Webs, Cell Assemblies, and Chunking in Neural Nets: Introduction

WAYNE A. WICKELGREN, *Columbia University, New York*

Abstract This introduction to Wickelgren (1992), describes a theory of idea representation and learning in the cerebral cortex and seven properties of Hebb's (1949) formulation of cell assemblies that have played a major role in all such neural net models.

Ideas are represented in the cerebral cortex by **webs** (innate cell assemblies), using sparse coding with sparse, all-or-none, innate linking. Recruiting a web to represent a new idea is called **chunking**. The innate links that bind the neurons of a web are basal dendritic synapses. Learning modifies the apical dendritic synapses that associate neurons in one web to neurons in another web.

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 or 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 of 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 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 Wickelgren (1992), 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 a subset of 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 of 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 of Wickelgren (1992) 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 span is that it provides a major source of motivation for the

Canadian Journal of Experimental Psychology, 1999, 53:1, 118-131

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 Wickelgren (1992).

Neural Representation of Ideas and Associations

Wickelgren (1992) 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 (Hebb, 1949).

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 directly connected or they are not.

CELL ASSEMBLIES

Hebb (1949) proposed that ideas are represented in the cerebral cortex by overlapping sets of neurons called cell assemblies. 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 of Hebb, 1949, p. 196). 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 (1967) 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 (1978) and Palm (1982) 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. Hebb's associative synaptic learning hypothesis became famous independent of its use in establishing cell assemblies.

Seventh, Hebb anticipated Miller's (1956) **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."

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 Wickelgren (1992) 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 nonnegative real number. In the cortical models of Wickelgren (1992), 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 the complete paper, Wickelgren (1992), 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 Wickelgren (1992). 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 Wickelgren (1992), 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.

ASSOCIATION OF IDEAS

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, Wickelgren (1992) does not use neural overlap, only learned synaptic association, as a basis for the association between cell assemblies.

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 Wickelgren (1992).

Only one type of learnable associative link between pyramidal neurons is assumed in Wickelgren (1992), and those associative links are not even 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 and modeled extensively are innate links that are presumed to bind neurons together into innate cell assemblies called webs.

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 synaptic 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 Wickelgren (1992), 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

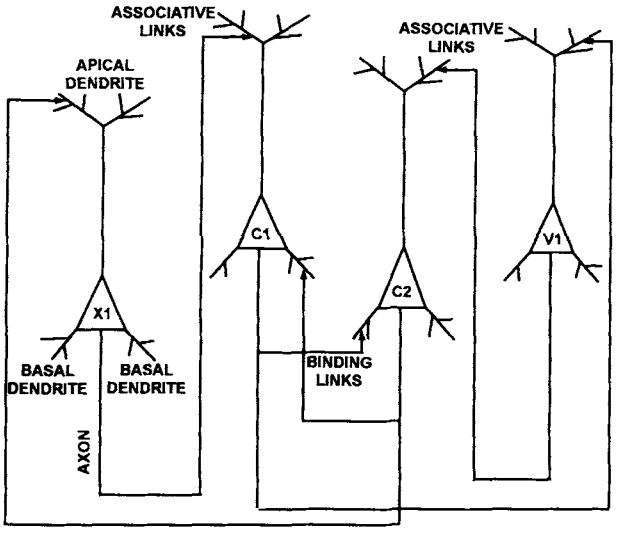

Figure 1. Apical vs. basal dendritic excitatory synapses among pyramidal neurons in the cerebral cortex. A cell assembly of 100-10,000 pyramidal neurons is assumed to represent an idea. Neurons C1 and C2 are two neurons belonging to the same cell assembly C. C1 tends to activate C2 and vice versa via 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 Figure 1, 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.

properties) of the synapses between the neurons in the net.

The theory presented in Wickelgren (1992) 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 Figure 1. Basal dendrites branch extensively near the cell body and receive input synapses from nearby pyramidal neurons (Braitenberg & Schuz, 1991). 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 (Braitenberg & Schuz, 1991). The derivations, calculations, and simulations of the properties of neural nets discussed in Wickelgren (1992) 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 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 the axon of pyramidal neurons (Abeles, 1991), 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 which has already been completely processed so that it does not interfere with processing 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 (Abeles, 1991), are in a good position to erase the prior state of excitation (depolarization) of a pyramidal neuron. It is plausible that at low to medium levels of activation basket cells regulate pyramidal activation thresholds with no erasure, but at higher levels of activation basket cells erase excitation.

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.

