structure which was performing similar cognitive functions in more primitive animals. For instance, analysis of visual input and integration of visual, auditory and tactile information is carried out in the various layers of the superior colliculus (or tectum) in primitive animals. The superior colliculus, in mammals, projects to the pulvinar complex in the thalamus and thence to the association areas of the occipital, parietal and temporal lobes, which carry out the same functions. But, in the evolution of the primate line, the visual input to the cortex via the collicular-pulvinar path plays a smaller and smaller role compared to the direct pathway from the retina to the LGN to V1. For instance, the strength of this secondary visual pathway can be assessed by considering the degree of blindness exhibited by animals in which V1 has been destroyed, so that they must rely wholly on this secondary pathway. Cats are not badly impaired by such a loss, monkeys much more so and humans lose all their sight except for a peculiar guessing skill known as 'blindsight'.[6]

The output of the cortex is more complex. As stated above, every part of the cortex is talking to its corresponding part of the thalamus. In addition, the principal motor output, the pyramidal tract, bypasses the thalamus and goes directly from area 4 to the spinal cord, resulting in an extremely fast command system. Area 4, in other words, has two outputs, the pyramidal tract and the posterior ventral lateral nucleus, VLp, of the thalamus with which it is reciprocally connected. Another output goes from vision-related frontal areas to a subcortical structure, the superior colliculus, and appears to be a motor output specifically for eye movements. A third group of output pathways goes from many cortical areas to the subcortical structures called

[5] For instance, the output nucleus of the globus pallidus, the internal segment, is estimated to contain merely 170,000 neurons (Shepherd 1990)
[6] It seems reasonable to conjecture that all the subcortical inputs to the thalamus play an important role in development in providing the initial seed that starts each area of the cortex moving towards its ultimate cognitive role, but that many of these inputs are not essential for the cognitive functions of the cortex in an adult
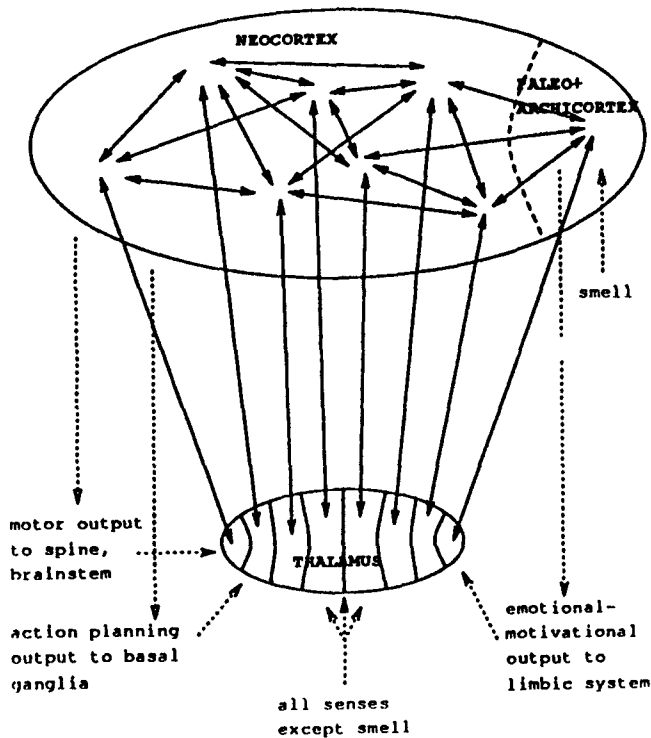
**Fig. 3.** Simplified schematic of cortical connections

the basal ganglia. These seem to be concerned with initiating complex actions. Finally the archicortex has a special output path to some subcortical structures concerned with emotional and motivational states. The picture which I have sketched is depicted schematically in Fig. 3, which may help the reader to put everything together.

Before discussing the role of the thalamus, two complications should be mentioned. As mentioned above, each cortical area is richly and continuously connected to a corresponding nucleus in the thalamus. The first complication is that some nuclei in the thalamus have no specific connections to the cortex, but only so-called diffuse or non-specific connections synapsing over large portions of the cortex (Jones 1985). These seem to play some global regulatory role. Moreover there are also diffuse pathways which go between nuclei of the thalamus which are already specifically connected to one area of the cortex, but which connect it to another area of cortex in a more 'diffuse', less point-to-point way. In addition to being diffuse, they have a different synaptic pattern: they are set up by pyramidal cells in layer V of the cortex instead of VI and are returned by thalamic neurons synapsing in layer I instead of layers III and IV (Jones 1985). For example, the primary visual area V1 has its specific connection with the LGN, but it is also connected diffusely to at least one nucleus in the pulvinar area of the thalamus (the layers involved in these connections are described in Weller and Kaas 1981).

