N.W. Milgram, C.M. MacLeod, + T.C. Petit (Eds.)

**Neuroplasticity, Learning, and Memory, pages 301–325**


# SITE FRAGILITY THEORY OF CHUNKING AND CONSOLIDATION IN A DISTRIBUTED ASSOCIATIVE MEMORY

Wayne A. Wickelgren

Department of Psychology
University of Oregon
Eugene, OR 97403

George Miller (1956) and those who further developed the idea of chunking as a learning process have produced a powerful new type of associative learning that goes substantially beyond the classical notions of associations of ideas, extant since Aristotle. In classical associative learning, two ideas activated contiguously in time had the connection between them strengthened, intuitively a horizontal association. In chunking, two ideas activated simultaneously in the mind recruit a new internal representative (node or nodes) to represent them and associations are strengthened from the constituent ideas to the new chunk idea and in the reverse direction.

Intuitively, chunking is a learned vertical association, with hierarchical structure similar to that found in the more genetically specified peripheral sensory and motor systems. Chunking is an important new type of learning for at least two reasons. First, chunking greatly reduces associative interference, by permitting associations to a chunk that are distinct from the associations to its constituents. Second, chunking permits high level representation of a complex idea that is as simple as the representation of the more elementary constituent ideas at their level.

In terrestrial biological minds, the mutations that produced learning by chunking appear to be those that produced the capacity for cognitive, as opposed to stimulus-response, thinking. Chunking permits the minds of birds and mammals to have mental maps or models of the world,

with mental entities representing objects and actions, not just stimuli and responses. Chunking permits us to have expectations of what our actions will accomplish, not just a strong urge to perform some response in a stimulus situation; the latter being, I believe, a fair description of the mind of a fish. I am basing these claims in part on the ideas and findings of Thorndike (1898), Tolman (1948), Bitterman (1969, 1975), and Razran (1971), but probably none of them would endorse all that I have just said concerning the difference between the minds of higher and lower vertebrates. One of my primary intellectual goals is to provide mathematical formulations of minds with and without chunking, to determine and compare the capacities of such minds more precisely.

Many of the ideas in this paper are incomplete and imprecise. Furthermore, my primary interest at present is theoretical cognitive science, not theoretical cognitive or physiological psychology. I hope some of these theoretical ideas will apply to real biological brains and the minds they make possible, and I will include a number of statements about the human mind and brain. However, my principal goal is to develop theories of possible minds, whether or not they correspond to any existing minds, though I will use what I have learned about real minds as the main stimulus for my thinking. One final warning: I will shift back and forth between statements about possible minds and statements about real minds and brains. This is ideal for theoretical exposition, so long as you remember that no careful attempt is being made in this paper to evaluate empirically any statements about real minds and brains.

## CHUNKS

George Miller (1956) invented the chunk, in the context of processing and short-term memory, as a unit of coding in the mind. Although Miller noted that a great deal of learning had gone into the formation of chunks, he did not attempt to explain the learning process that formed the internal representative of a chunk, which is my focus. Miller defined a processing strategy of "recoding", which is the use of an already learned chunk to represent a sequence (order set) of smaller chunks. Although we clearly have the ability to learn ordered sets, the manner

in which orderings are represented in an associative memory is beyond the scope of this paper. Briefly, I think that downward (implies) associations from chunks prime the unordered set of constituents, with the ordering of the constituents given by horizontal (lateral) associations among (context-sensitive) constituents (Wickelgren, 1969b, 1979b). In any case, here I am only concerned with the unordered set of constituents of a chunk.

## ASSOCIATIVE NETWORKS AND LINK TYPES

Throughout this paper I will work within a connectionist or network theoretical framework for describing a mind. Specifically, a mind is a digraph (directed graph), consisting of a set of nodes connected by directed links (that is, the link from A to B is distinct from the link in the reverse direction, from B to A). You should think of the mind discussed in this paper as an abstract model of the "association areas" of the human cerebral cortex with some of the nodes receiving specific sensory input (relayed through lower levels of the mind not modelled here) and all of the nodes receiving nonspecific arousal input from two arousal systems, the learning arousal system and the retrieval arousal system. Some of the nodes would output to lower levels of the mind as well, but I am not concerned with this, and you should assume that we can directly measure the activation output of each node in this mind.

Nodes in this mind are all of one type (excluding the arousal systems, which are considered external to this mind). In particular, there are no inhibitory nodes analogous to inhibitory neurons. Inhibitory functions will be modelled by inhibitory links between nodes. However, there are several types of links. Links are classified on four dimensions: conditionable or not, excitatory vs. inhibitory, specific vs. nonspecific, and implies vs. coimplies. Not all of the 16 possible link types exist, and I will concentrate on just two types in this paper: (a) conditionable excitatory specific implies links and (b) conditionable excitatory specific coimplies links. Hereafter, I will just refer to these as implies or coimplies links, with the other adjectives understood.

I will also make some use of two other link types,

unconditionable specific inhibitory links (referred to as inhibitory) and unconditionable nonspecific excitatory links (referred to as nonspecific), with the implies vs. coimplies dimension being irrelevant for these link types. The nonspecific links connect each node in the mind to the learning arousal system, not to each other. At a few points in the paper I will refer to another type of nonspecific link that might connect each node to the retrieval arousal system, but the properties of this system and its links to the mind are not discussed very much.

