Concepts in Neuroscience, Vol. 3, No. 1 (1992) 1–53

WEBS, CELL ASSEMBLIES, AND CHUNKING IN NEURAL NETS

WAYNE A. WICKELGREN
Psychology Department, Columbia University, New York, NY 10027, USA

Received May 18, 1992

ABSTRACT

This paper has three major foci: (a) describing a general, but incomplete, theory of the representation of ideas, learning, and thinking in the cerebral cortex, (b) describing some specific structural and dynamic models relevant to the theory, and (c) reporting the results of some analytic and empirical-mathematical investigations of their properties.

Ideas can be represented in the cerebral cortex by ***webs*** (innate cell assemblies), using sparse coding with sparse, all-or-none, innate linking. Each neuron connects to less than the square root of the number of neurons in the cortex. Recruiting a web to represent a new idea is referred to as ***chunking***. The innate links that bind together the neurons of a web are the basal dendritic synapses. Learning is modification of the apical dendritic synapses that associate neurons in one web to neurons in another web.

The ***minint*** (minimum internal connectivity) of a set of neurons is the minimum number of innate links that any neuron in the set receives from other neurons in the set. The ***maxext*** (maximum external connectivity) of a set of neurons is the maximum number of innate links that any neuron outside the set receives from neurons in the set. A ***web*** is a set of neurons whose minint is greater than its maxext. A web is an attractor state in a dynamical system composed of a binary-link neural net with a threshold activation function and the threshold set between the maxext and the minint. Webs are innate resonances of a neural net.

Thinking has four phases: two main phases — ***retrieval*** of familiar ideas and ***chunking*** of new ideas, with two subphases of each — ***selection*** and ***completion***. In the retrieval-selection phase, learned apical synapses select an initial set of active neurons in a module. In the retrieval-completion phase, activation converges on an asymptotic terminal set of active neurons that, ideally, is the familiar web having the greatest similarity to the initial set. If retrieval-completion fails, the chunking selection phase is entered in which all apical synapses (not just the learned ones) are used to select an initial set for chunking-completion. In the chunking-completion phase, the module converges on an asymptotic terminal set of neurons that, ideally, is a new web that comes to represent the set of constituent webs that produced the apical input to the module in the chunking-selection phase. To achieve this, at the end of chunking-completion, Hebbian contiguity conditioning strengthens all of the active apical input synapses to the neurons in the web, so that sufficiently large subsets of

these learned apical synapses will, on some future occasion, be able to select an initial set for retrieval that will reactivate the same web in retrieval-completion.

Some findings are: (a) There are essentially no webs in asymmetric regular random nets, but there are more webs than neurons in some symmetric regular random nets and in both asymmetric and symmetric proximity nets, where the linking probability between neurons declines with distance. (b) A model of activation dynamics in chunking-completion was devised that achieved 100% convergence on one of 597 webs in a 289-neuron symmetric proximity net from 1681 randomly selected initial states. Its distinctive features were a shelf-like threshold as a function of the number of active neurons in the module, the use of random noise in the threshold, and the assumption that neurons either fire twice or that synapses remain active for two consecutive time periods after receiving input. (c) For symmetric proximity nets, the retrieval completion obtained so far has been very fast and reasonably accurate. However, accuracy is well below that obtainable by an ideal decision maker with unlimited time to choose among ideas that are randomly selected sets of neurons. Improvements in retrieval-completion for webs can probably be obtained by changes in net structure and dynamics.

Keywords: Cell assembly, web, idea representation, thought, chunking, retrieval, neural net, cerebral cortex, pyramidal neurons, basal dendrites, apical dendrites, proximity net, symmetric connection, activation dynamics, attractor, cortical phases, selection, completion, neuromodulation, threshold control.

1. Mind: Ideas, Thoughts, and Thinking

This section describes some basic psychological concepts and principles that motivate the neural models described in the rest of the paper.

Assume that the mind contains a set of N ideas. ***Ideas*** are representational atoms, such as: a 45° line segment located 30′ to the left of the fovea, the letter 'p', pressing the tongue against the upper front teeth, a sound of 680 Hz, the word 'dog', the image of a particular dog, the concept of a particular individual dog, the concept of a member of the set of dogs, the concept of the set of all dogs, the proposition that dogs eat meat, etc.

I use "idea" very generally to mean any of the fundamental units of representation in any module of the mind, not just those modules concerned with the representation of concepts and propositions in semantic memory. Thus, any sensory of motor feature, segment, image, concept, proposition, action, or mental procedure is an idea.

Ideas have at least two states of ***activation*** in the mind, active and inactive, and there may be intermediate degrees of activation. Ideas also have various degrees of ***excitation***, which is the potential for future activation of an idea. Ideas with a high degree of excitation may already be active or may become active with a small amount of additional input excitation from associated ideas or the external world.

Only active ideas produce associative input that may add or subtract from the excitation or other ideas. Ideas with levels of excitation which are below the activation threshold do not provide associative input to other ideas. This is what it

means for an idea to be active, namely, that it provides the excitatory or inhibitory input to the excitation of other ideas.

In a two-state activation model, a ***thought*** is a set of activated ideas. In a continuous activation model, a ***thought*** is the N-dimensional activation vector that represents the activation of each of the N ideas in the mind. ***Attentional set*** (nonassociative short-term memory) is the N-dimensional excitation vector that represents the excitation of each of the N ideas in the mind.

Thinking is a sequence of thoughts. The successor thought is determined by a combination of sensory input, associative input from the ideas in the prior thought, and the persisting (decaying) excitation of each possible idea that results from prior sensory and associative input.

For the purposes of this paper, one may regard a set of active ideas as a conscious thought. In a larger context, I would probably not want to identify conscious thought with the entire set of active ideas, but only with the subset of active ideas concerned with semantic memory, language, imagery, and emotion.

The number of active ideas composing a thought is a very tiny portion of all the ideas that the mind contains in its long-term memory. The maximum number of ideas that may be simultaneously active in a thought is the ***attention span***. Sometimes it is alleged that attention span is on the order four or five ideas, whereas the total number of ideas in an adult human mind is surely in the millions or billions. I suspect that there are separate attention spans for parts of the mind, which I will refer to as modules. Most the this paper is concerned with modeling a single module of the mind, but I am not prepared to specify the nature of our limited attention capacity in either the mind or a module, beyond the principle that the number of active ideas is a tiny fraction of the number of ideas in the module.

The primary reason for mentioning limited attention spans is that it provides a major source of motivation for the chunking learning process. The chunking learning process recruits a new idea to represent each thought and strengthens associations in both directions between the new chunk idea and its constituents. Thus, the inventory of ideas in the mind does not remain constant over time, but rather increases due to chunking.

There are two primary reasons for chunking: First, chunking helps us to overcome the limited attention span of thought by permitting us to represent thoughts of arbitrary complexity of constituent structure by a single (chunk) idea. Second, chunking permits us to have associations to and from a chunk idea that are different from the associations to and from its constituent ideas. This is very important for minimizing associative interference.

We might want to assume that the number of ideas in the mind also decreases over time due to ***forgetting***. However, if forgetting is a continuous process, it may be better to assume that the number of ideas is steadily growing, but another property called the ***availability*** of an idea both increases and decreases. Forgetting and availability are not studied in this paper.

2. Neural Representation of Ideas and Associations

This paper develops one possible neural net representation of some of these psychological concepts and principles. The theory aims to model a group of nearby pyramidal neurons in the cerebral cortex. The most basic assumption is that an idea is represented by a set of strongly interconnected neurons similar to a Hebbian cell assembly [1].

Some terminology is as follows: ***Ideas*** are mental entities and the set of neurons representing an idea is a ***cell assembly***. ***Association*** is a mental, not a neural, relation between ideas, but, because of its mnemonic value, I will refer to the synapses on the apical dendrites of cortical pyramidal neurons as associative synapses. ***Neurons, synapses, connections***, and ***links*** are entities in the brain or in neural net models of the brain. A ***link*** is the set of synapses of a given type from one neuron to another. Thus, neuron-i can have at most one link of a given type to neuron-j, but that link may be composed of one or more synapses. ***Synapses*** and ***links*** can have many degrees of strength, though sometimes they are assumed to be all-or-none, that is, having only the values 1 or 0, respectively. ***Connections*** are always all-or-none, that is, two neurons are either connected or they are not.

2.1. *Cell assemblies*

Hebb proposed that ideas are represented in the cerebral cortex by overlapping sets of neurons called cell assemblies [1]. Hebb's definition of cell assemblies was not completely precise, but, implicitly or explicitly, Hebb's cell assemblies had seven properties, the first six of which have been important parts of many subsequent hypotheses concerning the neural representation of ideas. The present paper aims to develop the seventh property as well.

Some of these properties could be called structural in that they refer only to the graph-theoretic properties of a neural net, the types of sets of neurons that represent ideas and their synaptic interconnections. Other properties could be called dynamic in that they depend on assumptions concerning neural excitation, persistence, thresholds, and activation of neurons, as well as on structural properties of the net.

First, Hebb assumed ***overlapping set coding*** of ideas (see p. 196 of [1]). The same neuron could be a part of many different cell assemblies. Cell assemblies are overlapping sets of neurons — a structural property.

Second, Hebb implicitly assumed ***sparse coding*** of ideas, that is, any individual cell assembly contained a very small subset of all of the neurons in the cerebral cortex.

Third, Hebb assumed a structural ***integration*** property, that cell assemblies are sets of neurons with a relatively high density of excitatory synaptic interconnections. In nets with sparse connectivity, such as the cerebral cortex, most random sets of neurons cannot be cell assemblies and represent ideas, because they are not sufficiently densely interconnected by excitatory synapses.

Fourth, cell assemblies have a dynamic ***persistence*** property, that activation of a cell assembly will persist for a time via reverberatory feedback due to the high density of excitatory synapses among the neurons of the cell assembly.

Fifth, cell assemblies have a dynamic ***completion*** property, that activation of a large enough subset of a cell assembly results in activation of the complete cell assembly. Completion depends both on the structure of the connections among the neurons in a net and on the rules for activation dynamics of neurons. Legéndy [2] was the first to study the completion of cell assemblies within a precise mathematical model, referring to it as ***ignition*** of a cell assembly. Braitenberg [3] and Palm [4] made further important contributions, with Palm being the first to note that the ignition of cell assemblies is essentially the same property as pattern completion in an associative memory.

Sixth, there is the famous ***Hebbian learning*** postulate that correlated activation of two neurons strengthens any synapse between them. Hebbs' associative synaptic learning hypothesis became famous independent of its use in establishing cell assemblies.

Seventh, Hebb anticipated Miller's [5] ***chunking*** learning process for the representation of complex thoughts as unitary ideas. Hebb suggested that a new cell assembly T for an entire triangle emerges during the course of a phase sequence incorporating the activation of the three cell assemblies, a, b, and c, representing the three vertices of the triangle. Hebb emphasized that "The resulting superordinate system must be essentially a new one, by no means a sum or hooking together of a, b, and c".

2.2. *Excitation, activation, inhibition, and threshold of neurons*

The ***activation*** of a neuron is its output state, measured at its (presynaptic) axonal terminals, which are all assumed to be in the same state at time t. Activation is sometimes represented on a continuous scale, e.g. real numbers between 0 and 1, representing the neuron's rate of firing, but the models in this paper will assume all-or-none activation, namely, a spike or no-spike at time t.

The ***excitation*** of a neuron is its input state, measured at the cell body. Excitation is represented on a continuous scale by a non-negative real number. In the cortical models of this paper, excitation represents the summed dendritic potential at the cell body of a pyramidal neuron due to all excitatory synapses on that neuron. Total excitation is divided into a basal dendritic component and an apical dendritic component that may have different relative weighting and different rates of decay at different phases of thinking.

The models in this paper represent a very short-term memory at the level of the individual synapse by the assumption that each basal dendritic excitatory synapse makes its contribution to total excitation for two consecutive time steps after receiving input from its activated presynaptic neuron.

Inhibitory synapses are not explicitly represented as such in any neural model in this paper. Some types of ***inhibition*** may play a role in determining relative weightings of apical and basal dendritic excitation and decay rates. The primary way in which inhibition is represented most explicitly in these models is by setting the ***threshold*** for activation of a neuron. Greater inhibition raises the threshold for firing — outputting a spike. The simple threshold rule is used throughout the paper, namely: activation is 1 (spike) at time t, if and only if excitation equals or exceeds the threshold at time t; otherwise, activation is 0 (no spike).

Excitation (potential) and activation (spiking) of a neuron are different concepts from the excitation and activation of an idea represented by a set of neurons. The molar psychological concepts of excitation and activation of ideas may well be definable from the concepts of excitation and activation of neurons. However, the relation is not identity unless one subscribes to the specific neuron hypothesis that each idea is represented by a single neuron.

2.3. *Association of ideas*

2.3.1. *Similarity and contiguity*

Mental associations between ideas may derive in part from overlap in their cell assemblies (the similarity factor) and in part from strong links between the assemblies (the contiguity factor). It is important to remember that the psychological associative relation between ideas need not only be represented neurally by strengthened synapses between neurons in the associated cell assemblies, but may also be partially represented by the neural overlap of associated cell assemblies. However, this paper does not use neural overlap, only learned synaptic association, as a basis for the association between cell assemblies.

2.3.2. *Associative link types*

There is probably no more important unsolved issue in the study of the mind than the semantics of the association relation(s) between ideas. How many types of associations are there between ideas and what are they? I have grappled with this problem for decades and come to no firm conclusion. One possibility is that there are only two basic semantic types of association relations between ideas in the human cognitive mind: A is a constituent of B and B is a chunk of A. Perhaps both directions of the constituent relation established by chunking could be mediated by a single type of neural link. Perhaps there are multiple types of constituent associative links. In addition to constituent links, there may be learned sequential links for sequential activation of the constituents of a chunk idea representing an ordered set such as a procedure. These questions are beyond the scope of this paper.

Only one type of learnable associative link between pyramidal neurons will be assumed in this paper, and those associative links will not even be modeled, except in a very reduced way as the initial input to a set of neurons. The only excitatory links to be represented by link matrices in this paper and modeled extensively are

innate links that are presumed to bind neurons together into innate cell assemblies called webs.

2.4. *Link types, link matrices, and neuromodulation*

Neurons are known to be connected by synapses of different types — excitatory vs. inhibitory, but also several types of excitatory and inhibitory synapses. Dale's Law is that any given neuron has output synapses of only one type. Even if a neuron secretes the same mix of transmitters from each of its output synapses, from a functional standpoint, a synapse type is determined as much by the response of the postsynaptic neuron to the transmitter(s) as by the transmitter(s) secreted by the presynaptic neuron. If the postsynaptic response is different for different sites, then functionally the synapses are of different types.

There are at least two classes of neocortical neurons, pyramidal and stellate cells, with different neurotransmitters, and several types of subcortical neurons that send outputs to the neocortex. It is certain that cortical neurons have input synapses (inlinks) of more than one type, excitatory and inhibitory, and it is likely that many have two or more types of excitatory inlinks and two or more types of inhibitory inlinks.

In this paper, any single ***link matrix*** or ***connection matrix*** is restricted to representing synapses of a single type. Thus, in general, a neural net model may require more than a single link matrix to represent the strengths (or other properties) of the synapses between the neurons in the net.

The theory presented in this paper assumes two types of excitatory links between cortical pyramidal neurons: (a) innate ***binding links*** between pyramidal neurons in the same cell assembly that cause all the neurons in the set to be activated together and (b) learned ***associative links*** that represent associations between cell assemblies. Binding links are presumed to be excitatory synapses on the basal dendrites of pyramidal neurons, and associative links are presumed to be synapses on the apical dendrites of pyramidal neurons, as illustrated in Fig. 1. Basal dendrites branch extensively near the cell body and receive input synapses from nearby pyramidal neurons [6]. Apical dendrites project upward for some distance from the cell body, usually branching extensively in the uppermost layers of the cortex where they receive synaptic input both from nearby neurons and from neurons in remote areas of the cerebral cortex [6]. The derivations, calculations, and simulations of the properties of neural nets discussed in this paper use only the binding link matrix, but the theory assumes the existence of an independent associative link matrix as well.