Because of this second type of connection, it usually appears as though each nucleus in the thalamus is

connected to multiple cortical areas and vice versa. Whether or not a single thalamic nucleus is ever connected to several cortical areas in a *specific* way, i.e. with the thalamus synapsing in the middle layers IV and III (deeper part) of the cortex, does not seem to have been clearly settled. The simplest hypothesis is that the specific projections set up reciprocal maps between the whole of the cortex and the specific nuclei of the thalamus which are roughly one-to-one in each direction.[7] Alternately, each nucleus in the thalamus may communicate to one or more cortical areas via specific projections (see Graybiel and Berson 1981, for such a view).[8]

The second complication is that all pathways between the cortex and thalamus pass through a thin layer of cells on the surface of the thalamus known as the reticular complex of the thalamus, or RE thalamus. There the pathways in both directions excite RE cells, which in turn send inhibitory axons both to each other and back to the thalamus to the area of origin of the pathway. Although the RE neurons are inhibitory, experimental evidence (Steriade et al. 1986) shows that peaks of activity in a part of the RE thalamus occur at the same time as peaks of activity in the corresponding nuclei of the thalamus proper. For this reason, the mechanism by which the RE thalamus and the thalamus proper interact is not clear (Steriade and Llinas 1988, p. 712; Crick 1984, p. 4588; Sherman and Koch 1986, p. 12). In any case, as Crick says (Crick 1984, p. 4587): "If the thalamus is the gateway to the cortex, the reticular complex might be described as the guardian of the gateway."

## 3 The thalamus as a window on the world

What, then, is the function of the thalamus? Originally it was thought to be merely a passive relayer of sensory signals to the cortex proper. The small number of

---

[7] There is one well-known exception to this hypothesis: in small mammals, the somato-sensory and motor areas of the cortex can overlap or even coincide. Then two quite separate parts of the thalamus, VP and VL, project to the middle layers of overlaping parts of cortex. It is not known whether this is an isolated exception, or indicates a frequent pattern

[8] Several problems complicate this issue. One is that when very precise tracer experiments are carried out, it sometimes seems that the part of the thalamus projecting to a particular cortical area does not exactly coincide with a thalamic nucleus, and may even cross the boundary between two nuclei (Bender 1981, p. 677). Moreover, many papers report on connections without specifying the laminar pattern of cortical synapses. The real issue, it seems to me, is not whether the cortical areas and the thalamic nuclei correspond one-to-one, because further studies may suggest subdividing both areas and nuclei, but how *topographic* are the specific, reciprocal projections. In other words, one seeks corresponding parts of the cortex and thalamus between which these projections give a continuous map in both directions between the cortical surface and a set of 'rods' or 'columns' in the thalamus (see Jones 1985, Sect. 3.6.3). It may turn out that, at least in primates, both the cortex and the thalamus are divisible into pieces such that the specific projections set up a one-to-one correspondence between them consisting of such topographic projections, or this may fail in various ways

interneurons in the thalamus, the total lack of connections between nuclei within the thalamus or of any intra-thalamic axonal collaterals and the analysis of single-cell recordings all suggest that the thalamus does little or no computation by itself. But if it is merely a relay station, *a)* why do even association areas of the cortex receive thalamic input and *b)* why does the cortex reciprocate with a massive projection of fibres back to the thalamus? Both of these are biologically very expensive and even if they were built because of some phylogenetic quirk, they would decrease in size through selection if they weren't essential.

Stepping back from details for a minute, one can argue like this: phylogenetically, the *association* areas of cortex (those not closely connected to input or output circuits) developed by apparently replicating the structures and functionality inherited from the more primitive areas out of which more primitive brains were made. I want to ignore the motor end of the system for the time being, and concentrate on the sensory end. The original sensory cortex clearly acted as some kind of pattern analyzer taking its input from the thalamus, via layer IV. Assuming that evolution did not modify this plan, this suggests that other cortex also analyzes in some way the data presented to it in layer IV. If so, then the data in the corresponding nucleus of the thalamus will be that area of cortex's view of the world: it will carry a signal which will be sent to layer IV of cortex and analyzed as though this was the signal that some new sophisticated sense could deliver. The nuclei of the thalamus, from this view, are pseudo-sense organs with different views of the world, to be analyzed by the cortex. Each area of cortex is like a homunculus which has a certain narrow view of the world, in which it tries to remember patterns, recognizing familiar ones and lumping similar ones into categories.

In the 'higher' sensory areas this would mean that the thalamus sends them not the raw sensory data but a processed version of the input in which noise and irrelevant stuff have been dropped, and the interesting features are marked as such. For instance, in vision, a higher level representation might record a code for 'grass' in place of a whole area of intricately textured detail, producing a kind of *cartoon* version of the stimulus.[9] The psychological experiments of Bransford et al. (1972), which demonstrate that what we remember about a sentence is often what we thought or assumed was there, rather than what was really there, are consistent with the proposal that higher areas of the brain process a kind of rational reconstruction of the world rather than the raw data.

Still higher sensory areas, especially multi-modal areas integrating data from many senses, should process quite abstract data structures as their view of the world. The sort of structure I have in mind is a sort of geometrical net, with nodes corresponding to various objects or parts of objects, and with links expressing the geometrical relationship between them (e.g. 'to the left of', 'a part of'). This sort of structure was first proposed by Minsky (1975) and was elaborated by Winston (1975) and Marr (1982) (his so-called 3-D model) and has been investigated from a 'neural net' point of view in (Mjolness et al. 1988).