Previously, I assumed not two but three types of conditionable specific excitatory links -- up, down, and lateral (Wickelgren, 1979a). The correspondence is roughly as follows: Coimplies links will do the same job as the old up links, activating a node when the sum of the inputs from the link set that jointly coimplies the node exceeds an activation threshold. Coimplies links are for two-to-one or many-to-one associations. Implies links are for one-to-one associations and do the job of the old down links. At present, I assume that all excitatory lateral links are of the implies type, but some may be of the coimplies type.

Do not be misled by the word "implies" into assuming that if node A has an implies link to node B and A is strongly activated at time i that node B will necessarily be strongly activated at time i + 1. B will be activated above the threshold for possible inclusion in the next thought (consisting of all strongly activated nodes). However, a decision process (mediated by lateral inhibition) limits the next thought to the most strongly activated nodes within some limited attention span, and B may not make it. I am not prepared to provide a mathematical formulation of this decision process beyond this intuitive property, and indeed I will largely ignore all inhibitory links in this paper.

## RANDOM CONNECTIONS AND BINARY CHUNKING

In discussing mechanisms of chunking it is helpful to deal with the concrete case of chunking two nodes A and B into a site on some chunk node, which I will here call node (AB). Such binary chunking is the simplest case, but repeated binary chunking is sufficient to chunk sets of larger size, albeit with some form of binary syntactic

structure, such as ((AB)C), ((AB)(CD)), etc. There is also a strong probabilistic argument in favor of binary branching in (genetically) randomly connected associative networks where the total number of nodes is approximately the square of the average number of links per node. Throughout this paper, I will assume a concrete model with $10^8$ nodes and $10^4$ links per node. These exact numbers are not important, but it is important that the number of nodes in the mind is roughly the square of the number of links per node. Incidentally, the total number of neurons in the human cerebral cortex is roughly the square of the number of synapses per cortical neuron (Cragg, 1975; Pakkenberg, 1966).

Of course, we do not know that the chunking associative memory in humans is randomly connected. Part or all of it may be partitioned on the basis of the particular types of input or output connections to the more genetically specified sensory and motor modules of the mind. The transition from genetic to learned structuring of the connections of the mind may involve several steps or levels, even within chunking associative memory. Genetic guidance of chunking could easily result in the single-step chunking of sets of constituents larger than two. However, consideration of the case of binary chunking in genetically random associative networks will give us enough to chew on for the moment.

## LINKS, SITES, NODES: ACTIVATION, STRENGTH, FRAGILITY

A node has a set of input sites, with each site containing a set of input links to that node. Sites will be classified into three types: (a) a single implies link, (b) two coimplies links, or (c) one specific (implies or coimplies) link and one nonspecific link to the learning arousal system. For reasons that will be discussed later, these three types are called bound implies sites, bound coimplies sites, and free sites, respectively. Links, sites, and nodes all have a positive real-valued activation property. Links also have a positive real-valued strength property, and sites have a positive real-valued fragility property, with site fragility being some monotonic increasing function of the strength of the nonspecific link to that site from the learning arousal system. Sites with no nonspecific link are really sites with a very weak

nonspecific link (close to zero strength) and therefore close to zero fragility. For reasons that will be discussed later, fragility will serve as the theoretical measure of the degree of consolidation of a memory trace, with low-fragility representing high consolidation.

If node i has activation $x_i$ and the link from node i to node j has strength $z_{ij}$, then link ij has activation $x_i z_{ij}$. Greater link activation produces greater site activation for the site at which the link terminates, and greater site activation produces greater node activation. I do not wish to commit myself to any particular functions summing link activation to get site activation and summing site activation to get node activation. Obviously, there are advantages to assuming as much linear combination as possible with at least one nonlinear threshold parameter at the site, node, or both.

However, in this paper, I wish to consider the possibility that the summation of two coimplies link activations in a single site produces greater node activation than if the same coimplies link activations occurred in different sites. If this nonlinear property holds and if chunking could somehow bring two coimplies links to the same site, then if links a & b were in one site and links c & d were in another site on the same node, the node would be more strongly activated by (a&b) or (c&d) than by (a&c), (a&d), (b&c) or (b&d). I call this set of assumptions the site grouping hypothesis. Site grouping permits a single node to function more like a logical conjunction unit than it could with only link strengthening as a learning mechanism. I am not convinced this is either desirable or true of the cerebral cortex, but it is worth considering.

## CONTIGUITY CONDITIONING BY CROSS-CORRELATION

I make the standard assumption of contiguity conditioning of nodes (Hebb, 1949; Grossberg, 1967) that the strength of the link from node i to node j increases when the nodes are strongly activated at about the same time, more specifically when the activation of nodes i and j has a positive cross-correlation, typically with a temporal asymmetry ($\theta$) to reflect link delay times in transmission of activation and perhaps other factors.

Grossberg (1967) expresses this very elegantly by the equation:

$$\dot{z}_{ij} = -uz_{ij} + \beta x_i\,(t - \theta)x_j(t),$$

where $z_{ij}$ is the strength of the link from i to j, $\dot{z}_{ij}$ is its time derivative (rate of change of strength, $-uz_{ij}$ represents forgetting via an exponential decay of link strength (with which I disagree for long-term memory), and $\beta x_i(t - \theta)x_j(t)$ represents learning due to cross-correlation of the activation of node i at time $t - \theta$, $x_i(t - \theta)$, and the activation of node j at time t, $x_j(t)$.