The theory assumes two types of inhibitory links: (a) ***activation threshold control links*** and (b) ***excitation erasure links***. Threshold control links raise the threshold for activation, but do not cancel or erase the excitation of the pyramidal neuron caused by previous excitatory input to the dendrites. The general purpose of threshold control links is to keep the number of active neurons in a module within

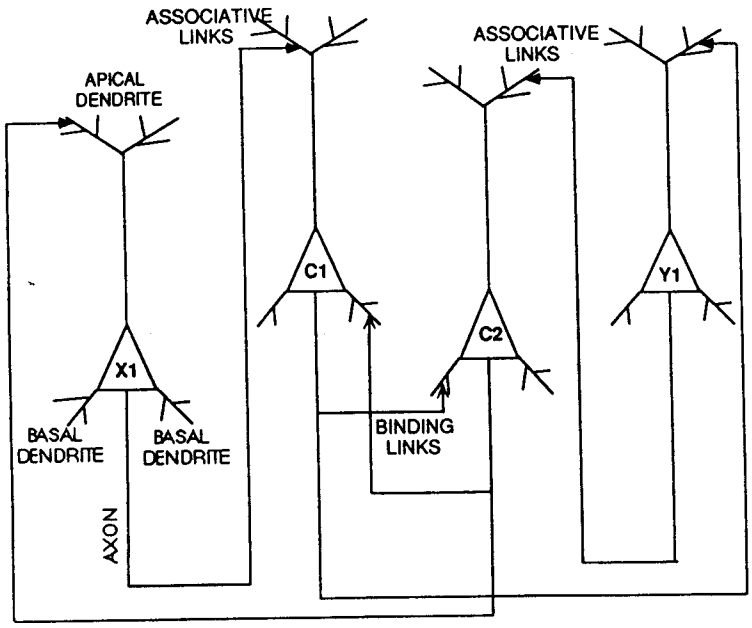

Fig. 1. Apical vs. basal dendritic excitatory synapses among pyramidal neurons in the cerebral cortex. A cell assembly of 100–10,000 pyramidal neurons is assumed to represent an idea. Neurons $C1$ and $C2$ are two neurons belonging to the same cell assembly C. $C1$ tends to activate $C2$ and vice versa via the innate binding links (synapses) that connect them into a cell assembly. The synapses on basal dendrites of pyramidal neurons are presumed to mediate these binding links among the neurons in each cell assembly. A set of cell assemblies that are activated in close temporal contiguity recruit a new chunk assembly to represent the entire set. $X1$ is one neuron from constituent assembly X, and $Y1$ is one neuron from constituent assembly Y. Via the chunking process, assembly C has come to represent the combination of assemblies X and Y. Assembly C is the chunk idea that represents the combination of the ideas represented by the constituent assemblies X and Y. The neural substrates of this chunk-constituent relation are the learned associative links from neurons such as $X1$ and $Y1$ in the constituent assemblies to the neurons in the chunk assembly C and the learned associative links in the reverse direction from the chunk neurons to the constituent neurons. Note that it is plausible to assume a high degree of symmetry in the binding synapses between individual neurons within a cell assembly — i.e. $C1$ links to $C2$ if $C2$ links to $C1$. Such binding symmetry may be useful in creating cell assemblies. However, it is not so plausible to assume a high degree of symmetry in the associative synapses between individual neurons in different assemblies, nor would this serve a useful purpose, though it is essential that there be symmetry at the module level for associative synapses, namely, that the constituent modules that send associative links to some chunk module should also receive associative links from the chunk module. In the figure, neuron $X1$ has an apical synapse on neuron $C1$, but $C1$ does not have an apical synapse on $X1$, though $C1$ may synapse on some other neuron in the X assembly. $C1$ is shown as synapsing on $Y1$, while $Y1$ does not synapse on $C1$. However, $C1$ and $C2$ are shown to have symmetric binding synapses.

the target range for the desired size of cell assemblies, not allowing the number of active neurons to decrease to zero or increase in an epileptic explosion.

What has been called reciprocal inhibition and lateral inhibition may both be implemented by control links, as can changes in threshold control with the phase of thought. Threshold control inhibition can often be modeled abstractly in the

dynamic laws of a neural net by making the threshold of neurons vary with the phase of thought and the total activation of neurons in the module.

Chandelier cells, which form presumably inhibitory synapses on the initial segments of the axon of pyramidal neurons [7], are in a perfect position to control the threshold for activation of a pyramidal neuron without affecting the state of depolarization of the cell body or dendritic tree, which carries the memory for the excitation of the pyramidal neuron.

By contrast, erasure links permanently cancel the effects of prior excitatory input to a pyramidal neuron. Excitation is presumed to decay passively over time in the absence of inhibition. Why would one want active inhibitory erasure? One wants erasure inhibition in cases where the idea represented by a set of pyramidal neurons has had its turn on the mental stage, and it is time to activate other ideas that were temporarily bypassed for processing and/or that are just now being excited. To reduce the noise level for idea recognition, it is desirable to clear off the desk that has already been processed so that it does not interfere with the processing of other ideas.

Erasure links serve the function of ***self-inhibition*** of an idea that has already been activated. At one phase of thought, an idea is assumed to inhibit itself so as to terminate the current thought and go on to the next thought. Self-inhibition is easily modeled in the dynamical laws without the need for explicit representation of inhibitory neurons or erasure inhibitory synapses in link matrices.

Basket cells, which form presumably inhibitory synapses on dendrites and cell bodies [7], are in a good position to erase the prior state of excitation (depolarization) of a pyramidal neuron.

The theory also assumes one or two neuromodulatory link types that modify the strengths of all excitatory links of a certain type in the module. Since such neuromodulation has a common effect on all links of a certain type and does not depend on the specific pair of neurons being linked (though it may vary with the strength of the link), such modulation can be abstractly modeled in dynamical laws and does not require matrix representation of each modulatory synapse.

2.5. *Apical vs. basal dendritic systems of synapses*

I got the idea of distinguishing the functions of apical and basal dendrites from Braitenberg's distinction between the A (apical) and B (basal) systems of synapses among cortical pyramidal neurons [8]. Braitenberg [6] considers this distinction to be similar to the distinction between an ametric and metric system of synaptic connections proposed by Palm and Braitenberg [9], where "metric" means that the probability of a synaptic connection decreases with increasing distance and "ametic" means that the probability of synaptic connection is independent of the distance between the neurons. While basal synapses are apparently entirely local and metric, some apical synapses are local and presumably metric while others are remote and presumably ametric. Thus, I will not identify apical with ametric.

Braitenberg [8] used the *A* system for binding together diffuse (global) cell assemblies. ***Diffuse cell assemblies*** are composed of neurons from many different modules (areas, regions) of the cerebral cortex. Braitenberg used part of the *B* system to bind together neurons in local cell assemblies. ***Local cell assemblies*** are composed entirely of neurons in the same module. Braitenberg used another part of the *B* system to associate all cell assemblies (diffuse or local).

Much later I discovered that Kohonen, Lehtio and Rovamo had distinguished the functions of the apical and basal dendritic systems in a manner closer to mine. However, I assume that only apical synapses can be modified and that basal synapses are innate and unmodifiable, whereas they assumed modifiable basal synapses and unmodifiable apical synapses [10]. Kohonen *et al.* assumed that the function of basal synapses is to associate the neurons representing parts of a pattern. They did not refer to the set of neurons representing a pattern as a cell assembly, but they wanted the basal synapses to function to complete the neural representation of any pattern starting from a subset, just as do cell assembly theorists.

Like Braitenberg [8], I distinguish between synapses that bind neurons into a cell assembly (binding synapses) and those that associate two assemblies (associative synapses), but I follow Kohonen *et al.* in assuming that cell assemblies are purely local, that only basal synapses bind neurons into a cell assembly, and that apical synapses deliver input from other modules.

However, in my model the only learning occurs at the input apical synapses. Apical synapses are considered to be the site for learned associations between cell assemblies. Associative input to apical dendrites may be considered analogous to sensory input for primary sensory areas of the cortex. Of course, thalamic sensory input is first delivered to spiny stellate neurons. According to Douglas and Martin [11], spiny stellates project to basal dendrites, rather than apical dendrites. If this is so, then sensory input to cortical pyramidal neurons must be handled differently from cortical associative input, and the analogy is flawed.

Finally, I assume that associations can be learned between cell assemblies in the same module as well as between cell assemblies in different modules. Thus, the model assumes that pyramidal neurons within a cortical module form apical synapses with other pyramidal neurons in the same module, as well as basal synapses. All of this is consistent with current knowledge of synaptic connections in the cerebral cortex [11].

In my model, basal dendritic synapses integrate neurons of the same cell assembly so that they will have the properties of persistence and completion in retrieval. Persistence means that, once all the neurons of a cell assembly are activated, they will remain active for a period of time until fatigue or specific inhibition terminates activation.

Completion in retrieval means that associative synaptic links from other ideas to a target idea need not be numerous enough and strong enough to activate all of the neurons in the cell assembly representing a target idea. Associative input needs only to activate a subset of the target assembly. This subset of the target assembly

is part of the initial set from which retrieval completion starts. The initial set will probably also contain activated (noise) neurons that do not belong to the target assembly. For completion to be successful, the initial set must be informationally sufficient to specify one and only one target assembly by having greater overlap with the target assembly than with any other assembly. The basal dendritic synapses then mediate the activation of the remaining neurons in the target assembly and the deactivation of the noise neurons.

Associative input to apical synapses activates initial sets of pyramidal neurons that serves as the starting point for both chunking and retrieval by the basal system, but the apical system is not investigated in this paper.

3. Coding of Thought

3.1. *Specific neuron coding — grandmother cells*

The simplest way for neurons to code ideas is specific neuron coding, so called in honor of Johannes Müller's doctrine of specific nerve energies, of which it is a simple generalization. This is what Horace Barlow and others have called the *grandmother cell* theory of coding in the brain, because it asserts that the internal representative of any idea, including a grandmother, is the activation of a single cell (neuron) [12]. Thinking of grandmother means that the grandmother neuron is firing at a high rate.

3.2. *Multiple neuron coding — giant neurons*

One possible alternative to specific neuron coding of ideas is multiple neuron coding of ideas, where a set of neurons represents an idea, but each neuron is only used in one cell assembly [13]. That is, the sets of neurons representing any two different ideas do not overlap — cell assemblies are non-overlapping. Multiple neuron coding of ideas can be called *giant neuron* coding, because the set of neurons encoding an idea acts much the same as if a single giant neuron were encoding that idea. The properties of multiple neuron coding are similar to specific neuron coding, but there are some important differences.

First, giant neuron coding has greater fault tolerance since the loss of one or a few neurons would presumably only slightly diminish the representation of an idea and the strength of its associations to other ideas, rather than completely abolishing an idea and all of its associations.

Second, each giant neuron has much greater input and output connectivity to other giant neurons than the connectivity of single neurons. If the average number of synapses per neuron is m, and g neurons are combined into each giant neuron, then, on the average, each giant neuron has mg input synapses and mg output synapses.

Third, greater connectivity is obtained at the expense of less representational capacity, since n neurons can code at most n/g ideas with giant neuron coding, whereas the maximum number of ideas is n with specific neuron coding.

If each giant neuron required only a single synapse from any other giant neuron in order to activate it, and if there are 10^{10} neurons in the cerebral cortex and 10^4 synapses per neuron, it would required at least 10^3 neurons per giant neuron to provide complete connectivity of every giant neuron with every other giant neuron. This would reduce the maximum number of representable ideas to 10^7, which might or might not be sufficient to represent all the thoughts a human being can have available at a point in time. However, such limited connectivity provides no fault tolerance for any given association, since it is carried by a single synapse.

To provide 100 synapses between each pair of giant neurons, 10^4 neurons are required for each giant neuron. This reduces the number of possible ideas to 10^6, which seems too small for human thinking. Multiple neuron coding seems unlikely to be the correct model for the representation of ideas in human cognitive thought. With giant neuron coding, it seems likely that there would not be enough idea representation capacity to chunk every thought into a single giant idea, and most thoughts could only be represented as sets of giant neurons. The arguments in favor of chunking given later in this paper argue against such a model, but, of course, I don't know what percentage of our thoughts get chunked. It makes an elegant model to assume that all thoughts get chunked automatically, though, in such a model, one assumes that most chunks see little subsequent use, and so the learned associative links that gave meaning to these chunks are forgotten. In any case, giant neurons have no role in the model developed in this paper, whose focus is to provide a mechanism for chunking.

3.3. *Overlapping set coding — cell assemblies*

More promising than multiple (non-overlapping set) coding is overlapping set coding of ideas. As with giant neuron coding, each idea is represented by a set of g neurons, but the sets for different ideas can overlap, perhaps extensively. Overlapping set coding was employed by Hebb [1] in his cell assembly model. However, Legéndy [2] was the first to develop systematically the theory of overlapping set coding and analyze its properties using powerful probabilistic methods.

Overlapping set coding has the same advantages as giant neuron coding with respect to fault tolerance and enhanced connectivity, but without the obligatory reduction in idea representational capacity [2,14].

3.3.1. *Sparse coding of ideas*

With extensive overlap in the representation of ideas, one might think there would be problems in discriminating different ideas based on proper subsets of the assemblies representing ideas. However, Legéndy [2], Palm [15–18], and Meunier, Yanai and Amari [19] have demonstrated that there can be a high degree of discriminability in the representation of different ideas with overlapping set coding.

As Palm [15–18] and Meunier *et al.* [19] show, the greatest number of discriminable cell assemblies is obtained by using ***sparse overlapping coding*** of ideas in

a neural net, that is, representing each idea by a small subset of all the neurons in the net. Sparse overlapping coding can represent as many or more ideas as there are neurons in the net, with a high degree of discriminability in the representation of different ideas. Sparse overlapping codes are essentially error-correcting codes for ideas that provide a high probability of determining which ideas was intended by choosing the idea with the greatest overlap with any activated set of neurons.

Some of the properties of overlapping set coding can be illustrated with a simple example. Consider a net with ten neurons (labeled 0, 1, ... ,9). Represent each idea by a subset of three neurons.

There are 120 different (unordered) sets of three neurons, which would allow us to code 120 different ideas, an order of magnitude more than the ten ideas that could be coded by specific neuron coding. However, if we were to try to use all of these different 3-neuron codes to represent 120 different ideas, no proper subset (of say two neurons) would be logically sufficient to communicate an intended idea. Thus, the system would have minimal ability to discriminate one idea from another in the presence of any noise in the form of deleted or added neurons. There is no error correction (fault tolerance) in the code.

So we give up on the possibility of making maximum use of the combinatorial possibilities of distributed coding and ask how many ideas can be represented by cell assemblies with three neurons each, such that any subset of two neurons is sufficient to identify uniquely which idea (set of three neurons) was "intended". This is not all of the error correction capability that one wants, but this is a toy example designed to communicate the basic idea. The answer is that one can represent 12 ideas with this degree of idea discriminability, by choosing the following 12 cell assemblies: (012), (034), (056), (078), (135), (146), (179), (236), (247), (258), (389), (459). This is slightly more ideas than can be represented with specific neuron coding, and the idea discriminability (error correction, fault tolerance) is better than for specific neuron coding.

For larger nets and larger cell assemblies, the representational capacity of sparse overlapping coding is probably also greater than specific neuron coding. Other considerations beyond idea discriminability may limit the number of represented ideas to be on the order of the number of neurons in the net, but this is beyond the scope of this paper.

3.3.2. *Thoughts*

In overlapping set coding, each idea is represented by a cell assembly. Each thought is represented by the union of its constituent cell assemblies. Thus, both thoughts and ideas are represented by sets of neurons.

3.3.3. *Chunking*

Since humans seem capable of thinking thoughts composed of ideas that are themselves thoughts composed of ideas, to no known limit, except total memory capacity,

we probably need a mechanism to prevent enormous variation in the size of the set representing each idea. Chunking is such a mechanism. As I envisage chunking, a new cell assembly is recruited to represent a thought (set of cell assemblies), with the new chunk assembly being of the same size as each constituent assembly. The meaning of the new chunk idea might be established in either or both of two ways: chunk-constituent overlap and chunk-constituent association.

