Handbook of Brain Theory, M. Arbib editor

Thalamus

David Mumford

Introduction

The thalamus is a subdivision of the brain of all mammals. It is situated at the top of the brainstem, in the middle of the inverted bowl formed by the cerebral hemispheres. It is shaped roughly like a pair of small eggs, oriented on the posterior-anterior axis and side by side, one in each hemisphere. Its cell count is approximately 2%–4% of the cortex.

The thalamus has a striking position in the flowchart of data in the brain: *essentially all input to the cortex is relayed through the thalamus*. The main exceptions are the diffuse projections of several brainstem nuclei carrying neuromodulators such as acetylcholine, but presumably not carrying detailed information-bearing signals such as those encoding sensory or motor data; the lateral olfactory tract that conveys the sense of smell to the cortex; and a connection from the amygdala to the prefrontal cortex, which duplicates a thalamic connection. The majority of the input to the cortex—visual, auditory, and somatosensory information; planning-tuning-motor output of the basal ganglia and cerebellum; and emotional-motivational output of the mammillary body—reaches the cortex exclusively through the thalamus. The thalamus is thus the principal gateway to the cortex, the cortex's window on the world at every level. What is equally striking is that the thalamocortical pathways are reciprocated by feedback pathways from the cortex back to the thalamus, forming a massive system of local loops between the thalamus and the entire cortex. For instance, Sherman and Koch (1986) estimate that in the cat, there are approximately 10^6 fibers from the *lateral geniculate nucleus* (LGN), the visual area of the thalamus, to the visual cortex but 10^7 fibers in the reverse direction. In other words, there are approximately 10 times more feedback than feedforward paths.

As knowledge of the anatomy and connections of the thalamus developed, the initial belief was that the principal function of the thalamus was simply to relay information from subcortical structures to the cortex. This view was reinforced by the simplicity of the internal circuitry of the thalamus and the apparent faithfulness of transmission shown by neurophysiological studies. Most of the cells in the thalamus are excited by subcortical input and send their output directly to the cortex, with no collaterals to other thalamic cells (see below for a more detailed description). This finding suggests that the thalamus is one of the simplest structures in the brain, and one that hardly requires modeling.

The central question concerning the thalamus is, however, what is the use of the massive feedback pathways from every area of the brain to its thalamic input nucleus? As Jones (1985:819) puts it: "Can it be that such a highly organized and numerically dense projection has virtually no functional significance? One doubts it. The very anatomical precision speaks against it. Every dorsal thalamic nucleus receives fibers back from all cortical areas to which it projects." Jones noted the central puzzle of the thalamus from a modeling perspective. The existence of these feedback pathways makes it evident that the thalamus plays an essential cognitive role, engaging in a dialogue with the cortex in which some information is being computed or combined, but what information?

Anatomy of the Thalamus

The thalamus is not a homogeneous mass of neurons, but a collection of smaller nuclei. In humans, there are approximately 50 of these nuclei in each hemisphere, with some subdivisions much clearer than others. An exhaustive survey of our knowledge of the nuclei and their connections (through 1984) is provided by Jones (1985).

Most of the nuclei are called *specific nuclei*. These nuclei connect in an ordered topological pattern, one nucleus to one area of the cortex. The topography of the projection from these nuclei to their cortical target area varies somewhat, but tends to follow what Jones calls a *rod-to-column* pattern of projection. In other words, a thalamic nucleus is divided into a family of disjoint *rods*, each of which projects to a column of cortical tissue slicing perpendicularly through the cortical plate from layer 1 to layer 6 (cf. Jones, 1985: figure 3.20, p. 126; figure 3.22, p. 129; and p. 811). This tendency has many variations, but specific thalamic relay cells seem to be constrained to synapse in specific cortical columns to preserve the relationship of their data with parallel data streams. The best known example, on which the largest amount of research has been done, is the projection of the LGN in the thalamus to cortical area V1, the primary visual area. This projection preserves the two-dimensional retinal layout of the visual image, through the LGN and onto the cortical surface, preserving in addition the separation of the signals from the two eyes onto distinct *ocular dominance columns* in area V1.