Elegant theoretical work, such as that of Grossberg, demonstrates that there is much to be learned by careful study of contiguity conditioning in the context of varying assumptions about other aspects of network minds. Although it has not been formulated mathematically, my previous theory of chunking (Wickelgren, 1979a) describes a network mind in which chunking, as well as conventional association of ideas, can occur via the contiguity conditioning learning mechanism. In the present theory, both implies and coimplies links are assumed to be strengthened by some cross-correlation type of contiguity conditioning. However, as mentioned previously, changes in link strength via contiguity conditioning may not be the only mechanism mediating chunking. Chunking may also group two or more coimplies links into a common site on the target node.

## FREE AND BOUND SITES

Recall that I classified the sites of the mind into three types: bound implies sites, bound coimplies sites, and free sites. For the moment collapse the first two types into one type. Thus, there is a partition of all of the sites of the mind into two subsets: free and bound. A free site has one specific link with low strength and one nonspecific link with high strength. Activation of a free site requires input activation of both the high strength nonspecific link and the low strength specific link. If a free site is consistently activated at about the same time as its node is activated, activation of the free site becomes a useful predictor of activation of the entire node, and the strength of the specific link to that free site is increased via the cross-correlation learning mechanism. Although the nonspecific link to that site contributed substantially to site activation, its

activation is random with respect to events to be represented by the mind and nonspecific links are assumed not to be strengthened by contiguity conditioning. Indeed, when the specific link to a site is strengthened by learning, this weakens the nonspecific link. Thus, learning is assumed to strengthen the specific link at a site and weaken the nonspecific link at the same site, converting it from a free site to a bound site. Basically, the notion is that, at birth, each node (whose links are not entirely specified genetically) has a bunch of weak specific links to (free) sites on other nodes. Some of these weak specific links will prove to be predictive of activation of the nodes they connect to, thus becoming strong specific links. The sites with strong specific input links are said to be bound to those links.

The preceding paragraph only describes the process of converting a free site to a bound site with a single specific link, that is a 1-1 association. How do we get the 2-1 association necessary for chunking? We get them from having two sites activated in temporal contiguity with activation of the target node. By the process described in the preceding paragraph, this converts both free sites to bound sites. Then the site grouping mechanism takes over and collapses the two sites into a single bound site, or perhaps the specific links of each newly bound site send collateral links to the other newly bound site. In the two cases, one gets either one or two bound coimplies sites. Note that it is not necessary to assume that the specific links of free sites are of two types, implies and coimplies. There need be only one type of specific link to free sites. A bound implies site results from a learning event that was 1-1. A bound coimplies site results from a learning event that was 2-1 and the site grouping process that follows such a learning event.

## CHUNKING AND THE REVERSE LINK HYPOTHESIS

There is undoubtedly a considerable degree of genetic constraint on the randomness of neural connections in animals, even in the cerebral cortex, and strong cases can doubtless be made for many different types of nonrandomness in the connections of minds, from a cognitive science standpoint. However, in this paper, I will assume a mind in which each node has specific links to a random sample of

other nodes, with one exception. The exception is the reverse link hypothesis, that whenever node i connects to node j, node j connects to node i. The link from i to j may have different strength than the link from j to i, but there is always a structural link from j to i, whenever there is a link from i to j.

If nodes i and j connect to node k, but nodes i and j do not connect to each other (the latter being the typical case for the sort of mind envisioned in this paper), then chunking i and j by binding them to node k will also strengthen 1-1 (implies) associations from node k to node i and from node k to node j. Thus, chunking not only strengthens two coimplies links from the constituent nodes to the chunk node, it also strengthens two implies links from the chunk node to the constituent nodes. Logically, nodes i and j together coimply node k, while node k implies both nodes i and j. The sense of this is that node k represents the conjunction of nodes i and j, so the conjunction implies its constituents.

## NONSPECIFIC LINKS AND THE LEARNING AROUSAL SYSTEM

There are some relatively obvious questions concerning the mechanisms by which chunking could be accomplished in a network mind such as the nervous system. The first question is how do we know there is any node that receives specific links from both A and B nodes in a genetically random network? In my first theory of chunking (Wickelgren, 1969a), I assumed that some electrochemical gradient created by the simultaneous activation of nodes A and B caused them to grow links toward each other until they met, whereupon they would link to the nearest node. Such a long distance growth process would be difficult to engineer and is generally deemed unlikely to occur in the adult nervous system as a mechanism of learning.

A selectional theory of learning is far more plausible for the human mind and more practical to engineer in an artificial mind. One could also develop a model of chunking in which one or more interneurons were enslaved by the chunking process purely for the purpose of getting some node that indirectly (via a chain of interneurons) received input links from both A and B nodes, but I have little interest in doing this. Furthermore, as I have argued

before (Wickelgren, 1979a), the ratio of synapses to neurons in the human cerebral cortex is such that, while it is very unlikely that any set of three or more neurons synapse with a common (possible chunk) neuron, it is highly likely that any set of two neurons do synapse on some common neuron. In an arbitrary random network there is no guarantee of this, but, if the ratio of links to nodes is great enough, the probability can be made as close to one as you wish.

