

Binding in Working Memory and Long-term Memory: Towards an Integrated Model

Jaap M.J. Murre¹, Gezinus Wolters², & Antonino Raffone^{3,4}

*¹University of Amsterdam, ²Leiden University, ³University of Sunderland, ⁴PDL RIKEN
Brain Science Institute*

Abstract

Models of long-term and working memory assume various forms of binding processes. Long-term memory consolidation involves a process whereby memory representations are first bound by the hippocampus and certain surrounding areas. Then, a consolidation process is assumed whereby the binding role is transferred from the hippocampus to neocortical sites, a movement from hippocampal-cortical to cortico-cortical connectivity. Many models of working memory assume that high levels of neural synchrony or simultaneous firing rate represent its memory contents. Short-term binding is thus accomplished through firing rates and temporal correlations of firing that emerge from the complex interplay of existing connections and from the effects of the recent activation history (from ‘thinking’, planning, perception, setting of motor movements, etc.). In a few seconds this content can in principle be transferred to long-term memory, specifically to the hippocampus, via Hebbian plasticity. In this paper, we will present a binding perspective from two connectionist models developed by us: a model of binding in working memory and a model of trace binding in long-term memory consolidation. We will delineate four different neurodynamical binding mechanisms, describe and discuss these two models of short-term and long-term memory binding, and finally present ideas for an integrated model architecture of binding in memory.

Introduction

There is a long tradition in viewing memory as associative (Hebb, 1949; Hinton and Anderson, 1989; James, 1890; Raaijmakers & Shiffrin, 1981; Willshaw, Buneman, Longuet-Higgins, 1969), assigning it a fundamental role in retrieving some information on the basis of a cue: someone’s name springs to mind when we see her face or when we read a movie title we remember the story-line and principal actors. Especially in humans, memory is an active process. Bartlett (1932), for instance, showed that the retrieval process often takes the form of an active reconstruction. Also during storage, processes of schematization are at work that continue to influence the contents of memory over time. During initial encoding, the combined effects of attention, perception, interpretation, and context all influence what will be stored, limiting for example what search cues can later elicit the information (e.g., Tulving and Thompson, 1973). In this paper, we will focus on

what we believe are the most fundamental binding processes of associative encoding, storage, and retrieval.

The past twenty years have seen a giant leap in our knowledge about the neural mechanisms that transiently or permanently bind information in memory. We will review some of this work on the basis of existing connectionist models and propose an integrated model. First, we will delineate four different binding mechanisms that we believe are important in memory. Then we will present a network model of working memory based on synchronized neuronal activation patterns that can be actively maintained and manipulated in recurrent loops between posterior and prefrontal cortical areas. We will also review some work on the modeling of storage and consolidation in long-term memory. In the final section we will integrate both approaches.

Four neurodynamical binding mechanisms

Binding is the process that determines the structure of representations in the brain (also see Damasio, 1989a-b). During the binding process neurons become bound into a coherent representation corresponding to actual objects or concepts in the world. Here, we distinguish four neural binding processes that allow flexible compositional operations in perceptual, working memory and long-term memory representations. The processes do not exclude each other, but cooperate in various ways during encoding, storage, and retrieval.

Co-activation-based binding

In general, co-activation through simultaneously high-frequency firing of neurons does not imply binding. Nor does such a high firing rate imply a causal link between active neurons. However, when co-active neurons both exchange excitation and are part of a larger cell assembly (Hebb, 1949) their activation states are interdependent and functionally related. For instance, the activation of the complete cell assembly may be triggered by an external input to a subset of its neurons, due to mutual associative connections. Local and non-local (global) assemblies for co-active binding can be inscribed in brain activity, supported by intra-regional and inter-regional (long-range) connectivity between pyramidal cells or via cortico-thalamic loops.

In neural networks, these cell assemblies or neural co-activation dynamics have been formalized in terms of attractor dynamics (Hopfield, 1982; Amit, 1989). In attractor dynamics, after a transient period, self-reinforcing synaptic excitation reaches a stationary state (attractor) in which the assembly pattern of activation remains persistent. To clarify what we mean by this, consider as the small model in Figure 1, which demonstrates some design principles of the model by McClelland and Rumelhart (1981). In this model for context effects in letter perception, there are word nodes (artificial neurons) and letter nodes. When a four-letter word such as *LAP* is activated through its constituent letter nodes *L*, *A*, and *P*, this pattern of co-activated nodes remains stable over time because the letter nodes excite the word node. The word node in turn activates the letter nodes thus reinforcing and maintaining its own source input: The co-activated letter nodes and word node are bound into an attractor.