## 4 Active blackboards

But why, then, should there be a recurrent pathway from the cortex to the thalamus? For higher areas, this pathway seems to be the principal way to excite the corresponding nuclei in the thalamus. The cortical area, then, receives its primary, layer IV input, its 'view' of the world as we have argued, from a thalamic nucleus which it is exciting. Note that these recurrent pathways were apparently copied from those present in the primary sensory pathways of simpler brains. Thus the visual signal goes from the retina to the lateral geniculate nucleus, or LGN, of the thalamus, and from there to the cortex. But the LGN $\Rightarrow$ V1 pathway is returned by an equally massive projection V1 $\Rightarrow$ LGN which is certainly not needed for supplying the visual signal. *Is there some computational device which not only functions as input to be read by the computer but on which the computer also writes?*

A key idea in early AI work on speech perception (the CMU work on HEARSAY, cf. Erman et al. 1980) as well as the psychological model embodied in 'Pandemonium' (see Selfridge 1959 or Lindsey and Norman 1977, p. 259), is that of the blackboard. When several sources of expertise, e.g. several different constraints, must be brought to bear on a problem, it is natural to try to carry out the computation in parallel, in independent streams, one devoted to working out the consequences of each source of expertise. These modules must, however, coordinate their work and the simplest way to do this is to have a common blackboard visible to each module, on which they write from time to time their suggestions or conclusions. Similarly, if some constraint or algorithm must be applied multiple times, it is natural to keep the running result on a blackboard, and the computational module must simply keep checking the blackboard and make small modifications to implement local constraints, or repeat some algorithm until satisfied with the result.

My proposal is that the thalamus is something like a blackboard. To use the thalamus as a blackboard means that the cortex must write on as well as read from this blackboard. Thus the thalamo-cortical fibers convey to the cortex the current picture of those aspects of the world with which that area of the cortex is concerned, distributing this data via their axonal arborizations locally in the cortex. The cortico-thalamic fibers convey to the thalamus proposed additions and revisions to this picture arrived at by many computations carried out in the cortex, which are integrated in the thalamus via the dendritic arbors of the thalamic neurons.

---

[9] By the metaphor of cartoon, we mean a data structure which is no longer a pixel-by-pixel record, but which is simplified to a list of areas, annotated by their features and boundaries

Think of the cortex as containing multiple experts with deep understanding of specific patterns and constraints usually present in the world: each expert makes guesses based on its knowledge and, while many of these guesses are compatible and presumably correct, some contradict others and decisions between them must be made. It is these decisions which I suggest are made by a kind of voting, taking place in the summation of stimuli in the dendrites of the thalamic cells.[10] The need and the techniques for such 'data fusion' in the case of vision, have been extensively studied in the recent monograph of Clark and Yuille (1990), and I am proposing that the thalamus implements something like the algorithms that they propose. Note that the thalamus not only integrates its multiple cortical inputs with each other, but also with whatever sub-cortical input, like sensory data, this nucleus receives.

But there are several ways in which the blackboard metaphor may be misleading. For one thing, a blackboard in a computer or in a Professor's office is a passive structure on which you merely write for communication: I have proposed that the thalamus plays an active role in synthesizing the results of calculations by various expert pattern recognizing modules in the cortex. Another major difference is that the computer's and the Professor's blackboard will store ideas indefinitely until erased. But the brain is a volatile computational structure, always reacting to new stimuli and its blackboards would get cluttered and unreadable unless they erased themselves. In other words, the current calculations of the brain must usually be completed within tens or hundreds of milliseconds or they become irrelevant, and the blackboard for such work must be actively refreshed by the senses or by cortical stimulus or it will fade. Thus the thalamus does not sustain its activity by itself, but can only project back to the cortex its integration of the data being sent down to it right now. When the brain wants to tuck some idea away for anything from minutes to years, it would not be a good idea to use these thalamic blackboards which are continuously bombarded with new ideas. These memory functions are accomplished by a different route and apparently require a complex interaction with non-neocortical areas: the hippocampus and the entorhinal cortex. Because of these differences, it seems better to call the thalamus an *active blackboard*, i.e. a blackboard which is volatile and continually presents the latest ideas, synthesized from multiple cortical sources.

Although thalamic activity normally disappears as soon as the neurons fire, there are indications that the thalamus has some mechanisms for maintaining traces of its activity over something like 100 ms. This idea comes from an analysis of the effect of calcium channels and has been linked with the possibility that the RE thalamus may play a major role in keeping the atten-

tion of an area of cortex focussed on some complex of ideas on a shorter time scale (Crick 1984, cf. Sect. 7 below). In other words, the RE thalamus might have a role not merely in gating but in sustaining cortical attention. The mechanisms for such an effect are still very speculative.

In the context of the blackboard metaphor, one can differentiate the role of the specific and the diffuse projections of the thalamus on the cortex. The specific projections are the ones I've been talking about so far: the link between each computational area of the cortex and the active blackboard which it reads and writes on. On the other hand, the diffuse projections from nuclei with specific connections to one cortical area can inform other areas of the cortex working on related sensory problems of what is the current hypothesis on this blackboard. For instance, blackboards in the pulvinar complex of the thalamus containing the optical flow field (data on the movement of the visual stimulus across the retina, thought to be computed in visual area MT) should be visible to areas concerned with figure/ground separation (possibly areas V2 and V4), so that motion clues can be used to distinguish figure and ground. This coordination might also be achieved by direct cortico-cortical pathways, so the brain seems to have two paths for coordinating different areas with overlapping concerns: through diffuse connections to the same nucleus of the thalamus or through cortico-cortical pathways.

Is the thalamus big enough to play such a role? An estimate which is often cited puts the ratio by weight or volume of the thalamus to the cortex at 2%.[11] I am suggesting that the cortex contains multiple independent 'experts' which analyze different aspects of each area's data and that their results are merely integrated in the thalamus. Moreover, I will propose in the second part of this paper that one of the computational burdens of the cortex is that it must translate its data into forms readable to both higher and lower areas, while the thalamus need only store the data in the coded form usable to the given area. Thus a size ratio of 50:1 doesn't seem unreasonable for such a function.