For two contiguously activated nodes to be chunked, they must have their links to the chunk node strengthened. The only link strengthening process we have assumed is contiguity conditioning. This requires the chunk node to be strongly activated at about the same time as the constituent nodes are strongly activated. Since, prior to chunking, the constituent nodes have only weak links to the chunk node, how does the chunk node get activated? I still like the basic mechanism described in Wickelgren (1979a) in which a (spontaneously active) learning arousal system provides strong nonspecific input to combine with the converging weak specific input to the (AB) chunk node from the constituent A and B nodes. By providing each free site with a strong nonspecific input link to compensate for the weak specific link, one could probably design a network (and the mammalian cerebral cortex may be one) in which input from two weak specific links is enough to activate the node strongly enough to trigger contiguity conditioning. A precise mathematical model is really important here, but the foregoing argument is intuitively persuasive.

Furthermore, we can postulate alternation of activation of learning vs. retrieval arousal systems, so that during the learning phase of mental functioning (occurring several times a second like the alpha rhythm) only free sites can be activated (Wickelgren, 1979a). Routtenberg (1968) presented considerable evidence to support the existence of two such arousal systems in the human brain, a limbic (hippocampal) arousal system and the more familiar reticular activating system. The former could serve as the learning arousal system and the latter the retrieval arousal system. I no longer think it is necessary to alternate learning and retrieval phases to permit free sites to activate their nodes in competition with existing strong links from constituent nodes, but it might be.

Nonspecific input would also assist the weak implies

links from the (AB) chunk node in activating the relevant sites on the constituent A and B nodes. A positive feedback loop is thus created between the chunk node and its constituents, which produces a relatively long period of paired activation of the chunk node and its constituents. This strengthens both upward coimplies and downward implies connections to the relevant sites by the plausible contiguity conditioning mechanism.

## FRAGILITY, CONSOLIDATION, UNLEARNING, DECAY AND AMNESIA

Although the learning event of contiguous activation can be accomplished in a second or less and I assume that the consequent increase in the strength of the specific link(s) occurs almost immediately thereafter, a long-lasting period of consolidation of this learning follows the learning event. The consolidation consists of the reduction in the strength of the nonspecific link(s) to the learning arousal system at the newly bound site(s). The nonspecific link has served its purpose of permitting a new site to be bound via the contiguity conditioning mechanism. Now that the site has been bound, a strong nonspecific link would only cause more rapid forgetting of the newly strengthened specific link. It is for this reason that the strength of the nonspecific link is also called the fragility of the site or, equivalently, the fragility of the newly strengthened specific link.

Once a free site has been bound, its association to the learning arousal system (fragility) begins to decrease, rapidly at first, then progressively more slowly over time. As fragility decreases, the probability of activation of the site by the learning arousal system and other random weak input decreases. Thus, there is less chance that the bound site will be activated without input from the specific link(s) it was bound to. Such uncorrelated activation is assumed to weaken the previously strengthened specific associations. The reduction in site fragility is, thus, a consolidation process which protects the memory traces (strengthened associations) from disruption by one kind of forgetting. This forgetting results from activation of the site without activation of the proper specific input links. This is a kind of (backward) unlearning, but, strangely enough, it behaves like a pure time decay process, because the events that drive the loss of trace strength are

unrelated to the events that produced the original learning.

Furthermore, as the trace consolidates, this decay slows down over time since learning, a prediction that has been overwhelmingly confirmed (Wickelgren, 1972, 1974). Finally, although several facts concerning human memory indicate that consolidation continues for years following learning, most of the consolidation occurs within the first few hours or days following learning.

Retrograde amnesia is loss of memory for events that occurred before some insult to the brain such as concussion, electroconvulsive shock, lesions of the hippocampus, etc. The same consolidation process can be used to explain the reduced susceptibility of older memories to retrograde amnesia. The theory also accounts for why subjects with retrograde amnesia show anterograde amnesia, a reduction in ability to learn new associations, the so-called amnesic syndrome (for an explanation, see Wickelgren, 1979a). Since the amnesic syndrome seems to apply precisely to learning that might be presumed to employ new chunking (Wickelgren, 1979a), the explanation of normal chunking, normal forgetting, and both retrograde and anterograde amnesia via a common mechanism is appealing. The evidence indicates that the long-term memories that are disrupted in the amnesic syndrome are located in the cerebral cortex, but that a neural circuit involving the cortex and the hippocampus (and perhaps other structures) is critically involved in the learning and consolidation processes.

## CONSOLIDATION, SITE RECYCLING, AND DENDRITIC SPINES

Apparently, virtually all excitatory synapses on mammalian cortical neurons are on dendritic spines. Thus, to apply the current theory to the mammalian cerebral cortex, let us assume that a site is a single spine or a set of nearby spines receiving one or more (specific) synapses from another cortical neuron and one or more (nonspecific) synapses from the learning arousal system. Although the nonspecific synapses start out stronger than the specific synapses at free sites, there is a sense in which the specific synapses are the genetically preferred synapses, because once the specific synapses are strengthened, this causes the nonspecific synapses to

weaken at the same site. This is not all implausible, and examples of just such a process were cited in Wickelgren (1979a). It is also possible that if the nonspecific synapses are on different spines from the specific synapses, then what consolidation does is somehow to protect the newly bound specific spine from the effects caused by input to nearby nonspecific spines. There are many ways this could be done.