APICAL VERSUS BASAL DENDRITIC SYSTEMS OF SYNAPSES

I got the idea of distinguishing the functions of apical and basal dendrites from Braitenberg's (1978) distinction between the A (apical) and B (basal) systems of synapses among cortical pyramidal neurons. Braitenberg and Schuz (1991) consider this distinction to be similar to the distinction between an ametric and metric system of synaptic connections proposed by Palm and Braitenberg (1979), where "metric" means that the probability of a synaptic connection decreases with increasing distance and "ametric" 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 (1978) 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 (1974) 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. Kohonen, Lehtio, and Rovamo 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 (1978), 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, Lehtio, and Rovamo 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 (1990), 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 (Douglas & Martin, 1990).

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 need only 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 serve as the starting point for both chunking and retrieval by the basal system, but the apical system is not investigated in Wickelgren (1992).

Coding of Thought

SPECIFIC NEURON CODING — GRANDMOTHER CELLS

The simplest way for neurons to code ideas is specific neuron coding, so called in honour 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) (Barlow, 1972). Thinking of grandmother means that the grandmother neuron is firing at a high rate.

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 (Feldman & Ballard, 1983). That is, the sets of neurons representing any two different ideas do not overlap — cell assemblies are nonoverlapping. 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 very 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 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 require 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 any 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 favour of chunking given in Wickelgren (1992) 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 Wickelgren (1992), whose focus is to provide a mechanism for chunking.

OVERLAPPING SET CODING — CELL ASSEMBLIES

More promising than multiple (nonoverlapping 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 (1949) in his cell assembly model. However, Legéndy (1967) 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 a reduction in idea representational capacity (Legéndy, 1967, 1968).

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 (1967), Palm (1980, 1986, 1987, 1991), and Meunier, Yanai, and Amari (1991) have demonstrated that there can be a high degree of discriminability in the representation of different ideas with overlapping set coding.

As Palm (1980, 1986, 1987, 1991), and Meunier et al. (1991) 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 idea 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 10

neurons (labelled 0,1, ... ,9). Represent each idea by a subset of three neurons.

There are 120 different (unordered) sets of 3 neurons, which would allow us to code 120 different ideas, an order of magnitude more than the 10 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 2 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 3 neurons each, such that any subset of 2 neurons is sufficient to identify uniquely which idea (set of 3 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 Wickelgren (1992).

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.

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 approximately 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 (1956) and Hebb (1949).

WEBS — INNATE CELL ASSEMBLIES

Sparse linking and innate versus learned cell assemblies. Using a graph-theoretic approach, Palm (1982) made a major advance in the precise formulation of the concept of overlapping cell assemblies. 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 (1991) has done extensive investigations of Hebbian learning. However, there is actually no role for learning in Palm's (1982) 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.

Palm (1986) 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. 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 (1986) has a clever suggestion for escaping this dilemma. He assumes that the neurons of the cerebral cortex are partitioned into modules, with the neurons having a high connection probability, about .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 (1975) 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 (<5,000) 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 (1991) mathematical-anatomical study of the connection probabilities of pyramidal neurons in the mouse cortex. They found that the connection probability of nearby pyramidal neurons was on the order of .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 1,000 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

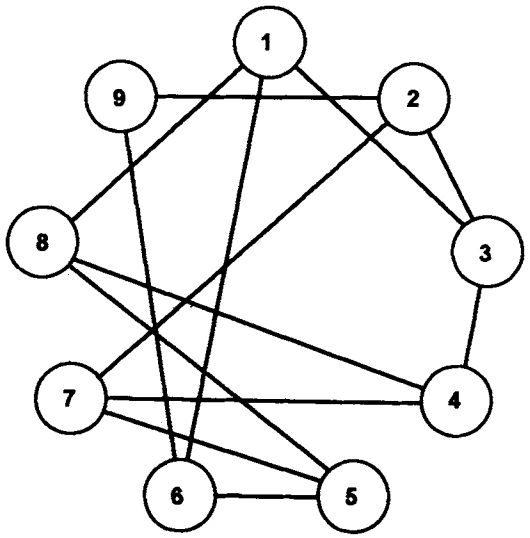

Figure 2. A tiny neural net with 9 neurons that contains the 6 webs shown in Table 1. This net is symmetric, and each line between neurons represents two links, one in each direction. Each web has a minint of 2 and a maxext of 1, so it will be an equilibrium state using a threshold activation function with a threshold of 1.5 active inlinks.

of the basal synapses that bind together the neurons within a cell assembly. Note that some basal synapses may be modifiable, and all basal synapses may be modifiable during development — indeed Wickelgren (1992) used the hypothesis of modifibility during development to provide a plausible neural mechanism for the development of a symmetric connection matrix for the web-defining basal synapses. What is essential for the theory is that there be some class of basal synapses that are unmodifiable after some stage of development.

Precise characterization of the role of learning versus innate structure in the mind and the brain is a central problem in psychology and neuropsychology. Legéndy (1967) 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 subassemblies 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 (after some stage of development).