## 5 Harth's theory

Ideas of the kind expressed in Sect. 4 can be found in the work of Luria (1969) who drew attention strongly to the parallel structures in the cortex and the thalamus and reciprocal connections between the two which occur on every level. But he did not invoke the specific computational metaphor of a blackboard. As far as I know, the only person to do so is Erich Harth, who developed that he called the Alopex theory of the interaction between the LGN, the visual area V1 and higher visual areas. In his popular book "Windows on the Mind" (Harth 1983), he expresses his theory like this:

---

[10] These calculations may also involve the interneurons in the thalamus, e.g. using microcircuits involving 3-way dendro-dendritic synapses with these interneurons. Because these interneurons are inhibitory, they also allow cortical cells to cast negative votes, i.e. to inhibit some thalamic cells

---

[11] e.g. 10 cubic cm. for the thalamus (bilaterally) to 500 cubic cm. for the grey matter of the cortex, or *very* roughly 2 million cells per nucleus in the thalamus to 100 million in an average cortical area

"*Recall that the part of thalamus that is concerned with vision, the LGN, preserves some of the character of the retina: activity is distributed over sheets of neurons which mirror the pattern of light falling on the retina. It is possible that corticofugal messages weave similar patterns on this inner retina, as Wolf Singer has called it. This is suggested by the fact that the fibres coming back from the cortex are about as numerous as those going in the opposite directions, and have the same spatial distribution over the sheet of neurons in the relay nucleus. Also there is evidence that the returning messages are feature specific; that is, they can enhance, select, and perhaps mimic sensory patterns. Another nit of evidence is the finding that in cats, activity in the LGN is heightened during REM sleep in which we are supposed to dream. Moreover, this activity was found to be similar in character to that evoked by real visual input.*

*I would like to suggest an extension of Singer's concept of an 'internal retina' to what I called an 'internal sketchpad'. The idea is that sensory patterns are laid down in the LGN by sensory input, but similar patterns may also be sketched there by higher centers. The LGN is a possible location for such a process, but certainly not the only place where this may occur.*"

Harth, with various coworkers, has gone much further and given a precise algorithm which they propose as a model of how the cortical feedback to the LGN might be computed. I quote from the abstract to his Science article (Harth et al. 1987):

"*The mammalian visual system has a hierarchic structure with extensive reciprocal connections. A model is proposed in which the feedback pathways serve to modify afferent sensory stimuli in ways that enhance and complete sensory input patterns, suppress irrelevant features, and generate quasi-sensory patterns when afferent stimulation is weak or absent. Such inversion of sensory coding and feature extraction can be achieved by optimization processes in which the scalar responses derived from high level neural analyzers are used as cost functions to modify the filter properties of more peripheral sensory relays. An optimization algorithm, Alopex, which is used in the model, is readily implemented with known neural circuitry.*"

This sketchpad hypothesis of Harth is similar to mine, but the Alopex theory goes much further in proposing a specific algorithm. I am suggesting more simply that the different parts of the cortex have many different computations to do and that the thalamus has an essential but relatively passive role, in integrating the reconstructions, schemes, ideas, etc. of each area and broadcasting them to this and other areas of the cortex. But in contrast to the Alopex theory, I am not proposing that the feedback loop between thalamus and cortex is part of any specific pattern recognition computation and, indeed, I want to ascribe to cortico-cortical loops versions of some of the computations Harth is interested in. These ideas will be developed in the second part of this paper.

## 6 The thalamus and the cerebellum

Now what is the role of the thalamus for the primary motor area, Brodmann's area 4? In all higher mammals, Area 4 is an easily distinguishable architectonic area which received no direct sensory input, but is responsible for initiating movement on the lowest, muscle-by-muscle, level. The evolution of mammals shows a clear progression in which motor control is shifted increas-ingly to the cortex and specifically to Area 4 which is the origin of a direct projection from cortex to motor neurons in the spinal cord.[12] This connection, the pyramidal tract, is set up by giant pyramidal cells (the cells of Betz) in layer V which control the muscles, with only a single synaptic relay in the spine. Whoever has his hand on the throttle, is in control and this pathway is clearly crucial in shifting control to the cortex, away from the more primitive sub-cortical motor systems, which are demoted to deal only with involuntary actions and simple requirements like balance.[13]

In addition, some of the axons of the pyramidal tract terminate in the brain stem, in the red nucleus, from which they project to a new section of the cerebellum, the neo-cerebellum consisting of the deep lying dentate nucleus and the lateral zones on the surface of the cerebellum. The output of the dentate nucleus projects primarily not to the brain stem or spine but up to the thalamus, to the nucleus VLp, and thence to the motor area.