First, when a chunk assembly is in the same module as one of its constituent assemblies, the chunk assembly may overlap (share neurons with) the constituent assemblies more than with a random cell assembly. This could provide information relevant to decoding a chunk into its constituents and to reactivating a familiar chunk from its constituents.

Second, and of more general importance, chunks are assumed to be associated to their constituents by learnable links in both directions. Thus, when a chunk assembly is activated to represent the prior thought, a Hebbian learning process is assumed to strengthen apical synapses connecting the neurons representing the constituent assemblies and the neurons representing the chunk assembly.

It is important to note that, both psychologically and neurally, chunking involves something more than associative learning. A new idea representative must be activated to represent a novel chunk. From the standpoint of traditional associative memory, this activation is a fundamentally new process that permits the learning of hierarchical (up and down) associations. Recruiting new idea representatives and hierarchical association was not a feature of traditional models of associative memory prior to the theoretical advances of psychologists such as Miller [5] and Hebb [1].

3.4. *Webs — innate cell assemblies*

3.4.1. *Sparse linking and innate vs. learned cell assemblies*

Using a graph-theoretic approach, Palm made a major advance in the precise formulation of the concept of overlapping cell assemblies [4]. A pure graph-theoretic approach would use only the two-valued (0 or 1) connection matrix that specifies which neurons synapse with which other neurons. In actual fact, Palm permitted a weighted (multivalued) link matrix, but rarely made use of more than two values.

Palm also followed Hebb in assuming that cell assemblies result from learned strengthening of connections between neurons that are contiguously activated, and Palm has done extensive investigations of Hebbian learning [18]. However, there is actually no role for learning in Palm's graph-theoretic definition of cell assemblies — a set of neurons either is or is not a cell assembly based on the current state of the connection matrix, which typically had only 0 and 1 entries [4].

Palm believes that the real connection matrix that defines cell assemblies is generated by a learning process, but he acknowledges the principal difficulty with this assumption, namely, that each cortical neuron only connects to a tiny fraction of all of the other neurons in the cerebral cortex [16]. In a fully connected net or in

a net where new connections can be established between any pair of neurons, cell assemblies can gradually develop as a function of learning in the manner envisaged by Hebb. However, in a relatively sparsely connected net, such as the cerebral cortex, which is currently assumed to have only a very limited capacity to form new connections, I think it is more reasonable to assume that cell assemblies are innate.

Palm has a clever suggestion for escaping this dilemma [16]. He assumes that the neurons of the cerebral cortex are partitioned into modules, with the neurons having a high connection probability, about 0.5, to each neuron in the same module, but a very low connection probability to any neuron in another module. Each pyramidal neuron in the human temporal and frontal cortex is estimated by Cragg [20] to have an average of about 40,000 synapses with other pyramidal neurons. If each assembly-module contained fewer than 10,000 neurons, the required number of binding synapses for each neuron (< 5000) would not unduly deplete the total number of synapses available for associating cell assemblies across different modules.

Assembly-modules are surely all local, that is, within a small compact region of the cortex, which means that if there are any global assemblies, they must be unions of local assemblies, which appears to be what Palm and Braitenberg assume. I think chunking removes the need for global cell assemblies, and I give some arguments against global assemblies as set unions of smaller assemblies in a later section. However, the resolution of this argument is largely irrelevant to the plausibility of Palm's assumption of assembly-modules in which connection probability is high enough to support learned cell assemblies.

The most negative evidence against learned cell assemblies is contained in Braitenberg and Shuz's recent mathematical-anatomical study of the connection probabilities of pyramidal neurons in the mouse cortex [6]. They found that the connection probability of nearby pyramidal neurons was on the order of 0.02. From my reading of Palm's work on this matter, I conclude that this is too small a connection probability to support learned cell assemblies, but I am not certain of this. Furthermore, I do not know the connection probability within clusters of 1000 or so nearby neurons in the human neocortex.

However, Palm's graph-theoretic definition of cell assemblies can also be interpreted as a definition of innate cell assemblies at least as easily as it can be interpreted as a definition of learned cell assemblies, and I do so interpret it. This does not assert that there is no learning in the cerebral cortex, which would be absurd, only that learning plays no role in which sets of neurons are potential cell assemblies, that is, sets of neurons with the potential to represent ideas.

It may be that only some of the cell assemblies with the potential to represent ideas ever actually get activated and come to represent ideas. Once activated, a cell assembly comes to represent an idea by Hebbian strengthening of the modifiable apical synapses on its pyramidal neurons from the neurons in the previously activated constituent assemblies and by Hebbian strengthening of the apical synapses from the chunk assembly to the neurons of its constituent assemblies. I currently assume that there is no modification of the basal synapses that bind together the

neurons within a cell assembly.

Precise characterization of the role of learning vs. innate structure in the mind and the brain is a central problem in psychology and neuropsychology Legéndy [2] began the process of emphasizing the role of innate structure in the definition of cell assemblies by using only innate weak links to bind together the neurons of "minor compacta", which are essentially sub-assemblies of larger cell assemblies called "major compacta". The present theory takes Legéndy's approach one step farther by using only innate synapses to define cell assemblies. Since this could be a mistake and since clear understanding of the brain includes knowing what is learned and what is innate, it is well to acknowledge explicitly the assumption that my cell assemblies use synapses that are innately equal in strength.

3.4.2. *Palm's definition of cell assemblies*

Although Palm's approach is primarily structural, Palm's actual definition of an assembly relied on activation dynamics, rather than being purely structural [4]. Some auxiliary definitions are necessary to define a Palm-assembly. A set X ***ignites*** a set Y, if activation of X eventually produces activation of Y. A ***persistent*** set is a set of neurons that, once activated, remains activated at a given threshold. An ***invariant*** set is a persistent set of neurons that does not recruit additional neurons to the activated set. A set X ***supports*** a set Y (X helps Y to be persistent), if Y is not persistent, but $X \cup Y$ is persistent. A ***Palm-assembly*** is an invariant set such that every persistent subset either supports or ignites the remainder of the neurons in the assembly (at a fixed threshold).

3.4.3. *Structural definition of cell assemblies*

The ***minint*** (minimum internal connectivity) of a set of neurons is the minimum number of innate links that any neuron in the set receives from other neurons in the set. The **maxext** (maximum external connectivity) of a set of neurons is the maximum number of innate links that any neuron outside the set receives from neurons in the set. A ***web*** is a set of neurons whose minint is greater than its maxext. "Web" is a short, elegant name for a cell assembly.

Figure 2 shows an example neural net with nine neurons and an average of 2.9 (bidirectional) links per neuron. Each line between neurons in Fig. 2 represents two links, one in each direction. Thus, the net is symmetric. The net in Fig. 2 contains the six webs listed in Table 1. For each web, the minint is 2, and the maxext is 1. Assume the threshold activation rule that a neuron is activated whenever its excitatory input exceeds a threshold. Set the activation threshold at some value between one and two active inlinks. Then, once any of these webs is activated, it will remain active, because at each time step each neuron in the web receives input from two other active neurons. In addition, no neuron outside the web will become active, because each outside neuron receives input from no more than a single active neuron. Thus, each of these webs is an invariant set. Furthermore, each of these

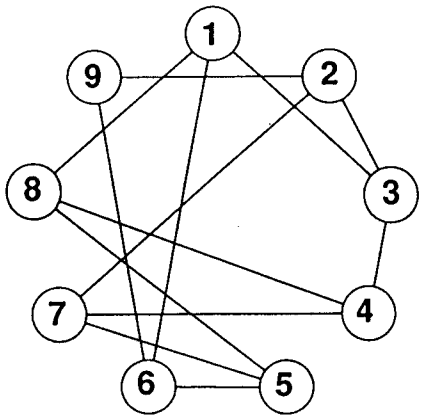

Fig. 2. A tiny neural net with nine neurons that contains the six webs shown in Table 1. This net is symmetric, and each line between neurons represents two links, one in each direction. Each web has a minint of 2 and a maxext of 1, so it will be an equilibrium state using a threshold activation function with a threshold of 1.5 active inlinks.

Table 1. Webs in 9-neuron net of Fig. 2.

Web	Neurons in web				
1	1	3	4	8	
2	1	5	6	8	
3	2	3	4	7	
4	4	5	7	8	
5	1	2	3	6	9
6	2	5	6	7	9

webs is a Palm-assembly. There are no persistent proper subsets of any of these webs.

You can check that each of the six alleged webs is indeed a web by putting your fingers on the neurons in a web, checking that each fingered-neuron receives at least two inlinks from other fingered-neurons, and that each of the unfingered-neurons receives one or fewer inlinks from the fingered neurons.

Note that the six webs overlap extensively, and, thus, each neuron belongs to more than one web. In fact, every neuron belongs to three different webs, except neuron-9, which belongs to two webs.

My definition of cell assemblies is purely structural. The advantages of a structural definition are: (a) It is clear from the properties of the connection matrix alone what kinds of sets are asserted to be cell assemblies. (b) It is easier to determine whether or not any given type of net has cell assemblies, and, if so, how many. One must then study how well any such structural definition fares in achieving the desired dynamic properties with different dynamical models.

The advantage of a dynamic definition is that it incorporates one or more desired dynamic properties into the definition of cell assemblies. One must then demonstrate that such cell assemblies exist and estimate the number for any given type of net with specified structure and dynamics. This is what Palm [4] refers to as "the main problem in the theory of cell assemblies". This problem is considerably simplified by using a purely structural definition, and, although he didn't say so, it is likely that Palm employed a purely structural definition when he determined the number of cell assemblies in various systematically constructed graphs [4]. I presume that definition was nearly equivalent to the one given here.

3.4.4. *Webs are equilibrium states (invariant sets)*

A web is an equilibrium state (invariant set) of a neural net with all-or-none links, a (noise-free) threshold activation function, and the threshold value between the maxext and the minint. That is, once a web is activated in such a dynamical system, it will remain activated and no other neurons will become activated. Webs are innate resonances of such a neural net. Each of the six webs in Fig. 2 is an equilibrium state with the activation threshold set at 1.5 (or anywhere between 1 and 2).

An equilibrium state is one in which one of Palm's invariant sets is activated. Thus, if Palm's definition had been simply that an assembly was an invariant set, then for the case of all-or-none links and a threshold activation function with a properly chosen threshold, the two definitions would have been equivalent. Informally, Palm's more complex definition serves to rule out as assemblies certain sets that are unions of cell assemblies and have very few links between any pair of sub-assemblies.

3.4.5. *Completion — basins of attraction, ignition, discriminability*

To my knowledge, everyone who has studied cell assemblies has wanted cell assemblies to have the dynamic property that a subset of the assembly has the capacity to activate the entire assembly. I too wish my cell assemblies, called webs, to have this property. However, I do not want to define webs by this dynamic property. I want to understand exactly what kinds of sets define a web purely structurally as a set of neurons whose connections to other neurons have certain properties. Then I want to find out whether such sets have the desired dynamic property that a sufficiently large subset of activated neurons from the cell assembly can activate the entire cell assembly, making some assumptions concerning the dynamics of the neural net.

If we assume an activation dynamics model in which each activated neuron fires twice (in two consecutive time periods), then each of the six webs for the net in Fig. 2 will be completed from an initially activated subset that is missing any one of the neurons in the web. That is, for any web with four neurons, any subset of three neurons suffices to ignite the entire web, and for any web with five neurons, any subset of four neurons will ignite the web. For each web with four neurons, two initial sets of two neurons will ignite the entire web and four initial sets of

two neurons will not. For the webs with five neurons, 45% of the subsets of three neurons will ignite the entire web and 55% will not.

For example, consider web 4 consisting of five neurons: 1, 2, 3, 6, and 9. If neurons 1, 3, and 9 are active at $t = 0$, then at $t = 1$, neurons 1 and 9 will activate neuron 6 and neurons 3 and 9 will activate neuron 2. By the assumption that activated neurons fire for two consecutive time steps after activation, neurons 1, 3, and 9 will also fire at $t = 1$, so the entire web is active at $t = 1$, and once the entire web is active it remains active and no additional neurons become active.

By contrast, if we take a different subset of three neurons from web 4 to be active at $t = 0$, namely, neurons 1, 6 and 9, no new neurons will be activated at $t = 1$ and only neuron 6 will have its threshold exceeded at $t = 1$, though neurons, 1, 6, and 9 will all fire at $t = 1$. However, at $t = 2$, only neuron 6 will fire and neurons 1 and 9 will cease firing. At $t = 3$, no neurons will fire, because no neuron received input from two or more active inlinks at either $t = 2$ or $t = 3$, so even the assumption that each neuron fires twice cannot prevent activity from dying out when the initially active subset is neurons 1, 6, and 9.

A ***basin of attraction*** for a web is the set of initial states which will ultimately lead to activation of the web as an equilibrium state, i.e. for which the web is an attractor. For larger nets, the basins of attraction around webs are much larger than in the previous example, but the problem of determining the extent of these basins beyond the equilibrium state is largely beyond the scope of this paper. A large basin of attraction around a web means that the idea represented by that web has a large range of generalization in terms of what sets of initially activated neurons will converge upon it.

A set of webs, all of which have large basins of attraction, has considerable coding redundancy that makes the ideas represented by the webs highly discriminable from each other. Associations from other ideas to such webs need not require learned synapses from the other idea to every neuron in the target web, but only to a (perhaps small) proper subset of the neurons in the target web; that is, if subsets of the web have the often-desired Hebbian completion property.

3.4.6. *Toward a structural definition of Palm-assemblies*

Palm's definition of cell assemblies prohibits persistent subsets that neither ignite nor support each other. For example, the following is a web by my definition, but not a Palm-assembly: a web of 60 neurons composed of two non-overlapping subwebs of 30 neurons that have no links from neurons in one subweb to those in the other subweb. Why would one want to disallow such sets as cell assemblies, if they are equilibrium states (invariant sets)? It is not clear to me that one does want to disallow such sets, in the definition of cell assemblies, and that is the main reason I did not exclude them from being webs.

Such "composite webs" and other webs that are not Palm-assemblies probably have much smaller basins of attraction than Palm-assemblies, so they would have

a narrower range of generalization in the initial sets that could converge on them. This probably makes composite webs less adequate for representing ideas than webs that are Palm-assemblies, since they would have a poorer degree of completion in retrieval starting from noisy, incomplete initial sets.

Of course, if this is so, then it is also less likely that such a non-Palm-web would ever get recruited in the chunking phase to represent an idea. In practice, I doubt that one needs to worry about webs that are not Palm-assemblies, because in neural net models of activation dynamics, I doubt that such sets get recruited very often to represent ideas, and this low frequency may well match what happens in real cortical networks. Note that in my model of chunking, recruiting a web to be a chunk idea requires that it be converged upon in the chunking phase of learning. I don't see how one could assume any other method of recruiting a web to represent an idea using innate binding links, except divine intervention.

If it proves necessary to rule out these webs, I would prefer a purely structural definition, one that requires only the connection matrix, not the connection matrix plus a model of activation dynamics. My inclination would be to place a requirement that all subsets of a certain size or greater have some minimum number of links to neurons in the complement subset of a web. We really want *all* large subsets of cell assemblies to have the capacity to ignite the entire cell assembly during completion phases, not just the persistent subsets. The ignition capacity depends on there being a sufficient number of links to the complementary subset. We also want large subsets to have a large number of internal links, so that they will persist long enough to ignite the entire web and thus be persistent. It isn't just the *persistent* subsets that we want to be able to ignite the remainder of the cell assembly, and to demand only a support relation seems far too weak. What we want is for *all large* subsets of assemblies both to be persistent and to ignite the entire assembly. However, my working assumption is that we ought not to incorporate these dynamic properties into the definition of cell assemblies, but rather define assemblies more simply in terms of minint and maxext connectivity and then study completion (ignition) in chunking and retrieval.