If a specific link becomes sufficiently weakened by forgetting, the consolidation process might reverse itself, recycling the site to the free state once more. There is also the more pessimistic version of this theory in which no site recycling is possible, and we gradually use up all of our free sites as we learn.

## LEARNING AND UNLEARNING VIA CROSS-CORRELATION

Both learning and unlearning can be obtained from the cross-correlation term of Grossberg's equation for contiguity conditioning provided we change the activation terms, $x_i$ and $x_j$, from absolute levels of activation to deviations in activation from some intermediate point. This permits negative contributions to link strength (unlearning) from the cross-correlation term whenever $x_i$ is high and $x_j$ is low (forward unlearning) and whenever $x_i$ is low and $x_j$ is high (backward unlearning). The effects of consolidation would then have to be reflected in a reduction of the $\beta$ cross-correlation parameter. This will reduce backward unlearning, but it will also reduce both forward unlearning and further learning at the same synapse.

The reduction in further strengthening of the same synapse is not in obvious conflict with the facts since learning is definitely subject to diminishing returns, and many theorists suspect that multiple-trial learning involves trace replication at different synapses more than trace strengthening at the same synapse. However, the reduction in forward unlearning does not appear to be in accord with the facts of human learning (Wickelgren, 1974). Furthermore, it is only the kind of neurally backward unlearning produced by nonspecific activation of the postsynaptic neuron that can be assumed to diminish with consolidation, since it is only that kind of unlearning that, behaviorally, appears to be a pure time decay process

(and not an unlearning process). In human forgetting, consolidation reduces the time decay factor and apparently not the unlearning factor, whether forward or backward, though the invariance of unlearning with time since original learning is a result in need of much further replication before we can be sure of it (Wickelgren, 1974). It is probably better to account for the reduction in forgetting due to time since learning (consolidation) by altering Grossberg's exponential decay term to one more in accord with the facts of long-term forgetting (Wickelgren, 1974). This is also more in accord with everyone's intuition (including Grossberg's) concerning the separation of learning and forgetting processes.

Nevertheless, it is interesting and worth remembering that modification of link strengths by cross-correlation of activation can be used to produce both forward and backward unlearning as well as learning. While I would be the last to downgrade intuitive verbal theory formulation, I am also a great admirer of mathematical theory formulation, in part because it can serve as a basis for, previously unsuspected, grand unifications of apparently disparate phenomena, such as the possible unification of learning and unlearning via the cross-correlation mechanism. Even though this particular unification may well be wrong for the mammalian brain, it is a fascinating possibility that would never have occurred to me without Grossberg's mathematical formulation of associative learning.

## SITE GROUPING AND LOCAL GROWTH

This section describes a speculative neural mechanism of site grouping following learning. You may have heard the old saying about how a little knowledge is a dangerous thing. You need to be warned that I have a little knowledge of the nervous system. Also, I do not think that old saying applies to me or to anyone else who is careful to encode the degree of support for an idea and something about the nature of that support. Of course if you are one of those people to whom the old saying does apply, please skip to the next section. For those of you coming along for the ride, it is time to fasten your seatbelt.

The first question that concerns me is whether there

is any plausible neural mechanism by which constituent links that coimply a chunk node could be grouped into a common site, nearby sites, or sites with some other kind of synergism that provided a superadditive combination of their input link strengths. The purpose of such site grouping is to make a chunk node represent something closer to a conjunction of its constituents, instead of an additive combination. Of course, it may be that chunks do respond additively to constituent input in the brain. Consideration of whether there is a plausible neural mechanism for conjunctive grouping is relevant to this issue and interesting in its own right.

One possibility is that the dendritic tree might grow and contract so as to keep all free sites on a connected subtree containing no bound sites. The bound subtree would also be connected and contain no free sites. The bound subtree might be proximal to the cell body and the free subtree distal, or the main dendritic trunk might divide near the cell body into a bound subtree and a free subtree. Either way, when free sites become bound, the entire portion of the free subtree between these newly bound sites and boundary with the bound subtree contracts so as to transfer the newly bound sites to the bound subtree. If two or more sites were bound at about the same time, both would be transferred to about the same place in the bound subtree and thus might have a conjunctive-like, superadditive combination in retrieval. Once the newly bound sites were pushed onto the bound subtree, the free subtree would grow back to approximately its previous size, perhaps observing some sort of constancy in the total number of free and bound sites or just the total number of free sites. When bound links decrease in strength to some low level, the terminals might either remain on the bound subtree or grow a very short distance to reconnect to a site on the free subtree, recycling the sites.

If all of this degeneration and regrowth of the dendritic tree seems implausible, consider the possibility of local presynaptic terminal growth. When two sites are bound at about the same time, they may set up some local electrochemical gradient in the intracellular space or within the portion of the dendritic tree that connects them. This gradient might direct the growth of an axonal branch from each terminal to the vicinity of the other terminal, where it might synapse with the postsynaptic

neuron in a nearby site, on the same spine perhaps or on adjacent spines.

Doubtless there are other possibilities for learned synaptic grouping that are as plausible as these or more plausible. Some sort of axonal or dendritic growth is probably required to achieve conjunctive-like superadditive site grouping, unless the entire dendritic tree of a chunk neuron is devoted to computing a single conjunction of two, three, or more synaptic inputs, with the rest of the 40,000 synapses per cortical neuron (Cragg, 1975) being wasted. However, the neural growth is of an extremely local kind that seems plausible. What is probably most exciting about such theories of site grouping is that they permit a neuron the potential to bind all of its sites in one-to-one or two-to-one combinations as desired. Furthermore, the number of remaining possible groupings of two or more free sites on a neuron degrades gracefully and minimally with increased binding of sites according to either theory. Recycling of decayed or unlearned previously bound sites and their specific links seems possible with either theory. All of this is of some importance in the distributed associative memory to be discussed next.