Palm's definition of cell assemblies. Although Palm's (1982) approach is primarily structural, his actual definition of an assembly relied on activation dynamics, rather than being

TABLE 1
Webs in 9-Neuron Net of Figure 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

purely structural. 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∪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).

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 Figure 2 represents two links, one in each direction. Thus, the net is symmetric. The net in Figure 2 contains the six webs listed in Table 1. For each web, the minint is two, and the maxext is one. 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 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 (1982) refered 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 (1982) employed a purely structural definition when he determined the number of cell assemblies in various systematically constructed graphs. I presume that definition was nearly equivalent to the one given here.

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 Figure 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 subassemblies.

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 sufficiently large 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 Figure 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 be 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 on 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 Wickelgren (1992). 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, many subsets of the web have the often-desired Hebbian completion property.

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 nonoverlapping 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, because 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 nonPalm-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 required 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. This 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 reasonably 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.

ICONIC, ELABORATIVE, ABSTRACTIVE, AND ELABSTRACTIVE CODING

In **iconic coding,** any simultaneously active set of neurons becomes a cell assembly, perhaps excluding a set that is 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 become so via learning) that they form a 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 (Anderson et al., 1977; Hopfield, 1982; Kohonen, 1972; Meunier, Yanai, & Amari, 1991; Palm, 1991). 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 (Kohonen, Lehtio, & Rovamo, 1974; Kohonen et al., 1977).

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 that 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 versus 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.

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 10 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 $160 = 10 \cdot 4^2$ 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 k^{th} level ideas would contain $s \cdot 4^{k-1}$ neurons. As discussed by Legéndy (1968), 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 (Wickelgren, 1979). 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 average, larger cell assemblies would have more potential input and output synapses than smaller cell assemblies, and more familiar (frequently 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 versus the largest cell assemblies, at least within a module.

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 representation 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 neither 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.

CELL ASSEMBLIES ARE LOCAL, NOT GLOBAL (DIFFUSE)

Braitenberg (1978) and Palm (1981) 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 is demonstrated in Wickelgren (1992), 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 to be 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.

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 (1978) 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. 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 .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 .1 and .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 nonoverlapping? 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 being 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 and perhaps basal (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 Wickelgren (1992) makes no distinction between these cases.

This chapter contains material from W. Wickelgren's article in Concepts in Neuroscience (1992) which is reproduced here with permission of World Scientific Publishing Co.

Send correspondence to Wayne Wickelgren, Department of Psychology, Columbia University, New York NY 10027.

References

Abeles, M. (1991). *Corticonics.* Cambridge: Cambridge University Press.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84,* 413-451.

Barlow, H. H. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception, 1,* 371-394.

Braitenberg, V. (1978). Cortical architectonics: General and areal. In M. A. B. Brazier & H. Petsche, (Eds.), *Architectonics of the cerebral cortex* (pp. 443-465). New York: Raven Press.

Braitenberg, V. (1978). Cell assemblies in the cerebral cortex. In R. Heim & G. Palm, (Eds.), *Theoretical approaches to complex systems* (pp. 171-188). Berlin: Springer-Verlag.

Braitenberg, V., & Schuz, A. (1991). *Anatomy of the cortex.* Berlin: Springer-Verlag.

Cragg, B. G. (1975). The density of synapses and neurons in normal, mentally defective and ageing human brains. *Brain, 98,* 81-90.

Douglas, R. J., & Martin, K. A. C. (1990). Neocortex. In G. M. Shepherd, (Ed.), *The synaptic organization of the brain* (pp. 389-438). New York: Oxford.

Feldman, J. A., & Ballard, D. H. (1983). Computing with connections. In A. Rosenfeld & J. Beck, (Eds.), *Human and machine vision* (pp. 107-155). New York: Academic Press.

Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences USA, 79,* 2554-2558.

Kohonen, T. (1972). Correlation matrix memories. *IEEE Transactions on Computers, C-21,* 353-359.

Kohonen, T., Lehtio, P., & Rovamo, J. (1974). Modelling of neural associative memory. *Annales Academia Scientiarum Fennica A, Medica, 167,* 1-18.

Kohonen, T., Lehtio, P., Rovamo, J., Hyvarinen, J., Bry, K., & Vainio, L. (1977). A principle of neural associative memory. *Neuroscience, 2,* 1065-1076.

Legéndy, C. R. (1967). On the scheme by which the human brain stores information. *Mathematical Biosciences, 1,* 555-597.

Legéndy, C. R. (1968). How large are Hebb's cell assemblies? In H. L. Oestreicher & D. R. Moore, (Eds.), *Cybernetic problems in bionics* (pp. 721-724). New York: Gordon & Breach.