3.5. *Iconic, elaborative, abstractive, and elabstractive coding*

In ***iconic coding***, any simultaneously active set of neurons becomes a cell assembly, provided that it is not a subset of some previously learned cell assembly. Iconic coding is possible for learned cell assemblies. However, as noted previously, for any set of neurons to be capable of forming a cell assembly without addition or subtraction of any neurons, any pair of neurons in the set must have a high probability of being synaptically connected or growing such synapses as a result of Hebbian learning. This does not appear reasonable for the cerebral cortex as a whole, but it might be possible within a module.

In ***elaborative coding***, the formation of a cell assembly to represent an idea involves the addition of "binding" neurons to the cell assembly that were not directly

activated by the sensory or associative input, but were strongly interconnected to those that were. With elaborative coding, any set of simultaneously active neurons in the same module can be incorporated as a subset of its representing cell assembly.

In ***abstractive coding***, the representation of an input set of simultaneously active neurons is by a *subset* of those neurons, a subset of neurons that are sufficiently densely interconnected (or which becomes so via learning) that they form cell assembly to represent the original input via this abstractive process.

Elabstractive coding is both elaborative and abstractive, so that the representative of an input set of activated neurons involves both the loss of some neurons from the original input set in the representing cell assembly and the addition of some new "binding" neurons in the cell assembly that were not in the input set.

It is not clear which of these types of coding Hebb assumed for his cell assemblies. Neural net models have often used iconic coding with fully or almost fully connected neural nets [21–23,19,18]. When iconic coding is used in conjunction with Hebbian learning and considerably less than full connectivity, completion of learned sets (patterns) from subsets is degraded, when the number of patterns to be learned becomes a substantial fraction of the number of neurons in the net [10,24].

Web theory assumes elabstractive coding, since it would be extraordinarily improbable for a randomly selected input set to be a web. With web coding, the total number of possible ideas is the number of webs, and the basin of attraction around each web represents, in some sense, the range of generalization of that idea. All active sets within one basin of attraction are represented by the attractor web of the basin.

Learned cell assemblies in fully connected nets also have basins of attraction beyond the cell assembly itself, but, with iconic coding, the net has the capability to define the "central" attractor states of the basins to be identical to input sets. In and of itself, this would appear to be an advantage for learned vs. innate assemblies, but it is my guess that innate assemblies have an advantage in representational capacity under conditions of sparse connectivity.

There is also an additional flexibility of learned assemblies which leads to an additional problem, namely, how different an input set must be before it becomes a new attractor cell assembly, rather than being coerced to the most similar existing cell assembly. Innate cell assemblies don't have this problem, because their basins of attraction are not modifiable. It is not clear whether this rigidity is an advantage or a disadvantage.

3.6. *Chunk assemblies are new webs, not set unions*

I share Hebb's bias that a chunk idea be represented by a new cell assembly, not merely the set union of the assemblies representing its constituents or the set union plus some additional relational or binding neurons. Either set union alternative assumes that the assemblies representing higher order chunks are substantially larger than the assemblies representing lower order chunks.

Though the human cerebral cortex is estimated to contain on the order of 10^{10} neurons, constructing complex cell assemblies to be unions of constituent assemblies is an exponential growth process. So, for example, if the basic ideas were represented by only ten neurons, and thoughts consisted of an attention span of four ideas, the second-level ideas would be represented by 40 neurons. This ignores set overlap, as we may until set sizes become a sizable fraction of all of the neurons in the net. Third-level ideas that are unions of second-level ideas would require $160 = 10 \times 4^2$ neurons. After reaching about level 15, this process requires the set union to be represented by all of the neurons in the human cerebral cortex.

If basic ideas are represented by s neurons, then the largest sets of kth level ideas would contain $s \times 4^{k-1}$ neurons. As discussed by Legéndy [14], for cell assemblies to be directly associated to each other, $s = 10^3$ to 10^4 neurons for each basic idea is a more reasonable guess than $s = 10$ for human cognitive minds. This limits human conceptual depth to about 10–13 levels using set unions to represent more complex ideas.

Human capacity for idea representation must be limited by total memory in any case, and it is possible that the hierarchy of human ideas is no more than 10–15 levels deep, though this seems unlikely. However, there are other problems with pure set union models of thought representation.

I have never seen a way to avoid intractable associative interference problems, if more complex ideas are represented by set unions of their constituent ideas [25]. It also seems difficult, perhaps impossible, to give even approximately equal importance in a thought to constituent ideas represented by vastly different numbers of neurons.

For these reasons, I believe a chunking process is needed by overlapping set coding to maintain the set size for idea representation within a reasonable range. Thoughts consisting of a union of say four ideas are chunked to an idea represented by a set of neurons no bigger on the average than any of the sets representing the constituent ideas.

Though the chunk assembly may or may not overlap more extensively with its constituent assemblies than with randomly chosen assemblies, there must be sufficient discriminability in the representation of any two cell assemblies to permit different ideas to have different associations.

The strength of association from idea A to idea B depends on the number and strength of apical synaptic links from neurons in A to neurons in B. On the average, larger cell assemblies would have more potential input and output synapses than smaller cell assemblies, and more familiar (frequency used) ideas may be represented by slightly larger cell assemblies. However, I think it is likely that the size of cell assemblies is restricted to a modest range, e.g. less than a factor of two in the number of neurons in the smallest vs. the largest cell assemblies.

Webs are especially suitable for representing ideas that are derived by chunking a set of constituent ideas, with the constituents also being chunks of sets of constituents, to an arbitrary and variable depth. The web theory of the representa-

tion of ideas places no limit on the hierarchical depth of chunking beyond the total memory capacity of the cerebral cortex, and maintaining about the same number of neurons in a chunk regardless of depth makes economical use of that memory capacity.

In the present model, the number of neurons in the web representing an idea is assumed either to increase nor decrease with its hierarchical depth. In particular, a web representing a set of constituent ideas is not the union of the webs representing the constituent ideas, though the chunk for the set may have greater overlap with any constituent idea that is in the same module as the chunk idea than with unrelated ideas in that module. All webs are roughly the same size, irrespective of their position in any kind of semantic hierarchy. It is possible that more frequently used ideas are represented by slightly larger webs, but I have not pursued this possibility.

3.7. *Cell assemblies are local, not global (diffuse)*

Braitenberg [8] and Palm [26] assume the existence of diffuse (global) cell assemblies, that is, assemblies with neurons from different modules possibly all over the cortex. Contrariwise, I assume that all cell assemblies are local, that is, confined to a single module, probably within a small region of that module. I do not envisage there being any cell assemblies that have neurons in more than one module.

Global cell assemblies are less plausible than local assemblies for at least three reasons: First, cell assemblies require a dense interconnection of their neurons, and nearby pyramidal neurons are known to have this property, while distant pyramidal neurons are not, and some special innate or learned long-distance guidance process would be required to achieve this.

Second, as will be demonstrated later in this paper, symmetry in synaptic connections on a neuron-to-neuron basis is very desirable (though not essential) in achieving cell assemblies by my definition of cell assemblies (which is very nearly identical to Palm's). As will be shown later, there is a neurally plausible mechanism for achieving symmetry for nearby neurons, but, once again, this mechanism is much less plausible for neurons in different modules.

Third, the activation dynamic process of converging on a cell assembly probably requires a number of time steps, and the synaptic delay time between modules is substantially larger than the synaptic delay time between nearby neurons in the same module. Besides local assemblies having faster convergence than global assemblies, there would be far greater problems in achieving synchronous activation of the neurons in a global assembly. Of course, synchronous activation may be unnecessary, though I have found it helpful in physiologically reasonable models of activation dynamics.

I avoid global cell assemblies by assuming that sets of cell assemblies, whether in the same or different modules, are chunked into a single assembly that may be in the same module as one or more of its constituents or may be in a different module from any of its constituents.

3.8. *Chunking between and within modules*

The cerebral cortex can doubtless be decomposed into modules, with the ideas in each module being in the same functional (semantic) category. Braitenberg advanced the elegant hypothesis that the human cerebral cortex is divided into square root compartments — that is, the 10^{10} cortical neurons are partitioned into 10^5 modules with 10^5 neurons in each [8]. Perhaps it is 10^4 modules with 10^6 neurons in each or 10^3 modules with 10^7 neurons in each or 10^2 modules with 10^8 neurons in each. In any case, the number of modules is not likely to be larger than the number of neurons in the average module.

I assume that there are about 10^3 neurons in each cell assembly and that the number of cell assemblies is approximately the same as the number of neurons (between 0.1 and 10 times as many). Assuming 10^{10} neurons in the cerebral cortex, this means that each neuron is a member of 100–10,000 different cell assemblies.

It is unlikely that cortical modules connect equally to all other cortical modules. More likely, the average module sends connections to between 0.1 and 0.001 of all the other modules.

I like to organize these modules into the following categories: sensory feature modules, motor feature modules, segment modules, object modules, concept modules, proposition modules, and procedure modules. A set of cell assemblies that represents edges and slits of different orientation, spatial frequency, and position might constitute a single visual feature module.

My current hypothesis is that it makes little difference to the chunking process whether the constituent assemblies associated to a chunk assembly come from the same or different modules. However, consideration of this question raises a number of difficult issues which I have not resolved: What is the definition of a module? How big are modules? Are modules overlapping or non-overlapping? Can more than a single idea be active in a module at one time, and, if so, under what circumstances and what are the consequences of this? Do different modules operate synchronously or asynchronously?

I am unsure whether there is a difference between chunking in which the chunk is in a different module from all of its constituents (remote chunking) and chunking in which the chunk is in the same module as one or more of its constituents (local chunking). Plausible examples of remote chunking are: a set of visual feature ideas form the constituents of a letter chunk or a set of words are constituents of a concept. It is not likely that letters are represented in the same module as visual features or that concepts are represented in the same module as words. Plausible examples of local chunking abound in semantic memory: "Commutative group" is a concept that is likely to be in the same module as its constituent concept "group", though perhaps in a different module from its constituent concept "commutative".

Although one can make the structural distinction as to whether or not a chunk is in the same module as any of its constituents, it is not clear what difference that makes to thinking. Pyramidal neurons make local apical (associative) synapses

as well as remote apical synapses, so local apical synapses can associate chunks to constituents in the same module in the same way as remote apical synapses associate chunks to remote constituents. The model for the dynamics of thinking presented in a later section makes no distinction between these cases.

4. How Many Webs?

Having a precise definition of what subnets of neurons constitute a cell assembly (web) simplifies study of what Palm [4] referred to as "the main problem of the theory of cell assemblies", namely, how many cell assemblies are there in various types of nets.

Although my conception of cell assemblies is closest to Palm [4,27], I did not generate nets in a systematic manner to have webs as Palm did. Rather, like Legéndy [2], whose two-tiered structure of cell assemblies is quite different from mine, I focused on random nets. I used greedy algorithms to find webs in particular nets and probability methods to estimate the total number of webs.

For cell assemblies to be feasible representative for ideas, there must be enough distinct cell assemblies to represent as many ideas as a human mind has potentially available for activation at any one time. We do not know how many ideas that is, but it seems likely to be greater than 10^5 and less than 3×10^{10}. The lower bound is roughly the vocabulary of words possessed by many educated individuals. The upper bound represents how many ideas a single person could learn at the rate of ten ideas a second for 100 years.

Can one be certain that idea capacity is within this range? No. If many chunks are learned for each thought, or thoughts occur at an average rate faster than ten per second, we could exceed the upper bound. If what seem in psychological experiments to be single chunks have no unitary representation, but are only composites drawn from a small set of basic ideas, then idea capacity could be below the lower bound. Neither alternative seems very likely, however. On the assumption that familiar images, words, concepts, propositions, actions, procedures, and plans are all single chunks at some level of representation (though they are also composites at a lower level of representation), the true idea capacity of the human mind is likely to be closer to the upper bound than to the lower bound.

To avoid confusion, please note that when I refer to a chunk as being at a higher level of representation than its constituents, I mean this in a functional sense, not necessarily a structural sense. The chunk may be in a different (higher) module than any of its constituents, as is doubtless the case for word chunks vs. their phonetic or graphic segmental constituents. However, at the highest levels of semantic memory, there are many cases where it is far more likely that a chunk idea such as "commutative group" is coded in the same module as at least one of its constituents, namely, the head noun "group". It is not clear that there is any important difference between intramodular chunking and intermodular chunking.

Since my current working hypothesis is that the asymptotic number of ideas in an adult human is in the order of the number of pyramidal neurons in the cerebral cortex, I want any distributed neural representation model to have at least 10% as many potential cell assemblies (webs) as there are neurons in the net and possibly 100% or more.

4.1. *Systematic nets*

Systematic nets are nets that are constructed according to some systematic plan with no random element in determining the connection matrix. For example, a fully connected net (complete graph), in which each of the n neurons has a synapse with each of the other $n-1$ neurons, is a systematic net.

Palm [4] constructed a number of systematic nets that had a reasonable number of cell assemblies by Palm's definition, which would be equivalent to my definition of webs in these cases.

4.2. *Asymmetric regular random nets*

In an ***asymmetric*** net, the existence of a link from neuron i to neuron j does not guarantee the existence of a link in the reverse direction from neuron j to neuron i. In an asymmetric random net, the conditional probability of a link from neuron j to neuron i given a link from neuron i to neuron j is identical to the unconditional probability of a link from neuron j to neuron i. By ***regular*** I mean that the random nets are constrained to have a constant number of inlinks to each neuron, selected independently and randomly for each neuron. That is, one picks an independent random sample of m other neurons from the net to provide the inlinks to each neuron in the net.

For a net of n neurons with m inlinks for each neuron, the probability that a random subset of size s will be a web with a minint $\geq y$ and a maxext $< y$ is:

$$P_{\text{web}} = P(\text{minint} \geq y) \cdot P(\text{maxext} < y) \ .$$

The expected number of webs in a net is P_{web} multiplied by the number of subsets of s neurons that can be drawn from a population of n neurons:

$$E_{\text{web}} = \text{Combinations}\,(n, s) \cdot P_{\text{web}} \ .$$

The probability that a particular neuron receives x inlinks from a set of s neurons is the probability that there are x neurons in the intersection of the set of s neurons with the inlink set of m neurons. For a population of n elements, the probability that there are x elements in the intersection of two independent randomly selected (without replacement) subsets of s and m elements is the hypergeometric probability:

$$P_{\text{hg}}(x, s, m, n) = \frac{s!\,(n-s)!\,m!\,(n-m)!}{n!\,x!\,(s-x)!\,(m-x)!\,(n+x-s-m)!} \ .$$

$P(\text{minint} \geq y)$ is calculated by determining the probability that any one neuron in the set of s neurons receives y or more inlinks from the other $s-1$ neurons in the set and then raising this to the s power to compute the probability that all s of the neurons in the set receive y or more inlinks from other neurons in the set. The probability that any one neuron receives y or more inlinks from other neurons in the set is independent of the probability that any other neuron receives y or more inlinks from other neurons in the set, because the m inlinks to each neuron are assumed to be selected independently of the selection of any other neuron's inlinks.

$$P(\text{minint} \geq y) = \left[1 - \sum_{x=0}^{y-1} P_{\text{hg}}(x, s-1, m, n-1)\right]^s .$$

$P(\text{maxext} < y)$ is calculated in a manner entirely analogous to $P(\text{minint} \geq y)$:

$$P(\text{maxext} < y) = \left[\sum_{x=0}^{y-1} P_{\text{hg}}(x, s, m, n-1)\right]^{n-s} .$$

The resulting formula for the expected number of webs is:

$$E_{\text{web}} = \frac{n!}{s!\,(n-s)!}\left[1 - \sum_{x=0}^{y-1} P_{\text{hg}}(x, s-1, m, n-1)\right]^s \cdot \left[\sum_{x=0}^{y-1} P_{\text{hg}}(x, s, m, n-1)\right]^{n-s} ,$$

$$E_{\text{web}} = \frac{n!}{s!\,(n-s)!}\left[1 - \sum_{x=0}^{y-1} \frac{(s-1)!\,(n-s)!\,m!\,(n-m-1)!}{(n-1)!\,x!\,(s-x-1)!\,(m-x)!\,(n+x-s-m)!}\right]^s$$

$$\cdot \left[\sum_{x=0}^{y-1} \frac{s!\,(n-s-1)!\,m!\,(n-m-1)!}{(n-1)!\,x!\,(s-x)!\,(m-x)!\,(n+x-s-m-1)!}\right]^{n-s} .$$

Using the above formula, I calculated E_{web} for a large number of parameter sets (x, s, m, n), with the number of neurons in the net (n) ranging from 10 to 10^8, the number of links per neuron (m) and the size of the webs ranging from substantially below $\sqrt{n}$ to substantially above $\sqrt{n}$, and used a one-dimensional hill-climbing program to estimate the value of minint (neural threshold) that yielded the maximum E_{web}. I also did a number of multidimensional hill-climbing runs in two- and three-dimensional parameter spaces starting from local maximums for E_{web}. I tried both the above formula and a Poisson approximation to the hypergeometric (to permit faster searches through the parameter spaces).