## DISTRIBUTED ASSOCIATIVE MEMORY

Previously, I have defined chunking in the context of the specific-node ("grandmother cell") theory of coding in associative memory, once defining a chunk idea to be a single node that represented a disjunction of conjunctions of constituent nodes (Wickelgren, 1969a) and once emphasizing only the "conjunctive" aspect by defining a chunk idea to be a single node representing an unordered set of constituents (Wickelgren, 1979a). The question of this section is, "What is the representative of an idea in the mind, whether a constituent idea or a chunk idea?" Network minds offer some interesting alternative answers to this question. I will briefly describe six different classes of idea coding systems. These six systems do not exhaust the possibilities for idea representation in network minds, nor are they even mutually exclusive. The human brain makes some use of at least two of these six.

The six classes of coding systems include two nonassociative systems: (a) <u>coding by temporal pattern</u> of

activation in any single node or small set of nodes and (b) coding by spatial pattern of activation in a small set of nodes (like the pattern of 0s and 1s in a von Neuman computer's memory registers). The human brain is known to make some use of temporal pattern coding in the more peripheral parts of the auditory nervous system, namely, periodicity information in pitch perception and phase information in localization. However, the brain converts both of these temporal pattern codes into some kind of "which set of neurons fire" code at higher levels, since the temporal spiking pattern of neurons at higher levels has no correlation with the auditory input in periodicity and phase. Spatial and temporal pattern coding may be used to some extent in motor control systems if there is any truth to the coupled oscillator theory (see Gallistel, 1980, for an insightful review).

However, there is every reason to believe that higher sensory, motor, and cognitive coding in the human mind uses some version of one of the four following classes of associative "which neuron fires" codes: (c) specific node coding (the grandmother cell theory), in which activation of a particular node represents thinking of an idea (your grandmother, for example), (d) overlapping set coding, in which thinking of an idea is represented by activation of a set of nodes that will generally overlap with the representation of different ideas (e) node activation function coding, in which the representation of an idea is a particular activation function defined over all of the nodes in the mind, and (f) link activation function coding, in which the representation of an idea is an activation function defined over the links. Note that (e) can be considered to be a generalization of (c) and (d), and (f) can be considered to be a generalization of (e).

The extreme version of specific node coding in which there is only one node for every idea is not very fault tolerant, so one probably wants to have several similarly linked nodes to represent each idea. As long as the set of nodes representing one idea do not overlap with the set of nodes representing another idea, I will classify this as a version of specific node coding because the properties appear to be quite similar.

Overlapping set coding is a discrete (all-or-none) version of node activation function coding, and both can be

used to represent distributed associative memories. In overlapping set coding, an idea is represented by a set of nodes that generally overlaps the node sets representing other ideas. The degree of distribution in the representation of an idea can vary enormously within each of these two classes. In the overlapping set coding system, the maximum size set for representing an idea might vary from two nodes to all of the nodes in the network. If one defines a special "don't care" value for activation or considers levels of activation that are close to zero to be "don't care" values, then similar wide variability in the degree of distribution of coding is possible in the node activation function coding system. Distributed memory versions of node and link activation function coding pose fascinating conceptual problems about which I need to think more. Overlapping set coding seemed more tractable for my first step into the distributed associative memory area, and so I will describe a theory of chunking in terms of coding by overlapping sets of nodes.

I think it is of some interest that it was possible to present most of the ideas about chunking that are in this paper without explicit adoption of either this distributed memory model or the specific node model. Finally, I should note that I am not at all convinced that human associative memory uses distributed as opposed to specific node coding. Randomly connected associative memories probably function better with distributed node coding, but when there is some "genetic" guidance to restrict the possible connections intelligently, specific node coding may be functionally superior. I chose to investigate chunking in a distributed memory context mainly because I had not done so before and wanted to become more familiar with the properties of distributed associative memories.

## DISTRIBUTED ASSOCIATIVE MEMORY VIA OVERLAPPING SET CODING

An idea is represented by a set of nodes, which I currently imagine to vary in size from a few hundred for a newly formed idea to around 10,000 for a highly familiar idea. Different ideas are represented by different, but overlapping, sets of nodes. Two constituent ideas specify a new chunk idea when specific links from the two sets of nodes representing the constituent ideas help

activate some free sites in a set of other nodes that will become the set representing the chunk idea. Those links ending on the chunk nodes are strengthened by contiguity conditioning, binding the sites on the chunk node that they end on. Because nodes are always connected in both directions, the reverse, chunk to constituent, links also get strengthened and bind their sites on the constituent nodes. So an idea is represented by a set of nodes and the various sites on a node represent many different ideas that may have no conceptual relation to each other. A node represents a random collection of ideas.