Here is our scenario of what these new structures are doing. The cortex, through complex integrated actions of all lobes, 'decides' to make a certain movement. This command winds up in the motor area, which does two things: it writes the motor command on its blackboard VLp, and it sends it off down the pyramidal tract. Apparently this message can be either imperative or tentative. The imperative mode is reserved for commands which brook no delay and which should be done instantly, even if awkwardly: these take about 7 ms from motor area activity to muscle response (Evarts 1973). Most commands get caught at the red nucleus, and in the local spinal circuits and don't start muscle action immediately. In fact, experimental recordings show delays of around 100 ms between pyramidal tract activity and muscle response.[14]

The cerebellum meanwhile analyzes this motor command and does what it has done in all vertebrates: it modifies and specifies in detail what combination of forces for what periods of time in which muscle groups best carry out this command. Then the cerebellum writes this prescription on VLp, for the motor area to

---

[12] Area 6, the *pre-motor* area, also projects directly to the cord via the pyramidal tract. The hierarchy of motor areas will be discussed in the second part of this paper

[13] It would be nice to be able to say not merely that mammals are unique in having a massive cortex directly controlling their muscles via the pyramidal tract, but also that mutations creating the pyramidal tract distinguish mammals from reptiles. In fact, the homologies between cortical structures in mammals and other structures in reptiles are complex, and, by one theory, part of the mammalian cortex evolved from the reptilian dorsal striatum which does have a motor output, albeit with relays in the brainstem

[14] An extensive series of papers by Sasaki and Gemba (cf. Gemba et al. 1981) have recorded the difference in field potentials between layer I and layer VI in area 4 during the performance of trained movements. They interpret surface negative, depth positive potentials as due to currents in the apical dendrites of superficial pyramidal cells and the opposite potential as due to currents in deep pyramidal cells, such as the Betz cells. The latter typically start up about 100 ms before the muscle cells fire. Single cell recordings due to (Georgopoulos et al. 1989) show roughly the same delay

read. This allows the cortex to use the carefully learned muscle programs expressed in the synaptic weights of the cerebellum. In the next few milliseconds, the commands sent down the pyramidal tract strengthen or are somehow modulated to say, *now do it*, and a polished movement is executed. Of course, more primitive parts of the cerebellum will also monitor the on-going movement and modify it by more direct paths, which supplement the cortico-spinal path. In all of this, VLp is playing the role of a low-level motor blackboard, in which projected and on-going movements are encoded in terms of specific forces to be exerted by specific muscle groups. Although the cerebellar input to VLp will be more informative in terms of what forces will accomplish the required movement most efficiently and smoothly, the cortex has available to it more highly processed sensory data that may lead to modifying the specifications for the movement. The blackboard VLp will integrate these needs.

Luria (1969, English edition, p. 55) has a quite similar analysis of these circuits, though he doesn't use the word blackboard of course:

> "The principle of feedback is applied quite differently in the activity of that part of the cortex responsible for the organization, programming and execution of voluntary motor activity, for in this realm it becomes the main source of information on the effects of the movements and actions performed. The physiological role of the motor cortex essentially consists of matching the "assigned program" of a motor act, formed mainly on the basis of the analytical and integrative cortical activity of the posterior divisions of the hemispheres, with the actual course of its performance, i.e. in detecting signals of success and signals of error (agreement or disagreement between the program and the performance) and in making the required corrections at the right time in the course of the actions. In view of what has been said, it will be apparent that both the centrifugal and centripetal (responsible for feedback) chains of relays of impulses, connecting the motor cortex to the subcortical formations, are included in the extrapyramidal system's of the brain, which are known to be of essential importance to the coordination of voluntary movement."

The hypothesis that Area 4 and VLp perform incremental calculations converging step by step to the precise muscular act to be performed is consistent with the startling results of Georgopoulos et al. (1989) from single cell recordings in Area 4. He found not only that the pattern of excitation in Area 4 at the time of an arm movement correlated closely with the direction of arm movement in each repetition of his experiment, but that in the 100 ms period before arm movement, the pattern of excitation in Area 4 built up in a definite sequence which can be interpreted as forming mental images of arm movements intermediate between reaching straight ahead and reaching in the direction now desired. It was exactly as if the required muscular commands were being computed in stages, starting from simpler ones for which a template was known and making incremental modifications, using the VLp as blackboard to record the current proposed arm movement.

A specific prediction that I would like to make is that, like the auditory areas, the low level motor blackboards must have a certain amount of temporal buffering. If the data structure for Area 4 is the sequence of

muscular commands over the next second or so, then, assuming the animal's current plan is not interrupted, there should be a correlation between neuronal activity in Area 4 at a given time and the action taken after moderate time lags, as well as the action taken immediately afterwards. The temporal buffering could be done geometrically with different strips of neurons, or it could be done by a subtler in-place coding (see Sect. 9).

## 7 The thalamus and attention

The most widely discussed alternative suggestion about the function of the thalamus is that, in addition to relaying data, the thalamus gates it in some way. For example, it may be used in focussing attention on some part of the stimulus, or blanking out other parts of the stimulus (e.g. retinal input during a saccade). This theory was developed at length by Crick (1984), correlating it with Treisman's experiments suggesting the mind had an internal 'spotlight of attention' that could be moved around the visual field without actually moving the eye (Treisman 1988). In particular, he suggests that the position of the RE thalamus, smack in the middle of the pathway to the cortex, makes it the logical candidate for implementing a focus of attention. The exact mechanism he proposes is rather subtle, involving an unusual property of thalamic neurons, related to calcium channels which cause 'low threshold spikes' (Jahnsen and Llinas 1984), and has been disputed by others (Sherman and Koch 1986). Nonetheless, the idea that the RE thalamus in some way gates the flow of data from the thalamus to the cortex is very plausible, given its location and is quite compatible with the thalamus being a blackboard.