I never found a value of E_{web} greater than 0.1. There do not appear to be any webs in most asymmetric regular random nets. The expected number of webs is so far below the number of neurons in the net — indeed it appears to be below 1 — that asymmetric regular random nets do not seem to be suitable architectures for

representing ideas by cell assemblies, assuming that a cell assembly has something like the structure of a web.

4.3. *Symmetric regular random nets*

Searching for a more propitious type of random net in which to find webs, I turned my attention to symmetric nets. Unfortunately, I was unable to derive an analytic expression for the expected number of webs in symmetric regular random nets, so I was forced to generate some (approximately) symmetric regular nets of tractable size and search for webs in them. Using MatLab, I generated four of them: ($n = 50, m = 7$), ($n = 100, m = 9$), ($n = 216, m = 13$), and ($n = 300, m = 15$), where n = number of neurons in the net and m = number of inlinks per neuron.

I devised a greedy algorithm embodied in a MatLab program that was very successful in finding webs, limiting the maximum numbers of neurons in a web to the lesser of 50 neurons or half the number of neurons in the net. The greedy algorithm started with a set of three randomly picked neurons as the active set, and at each time step added two neurons that had the maximum linking to the active set and deleted one currently active neuron with the fewest links to the other neurons of the active set. Then the active set was tested to see if it was either a web or had reached the maximum allowable size. Positive results of the test terminated the search either in a web or in failure to find a web, and the process was repeated. There are many other greedy algorithms based on maximizing the internal linking of the active set and minimizing the external linking. Neurally plausible activation dynamic models are also greedy algorithms for finding webs, and in later studies I used such activation models that worked even better than this implausible one.

For the 50-neuron net, 898 of 1000 trials terminated in a web. For the 100-neuron net, 909 of 914 trials terminated in a web. For the 216-neuron net, 253 of 257 trials terminated in a web. For the 300-neuron net, 454 of 466 trials terminated in a web. The number of different webs found were 121 for $n = 50$, 301 for $n = 100$, 213 for $n = 216$, and 381 for $n = 300$.

4.3.1. *Estimating the number of webs in a net*

I did not continue to use as many trials to search for webs in the larger nets because it was much more time consuming to search large nets and because I had worked out a maximum likelihood estimation technique to estimate the number of webs in a net (in the target size range) from a sample of modest size.

An urn model was used to estimate the number of webs in the net. Assume that there are w webs in a particular net labeled $1 \ldots w$. On each trial you reach into the urn, pick out a web, record its number, see if it matches any number previously picked, record a "0" if it matches a prior web (old web) and a "1" if it does not match (new web), and then put the web back into the urn. One writes an expression for the probability of obtaining the exact sequence of 0s and 1s that one obtains. This probability is a function of the parameter w, the number of webs in the urn.

Obviously, only trials on which some web (old or new) is obtained are included in this sequence. That is, failures to find a web are discarded.

Let $R = [r_k]$ be the sequence of numbers of prior different webs on which repetitions of old webs were sampled. Let $r = |R|$ be the number of old webs sampled (the number of 0s in the sample sequence). Let t be the number of trials (excluding fails), i.e. t is the total length of the sequence of 0s and 1s. Let $b = (t - r)$ be the number of new webs (the number of 1s in the sample sequence). For example, if the sequence of new and old webs sampled was 1101001110, then $R = [2,3,3,6]$ and $r = 4$ repetitions of previously sampled webs in t samples.

Let P be the probability (likelihood) of obtaining the sequence of repetitions R in t trials with w webs in the net (urn).

$$P = \prod_{i=0}^{b-1} \left(\frac{w-i}{w}\right) \prod_{k=1}^{r} \left(\frac{r_k}{w}\right) ,$$

$$P = w^{-t} \prod_{i=0}^{b-1} (w-i) \prod_{k=1}^{r} (r_k) .$$

As far as the maximum likelihood estimation of the parameter w is concerned, the product of the r_k terms is irrelevant, since it is a constant independent of the value of w. Thus, I wrote a Maple program to search for the value of w that maximized:

$$w^{-t} \prod_{i=0}^{b-1} (w-i) .$$

The values so obtained for each value of b from the various nets constitute the maximum likelihood estimates of w, the number of webs in each net (within the target size range). As shown in Table 2, the estimated number of such webs was 121 in the 50-neuron net, 319 in the 100-neuron net, 710 in the 216-neuron net, and 1253 in the 300-neuron net. The ratio of the number of webs to the number of neurons in the net increased from 2.4 for the 50-neuron net to 4.2 for the 300-neuron net.

Table 2. Number of webs in regular symmetric nets.

Neurons (n)	Inlinks per neuron (m)	Trials (t)	Webs found (b)	Estimated no. webs (w)	Webs per neuron
50	7	898	121	121	2.4
100	9	909	301	319	3.2
216	13	253	213	710	3.3
300	15	454	381	1253	4.2

In short, the number of webs exceeded the number of neurons in symmetric regular random nets, and the ratio of webs to neurons appears to be increasing

with the size of the neural net. Symmetric random nets contain enough webs to represent as many or more ideas than can be represented by specific neuron coding. Furthermore, this representation capacity can be achieved with a connectivity parameter that is characteristic of the cerebral cortex — i.e. the number of synapses per neuron was, it all cases, less than the square root of the number of neurons.

If local basal or apical connections within a module are ametric, but symmetric, then the present findings indicate that there are a plenty of innate local cell assemblies to represent ideas.

It seems likely that remote (long-distance) ametric apical synaptic connections are asymmetric on a neuron-to-neuron basis, though they are likely to be symmetric on a module basis. Thus, the apparent necessity for symmetry in neural connections for ametric regular random nets to have any webs at all argues against the global (diffuse) cell assemblies envisaged by Braitenberg and Palm [8,26,6], unless remote connections are guided, either genetically or by learning, to achieve symmetry or some other systematic connection structure.

It is more likely that local apical or basal synaptic connections are metric, that is, connection probability declines with the distance between neurons. Both asymmetric and symmetric models of net structure based on this assumption are investigated in the following section on proximity nets.

4.4. *Proximity nets*

Proximity nets organize all nodes in a net into a k-dimensional structure, and the linking probability, l_{ij}, between any two nodes i and j is a function of some distance measure between i and j. The net might be symmetric or asymmetric. One can use either lattice packing or hex packing of neurons. With lattice packing, neurons are assumed to be located at each of the lattice points for all possible integer values of each dimension. For example, a two-dimensional neural net with 17 possible values of the x and y dimensions has $17 \times 17 = 289$ neurons, one at each of the 289 possible lattice points. Each neuron can be represented by its location in this space by two coordinates (x, y), where x and y are integers from 0 to 16. Alternatively, one can use hexagonal packing of the neurons, representing neurons in odd rows at odd values of x from 1 to 33 and neurons in even rows at even values of x from 0 to 32. Hex packing of neurons is more plausible than lattice packing, but lattice packing is a bit simpler to work with, and so I used lattice packing in this case.

A three-dimensional net is most plausible for the cerebral cortex with a depth (z) dimension that is very small in relation to the two horizontal surface (x and y) dimensions. The z dimension corresponds to the depth of the neuron within the 6-layer (thin) neocortical mantle. For neurons of the same type (role), one would want to assume that there was a high linking probability for neurons with different z values at the same (x, y) location and a decrease in linking probability with increasing distance in the (x, y) locations. A simple and reasonably plausible model is for linking probability to be independent of the z coordinate. I wasn't up

to constructing three-dimensional proximity nets, because, to avoid edge effects in small nets, one needs to make all dimensions cyclic. So I settled for two-dimensional (toroidal) nets with lattice packing — challenge enough for me.

Asymmetric proximity nets are not constrained to be symmetric beyond the constraint induced by the proximity metric. However, a suitably constructed asymmetric proximity net will have a degree of symmetry that is intermediate between that of an asymmetric random net and a symmetric random net. By itself, that would make the likelihood of webs also intermediate. However, proximity nets have a much higher probability of short cycles of synapses among small sets of three, four, five or six neurons than either symmetric or asymmetric random nets with the same total number of synapses. This could act to increase the likelihood of webs.

I have no analytic solutions for the expected number of webs in any type of proximity net. However, I constructed a number of one- and two-dimensional proximity nets to determine whether proximity nets had webs and whether the ratio of webs to neurons was large enough to make it seem plausible that webs might be the representatives of ideas in cortical modules organized as proximity nets. Since any "empirical-mathematical" investigation of this sort necessarily covers only a very tiny portion of the infinite space of all possible proximity nets, these results can only set lower bounds on the ratio of webs to neurons in proximity nets. Nevertheless, since the lower bounds I obtained for certain types of proximity nets indicated that there are often more webs than neurons, this is sufficient to show that proximity nets can have a sufficient number of webs to represent ideas.

To eliminate edge effects, I made all my one-dimensional proximity nets cyclic and all my two-dimensional proximity nets toroidal (donut shaped). For example, in a one-dimensional 50-neuron cyclic net, neuron-1's left-hand neighbor is neuron-50 and its right-hand neighbor is neuron-2. If each of the dimensions in a two-dimensional net is cyclic, one obtains a torus.

4.4.1. *Asymmetric proximity nets*

I began with a one-dimensional cyclic proximity net with 50 neurons and found no webs at all when I used seven links per neuron, randomly distributed in a band of 8, 10, 14, or 30 neurons around the neuron being linked to. The key to this failure turned out to be that I made the linking probability a simple step function: zero outside the proximity band and independent of distance within the band. Whether the linking probability was high or low within the band and regardless of the width of the band, no webs were obtained.

As soon as I ramped the probability so that linking probability decreased more gradually with distance from the target neuron, rather than abruptly in a single step, I obtained webs — about 12 webs in a 50-neuron net with 7 links/neuron, for a ratio of 0.24 webs/neuron. This is a lower bound and probably one can get a higher ratio for one-dimensional proximity nets, but 0.24 is certainly a high enough ratio to make webs plausible candidates for idea representatives. Web size averaged

9 neurons over a range from 6 to 13. Each neuron was a member of at least one web and the average neuron belonged to two different webs. As might be expected, each web consisted of a band of immediately adjacent neurons, with no neurons left out, e.g. neurons 49, 50, 1, 2, 3, 4, and 5 might constitute a web, but not neurons 49, 1, 5, 23, 24, 31, and 38.

With two-dimensional proximity nets, it was again necessary to have a region around the target neuron where linking probability decreased gradually, rather than abruptly, to zero with increased distances. Nets with only two levels of linking probability, high and zero, had no webs.

In the case of symmetric random nets, it was important to demonstrate that the webs/neurons ratio did not decrease with an increasing number of neurons in the net — indeed, it appeared to increase. By contrast, for proximity nets, it is highly likely that the webs/neurons ratio remains essentially constant with net size for all but very small nets. To understand why this is so, we need some terminology. Let the ***net width*** of a cyclic net with K neurons or a K by K square toroidal net be K, the number of lattice "steps" along a dimension needed to circumnavigate the net. Define the ***web width*** or ***set width*** to be the Euclidean distance between the most distant neurons in the web or set.

The ratio of the number of webs to the number of neurons in the net might change with the size of the webs and the web width. However, once the width of a net is two or three times the maximum web width, there is probably no further change in the ratio of the number of webs to the number of neurons in the net with increases in the size of the net, because neurons that are remote from a set of neurons cannot send or receive links from that set in a proximity net. Thus, remote neurons cannot influence the probability that a set is a web for sets whose width is small in relation to net width. If threshold control limits the number of active neurons to some relatively small number, then only sets with relatively small set width can be webs in proximity nets. This means that the webs/neurons ratio should remain approximately constant with increasing net size once the net width is two or three times the maximum web width. Certainly, the webs/neuron ratio should not decrease with increasing net size.

I did two small experimental tests of this hypothesis. First, I compared 50- and 100-neuron one-dimensional cyclic nets with the same function relating linking probability to distance and the same average of 7 links/neuron, with nonzero probability of linking only within a radius of six lattice steps on either side of the target neuron. Using a greedy algorithm starting from random sets of three neurons with the number of starting sets (trials) proportional to the number of neurons in the net (50 and 100, respectively), I found 12 webs in the 50-neuron net for a webs/neurons ratio of 0.24 and 24 webs in the 100-neuron net for a webs/neurons ratio of 0.24.

Second, I compared $9 \times 9 = 81$ and $15 \times 15 = 225$ two-dimensional proximity nets, each having an average of 9 links per neuron within a Euclidean distance of less than 2.5 lattice steps from the target neuron. Using a greedy algorithm and starting sets of neurons consisting of 2 by 2, 3 by 3, and 4 by 4 squares (4, 9, and 16

neurons, respectively), I searched for webs in each net with all 81 possible starting square blocks of neurons in the 81-neuron net and all 225 possible such starting blocks of neurons in the 225-neuron net. I found 43 webs with 24 or fewer neurons in the 81-neuron net for a webs/neurons ratio of 0.53 and 148 webs with 24 or fewer neurons in the 225-neuron net for a webs/neurons ratio of 0.66.

These results are consistent with the hypothesis that the webs/neurons ratio does not decrease with the number of neurons in the net, all other factors held constant (except the exact pattern of connections in the nets), but obviously more stringent tests are necessary to be completely sure of this very plausible hypothesis. Furthermore, there was a moderate increase in the webs/neurons ratio for the two-dimensional nets studied, rather than complete invariance. This may be because the smaller net was not sufficiently large in relation to web width or a random error. In any case, what is important is that the web/neurons ratio does not decrease with increasing net size, not whether it is invariant or increasing.

One focus of this brief study of asymmetric proximity nets was in heuristic exploration of various distance functions for link probability in the effort to find link probability functions that maximized the webs/neurons ratio. The most successful case was the largest net I studied: a two-dimensional proximity net with $17 \times 17 = 289$ neurons and an average of 70 links per neuron. This within-module linking frequency is larger than the square root of the number of neurons in the net, but this is of no significance for proximity nets if the webs/neurons ratio is invariant with the number of neurons beyond some net width. The 289 neurons can be considered to be a subnet of a much larger net for which the 70 links per neuron would be smaller than the square root of the number of neurons.

In the 289-neuron net with 70 links/neuron, I found 797 webs with sizes ranging from 25 to 139, for a webs/neurons ratio of 2.76, that is, almost three times as many webs as neurons. Moreover, this is an underestimate of the true ratio for this net, since I use heuristic greedy algorithms to search for webs and can only afford to search for a limited amount of time in each case. Also, I always place lower and upper bounds, in this case 25 and 140, respectively, on the size of the sets I investigate to find webs.

Since the maximum webs/neurons ratio was obtained for the case where I had, by far, the largest webs, it would appear that the webs/neurons ratio does not decline for large webs. Since cortical cell assemblies are likely to contain between 50 and 10,000 neurons, it is reassuring that the webs/neurons ratio does not appear to decrease with increasing web size or module (net) size. These results are quite encouraging for the prospects of ideas being represented by webs in the cerebral cortex.

4.4.2. *Symmetric proximity nets*

Although asymmetric proximity nets have greater symmetry than asymmetric random regular nets, they are far from being completely symmetric. Since symmetry

was critical for producing webs in random regular nets, it is likely that symmetry would be beneficial for producing webs in proximity nets. (However, it is not clear if one wants nets with greater webs/neurons ratios than were obtained for asymmetric proximity nets.) The largest webs/neurons ratio I found for a symmetric proximity net was 2.07 (597 neurons in a 289-neuron net), but that was while restricting the size of webs to between 31 and 85, which is less than half the size range for the asymmetric case that produced the larger ratio of 2.76. For either symmetric or asymmetric nets, the important finding vis-a-vis number of webs is that one can construct random proximity nets such that the number of webs exceeds the number of neurons, even when one requires web size to be within some restricted range.