## IDEA INTEGRATION IN LEARNING AND IDEA COMPLETION IN RETRIEVAL

Although a chunk node must receive inputs from two weak specific links to be activated and bind its respective input sites, nothing guarantees that these two links will be one from one constituent idea set and one from the other. Both may be from different nodes in the same constituent idea set. Initially, I thought it was a problem that the chunking mechanism of the theory would chunk pairs of nodes that were members of the same idea set (intrachunking) as well as chunking pairs of nodes that were members of different idea sets (crosschunking). Then I realized that intrachunking might serve a very useful function, similar to Hebb's (1949) cell assemblies, that of integrating the nodes of an idea set.

Some definitions are useful for a precise explanation of the role of idea integration in thinking:
(D1) A thought is a set of idea sets that are simultaneously activated.
(D2) A initial subset is the subset of an idea set activated by the last thought.
(D3) The completion subset is the subset of an idea set generated asymptotically from the initial set by intraidea links adding nodes to the activated subset of the idea set until no further nodes can be added via intraidea links. The completion of some initial subsets might be the entire idea set, but the completion of other initial subsets might be less than the entire idea set. The completion function maps initial subsets to completion subsets.

Of course, you should note that this is not a

completely precise model, as we need to deal with the intra- vs. inter-idea retrieval problem, that each node will have many strong interidea links besides those intraidea links involved in idea integration. Indeed, these interidea links are essential for activating the next thought, as in Hebb's phase sequence. My current semiprecise working hypothesis is that thinking consists of a cycle of phases that repeats over and over. If we start the cycle with the last thought's having activated an initial subset of some current idea sets, then the cycle has three phases as follows: initiation (inhibition of the last thought and activation of the initial idea sets for the current thought), completion (activation of the completion of some number of these initial idea sets up to a limit set by the attention span), and chunking (activation of new nodes, and link strengthening to further integrate existing chunks and to form new chunk ideas). Initiation and completion are two phases of the retrieval process, while chunking is a learning process.

## IDEA DISCRIMINATION

Consider the following idea discrimination problem: How big an initial subset of nodes in one idea set must be activated (by the prior thought) to uniquely specify that particular idea set? This form of the question demands some further clarification. First of all, I am not concerned with all of the complexities of the actual idea retrieval process in this problem. For example, I am not concerned with whether the completion of an initial set is the entire idea set. Indeed, I am not concerned with any aspect of the actual activation of nodes. I am asking only about how big a subset of some particular idea set is necessary in a logical sense to distinguish this idea set from any other idea set encoded in the mind. That is to say, you are to assume that there exists some number (N) of idea sets in a particular mind, with all N idea sets known completely by an omniscient observer. You give the observer a subset of x nodes all of which are guaranteed to be from one idea set, and the question is, "With what probability (P) will the omniscient observer have enough information to determine uniquely which set that is?"

P should be very close to unity for good idea discrimination. That is, we want our initial set to

specify a unique idea with high probability, at least logically, since otherwise there is no possible retrieval mechanism that we could adjoin to this theory of idea coding to make it work properly. Just because it seems plenty high enough and works out conveniently, let's take P=.9999 and see how big the size of the initial set X needs to be to achieve this P for $n=10^8$ nodes in the mind, $d=10^4$ nodes/idea, and $N=10^8$ idea sets in the mind. The answer is that x=3 gives P=.9999 that no other idea in this mind also contains all 3 members of the initial set X! For $N=10^{12}$ ideas, x must be 4! For $N=10^{16}$, you need x=5. These are very small numbers.

Since specific node (grandmother cell) coding can only get a maximum of $10^8$ ideas coded by $10^8$ nodes (though x=1 for P=1), we are clearly able to realize an enormous increase in idea coding capacity with overlapping set encoding at a very modest cost in the logical discriminability of ideas. Note also that the idea discrimination capacity of overlapping set coding is so great that there is no incentive on these grounds to use more than two values (on or off) of node activation in idea coding. Multiple or continuous values of node activation may play some useful role in learning and retrieval dynamics, but they are certainly unnecessary for the coding of ideas in network minds.

## REDUNDANCY, DIMINISHING RETURNS, AND SPACING EFFECTS IN CHUNKING

With overlapping set coding, when constituents A and B are chunked, sites on more than one chunk node are assumed to have their input associations for A and B strengthened, so that the AB idea is represented by a set of nodes. Subsequent experience with the AB pair is presumed to result in associating A and B to more chunk sites, enriching the redundancy of representation of the AB idea. However, while it doubtless makes sense to have more frequently used ideas represented by larger sets, it is probably not useful to add chunk sites in direct proportion to the learning time or the number of learning trials. Furthermore, we know that, by virtually any commonly used measure of memory strength, the rate of learning eventually slows down as a function of trials or time--the law of diminishing returns in overlearning.

Indeed, although there is a variable period of time after exposure to material before a person settles on an encoding and gets that first huge learning increment, which I presume to reflect the initial chunking, after that initial chunking, the rate of further chunking appear to decrease monotonically as strength increases. Of course, after some time elapses, there is a reduction in the strength of chunking due to forgetting, which permits a greater amount of chunking to occur after a longer spacing interval between learning trials (see Wickelgren, 1981, p.39-40 for a brief review). This greater amount of learning (here presumed to be chunking) after greater spacing between learning trials usually more than compensates for the greater amount of forgetting that also occurs, producing the familiar benefits of spaced over massed practice.