What seems most implausible to me in the theory that gating and attention are the primary uses of the cortico-thalamic projection, is why the need to gate this flow of data would require such a massive projection, at least as big as the thalamo-cortical projection. Bit for bit, it would seem that transmission of data requires more bandwidth than the selection of part of the data. Moreover, this theory ignores the fact that the cortex needs to write most of this data on the non-sensory nuclei of the thalamus before it can read it, and the data doesn't automatically stay around, i.e. the thalamus doesn't have loops so that it can maintain a state of excitation. Why go to all this trouble to send data down to the thalamus, as well as sending the gating signals, so that a subset of this data will echo back? This operation makes more sense if, at each stage in a calculation, writing and reading from the thalamus both have roles, as they would if it was serving as a blackboard, synthesizing the cortical results via the dendritic arbors of the thalamic neurons, and distributing them back to the local cortical area via the spread of the thalamo-cortical axons.

Let's look at some numbers to bring this home. Unfortunately, quantitative neurobiology is not in vogue and most papers avoid estimating the numbers of neurons and axons in the structures and pathways

discussed. However, for the LGN, Sherman and Koch's article in (Shepherd 1990) gives some figures for the so-called A-laminae of the LGN in the cat. There are two pathways between the retina and the cortex: the X pathway (homologous to the P pathway in primates) concerned with shape and color, and the Y pathway (homologous to the M pathway) concerned with motion. The X pathway has 90,000 axons from the retina to the LGN, which synapse on some 175,000 relay cells, while the Y pathway has 10,000 axons synapsing on 125,000 relay cells (all figures should be considered as ±20% or so). The cortico-geniculate pathway, on the other hand, contains 4,000,000 axons synapsing on the X and Y relay cells in the A-lamina (how many on each is not known). Thus the cortical input to the LGN is about 40 times bigger than the retinal input, and 13 times bigger than the reciprocal LGN to cortex pathway. Even allowing largely for multiple synapses of the retinal axons, they estimate that only 10–20% of the synapses on the LGN cells arise from retinal axons, while 80–90% arise from cortical axons (Shepherd 1990, p. 264 and p. 278). I suggest that the only way to make sense of these figures is that most of the data in the LGN is calculated not directly from the retinal input, but via one or more passes through the geniculo-cortical loop, this data representing the visual input with considerable image processing added (see Sect. 8).

The attention theory has also been proposed by Ojemann (cf. his review article, 1983). Ojemann has carried out experiments stimulating varous thalamic nuclei in awake humans in the course of operations in which various subcortical structures are being surgically destroyed. He proposes that the thalamus is responsible for a 'specific alerting response'. Specifically, he observes that suitable thalamic stimulation can i) improve verbal memory if given at the time of presentation of the item to be remembered and ii) increase rate of response and number of errors if given at the time of recall. Moreover, such stimulation can also cause persevervation, both on the first syllable of a word being pronounced, or on an earlier response which is not correct for the next task. Let me point out, however, that these results are also compatible with the blackboard role for the thalamus: if the thalamus is a blackboard, stimulating it could have the effect (a) of highlighting one of the data items represented on it, resulting in better memory for the item, and (b) of hindering the power of cortico-thalamic pathways to revise and update the data in the thalamus appropriately, causing persevervation and the other types of error Ojemann observes.

## 8 Possible tests

This proposal admits some straightforward experimental tests. The most unambiguous corroboration would be to demonstrate an effect of the cortico-thalamic projection on the LGN. If the LGN serves as a blackboard, its state should be determined not merely by retinal stimulation, but by the ongoing analysis of this

stimulation by the corresponding area of the cortex, V1 or Brodmann's Area 17. Both the LGN and V1 have two classes of cells: fast-responding ones concerned with motion (called Y in cats and M in primates) and slower, sustained-response cells concerned with relatively static shapes (called X in cats and P in primates). One simple class of LGN neurons of the second type are the black-white center-surround opponent cells, which respond in a sustained way to the difference of the amount of light in a roughly circular central field minus that in a roughly circular annulus around it (or vice versa). Now for these cells, V1 seems concerned with using this data to find edges and lines in the retinal image, computing their orientation, and how far they continue straight (as in the responses of so-called end-stopped cells).

In the real world, this is a non-trivial operation because of noise, shading, texture, etc. What I predict, therefore, is that over some cycle of maybe 50 milliseconds, the picture held by the LGN will improve, noise being removed, edges and lines being sharpened, filled in for instance where the veins on the retina cross them. Part of the image is, of course, changing, and the motion system concerned with this should be making its own kind of improvements to the LGN responses. But for the relatively static parts of the retinal image, i.e. objects which are still when the eye is still, or objects being tracked when the eye is in tracking mode, I would conjecture that the sustained response cells alter their rate of firing in this 50 ms window to mimic what would happen if the retinal signal were improved, much like the image processing which astronomers perform on satellite images. To test this, the first requirement is not to use the extremely simple stimuli typical of these experiments: bars, dots, sine-waves, etc. but more realistic abstractions of real world data: bars with white or other noise superimposed, edges with small gaps, etc. Secondly, the firing of the LGN neurons in the response period should not be averaged, but counted separately in each 5 or 10 ms interval following the stimulus. For an example, see Fig. 4, where I conjecture that the response to the bar with blurry spot will start out less than that to the whole bar, but build up, when cortical feedback kicks in, to the same as the response to the full bar.