Meunier, C., Yanai, H.-F., & Amari, S.-I. (1991). Sparsely coded associative memories: Capacity and dynamical properties. *Network, 2,* 469-487.

Miller, G. A. (1956). The magical number seven, plus or minus

two: some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Palm, G. (1980). On associative memory. *Biological Cybernetics, 36,* 19-31.

Palm, G. (1981). Towards a theory of cell assemblies. *Biological Cybernetics, 39,* 181-194.

Palm, G. (1982). *Neural assemblies.* Berlin: Springer-Verlag.

Palm, G. (1986). Associative networks and cell assemblies. In G. Palm & A. Aertsen, (Eds.), *Brain theory,* p. 211. Berlin: Springer.

Palm, G. (1987). Computing with neural networks. *Science, 235,* 1227-1228.

Palm, G. (1991). Memory capacities of local rules for synaptic modification. *Concepts In NeuroScience, 2,* 97-128.

Palm, G., & Braitenberg, V. (1979). Tentative contributions of neuroanatomy to nerve net theories. In R. Trappl, G. J. Klir, & L. Ricciardi (Eds.), *Progress in cybernetics and system research* (pp. 369-374). New York: Wiley.

Wickelgren, W. A. (1979). Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychological Review, 86,* 44-60.

Wickelgren, W. A. (1992). Webs, cell assemblies, and chunking in neural nets. *Concepts in Neuroscience, 3,* 1-53.

Sommaire

Ceci est une introduction à Wickelgren (1992), décrivant une théorie de la représentation d'une idée et de l'apprentissage dans le cortex cérébral et la formulation des sept propriétés de Hebb (1949) sur les ensembles de cellules qui ont joué un rôle prépondérant de la représentation des idées, de l'apprentissage et de la pensée dans les modèles de réseaux neuronaux.

Les idées sont représentées dans le cortex cérébral par des **entrelacs** (ensembles des cellules innées), utilisant un codage peu dense, avec un enchaînement inné, rare si ce n'est pas inexistant. L'excitation d'un entrelacs pour représenter une nouvelle idée est désignée par le terme de **composition en modules compilés.** Les liens innés qui lient les neurones d'un entrelacs sont des synapses dendritiques basales. L'apprentissage modifie les synapses dendritiques apicales qui associent des neurones d'un entrelacs à des neurones d'un autre entrelacs.

Le **minint** (ou connectivité interne minimale) d'un groupe de neurones est le nombre minimal de liens innés que n'importe lequel des neurones du groupe reçoit de la part d'autres neurones du groupe. Le **maxext** (ou connectivité externe maximale) d'un groupe de neurones est le nombre maximal de liens innés que n'importe lequel des neurones se trouvant hors du groupe reçoit d'autres neurones se trouvant dans le groupe. Un **entrelacs** est un groupe de neurones dont le minint est plus grand que son maxext. Un entrelacs assume le rôle d'attracteur dans un système dynamique composé d'un réseau neuronal à lien binaire ayant un seuil de fonction d'activation et dont le seuil de consigne se trouve entre le maxext et le minint.

La pensée comprend quatre phases: deux phases principales - le **repérage** d'idées familières et la **composition en modules** d'idées nouvelles, et deux phases sous-jacentes à chacune soit: les processus de **sélection** et **d'accomplissement.** Dans la phase de sélection-repérage, les synapses apicales d'apprentissage sélectionnent un premier ensemble de neurones actifs dans un module. Dans la phase de repérage et d'accomplissement, l'activation converge sur un ensemble de terminaison asymptotique de neurones actifs qui, idéalement, constitue l'entrelacs commun ayant la plus grande similarité avec le premier ensemble. Si la phase de repérage-accomplissement échoue, la phase de sélection-composition en modules entre en jeu et toutes les synapses apicales (non pas seulement celles qui ont été apprises) sont utilisées pour sélectionner un ensemble premier pour la phase d'accomplissement et de composition en modules. Dans la phase de composition en modules-accomplissement, le module converge sur un ensemble de terminaison asymptotique de neurones qui, idéalement, constitue un nouvel entrelacs représentant l'ensemble de toutes les composantes des entrelacs qui ont produit l'entrée apicale au module au cours de la phase de sélection. Pour faire cela, dans la phase d'accomplissement et de composition en modules, le conditionnement de la contiguïté hebbienne, renforce toutes les entrées des synapses actives apicales vers les neurones dans l'entrelacs afin que des sous-ensembles assez grands de ces synapses apicales apprises puissent, à l'avenir, être capable de sélectionner un premier ensemble pour la recherche, ce qui réactivera le même entrelacs lors de la phase de repérage et d'accomplissement.