In addition to the specific nuclei, there are also *nonspecific* nuclei which project diffusely, often to the entire cortex. They play various kinds of regulatory roles and will not be discussed here. These specific and nonspecific nuclei make up the dorsal thalamus. In addition, there are several structures called the *ventral thalamus*, including the reticular thalamic nucleus (RE) and perigeniculate nucleus. These structures form a thin layer of cells covering the anterior, dorsal, and lateral surfaces of the thalamus (the perigeniculate over the LGN) through which all thalamocortical and corticothalamic fibers must pass and which sends inhibitory projections back to the thalamus. From an anatomical point of view, Crick (1984) states that, "If the thalamus is the gateway to the cortex, the RE might be described as the guardian of the gateway."

The principal cell type in the specific nuclei of the thalamus is the medium to large excitatory cells known as relay cells. They make up approximately 65%–80% of all cells. Their axons go directly to the cortex, giving off no local collaterals, except on cells in the reticular nucleus as they pass through this structure. These axons synapse principally in the cortex in layer 4 or deep layer 3, the standard input layers of the cortex. Some reports show a small group of small excitatory cells which project more diffusely, possibly to several areas of the cortex, synapsing principally in layer 1 as well as in layer 6 (Jones, 1985: 97, 158). The remaining cells are inhibitory GABAergic interneurons which provide the only intrathalamic circuitry (for cell counts, see Jones, 1985:166–167). They synapse on the relay cells and on each other. Figure 1 shows a synopsis of these circuits.

The thalamic relay cells do not *always* relay information faithfully as it comes in. In drowsiness, in non–rapid eye movement (REM) sleep, or after the administration of various laboratory preparations, these cells go into an oscillatory mode in which they alternate between short, high-frequency bursts and extended periods of hyperpolarization, repeating at a frequency of 7–14 Hz. This oscillatory mode is a key property of thalamic relay cells, but it is not clear whether it has any cogni-

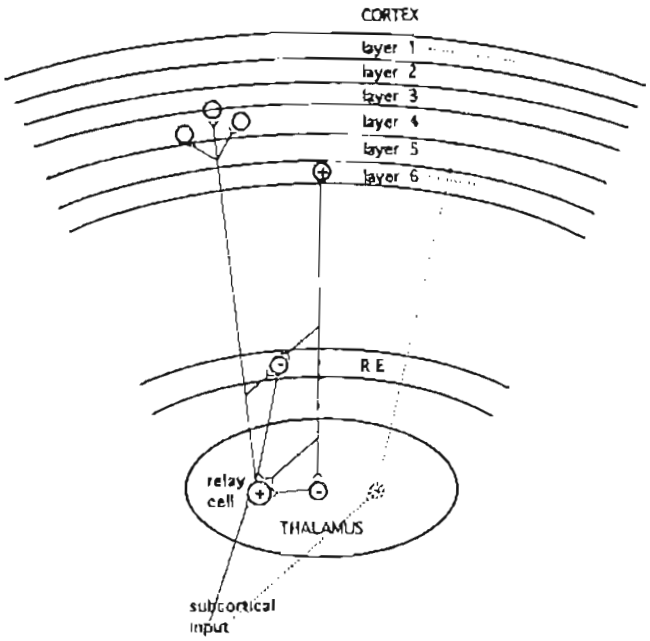

Figure 1. Simplified diagram of the neurons of the thalamus and their principal connections. The signs indicate which neurons are excitatory and which are inhibitory.

tive significance (see THALAMOCORTICAL OSCILLATIONS IN SLEEP AND WAKEFULNESS).

Gating and Selective Attention Through Feedback

Perhaps the most widespread belief about the role of feedback is that cortical feedback gates thalamic transmission of subcortical data; hence, it allows the cortex to attend to part of these data selectively. Such ideas were suggested by Singer (1977), Crick (1984), Sherman and Koch (1986), Koch (1987), and Desimone et al. (1990), among others.