A more speculative proposal is that the increase in the number of X-relay cells in the LGN versus the number of axons of X-ganglion cells in the retina is due to the need, for stereo fusion, of constructing shifted versions of the raw input. In other words, the cortex seeks to compare the signal from the left and right retinas. Because of the geometry of stereo vision, these often match up closely after a horizontal shift (whose size is a function of the distance to the viewed surface and the vergence of the two eyes), and the proposal is that such a shift may be physically realized in the responses of some of the LGN cells during the process of fusion of the images from the two eyes. To test this, one need only present image pairs to two paralyzed eyes with varying degrees of disparity, and observe the time course of response of LGN X-relay cells: the possibility
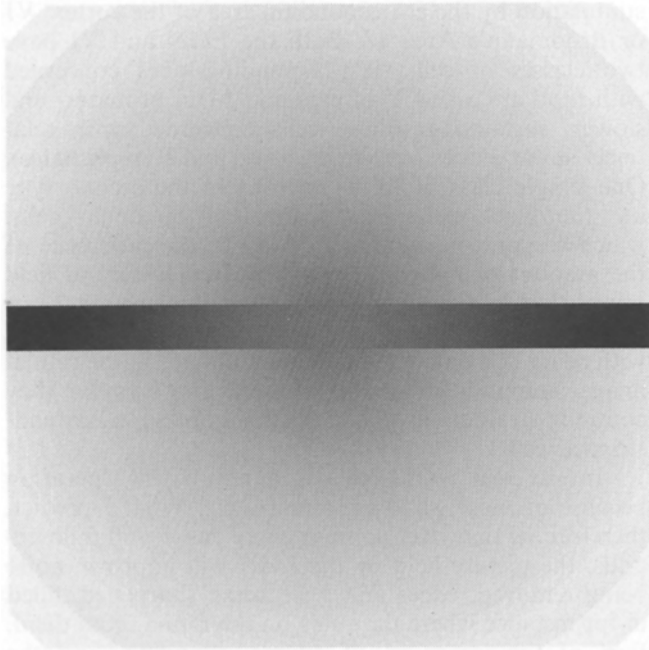
**Fig. 4.** Stimulus for test of LGN response

is that their receptive fields will sometimes shift to achieve better registration of the two images.

Another area in which we may test our theory of active blackboards is by examining the responses of thalamic cells which are not principally driven by sub-cortical input. These cells receive most of their input from the cortex and I would propose that their responses will indicate clearly what that area of cortex is concerned with. A good example is the inferior pulvinar nucleus which is reciprocally and specifically (i.e. synapsing in layer IV or deep III) connected to visual area V2. A conjecture is that while V1 is concerned with identifying small pieces of edges, their orientation and motion and possibly curvature, V2 is assembling this data into a global picture of the scene. More specifically, I mean tracing long edges, finding regions of coherent color and texture and deducing from this a segmentation of the scene into individual objects. In psychology, this includes what is called figure-ground separation and it involves crucially the principles of Gestalt psychology. Accomplishing this segregation properly involves integrating the data from stereo vision (because individual objects have continuous disparities without jumps) and from motion (because individual objects move coherently). Neurophysiological data suggests that V2 does at least some of this

---

[15] As Terry Deacon pointed out to me, the rate of phoneme production is not necessarily the rate at which new data appears in the cortex. Although speech is described by linguists as a sequence of phonemes, in the sound itself the phonemes overlap, i.e. the time intervals in which each phoneme affects the sound overlap. If subcortical structures extract features which encode the clues for all the overlapping phonemes at each instant, then new data appears in the cortex roughly with each new *syllable*, not each new phoneme. This suggests new data every 100–150 ms

(Van der Heydt and Peterhans 1989). Therefore, I propose that neurons in Pli will record boundaries between objects in a scene, no matter how they are marked (or obscured) in the raw image. Moreover, the areas of a scene which are part of a single object must somehow be marked as such in order that the shape of the object can be analyzed, leading to its identification. This is an operation called 'coloring' by Ullman (1984), without which the regions corresponding to individual objects cannot be dealt with as units. One plausible conjecture is that the responses of some neurons will be locked on to particular objects, so that, when the scene is shifted in front of the eyes, the neuron will fire so long as its receptive field overlaps (or is contained in) the visible surface of the object and will drop off as soon as the region moves away.

## 9 Temporal sequencing

Some of the thalamic active blackboards deal with relatively static information and some deal with rapidly changing information, that is only relevant for less than 100 ms say. Thus higher level conclusions, about the geometry of your immediate surround and the objects and the people there, will change only slowly, provided nothing is in rapid motion. But data about a speech signal is superceded by the next phoneme with 50–100 ms or so.[15] For those blackboards dealing with rapidly evolving data whose temporal pattern is essential for its classification, such as the lower level auditory and motor blackboards, it is essential to maintain a certain amount of temporal buffering: i.e. to keep at any instant the description of the input over some fixed period up to the present. Such a prediction should be easy to test. There are several ways to set up such a buffer: e.g. *a)* write the new signal cyclically into a family of neurons, *b)* write the latest data always in the same place but shift earlier data along some line, or *c)* write the data on top of itself with some 'in-place' coding. Another possibility is that *d)* the thalamic-cortical loop itself is employed in doing this buffering, so that aspects of the time-delayed signal are fed back by cortico-thalamic fibres and recur once or several times in thalamic activity. In cases *a)*, *b)* and *d)*, careful measurement of the time lags between stimulus and neuron response in primary or secondary auditory cortex, or in the medial geniculate nucleus, MGN, of the thalamus should reveal such buffering, especially if multi-neuron recordings are made. In all cases, there should be some correlation between activity in auditory cortex at a given time and the stimulus presented some hundreds of milliseconds earlier.