There were three primary reasons why I investigated symmetric proximity nets: First, I wanted to verify what seemed intuitively likely — that symmetric proximity nets can achieve webs/neurons ratios greater than one with a smaller number of links per neuron. This was immediately apparent. For example, the symmetric proximity net with the largest webs/neurons ratio had an average of only 14 links per neuron, whereas the asymmetric proximity net with the largest webs/neurons ratio had an average of 70 links per neuron — five times as many!

Second, I had started using neurally plausible activation models to search for webs, and I suspected that I could get a higher probability of converging on a web and a shorter time for convergence with symmetric proximity nets than with asymmetric nets. Informally, this was also overwhelmingly clear, but this result is totally confounded with the different types of activation dynamics models that I used in the two cases. I was striving to reach 100% convergence on webs in a modest number of time steps. All I can say is that introducing symmetry seemed to be a very helpful step toward this goal, but there were confounding changes in the dynamics models also.

Third, I suspected that webs in symmetric nets would be more discriminable from each other in retrieval than webs in asymmetric nets, as judged by the degree of completion starting from (noisy) subsets of these webs. Certainly, I achieved much more satisfactory retrieval-completion performance for symmetric than for asymmetric proximity nets, but this conclusion is also confounded with the evolution of my activation dynamics models.

4.4.3. *Plausible neural mechanism to achieve symmetric linking*

Perhaps the most important idea I had concerning symmetry in proximity nets is a plausible neural mechanism for generating symmetric proximity nets. Like many others, I have been impressed with the usefulness of symmetry in neural net models, but I could not think of a plausible neural mechanism to justify the symmetry assumption. To construct symmetric proximity nets, I generated two alternative ways, one of which I recognized was highly plausible neurally. The neurally implausible way is to generate an asymmetric proximity net and then add the necessary symmetrizing links.

The neurally plausible way to generate a symmetric proximity net is to generate an asymmetric proximity net and then *delete* all the asymmetric links. When I thought of this mechanism, I had a eureka experience, because it is known that in neural development, many synapses are formed that later disappear. Wherever it is advantageous for the nervous system to have symmetric connections, there is a plausible natural selection method to accomplish this, namely, forming lots of quasi-random connections and then deleting those connections that do not, by chance, form positive feedback cycles of symmetric linkages with other neurons. What was critical for the generation of this idea was to be working with neurally plausible proximity nets. Proximity nets have a reasonably high probability of forming symmetric linkages "randomly" (due to the proximity metric) even without forcing symmetry as a constraint, and there needs to be at least a moderate number of synapses left after deleting the asymmetric ones.

It should be noted that this neurally plausible mechanism for obtaining symmetric linking applies to proximity nets, which are plausible models for the basal dendritic synapses that, according to web theory, bind pyramidal neurons into cell assemblies. Proximity nets are also plausible models of apical dendritic synapses within a module, but not between modules. In short, the binding synapses and the within-module (local) associative connections may be symmetrical by this mechanism, but it does not seem reasonable (to me at this time) to imagine that between-module (remote) associative connections could be made symmetrical by this mechanism.

4.4.4. *Symmetric connection of modules*

Fortunately, it is only the basal synapses, which bind together the neurons of a web, for which symmetry of connections at the level of neurons is a dynamic asset. Because chunks need to be associated to their constituents in both directions, we probably want *symmetry at the module level* for remote (long-distance) associative connections; that is, whenever neurons in module-A send axons to form synapses in module-B, we want neurons in module-B to send axons to form synapses in module-A, and a strong tendency in that direction has often been observed in the cerebral cortex. However, so long as there are sufficiently large number of connections in each direction between two modules and sufficiently large cell assemblies, it is not necessary for there to be connection symmetry at the level of individual neurons. Chunk-A can be associated to constituent-B without there being an above chance level of symmetry in the connections of the neurons in chunk-A to the neurons in chunk-B.

5. Dynamics of Thought

There are different purposes of neural modeling. If one wants to model the physiological mechanism by which lateral inhibition sets neural thresholds for activation, it is probably essential to use a synaptic matrix representation of the inhibitory

links. If one wants to model the physiological mechanism by which neuromodulation could make all apical synapses equally strong or gate on or off apical dendritic input to the cell body, then one needs to use a fairly detailed model of the apical dendritic tree, apical spines, transmitters, receptors, etc. However, if one wants to model the functional effects of such "whole-neuron" modulation, then a plausible and simple abstract model is to suppress explicit representation of the modulatory neurons, their transmitters, their synaptic or nonsynaptic effects on pyramidal dendritic trees, etc. and, instead, just incorporate their presumed functional effects into the laws of the neural dynamics of the excitatory connections.

5.1. *Lateral inhibition and threshold control*

One simple example of this is to model lateral inhibition by a dynamical law that makes the ***threshold***, h, of individual neurons in a neural module a monotonically increasing function of ***module activation*** Y (the summed activation of all the neurons in the module). I call this the ***thresh function***. The thresh function might be a linear, logarithmic, or power function, or it might be more complicated. In some simulations, I used a thresh function that increased as a simple linear function of total module activation:

$$h = b + cY \ ,$$

where $Y = \sum_{i=1}^{n} y_i$, y_i = activation of neuron i, n = number of neurons in the module, b is the base threshold for zero module activation (the intercept parameter of the thresh function), and c is the gain parameter that reflects the rate of increase in threshold with module activation.

Both the intercept and gain parameters of the thresh function may be variable mental state parameters that regulate thinking and learning. In particular, I assume that these parameters vary with the phase of thinking.

Milner [28] recognized the importance of controlling the positive feedback activation process in Hebb's [1] model of the cerebral cortex with a lateral inhibition mechanism that increases the activation threshold of neuron as a function of the total activation of the neural net. Increasing thresholds as a function of activation provides the negative feedback control of total activation that prevents the extremes of having all neurons active (epilepsy) or no neurons active (quiescence).

Legéndy [2] observed that if the sole function of lateral inhibition is threshold control as a function of activation, then one can model this function directly without representing individual inhibitory neurons and synapses and their dynamics — an enormous simplification. Legéndy suggested that the dynamics of the threshold control mechanism are reflected in EEG potentials and that the mechanism of achieving threshold control was in reticular formation. Braitenberg and Palm developed the concept of threshold control as a mechanism for restricting activity to single cell assemblies of different sizes or permitting activation of several cell assemblies without activation spreading to all the neurons in the net [3,4,27,29].

An idea related to threshold control by inhibitory subnets is Grossberg's concept of pattern normalization, in which an inhibitory subnet regulates the total activity of a neural module so as not to saturate to high intensities input to the module [30–32]. Normalization is a kind of neural adaptation to maintain total activation within certain bounds despite differences in input either from other modules or other neurons in the same module.

5.2. *Braitenberg's two-phase pump of thought*

Explicitly or implicitly, all models of cell assemblies assume that a conscious thought is composed of a set of activated ideas, with an activated idea being simply an activated cell assembly.

Braitenberg [3] generated the idea that threshold control could be used to control thinking beyond merely keeping the total number of activated neurons intermediate between quiescence and epilepsy.

Braitenberg suggested that the threshold control mechanism might have an alternating cycle of low and high threshold values. When the threshold was low, several cell assemblies could be active simultaneously, but when the threshold was raised, only a single cell assembly could remain active.

Braitenberg conjectured that shortly after receiving some perceptual input, the threshold might be set low to recruit a variety of cell assemblies, then raised to permit only the most strongly connected assembly to survive, then lowered to bring in new assemblies, then raised to select only one for survival, etc. Braitenberg suggested that active neurons might exhibit adaptation or fatigue, raising their threshold, and leading to an increased probability for new cell assemblies to enter the next thought. Braitenberg referred to the threshold control mechanism as the "pump of thought".

5.3. *Neuromodulators and the phases of thinking*

Following Braitenberg, the present theory assumes that thinking in any module of the cerebral cortex has different phases, resulting from alteration of the neuromodulatory parameters that control the dynamics of cortical functioning. There are a number of differences from Braitenberg's theory. There are four phases, and no phase or group of phases corresponds to either of Braitenberg's phases. Phases do not always progress in the same cycle. Finally, neuromodulation consists of more than threshold control of the pyramidal neurons.

Web theory postulates: (a) threshold control, (b) self-inhibition of neurons to terminate a thought, which is accomplished by erasure inhibition on dendrites, (c) apical dendritic gating, which regulates the amount of excitatory input to the pyramidal cell body from the apical dendritic tree, and (d) apical synaptic switching, which flips between states — a first state where apical synapses vary in strength depending on the degree of learning and forgetting vs. a second state where all apical synapses have equal strength.

The obvious place for inhibitory neurons to achieve threshold control is on the initial segment of the axon or else on the cell body, and chandelier cells are a plausible candidate, as already mentioned.

Self-inhibition could be achieved either by inhibiting the cell body or initial segment long enough for excitatory generator potentials to dissipate passively or by erasure inhibition of the generator potentials everywhere on the dendritic tree. As previously mentioned, basket cells have the proper synaptic locations on pyramidal cells for erasure inhibition.

The obvious place for apical dendritic gating control is on the main apical dendritic shafts near the pyramidal cell body, achieved either by synaptic or nonsynaptic modulation of the transmission of dendritic generator potentials to the cell body.

The obvious place for apical synaptic switching is all over the apical dendritic tree, and the most likely mechanism would be an nonsynaptic neuromodulatory system, provided it could change state fast enough.

According to the theory, some neuromodulation has multiple gradations and some has only two states with relatively rapid transitions between them. Phases are defined by state changes in discrete control parameters and reversals of direction in graded control parameters.

The theory assumes four phases: two main phases — ***retrieval*** of familiar ideas and ***chunking*** of new ideas, and two subphases of each — ***selection*** and ***completion***. I suspect that different cortical modules need not be in the same phase at the same time, but this paper only deals with a single cortical module. Figure 3 illustrates the main ideas described in more detail in the sections below. Note that: (a) the two selection phases are presumed to be of approximately equal duration, (b) the selection phases are shorter than the completion phases, and (c) chunking completion is longer than retrieval completion. For clarity, Fig. 3 under-represents what I presume to be the disparities in phase duration.

5.3.1. *Retrieval selection phase*

We begin with a cortical module in the ***retrieval-selection*** phase. Self-inhibition of the last active web in the module is triggered at the beginning of the retrieval-selection phase to clear out the idea previously active in the module. Self-inhibition erases all the basal dendritic generator potentials, so none of the basal dendritic "results" of the prior completion phase is present to contaminate the "recognition" of the new apical "input" to the module. It may also erase the apical dendritic generator potentials of only the previously active neurons, but I am unsure of this.

During the entire retrieval-selection phase, apical synapses are modulated so that only learned synapses are functional. That is, weak (decayed, unlearned, or never-learned) synapses do not contribute significantly to apical generator potentials. This is appropriate for the retrieval phase, whose function is to recognize whether the current apical input is sufficiently similar to a pattern of previous input to categorize that input as signaling a familiar idea represented by a web in that module.

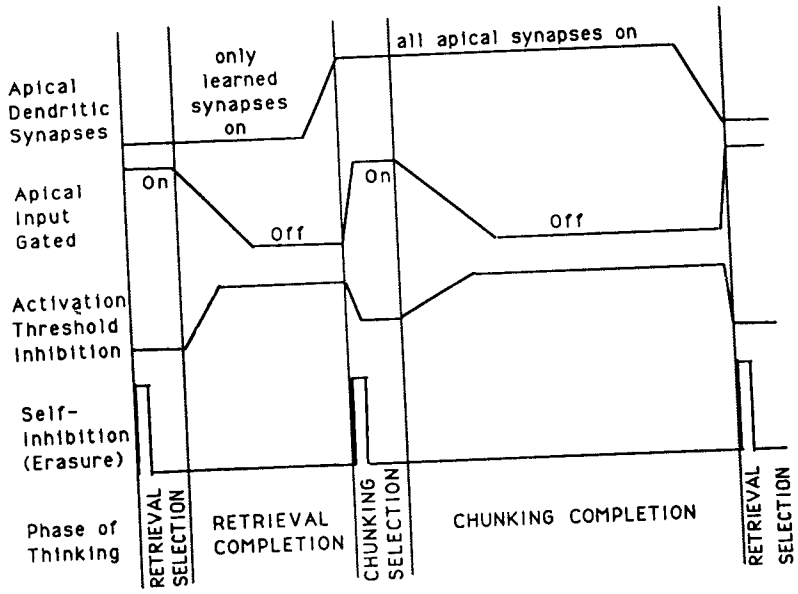

Fig. 3. Four phases of thinking in the cerebral cortex. During retrieval-selection, the basal dendritic potentials in the model are first erased (self-inhibition) and then new input from outside the module is gated on from the apical synapses to the cell body, though only the strong (learned) apical synapses are functional. During retrieval-completion, apical input is gradually turned off and the basal synapses attempt to find the web that is most similar to the initial set of neurons activated during retrieval-selection. If retrieval-completion succeeds, the module returns to the retrieval-selection phase to process the next thought. If retrieval-completion fails, the input is assumed to be a new chunk and the module enters the chunking-selection phase. In chunking-selection, basal dendritic potentials are again erased and apical input is again turned on, but this time all apical synapses, learned or unlearned, are enabled and made equally strong. An initial set of neurons is activated by the apical input, which is followed by the chunking-completion phase that attempts to recruit a new web to represent the input. The threshold-control inhibition is set to result in activation of a number of neurons in the module that is in the appropriate range for the desired size of webs (innate cell assemblies).

Apical dendritic input to the cell body is gated on, since the module "wants" to process new input in that phase to activate an ***initial set*** of neurons within itself that represents that input.

Note that self-inhibition does not erase apical dendritic generator potentials, except possibly for the neurons in the previously active cell assembly. I assume that the decay rate for apical synaptic potentials varies by orders of magnitude over different modules of the cerebral cortex, with sensory feature modules having rapid decay, segment modules having somewhat slower decay, concept modules still slower decay, proposition modules still slower decay, and high-level schemata and procedural modules having still slower decay. The slower the rate of decay of the apical generator potentials in a module, the longer the time window over which the module integrates input to determine initial sets in the selection phases.

The activation threshold is set so that some number of strong apical input synapses must be active on each neuron to activate the pyramidal neuron. It is

an open question in my mind what that number is, and it is beyond the scope of this paper. The threshold in selection phases may increase with increasing levels of apical input excitation of the module.

In addition to the general purpose of threshold control to keep the number of active neurons in a module within some target range, a specific purpose of threshold control in the selection phases is to defer processing of the apical input to some neurons until the processing of competitively greater apical input to other neurons is finished. For example, an action may require body parts to be in a sequence of different positions. The neurons that put body parts in each of the positions in the sequence may well be excited simultaneously, but output activation must necessarily be sequential. Some excited thoughts may need to wait their turn for activation until more strongly excited thoughts have been activated.

5.3.2. *Retrieval-completion phase*

Having activated an initial set of neurons within the module to represent the "perceptual" input to the module, the module now enters the ***retrieval-completion*** phase. The purpose of the retrieval-completion phase is to map the initial set into a web of activated neurons. The origin of the term "completion" is that the initial set contains a subset of the "target" web. When webs in other modules became associated to the target web in this module, a proper subset of their neurons had synapses on a proper subset of the neurons composing the target web. These apical synapses were strengthened by Hebbian contiguity conditioning. In the retrieval-selection phase, these strong synapses provide the input which activates the initial subset of the target web. The function of the retrieval-completion phase is to "completely" activate all the neurons in the web starting from a proper subset.