To distinguish this model from others, it is useful to describe it mathematically. Consider the visual pathway from the LGN to area V1, and let $I(x, y, t)$ represent the visual signal incident on the retina as a function of coordinates on the retina and time. The ganglion cell output of the retina has been modeled as a filtered and rectified function of I: $J_\alpha = (I * F_\alpha)_+$, for a set of filters F_α, one for each class of ganglion cells. If the LGN was a faithful relay station, J_α would also model the activity K_α of the LGN relay cells. The gating hypothesis is, roughly, that instead, the relay cell activity is modeled by $K_\alpha(x, y, t) = w_\alpha(x, y, t) \cdot J_\alpha(x, y, t)$ where w_α represents weights that selectively enhance or suppress parts of the signal. The idea is that weights w_α which gate the strength of the signal as it passes through the relay cells can be set up either (1) by excitation of the LGN relay cells on the distal parts of their dendrites by direct corticothalamic feedback, possibly using *N*-methyl-D-aspartate (NMDA) channel mechanisms (Koch, 1987), or (2) by inhibition through an intermediate inhibitory cell in RE or LGN. These mechanisms should be contrasted with possible, much more complex, transformations of the signal $J_\alpha \to K_\alpha$ that the LGN might perform as a result of cortical feedback (possibly after several cycles of sending signals to the cortex, then back, then to the cortex again, etc.).

Evidence for such gating was discovered by Singer and Schmielau (1976) in their study of LGN relay cell responses to binocular stimuli. The LGN relay cells are, to first approximation, monocular cells. Different layers of the LGN respond to different eyes. In general, there is an interaction between LGN relay cells whose receptive fields overlap to inhibit each other, e.g., center-surround cells with adjacent receptive fields, *on* and *off* cells with the same receptive fields, and cells in the sustained versus the transient pathway inhibit each other. This observation is interpreted by Singer (1977) as a way of increasing the signal-to-noise ratio in the relaying process. When signals from opposite eyes are compared, the signals will differ by a left-right *disparity shift* whose size depends on the distance between the plane of fixation and the visible surface in that direction. What Singer and Schmielau discovered was that relay cells in laminae A and A1 of the cat LGN, responding to ispilateral and contralateral retinas, inhibit each other as noted above unless the visible surface lies in the fixation plane. In that case, their responses are enhanced. Moreover, whereas this inhibition between left and right responsive relay cells is mediated by LGN inhibitory cells, the suppression of the inhibition and its replacement by an enhancement when the visible surface is part of the fixation plane is caused by *corticofugal* signals. The effect is to highlight objects on the fixation plane and suppress nearer and farther objects whose binocular signals are out of registration. Their article contains a proposal for circuitry underlying this effect.

An influential model incorporating this idea of gating was proposed by Crick (1984). To explain the latency of low-level visual responses, Anne Treisman and others proposed that some tasks can be performed in parallel on the entire visual field, while others can be performed only in one small *window of attention* at a time. In Crick's model, these windows are created by the reticular nucleus RE suppressing the relay cells, except for those within the window of attention. He proposes that temporary cell assemblies are then created, like a buffer for the subimage "in the *searchlight*," within which further processing will occur. The mechanisms for such gating of visual signals are discussed in Sherman and Koch (1986).

Active Roles for the Thalamic Buffer

Two objections to the hypothesis described in the last section are that (a) simply suppressing or enhancing different parts of the subcortical signal would not seem to require such a massive feedback pathway, and (b) some thalamic nuclei do not receive major subcortical input, yet they still are reciprocally connected to the cortex with massive pathways.

To my knowledge, the first person to propose a more complex role for the LGN was Harth. Beginning in 1974, he developed his ALOPEX neural net theory for visual processing and especially for corticothalamic feedback. As described in Harth, Unnikrishnan, and Pandya (1987): "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." Although formulated as a time-varying gate, as in the previous section, this theory is an iterative one in which many signals traverse the thalamocortical loop, optimizing by SIMULATED ANNEALING (q.v.) an objective function which seeks to enhance remembered patterns partially or noisily present in the input. Mathematically, the key point is that if we break up the time interval within each fixation into a sequence of times $\{t_n\}$, then

$$K_\alpha(x, y, t_n) = F(J_\alpha(x, y, t_n), \ldots, K_\beta(u, v, t_{n-1}), \ldots)$$

where F represents the iteration of some algorithm combining retinal input with feedback. The idea that an iterative algorithm is carried out in the thalamocortical loop has received

interesting experimental confirmation in the oscillations observed in Ribary et al. (1991).