Moreover, the Y/M-cells in the visual pathway deal with motion, and their signal is also often evolving rapidly. In fact, an obstacle to the blackboard hypothesis that is often raised is how the LGN can be used as a blackboard when the visual signal is changing so fast. Now the brain has evolved a special tracking mode of eye motion precisely to keep a uniformly moving object nearly stationary on the retina, but this usually

works for only one object at a time and motion may not be uniform, nor the object rigid. I'd like to propose that this is exactly why, in the Y/M-pathway, there is such a large increase in number of LGN cells to number of retinal cells (estimated at 12:1, see Sect. 7). The extra LGN cells are available for temporal buffering, storing the motion history during a single fixation of the eye. Making a more precise prediction requires a specific hypothesis of how temporal buffering is done: hopefully the same mechanism is used by mammals in the visual, auditory and motor domains, but I don't want to make a guess.

# References

Bender DB (1981) Retinotopic organization of the macaque pulvinar. J Neurophys 46:672–693

Blinkov SM, Glazer II (1968) The human brain in figures and tables. Basic Books, New York

Bransford J, Barclay JB, Franks J (1972) Sentence memory: a constructure vs. interpretive approach. Cogn Psychol 3:193–209

Brodal A (1981) Neurological anatomy. Oxford University Press, Oxford

Cherniak C (1990) The bounded brain. J Cogn Neurosci 2:58–68

Clark J, Yuille A (1990) Data fusion for sensory information processing systems. Kluwer Academic Press, Amsterdam

Crick F (1984) Function of the thalamic reticular complex: the searchlight hypothesis. Proc Natl Acad Sci 81:4586–4590

Erman LD, Hayes-Roth F, Lesser VR, Reddy R (1980) The HEARSAY-II speech understanding system. Comput Surv 12:213–253

Evarts EV (1973) Motor cortex reflexes associated with learned movement. Science 179:501–503

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in primate cerebral cortex. Cerebral Cortex: (to be published)

Gemba H, Hashimoto S, Sasaki K (1981) Cortical field potentials preceding visually initiated hand movements in monkeys. Exp Brain Res 42:435–441

Georgopoulos AP, Lurito JT, Petrides M, Schwartz AB, Massey JT (1989) Mental rotation of the neuronal population vector. Science 243:234–236

Graybiel AM, Berson DM (1981) Families of related cortical areas in the extrastriate visual system. In: Cortical sensory organization, vol 2. Humana Press, Clifton, NJ, pp 103–120

Harth E (1983) Windows on the mind. Quill, New York

Harth E, Unnikrishnan KP, Pandya AS (1987) The inversion of sensory processing by feedback pathways: a model of visual cognitive functions. Science 1987:184–187

Jahnsen H, Llinas R (1984) Electrophysiological properties of guinea-pig thalamic neurons: an in vitro study. J Physiol London 349:205–247

Jones EB (1985) The thalamus. Plenum Press, New York

Lindsey P, Norman D (1977) Human information processing. Academic Press, New York

Luria AR (1969) Higher cognitive functions in man, 2nd edn. Moscow University Press, Moscow (English edn 1980, Basic Books, New York)

Marr D (1982) Vision. Freeman, San Francisco

Minsky M (1975) A framework for representing knowledge. In: Winston P (ed) The psychology of computer vision. McGraw-Hill, New York

Mjolness E, Gindi G, Anandan P (1988) Optimization in model matching and perceptual organization. Research report YaleU/DCS/RR-634

Ojemann G (1983) Brain organization for language from the perspective of electrical stimulation mapping. Behav Brain Sci 2:189–206

Rockel AJ, Hiorns RW, Powell TPS (1980) The basic uniformity in structure of the neocortex. Brain 103:221–244

Selfridge O (1959) Pandemonium: a paradigm for learning. In: Symposium on the Mechanization of Thought Processes. HM Stationary Office, London

Shepherd G (1990) The synaptic organization of the brain, 3rd edn. Oxford University Press, Oxford.

Sherman SM, Koch C (1986) The control of retinogeniculate transmission in the mammalian lateral geniculate nucleus. Exp Brain Res 63:1–20

Steriade M, Llinas RR (1988) The functional states of the thalamus and the associated neuronal interplay. Physiol Rev 68:649–742

Steriade M, Domich L, Oakson G (1986) Reticularis thalami neurons revisited: activity changes during shifts in states of vigilance. J Neurosci 6:68–81

Treisman A (1988) Features and objects. Q J Exp Psychol 40A: 1988

Ullman S (1984) Visual routines. Cognition 18:97–159

Van der Heydt R, Peterhans E (1989) Cortical contour mechanisms and geometrical illusions. In: Lam DM, Gilbert CD (eds) Neural mechanisms of visual perception. Gulf, Houston, Texas

Weller RE, Kaas JH (1981) Cortical and subcortical connections of visual cortex in primates. In: Cortical sensory organization, vol 2. Humana Press, Clifton, NJ, pp 121–156

Winston P (1975) Learning structural descriptions from examples. In: Winston (ed) The psychology of computer vision. McGraw-Hill, New York

Dr David Mumford
Mathematics Department
Harvard University
1 Oxford Street
Cambridge, MA 02138
USA