However, reality must be more complex than this. In order to be able to represent as many ideas as specific neuron coding, web coding must code ideas by means of ***overlapping sets*** of neurons. Thus, every neuron is a member of many different webs representing many different ideas with many different input and output associations. If this is true, then webs in other modules must have strong synapses to neurons in the target module that are not in the target set, by virtue of the associations that many of their neurons have in their roles as members of other webs. This means that in the retrieval-selection phase, the initial set may well contain activated neurons that are not members of the target web. The retrieval-selection phase is ***noisy***, and another function of the retrieval-completion phase is to eliminate the noise neurons in the initial set, as well as to "complete" activation of the target web of neurons.

Apical dendritic input is only gradually turned off over several time steps during retrieval-completion. This permits the initial (input) set of activated neurons to exert a greater influence on the web finally converged upon than would be the case if the initial input was limited simply to selecting the initial set of activated neurons in the retrieval-completion phase. If the initial set were a noiseless proper subset of

the target web, then it would be functionally ideal to retain apical dendritic input potential at full strength throughout the retrieval-completion phase, acting as a "hard constraint" on which web was finally converged upon. However, when there is noise in the initial set, it is probably necessary to fade out the apical input to eliminate the noise neurons from the active set.

The activation threshold is low at the beginning of the retrieval completion phase and gradually rises to the level appropriate for selecting a web of appropriate size for representing ideas. Figure 3 displays an oversimplified representation of the actual neural threshold, which can be increased by the threshold control mechanism whenever the number of active neurons increases beyond some upper bound or decreases below some lower bound.

Furthermore, there may be some noise in the threshold. Figure 3 represents the average threshold when the number of active neurons in the module is within a certain target range.

While apical dendritic input plays a key role in the initial dynamics of the retrieval-completion phase, it is the innate connectivity of the basal dendritic synapses within the module that produces the "asymptotic" active set of neurons starting from the initial set. When the initial set is sufficiently similar to some previously learned web, successful retrieval occurs and the retrieval-completion phase converges to an asymptotic active set that is the target web. Recall that webs are defined solely by the innate link matrix of the basal synaptic system. By the time the apical input has been completely switched off, the result of the apical input has been to put the module into a basin of attraction for one of its webs. From that point on, the retrieval-completion phase progressively moves the set of active neurons along a trajectory that ends at the attractor web. Furthermore, this convergence on a web must be accomplished relatively speedily, presumably within a fraction of a second in many cases. This is the scenario for successful retrieval in web theory.

When retrieval succeeds in converging on a web, the web remains active for several time steps, and this prolonged activation results in Hebbian contiguity conditioning of its recently activated apical dendritic synapses, though this is outside the scope of the present paper. When the same set of neurons fires for a long enough period of time, web theory postulates that a familiarity recognition mechanism is triggered that results in our feeling of familiarity when the module is in the retrieval-completion state. This familiarity mechanism also eventually triggers the self-inhibition mechanism that erases basal dendritic potentials and thereby suppresses the activated neurons in the module to permit activation of a new thought in the module. In the case of successful retrieval, the retrieval-completion phase is followed by another retrieval-selection phase, where previously unprocessed and new apical input from other modules is passed to the cell body for processing and possible recognition.

Retrieval attempts are not allows successful. The input may actually be novel or functionally novel due to forgetting, or the input to the module may be information-

ally too poor or too noisy. All of these cases can be thought of as having initial sets that contain too few neurons in common with any previously activated web and/or too many noise neurons. In these cases, web theory assumes that the retrieval-completion phase fails, within some time limit, to converge on any web. In some of these cases, activation within the module may die out (zero neurons active in the module), because the neurons in the initial set were too few and/or not sufficiently well connected within the basal dendritic system of synapses. Whether activity dies out or merely fails to converge within the time limit, the familiarity recognition mechanism is not triggered within the time limit, and the retrieval-completion phase is succeeded by the chunking-selection phase.

5.3.3. *Chunking-selection phase*

When retrieval fails, basal dendritic potentials are erased by self-inhibition and activation of neurons in the module temporarily ceases. Apical dendritic excitation is not erased. Then apical dendritic input is once again turned on for another try at classifying this input, but now as a new idea, not an old one. Instead of just the learned apical synapses being enabled, now all apical input synapses — learned and unlearned — are enabled and made strong. The activation threshold of the neuron is set so that some number of active apical input synapses are sufficient to activate a pyramidal neuron. Figure 3 shows the threshold being somewhat higher during chunking-selection than during retrieval-selection to compensate for the larger number of enabled apical synapses, but I have no very good reason to assume this. In any case, a new initial set is activated, which might be either larger or smaller than the initial set resulting from the retrieval-selection phase, depending on the relative threshold in the two selection phases. In any case, the initial sets would generally be somewhat different.

5.3.4. *Chunking-completion phase*

The chunking-completion phase is much like the retrieval-completion phase in that its function is to converge on a web starting from an initial set. The principal difference is that, in chunking, the initial set is generally much less similar to any web of the module than is the initial set in retrieval of familiar webs. This means that the time to converge on a web is greater on the average.

Apical input is gradually gated off. The only reason not to turn the apical input off immediately is to induce greater overlap in the neurons composing the initial set and the eventually activated web. It is not clear to me how important this is.

The activation threshold remains below its target level for a longer time than in retrieval-completion to prevent activity from dying out in the module. With an initially lowered threshold, the neural net evolves to states consisting of sets of active neurons that have, on the average, greater and greater internal connectivity. As this happen, the threshold is increased to prevent an epileptic explosion in the number of active neurons and to purge neurons with the least connectivity to the

other active neurons.

Eventually, if chunking is successful, the system converges on a web, which is a set of active neurons whose minimum internal connectivity exceeds its maximum external connectivity.

The module remains in the state of an activated web for a period of time during which Hebbian contiguity condition strengthens all of the active apical input synapses to the neurons in the web. The result of this learning is that, on some future occasion, sufficiently large subsets of these learned apical synapses will select an initial set in retrieval that will reactivate the same web in the retrieval-completion phase. Note that it is the active synapses on the web set that are learned, not the active synapses on the initial set.

Neural networks are dynamical systems whose dynamics depend both on the link matrix (matrices) and the dynamical laws and parameters of the neurons and synapses. In general, even a neural net with all-or-none activation need not converge on any single set of active neurons. It may enter a limit cycle with two or more activation states recurring in a cycle. With random noise added to certain variables or parameters, it is possible to wander about and never converge on either a single state or a cycle. Finally, even if the system does converge on a single set of active neurons, it need not be one of those very rare and special sets that are webs.

For a neural net to converge reliably on webs, it is necessary that the neural net have webs. Webs are defined by purely structural properties of the link matrix, but many link matrices have no webs at all, others have too few webs to have adequate idea representational capacity, and (I am less sure of this) still others may have too many webs for ideal levels of generalization and discrimination of ideas.

For a neural net to converge reliably on webs, its dynamics must also be suitably defined. Long I wandered in a gigantic space of laws and parameters that waxed with my imagination and waned with the frequent experience of numerous simulations gone sour. Finally, I found a link matrix rich in webs and a set of dynamical laws and parameters that converged on these webs with 100% reliability over a run of 1681 randomly selected initial sets. After many strikeouts, there was joy in Mudville. The dynamic model that achieved this joyful result will be described in the next section.

6. Chunking-Completion Simulation

My goal was to find a net that had more webs than neurons, restricting the size range of webs to less than a factor of 3, and to find a physiologically plausible model of neural activation dynamics that converged on a web almost all of the time. It was not very difficult to find nets with more webs than neurons, though the more you restrict the possible range of web sizes, the more difficult it is. However, my initial models of activation dynamics converged on webs less than half the time and getting this percentage close to 100% was very difficult. Furthermore, there is some interaction between net structure and the ideal model of activation dynamics.

Eventually, I succeeded in finding a neurally plausible model of activation dynamics that converged on webs 100% of the time in 1681 trials in a neurally plausible symmetric proximity net with 289 neurons that had at least 597 webs in a size range from 30 to 84 neurons. Of course, cortical modules are much larger than 289 neurons and the proximity net was only two-dimensional, not three-dimensional, but edge effects were eliminated by wrapping the net around into a toroid and the maximum web size was small enough to avoid webs that wrapped around the toroidal net. In short, I believe that except for the likely possibility that real cortical webs might be larger than 30–84 neurons, this simulation of chunking was a reasonable model of what I imagine real chunking to be like in a local region of a real cortical module.

This was never intended to be a systematic study of different types of nets and different models of activation dynamics. The spaces of each are vast. My goal was to find a combination of a net and a dynamics model that worked and that satisfied my criteria for neural plausibility. I did find such a combination, and I will describe that combination in this section along with some of the tentative intuitive conclusions I came to about the net structure and activation dynamics in the process. However, there are many more criteria that a good net-dynamics combination should satisfy than are satisfied by the combination described here. For one, as will be pointed out in the retrieval-completion section, the reliability of retrieval-completion of these webs is less than I think it can be.

6.1. *Symmetric proximity link matrix*

The net that produced 100% convergence on 597 different webs in 1681 trials was a two-dimensional symmetric proximity net with 289 neurons arranged at the lattice points of a 17 by 17 grid with each dimension being cyclic, so that the closest neighbors of neuron (1,1) are neurons (1,17), (1,2), (17,1), and (2,1). Note that the foregoing labeling is "clock-like", with 17 being equivalent to 0 in modular arithmetic labeling.

The net was constructed in two steps. First, an asymmetric proximity net was generated with the following inlinking probability metric around each neuron: Distance between neurons in Euclidean, with the unit being the distance between two adjacent lattice points, e.g. (1,1) and (1,2). The inlinking probability falls linearly at the rate of 0.15 per unit distance from (a hypothetical) 0.9 at 0 distance from the center neuron, but there is no self-linking, and inlinking probability is set to zero beyond a distance of 5 from the center neuron. Thus, the actual inlinking probabilities are: 0 at $d = 0$, 0.75 at $d = 1$, 0.6 at $d = 2$, 0.45 at $d = 3$, 0.3 at $d = 4$, 0.15 at $d = 5$, and 0 at $d > 5$. Of course, there are also neurons at noninteger distances.

After the asymmetric proximity net was generated, all the asymmetric links were deleted. This resulted in a symmetric proximity link matrix with an average of 14 links per neuron, which is less than the square root of the number of neurons in

even this very small net (bycerebral cortex standards). With innate web coding of ideas, sparse connectivity suffices to yield lots of webs.

6.2. *Activation dynamics model*

The model of activation dynamics in the chunking-completion phase that achieved 100% convergence to webs in 1681 attempts is described in this section.

6.2.1. *Synchronous discrete-time updating*

Time is discrete, and all neural and synaptic parameters are updated synchronously at each time step. While asynchronous updating is probably more reasonable physiologically, it is computationally very difficult to do asynchronous updating properly.

The easy ways that I have seen asynchronous updating done are much poorer approximations to neural reality than synchronous updating. In the beginning, I tried various forms of asynchronous updating similar to that of Hopfield [22]. They worked quite well to produce convergence on attractors, but the more I thought about the validity of updating every outlink of a neuron when I updated the neuron (randomly, cyclically, or however), the more unreasonable it seemed, since the true time between updating different neurons in an asynchronous scheme is tiny fractions of a millisecond, several orders of magnitude shorter than real synaptic delays.

Synchronous updating cannot pretend to be more than an approximation to reality, but at least it is a reasonable approximation, with the time step being the average synaptic delay time. For simulation of a small part of a single cortical module with all connections being local, this approximation is especially plausible.

6.2.2. *All-or-none threshold activation*

Activation of a neuron is all-or-none — i.e. spike or no spike at time t. A neuron is activated at time t, if excitation at time t equals or exceeds the threshold at time t.

6.2.3. *Excitation — apical and basal*

Excitation (potential) of a neuron at time t is a continuous property at time t, which equals the sum of apical input excitation and basal weblink excitation. Apical input excitation is from outside the 289-neuron net, whereas basal excitation comes from feedback connections within the 289-neuron net, as specified by the symmetric proximity links matrix described above.

Apical input excitation of a randomly selected set of 40 neurons (out of the 289 neurons in net) is set at a level equal to six active basal inlinks at time $t = 0$. Note that I am here referring to apical excitation, but measured using the excitation produced by one active basal inlink as the unit of measurement. This is above the lowered threshold at $t = 0$, and these 40 neurons constitute the initial set activated in the chunking-selection phase. Apical input is zero for all other neurons at $t = 0$ and throughout the chunking-completion phase. Apical excitation for the initially

selected set of 40 neurons decays at the rate of 1.2 basal inlinks per time step, so that the apical input contribution to excitation for the selected neurons is zero at $t = 4$ in the chunking-completion phase.

Basal excitation at time t equals the number of active inlinks at time t, and has no persistence. That is, basal excitation at time t is measured simply by the number of active inlinks from other neurons within the 289-neuron net, and basal excitation is the only source of excitation after apical input excitation drops to zero at time $t = 4$.

6.2.4. *Basal inlink persistence*

However, while overall basal excitation of each neuron-i (measured say at the cell body) has no persistence, each active inlink-ij to neuron-i does have a persistence of excitation, remaining in the active state for exactly two time periods (t and $t+1$) when neuron-j is active at time t. There are at least two possible neural mechanisms for this — either a neuron emits a train of two spikes (at t and $t+1$) whenever its excitation exceeds its threshold, or basal dendritic spines remain in the active state for two time periods from a single presynaptic spike. In the present dynamics model, the two mechanisms produce essentially equivalent network performance. Whatever the mechanism, ***basal inlink persistence*** is functionally important, because it greatly reduces the tendency of neural dynamic systems to get into limit cycles of length-2, which is the most common alternative to converging on an attractor, in my experience.

6.2.5. *Plateau thresh function*

The thresh function used in the dynamic model that achieved 100% convergence to webs in 1681 attempts was monotonic nondecreasing with the number of active neurons in the module (Y), but more complex than a simple linear function. The ***plateau thresh function*** was composed of three different linear segments, a low constant threshold ($h = 1.9$ active basal synapses) over a range from 0 to 29 active neurons, a medium constant threshold ($h = 6.5$) over a range from 30 to 84 active neurons, and an abrupt increase in h at 85 active neurons ($h = 11$) with a linear increasing h function after that ($h = 11 + .07\{Y - 85\}$, where $\{x\} = \max(x, 0)$, i.e. $\{x\} = 0$ for negative values of x and $\{x\} = x$ for positive values of x).

What this plateau thresh function accomplished was to keep the number of active neurons within the range of 30 to 84 almost all of the time and virtually guarantee that any asymptotic equilibrium state would be a web with a minint ≥ 7 and a maxext ≤ 6 and a size between 30 and 84 neurons. Whenever the active set contained 85 or more neurons, the steep increase in threshold would purge the least well-linked members of the active set, pushing the active set size back below the upper cutoff. Whenever the active set contained fewer than 30 neurons, the sharply lowered threshold would recruit more neurons to the active set, pushing the active set size above the lower cutoff. Actual dynamics were complicated somewhat by

other factors to be discussed later, but the above description is correct to a first approximation.

6.2.6. *Noisy thresholds*

A uniformly distributed random noise component was added to the neural threshold at each time step. The uniform noise distribution was over the range of ± 0.6 active basal inlinks at almost all time steps, except that after $t = 43$, whenever t was a multiple of 11 (e.g. 44, 55, 66, ...), the range was ± 2.5 basal inlinks. The noise increment or decrement to the threshold was perfectly correlated over all 289 neurons in the net, as if it came from a common source, e.g. a set of inhibitory neurons that summed pyramidal neuron activation in this cortical region and inhibited pyramidal cells according to the thresh function with this added noise.

Noise was added, not for increased realism, but because it was my impression that noise was functional in escaping from limit cycles and nonweb attractors such as states where minint = maxext. There is some similarity in both mechanism and purpose between noise and the temperature parameter of simulated annealing [33], but the analogy is far from perfect. In both cases, the purpose is to increase the probability that the dynamical system will not get stuck in less desirable states or cycles. However, rather than have maximum noise in the beginning, noise was constant, except that if the net had not reached an equilibrium state by $t = 43$, there was a large increase in the noise every 11 time steps in the attempt to shove the system into a new region of the state space. In the beginning, there is no need for noise in this type of model of activation dynamics. The need arises later, particularly in cases where the system gets stuck in a cycle. Noise is also helpful in pushing the system away from weak attractors such as states where minint = maxext. Perfectly correlated noise was used because it seemed to work better than uncorrelated noise in getting out of undesirable states, which is reasonable, since correlated noise provides a bigger push than uncorrelated noise.