Experimental evidence that the LGN is doing much more than gating comes from Murphy and Sillito (1987). They report that most relay cells in the cat LGN are *end-stopped*, i.e., they respond to moving bars of a certain length, but their response drops off markedly when the length of the bar is increased. However, if cortical feedback is removed by destruction of those visual areas projecting to the LGN, the end stopping ceases. This observation is puzzling because the cells in the cortex which project to the LGN do not show end stopping, but the result is that the feedback has strong, complex inhibitory effects. One possible interpretation is that, under cortical feedback, the LGN cells become more narrowly tuned to specific types of local features, e.g., bars of a particular length or curved bars.

To clarify the various models for active roles of the thalamus, I would like to distinguish two distinct ideas in Harth's model. One is the concept of generating completed sensory patterns from memory when the actual stimulus is noisy and incomplete. The roots of this idea go back at least as far as MacKay (1956), who proposed that feedback in general may represent a process of actively creating from memory synthetic patterns which try to match as closely as possible the current stimulus. According to this theory, the feedback signal was the pattern synthesized from memory, and the feedforward signal contained features of the stimulus or of the difference between the stimulus and the feedback (the *residual*). MacKay's ideas were developed from a Bayesian statistical perspective in the *Pattern Theory* of Grenander (1976–81, 1994). An influential neural net model of this type is the ADAPTIVE RESONANCE THEORY (ART) (q.v.) of Carpenter and Grossberg (1987). Pece (1992) developed a related theory in which thalamocortical feedback implements MacKay's ideas. He proposes that the area V1 to LGN pathway carries *negative* feedback so that after iteration, area V1 converges to a pattern of activity whose feedback closely matches the most salient visual patterns in the stimulus and then cancels the retinal traces in the LGN. We call this type of algorithm *feedback-pattern-synthesis*.

A second computational idea in Harth's theory is that the LGN is an internal sketchpad, or an *active blackboard*, on which various patterns can be written, misleading patterns can be supressed, and a best reconstruction can be generated. Evidence for this hypothesis comes from Sillito et al. (1994). They find that cortical feedback synchronizes the firing of specific sets of LGN relay cells, namely those responding to a common feature, like parts of the same edge. This type of function is a kind of enhanced image processing, which can be considered independently of the idea of feedback-pattern-synthesis. The concept of a blackboard was first introduced in computer science in the HEARSAY speech project at Carnegie Mellon University, in the 1970s; it refers to a small shared common memory on which multiple experts can read and write, possibly combining their results by the convergence of weak pieces of evidence or one expert vetoing another if their conclusions conflict (see DISTRIBUTED ARTIFICIAL INTELLIGENCE).

In Mumford (1991, 1992, 1994), an integrated theory for the corticothalamic and corticocortical feedback loops was proposed in which the active blackboard image processing role is assigned to the thalamocortical feedback loop, while feedback-pattern-synthesis is the role assigned to the corticocortical feedback loops. The activity of each specific nucleus in the thalamus is assumed to act as a blackboard in representing the current view of the world for those areas of the cortex to which it is connected. The thalamic buffers connected to primary and secondary sensory areas and to associational and multimodal areas will contain progressively more abstract representations of some aspect of the world. Each view is based on data coming from connections to the external world through subcortical pathways and on data computed in the cortex and written on the thalamus by cortical feedback. These top-down data may be used to enhance the bottom-up signals, to reconstruct missing data, or to externalize for further processing views of the world created purely by mental imagery. For instance, the actual retinal signals are both noisy and complex, with multiple physical effects creating a highly coded, but incomplete view of three-dimensional objects and their illumination. The cortex must disentangle the effects of lighting, texture, shape, and depth. The hypothesis is that, instead of copying the image into successive buffers in a feedforward architecture as various remembered patterns are identified and used to construct the world scene behind the viewed image, small numbers of buffers in the thalamic nuclei are used to combine the reconstructions made by various cortical experts. Many cortical experts can search independently for a large variety of patterns in the image, sending them all to the thalamus, where a kind of voting takes place by summation in the dendritic arbors of the relay cells and by inhibition through the interneurons. Thus, the active thalamic blackboard can be used to decide which representation is the most successful, rejecting the weaker matches. The resulting pattern of activity is then sent back to the cortex as an enhanced view of the world.