By what mechanism might the rate of chunking be reduced with increasing total strength of association from A and B to AB chunk sites? Assume that the familiar lateral inhibition mechanism constrains the total sum of activation of all nodes, as in Milner (1957)--perhaps some kind of conservation of activation law or in any case an upper bound on total activation. Chunking might occur when the active A and B nodes are less strongly associated to each other via chunk nodes and thus not as strongly activated as they would be if they were more strongly chunked. When the A and B nodes are less strongly activated and there are fewer AB chunk nodes activated, there is less total activation, less lateral inhibition, and thus more chance for new AB chunk sites and nodes to become activated. As the number of AB chunk nodes increases, this probability of activating and thereby specifying, new AB chunk sites goes down, producing the diminishing returns in chunking. Since recently chunked AB sites are strongly linked to both A and B nodes and to the learning arousal system, such recently chunked AB nodes might be hyperactive, further reducing the probability of new chunking. This provides an even greater benefit to the spacing of learning trials in that spacing allows both consolidation and forgetting to occur. Please note that this is a quantitative argument that requires more than verbal logic for adequate demonstration, and the present argument is hardly more than superficial handwaving. This is only the germ of an idea. Some of us are easily infected by idea germs.

## HIERARCHICAL LEARNING SUCH AS MATHEMATICS

In shallow learning, we chunk unrelated AB, CD, EF, etc. pairs. In deep learning, such as mathematics, we may chunk AB, then (AB)C, then ((AB)C)D, and so on sometimes to great depth. Though the AB chunk may have a strong trace immediately after initial learning, the hyperactivity of the recently chunked AB nodes may limit additional chunking, such as (AB)C, that uses AB as a constituent. The explanation is the same as for the spacing effect in further learning of AB. Although substantial forgetting of an AB chunk occurs from one session to the next, requiring time in review, such spacing may be optimal for new chunking that builds on the AB chunk. In addition, such review benefits the learning and retention of AB.

Of course, it is dangerous to prescribe educational practice from even a well-verified theory. Subject to that warning, the present theory suggests that there might be advantages in mathematics teaching to increasing the number of different topics covered in one session and restricting the depth of learning about each topic, reviewing each topic and adding another layer of learning in the next session. Since review takes time, this undoubtedly means that a smaller number of concepts and propositions could be covered in a term, but those that were presented might be vastly better learned and remembered in subsequent terms. Of course, this one-layer-at-a-time approach is more disorganized within a single session than presenting layer after layer of definitions, lemmas, and propositions, culminating in some beautiful theorem, in one continuous stretch of time. But the beautiful view from the top of the mountain is missed if one gets lost on the way up, and that happens all too frequently. Following the one-layer-at-a-time approach, it might take a few more days to get to the first beautiful view, but, except for the time lost to review, the number of beautiful views need not be drastically reduced, and the number of students getting the views might be increased.

The prediction that spacing between AB learning and (AB)C learning will be beneficial is novel, and, to my knowledge, there is no relevant experimental evidence. However, it agrees with my intuition that you cannot cram mathematics learning into as short a time as you can

shallower subjects. Spaced study may be even more important to hierarchically deep learning than to shallow learning, though students of any subject should be told to to study at least a little every day, because crammed knowledge is poorly learned and quickly forgotten. In learning, it is wise to revere the turtle.

REFERENCES

Bitterman ME (1969). Thorndike and the problem of animal intelligence. Amer Psychol 24:444-453.

Bitterman ME (1975). The comparative analysis of learning. Science 188:699-709.

Cragg BG (1975). The density of synapses and neurons in normal, mentally defective, and aging human brains. Brain 98:81-90.

Gallistel CR (1980). "The Organization of Action: A New Synthesis." Hillsdale, NJ: Erlbaum, p 432.

Grossberg S (1967). Nonlinear difference-differential equations in prediction and learning theory. Proc Natl Acad Sci USA 58:1329-1334.

Hebb DO (1949). "The Organization of Behavior." New York: Wiley, p 335.

Miller GA (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63:81-97.

Milner PM (1957). The cell assembly: Mark II. Psychol Rev 64:242-252.

Pakkenberg H (1966). The number of nerve cells in the cerebral cortex of man. J Comp Neurol 128:17-20.

Razran G (1971). "Mind in Evolution." Boston: Houghton Mifflin, p 430.

Routtenberg A (1968). The two-arousal hypothesis: Reticular formation and limbic system. Psychol Rev 75: 51-80.

Thorndike EL (1898). Animal intelligence: An experimental study of the associative process in animals. Psychol Rev Monograph Supplements 2 (No. 8).

Tolman EC (1948). Cognitive maps in rats and men. Psychol Rev 55:189-208.

Wickelgren WA (1969a). Learned specification of concept neurons. Bull Math Biophysics 31:123-142.

Wickelgren WA (1969b). Context-sensitive coding, associative memory, and serial order in (speech) behavior. Psychol Rev 76:1-15.

Wickelgren WA (1972). Trace resistance and the decay of long-term memory. J Math Psychol 9:418-455.
Wickelgren WA (1974). Single-trace fragility theory of memory dynamics. Memory & Cognition 2:775-780.
Wickelgren WA (1976). Network strength theory of storage and retrieval dynamics. Psychol Rev 83:466-478.
Wickelgren WA (1979a). 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. Psych Rev 86:44-60.
Wickelgren WA (1979b). "Cognitive Psychology." Englewood Cliffs, NJ: Prentice-Hall, p 436.
Wickelgren WA (1981). Human learning and memory. Annual Rev Psychol 32:21-52.