6.2.7. *Initial thresholds reduced*

Although one probably could rely on the above thresh function from the beginning, it seemed to speed convergence to reduce the threshold a little at the beginning of the chunking-completion phase. In the most successful simulation, this amounted to a reduction of 1.6 active basal inlinks at $t = 0$ linearly approaching zero at the rate of 0.3 inlink per time step, and thus reaching zero by $t = 6$.

6.2.8. *Stopping criterion — high similarity of activation state*

How does a cortical module "know" when it has reached a web and "ought" to end the chunking phase? It is trivial to write a computer program that tests for whether the set of active neurons is or is not a web. I wrote greedy algorithms to find webs that hill-climbed on the minint and did a web check on every set of active neurons the system passed through. That works nicely for estimating the number of webs in

a net. If we view a neural net as having the power of a Turing Machine, we might just assume that cortical neural nets do that. Intuitively, checking for whether the active set is a web is much more plausible for a neural net than the backpropagation learning algorithm. But I don't regard that as a very good argument is favor of incorporating such web checks in a model of activation dynamics.

Instead, I assumed that stopping depended on computing the similarity of the last two states, computing a running average of the current similarity weighted 0.75 and the prior average similarity weighted 0.25, and stopping when this average similarity exceeded 0.999. These are not necessarily the ideal parameters, but they worked well in stopping on webs and not stopping on nonwebs. Functionally, these parameters stopped the system when the same set of neurons was active on four or five consecutive time steps.

How plausible is it to assume that cortical modules compute the similarity of which neurons are active in adjacent time intervals? It can be done locally by having neurons that respond to changes in rate of firing, and then it can be summed more globally over the entire module or a region thereof. We know that recognition of a change in stimulation is something that can be done by nervous systems much simpler than the cerebral cortex, so it seems like a recognition of activation similarity is a plausible stopping criterion.

6.3. *Simulation results*

The simulation stopped on a web 100% of the time from 1681 different random initial states consisting of 40 active neurons. Of the 1681 terminal web states, there were 597 different webs ranging in size from 30 to 84 neurons. Figure 4 plots the average threshold as a function of the number of active neurons, using the thresh function for this simulation without the random noise and ignoring the reduction in threshold over the first six time steps. Figure 4 also plots the minint for the terminal webs as a function of the number of neurons in the web. Obviously, the dynamical system was successful limiting terminal states to webs with minints of 7 and maxexts of 6 (rarely less).

The average number of time steps to reach the stopping criterion was 25 and the median was 20. The median time to reach the terminal web was about 16. If each time step represents a synaptic delay 3 ms, this means that median chunking-completion time was $20 \times 3 = 60$ ms. Seventy-five per cent of the time chunking finished in under 28 time steps — 84 ms, and 90% of the time it finished in 46 time steps — 138 ms.

7. Retrieval-Completion Findings

Perhaps the most common test of the adequacy of learned cell assemblies is what I call the completion test of retrieval of cell assemblies starting with an initial subset of each target assembly. Of course, one must define some model of retrieval dynamics, which may or may not be the same as the dynamics of the neural net during learning.

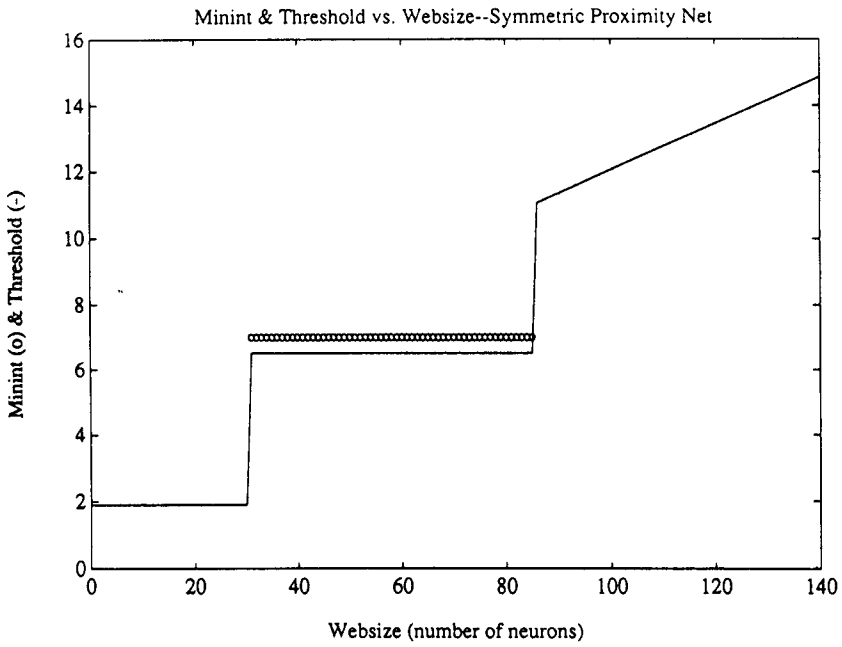

Fig. 4. The threshold and the minint (minimum internal connectivity) for each of the webs as a function of websize (the number of active neurons). All webs that were attractor states for this net with these dynamics had a minint of 7 and a maxext of no more than 6, usually exactly 6. The threshold function that achieved this and pushed the number of active neurons back into the desired range from 30 to 84, whenever that number left the range, was a plateau function with an abrupt decrease from 30 to 29 active neuron and an abrupt increase from 84 to 85 active neurons. From 30 to 84, the value of the threshold was 6.5, halfway between the minint and maxext. The gradual increase in threshold after 85 may or may not be necessary, but it served as additional insurance against an epileptic breakout.

The model of retrieval dynamics must include a stopping criterion, which might be to stop after a fixed number of time steps and output whatever the active is at that point or to wait until convergence has been obtained, as measured by the similarity of the set of active neurons over several successive time steps. Having a model of retrieval dynamics, the connection matrix for the neural net, and the set of neurons composing each cell assembly, one then takes a random sample of some percentage of the neurons in each cell assembly as the set of initially active neurons and records the terminal set of active neurons after the system stops.

For many purposes, the best measure of success in retrieval-completion is provided by the symmetric difference, D, between the terminal set and the target assembly. The symmetric difference of two sets of neurons A and B is the sum of the number of neurons in the differences between the two sets, $|A - B| + |B - A|$, divided by the number of neurons in the union of A and B, $|A \cup B|$. $|A - B|$ is the number of the neurons in A that are not in B, while $|B - A|$ is the number of neurons in B that are not in A. The symmetric difference, D, is $1 - S$, where S is the symmetric similarity. The symmetric similarity is the number of neurons in the

intersection of A and B divided by the number of neurons in the union of A and B.

$$D = \frac{|A - B| + |B - A|}{|A \cup B|} = 1 - \frac{|A \cap B|}{|A \cup B|} .$$

I performed a retrieval-completion test starting with initial subsets of 50% of the neurons in the target web. Target webs were from the set of 597 webs (innate assemblies) obtained by the best model of chunking-completion dynamics. I used the same basic model of retrieval-completion dynamics that was used for chunking-completion dynamics, but the parameters were slightly different.

The best set of retrieval parameters achieved a symmetric difference of $D = 0.07$ on a sample of 597 trials (once for each of the 597 webs). This means that the overlap of the terminal set and the target web contained 93% of the neurons in their union. On 30% of the trials, the terminal set was identical to the target set, on 20% of the trials, it was a subset of the target web, and on 35% of the trials, it was a superset of the target web. On the remaining 15% of the trials, the terminal set overlapped the target web extensively, but was neither a subset, nor a superset.

The average number of time steps to converge to the terminal set was 8.6 time steps, which is considerably faster than the average convergence time in chunking-completion (25 time steps).

I also studied retrieval-completion for 410 webs in a different symmetric proximity net that had 289 neurons and an average of 32 links per neuron. In this case, I varied the percentage of neurons from the target web that were in the initial set over the range from 20% to 70%. The symmetric differences were $D = 0.18$ for 20% overlap of initial set and target web, $D = 0.12$ for 30%, $D = 0.06$ for 40%, $D = 0.03$ for 50%, and $D = 0.02$ for 70%. Note that retrieval-completion was better for the symmetric proximity net with greater link density. I think that link density is the main factor, but I cannot completely rule out differences in retrieval dynamics.

In this second case, for the ideal decision maker with random sets, the symmetric difference should be essentially 0 when 20% of the neurons in the target web are activated at the onset of retrieval. Since the best dynamical system for retrieval was largely identical to the dynamical system that was used in finding the 410 webs, my guess is that the disparity between real and ideal performance in retrieval is largely due to the fact that webs are nonrandom sets of neurons rather than to any deficiency in the dynamical system for retrieval, but I am not certain of this.

It may be possible for web coding to provide almost as good discriminability of ideas as randomly selected sets of neurons, and indeed I have some indication of this is a set of webs from a 1000-neuron symmetric regular random net. However, for reasons of neural plausibility I am currently concentrating on proximity nets.

This brief study of retrieval was only meant to suggest the likelihood that satisfactory retrieval-completion can be obtained for webs in symmetric proximity nets. A more extensive study is needed with more extensive variation of the models and parameters of retrieval-completion dynamics, variation of net structure and target web size to determine their effects on retrieval-completion, systematic variation of

the percentage of target set neurons in the initial set, and study of the effects of noise in the initial set (activated neurons not from the target set). Finally, one should formulate and investigate the properties of plausible models of the apical dendritic connections within and between modules to determine realistic values for the number of target and noise neurons that are active in the initial sets, which are the product of the retrieval-selection and chunking-selection phases.

8. Learning

Although I assume that the number of ideas actually represented in any one human mind is finite, I also assume that the number of possible ideas that a human mind could represent is infinite. Using overlapping set coding of ideas, whether learned cell assemblies or innate webs, this means that learning must be involved in the coding of ideas. Where is the learning in web theory? I have already said where it is, but the focus of this paper is so much on the innate aspects of web theory that I think the role of learning needs to be summarized in a section of its own.

In web theory, the modifiable synapses are the apical dendritic synapses on pyramidal neurons which are presumed to encode the associations between ideas. The basal dendritic synapses are assumed to bind together the neurons that make up webs, and these basal synapses are assumed to be innate and unmodifiable. Web theory might be modified to accommodate a modest role for learning in the basal synapses that are presumed to bind together the neurons of each web, but the primary function of learning is to associate webs, not to integrate the neurons within a web.

Web theory assumes that the potential representatives of ideas, the webs, are already there, with integrating basal connections in place, prior to associative learning. But, prior to learning, the webs do not represent anything, they have no meaning (no semantics), and they serve no function in the mind or for the organism, because they have no differentially strong input or output apical associations to other neurons in the cerebral cortex or in subcortical areas of the nervous system.

When a web is first activated by the chunking-completion process, the neurons of that newly activated web acquire strong input and output links to the neurons in other webs that were recently active and that have connections to the newly activated web. This gives meaning to the newly activated web. Whenever that web is activated again in retrieval, if the prior or subsequent context is slightly different, then the web in question will have its meaning modified somewhat by the acquisition of associations to the ideas in the slightly different context. I assume that all cortical ideas have the potential to represent classes of events, not just single events.

9. Conclusions

In symmetric nets and in proximity nets, there are a number of reasons for thinking that webs are well suited to be the representatives of ideas: First, there are lots

of webs — probably more than enough in either proximity nets or symmetric nets to represent all the ideas that humans can represent. Second, webs are equilibrium states in reasonable dynamical network activation models. Achieving 100% convergence on webs in a 289-neuron net with a particular dynamics model shows that, in this case, the webs had large basins of attraction that appeared to largely or completely span the entire space of initial sets of the size used. Achieving this perfect convergence on webs with an average of 25 time-steps also encourages one to believe that this model of idea representation may have some validity. Third, retrieval-completion of webs starting from a proper subset of a web is very fast and reasonably accurate. There is also little doubt that higher levels of retrieval-completion accuracy can be achieved than I have so far obtained. Fourth, findings on web density, chunking, and retrieval in proximity nets seem certain to generalize to larger nets with the same size webs, and there is a good chance that they will generalize to larger webs in larger nets. Finally, there is a neurally plausible way to generate a symmetric proximity net, which is to generate an asymmetric proximity net and then *delete* all of the asymmetric links.

References

[1] D. O. Hebb, *Organization of Behavior* (Wiley, New York, 1949).
[2] C. R. Legéndy, *Math. Biosci.* **1**, 555 (1967).
[3] V. Braitenberg, in *Theoretical Approaches to Complex Systems*, ed. by R. Heim and G. Palm (Springer-Verlag, Berlin, 1978) p. 171.
[4] G. Palm, *Neural Assemblies* (Springer-Verlag, Berlin, 1982).
[5] G. A. Miller, *Psychol. Rev.* **63**, 81 (1956).
[6] V. Braitenberg and A. Schuz, *Anatomy of the Cortex* (Springer-Verlag, Berlin, 1991).
[7] M. Abeles, *Corticonics* (Cambridge University Press, Cambridge, UK, 1991).
[8] V. Braitenberg, in *Architectonics of the Cerebral Cortex*, ed. by M. A. B. Brazier and H. Petsche (Raven Press, New York, 1978) p. 443.
[9] G. Palm and V. Braitenberg, in *Progress in Cybernetics and System Research*, ed. by R. Trappl, G. J. Klir and L. Ricciardi (Wiley, New York, 1979) p. 369.
[10] T. Kohonen, P. Lehtio and J. Rovamo, *Ann. Acad. Sci. Fenn. A* **167**, 1 (1974).
[11] R. J. Douglas and K. A. C. Martin, in *The Synaptic Organization of the Brain*, ed. by G. M. Shepherd (Oxford University Press, New York, 1990) p. 389.
[12] H. H. Barlow, *Perception* **1**, 371 (1972).
[13] J. A. Feldman and D. H. Ballard, in *Human and Machine Vision*, ed. by A. Rosenfeld and J. Beck (Academic Press, New York, 1983) p. 107.
[14] C. R. Legéndy, in *Cybernetic Problems in Bionics*, ed. by H. L. Oestreicher and D. R. Moore (Gordon & Breach, New York, 1968) p. 721.
[15] G. Palm, *Biol. Cybern.* **36**, 19 (1980).
[16] G. Palm, in *Brain Theory*, ed. by G. Palm and A. Aertsen (Springer, Berlin, 1986) p. 211.
[17] G. Palm, *Science* **235**, 1227 (1987).
[18] G. Palm, *Concepts Neurosci.* **2**, 97 (1991).
[19] C. Meunier, H.-F. Yanai and S.-T. Amari, *Network* **2**, 469 (1991).
[20] B. G. Cragg, *Brain* **98**, 81 (1975).
[21] J. A. Anderson, J. W. Silverstein, S. A. Ritz and R. S. Jones, *Psychol. Rev.* **84**, 413 (1977).

[22] J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
[23] T. Kohonen, *IEEE Trans. Comput.* **C-21**, 353 (1972).
[24] T. Kohonen, P. Lehtiö, J. Rovamo, J. Hyvärinen, K. Bry and L. Vainio, *Neuroscience* **2**, 1065 (1977).
[25] W. A. Wickelgren, *Psychol. Rev.* **86**, 44 (1979).
[26] G. Palm, *Biol. Cybern.* **39**, 181 (1981).
[27] G. Palm, *Concepts Neurosci.* **1**, 133 (1990).
[28] P. M. Milner, *Psychol. Rev.* **64**, 242 (1957).
[29] G. Palm, in *Real Brains — Artificial Minds*, ed. by J. L. Casti and A. Karlqvist (Elsevier, New York, 1987) p. 165.
[30] S. Grossberg, *J. Theor. Biol.* **27**, 291 (1970).
[31] S. Grossberg, *Kybernetik* **10**, 49 (1972).
[32] S. Grossberg, *Stud. Appl. Math.* **52**, 213 (1973).
[33] S. Kirkpatrick, J. C. D. Gelatt and M. P. Vecchi, *Science* **220**, 671 (1983).