The function of corticothalamic feedback remains a matter of speculation. It is hoped, however, that these speculations will stimulate another generation of more sophisticated experiments, with more complex stimuli, that will enable us to see further.

Road Maps: Mammalian Brain Regions; Vision
Related Reading: Electrolocation; Selective Visual Attention; Stereo Correspondence and Neural Networks; Visual Schemas in Object Recognition and Scene Analysis

References

Carpenter, G., and Grossberg, S., 1987, A massively parallel architecture for a self-organizing neural pattern recognition machine, *Comput. Vis. Graph. Image Proc.*, 37:54–115.

Crick, F., 1984, Function of the thalamic reticular complex: The searchlight hypothesis, *Proc. Natl. Acad. Sci. USA*, 81:4586–4590.

Desimone, R., Wessinger, M., Thomas, L., and Schneider, W., 1990, Attentional control of visual perception: Cortical and subcortical mechanisms, *Cold Spring Harbor Symp. Quant. Biol.*, 55:963–971.

Grenander, U., 1976–81, *Lectures in Pattern Theory I–III*, New York: Springer-Verlag.

Grenander, U., 1994, *Pattern Theory*, New York: Oxford University Press.

Harth, E., Unnikrishnan, K. P., and Pandya, A. S., 1987, The inversion of sensory processing by feedback pathways: A model of visual cognitive functions, *Science*, 1987:184–187.

Jones, E. G., 1985, *The Thalamus*, New York: Plenum. ◆

Koch, C., 1987, The action of the corticofugal pathway on sensory thalamic nuclei: A hypothesis, *Brain Res.*, 23:399–406.

MacKay, D., 1956, The epistemological problem for automata, in *Automata Studies* (C. E. Shannon and J. McCarthy, Eds.), Princeton, NJ: Princeton University Press, pp. 235–251.

Mumford, D., 1991, On the computational architecture of the neocortex, pt. I, The role of the thalamo-cortical loop, *Biol. Cybern.*, 65:135–145.

Mumford, D., 1992, On the computational architecture of the neocortex, pt. II, The role of the cortico-cortical loop, *Biol. Cybern.*, 66:241–251.

Mumford, D., 1994, Neuronal architectures for pattern-theoretic problems, in *Large Scale Neuronal Models of the Brain* (C. Koch, Ed.), Cambridge, MA: MIT Press, pp. 125–152.

Murphy, P., and Sillito, A., 1987, Corticofugal feedback influences the generation of length tuning in the visual pathway, *Nature*, 329:727–729.

Pece, A. E. C., 1992, Redundancy reduction of a Gabor representation: A possible computational role for feedback from primary visual cortex to lateral geniculate nucleus, in *Artificial Neural Nets* (I. Aleksander and J. Taylor, Eds.), Amsterdam: Elsevier Science.

Ribary, U., Ioannides, A., Singh, K., Hasson, R., Bolton, J., Lado, F., Mogilner, A., and Llinás, R., 1991, Magnetic field tomography of coherent thalamocortical 40-Hz oscillations in humans, *Proc. Natl. Acad. Sci. USA*, 88:11037–11041.

Sherman, M., and Koch, C., 1986, The control of retinogeniculate transmission in the mammalian LGN, *Exp. Brain Res.*, 63:1–20.

Sillito, A., Jones, H., Gerstein, G., and West, D., 1994, Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex, *Nature*, 369:479–482.

Singer, W., 1977, Control of thalamic transmission by corticofugal and ascending reticular pathways in the visual system, *Physiol. Rev.*, 57:386–419.

Singer, W., and Schmielau, F., 1976, The effect of reticular stimulation on binocular inhibition in the cat LGN, *Exp. Brain Res.*, 14:210–226.

Time Complexity of Learning

J. Stephen Judd

Introduction

This article delves into questions about learning, specifically the number of computing steps required to support simple associative memory in neural networks, and includes comments on mistake bounds.

Unfortunately, many of the learning algorithms reported in the literature are often very slow even for the small network sizes (tens or hundreds of nodes) used in experiments; they have all been unacceptably slow in large networks. It is clear that we need to be able to scale up our applications to much bigger networks, and so we need to understand the cause of these learning difficulties and how to deal with large sizes.

The type of learning investigated here is known as supervised learning. In this paradigm, input patterns (called *stimuli*) are presented to a machine paired with their desired output patterns (called *responses*). The object of the learning machine is to remember the associations presented during a training phase so that in a future retrieval phase the machine will be able to emit the associated response for a given stimulus.

It is assumed that the networks can change their behavior (by changing their weights); in the work reported here, this change does not involve altering the connectivity structure.

A set of (stimulus, response) *pairs*, or SRpairs, will herein be called a *task*. When a small task is drawn randomly from a large set of possible pairs, the literature usually calls it a *sample*.

An *architecture* specifies the input lines, the connectivity from each node to others, and which nodes will be network outputs. It includes all data about a circuit except what functions the nodes perform. In most of this article, we consider only feedforward networks for which a stimulus fields a unique response.

Each node in a network is designed to compute one of a certain family of *node functions*. Typical examples pass the weighted sum of inputs through a step function or sigmoid function.

A *configuration*, $F = \{f_1, f_2, \ldots, f_n\}$, of a network is an assignment of one function from the node function set to each node in the architecture, to specify what that node computes. An architecture, A, and a configuration, F, together define a mapping from the space of stimuli to the space of responses. This mapping describes how the network will behave during retrieval.

A goal of neural network learning research has been to find a "learning rule" that each network node can follow to adjust its weights, i.e., to find a configuration such that the retrieval behavior of the whole network eventually implements some desired mapping from stimuli to responses. It was hoped that a learning rule, and especially some biologically plausible learning rule, would work for any network design. Many researchers have developed candidates for such a learning algorithm, as this book attests, but much of this article reports studies where the biological nature is sacrificed, in recognition of the fact that the form of the learned representations may be as important as how they are obtained.

There are several measures of the difficulty of learning: one is the amount of time it takes to learn, another is the amount of data it takes to learn, and another is the number of mistakes that will be made during a learning process. For each of these questions, there are other issues that have to be specified: how "neural" the algorithms are, how exactly correct they need to be, how dependable they need to be, how they get their data, what they are allowed to manipulate, and how helpful the teacher is. This article deals with some questions regarding time and mistakes; see VAPNIK-CHERVONENKIS DIMENSION OF NEURAL NETWORKS for some data complexity issues.

Neural Algorithms

The original perceptron had the impact it did because its learning rule was deemed to be "biologically plausible." This attribute is still a hallmark of many learning schemes. Typically, a sample of data is collected and then repeatedly presented to the machine while it incrementally alters its hypothesis toward the correct one.

Rosenblatt (1961) and others proved a theorem stating that the various perceptron learning rules eventually converge to correct weights if such weights do exist (i.e., if the task is linearly separable). This development demonstrated that the perceptron would learn in finite time, but it gave no scaling information. The scaling issues are with respect to s, the number of input lines in the stimulus vector.

Muroga (1965) showed that there are linearly separable functions whose weights are approximately as large as 2^s. Thus, even when the function is performable, it will take the various perceptron learning rules $\Omega(2^s)$ adjustments before getting acceptable weights. Hampson and Volper (1986) extended the argument to the average case (as opposed to the worst case) and derived a bound of $\Omega(1.4^s)$. (The Ω notation is like the O notation in that it makes no claim about the constant multiplier, but whereas $O(f(n))$ says that the scaling is no worse than $f(n)$, $\Omega(f(n))$ claims that the scaling is at least as bad as $f(n)$.)

Tesauro (1987) studied time-scaling issues in some simple families of multilayered networks and measured learning time