Insert Figure 1 about here

In general, we define two co-activated neurons to be bound if they (i) are part of a neural assembly (attractor) and thus reinforce each others activation state, either directly (e.g., letter and word nodes in Figure 1) or indirectly (the letter nodes, via a word node), or if they (ii) are activated by some external source, for example, a visual input. In the latter case, it is the input pattern that causes node co-activation. In the visual system, this input-driven co-activation is achieved via feedforward converging connections (Felleman and van Essen, 1991). This bottom-up binding will not persist when the input source is removed, but plays a crucial role in the ignition of attractor dynamics at a higher representational stage. Attractor-driven binding, by contrast, will persist over time also past the offset of external input.

Both feedforward and attractor-based forms of co-active binding are involved in cortical information processing. An interesting fMRI study conducted by Courtney et al. (1997) showed a progressive increase in selectivity and stability (persistence) of neural activity from posterior (occipital) to more anterior (prefrontal) cortical areas involved in visual processing. Co-activation may initially be based on feed-forward input volleys in more posterior visual cortical areas with progressively dominant attractor dynamics in high-level visual processing areas and prefrontal cortex.

Synchrony-based binding

Perceptual binding based on co-activation or firing rates has some limitations that may be overcome by a richer short-term binding mechanism. In particular, it has difficulty explaining how multiple objects may be segregated in a given visual scene (von der Malsburg, 1981). The same segregation problem may be observed in a working memory representation of multiple patterns (Luck and Vogel, 1997). For example, if six neurons are co-activated, it will not be clear whether or not they represent one, two or more objects or events, whereas if the spiking times and synchronization/desynchronization patterns between the six neurons are considered, there may be two or more groups of neurons with high within-group temporal correlations of firing but with low between-group correlations. In this way multiple objects can be represented simultaneously (von der Malsburg, 1981; Engel et al., 1992).

Insert Figure 2 about here

In Figure 2a, spiking of neurons 1, 2, and 5 is temporally correlated and so is that of neurons 3, 4, and 6. But spiking of two neurons from different groups, such as neuron 2 and 3, is only weakly correlated. In this form of binding, when two neurons are highly correlated in their time-resolved firing, there is a strong indication that they are bound into a common representation.

The mechanisms for synchronization-based binding can be similar to the mechanisms for co-active binding: (i) forming part of an attractor in the presence of fast synchronizing connections, or (ii) temporary synchronization by some outside source (e.g. correlated noise). However, faster synaptic processes of lateral excitation and inhibition are demanded. We will discuss this form of neural binding below with reference to perceptual and working memory representations.

Interactive or ecphoric binding

Interactive binding refers to the interaction of two or more partly overlapping or convergent representations. For example, two input streams to a given brain area may each activate a large number of neurons. The neurons that receive support from both streams are most strongly activated (see Figure 3). If there is suitably calibrated inhibition or a multiplicative combination of neural inputs, only the strongly activated neurons will remain activated. In neural networks, this is typically modeled by some form of inhibition that drives a competition process so that it selects the ‘winning’ neurons, or by a mechanism in which neurons remain active only if two (or more) inputs arrive in close succession. Interactive binding decorrelates the winning neurons from the losing neurons, resulting in a concise representation coded by a subset of the total set of activated neurons. For example, if a set of neurons represents a broad concept such as GROUND, selection of a subset (e.g., by combining it with a particular context input) may be viewed as a process of interpretation, for example, GROUND as ‘the top layer of a garden’. In a similar vein, the combination of goal-relevant top-down signals with a bottom-up perceptual input may produce a selective attention effect.

Insert Figure 3 about here

Interactive binding may work in tandem with other types of binding, as can be seen in the small network of Figure 1. The letter nodes deliver a broad range of co-activations to the word nodes, which through inhibitory processes are reduced to a single interpretation: the winning word node. Interactive binding can occur in two ways. One is to combine different streams multiplicatively. The other is to feed different successive inputs to the same receptive units. Funneling a rich stream of neural activations through a *single* node is an extreme form of interactive binding. It can nonetheless be a very powerful coding process that binds multiple subordinate representations into a single superordinate representation (Page, 2000).

Interactive binding can explain Tulving’s ‘ecphoric’ processes that are responsible for encoding specificity. Thus, if during list learning, a word pair COFFEE – GROUND is presented, COFFEE will later be able to function as a retrieval cue for GROUND. In fact the newly learned association may be even more effective than pre-existing associations like COLD (Tulving and Thompson, 1973). This can be understood in terms of the interactive binding whereby from the many neurons activated by COFFEE and GROUND only a small set is selected and associated, creating a highly specific episodic representation.

A similar mechanism is operative during any type of retrieval that involves a reference to some context. This pretty much includes all forms of explicit memory. Modelers have noticed time and again that in order to have a context cue C retrieve only the items encoded in context C , but not those encoded in other possibly similar contexts C_1 , C_2 , etc., it is necessary to have a very selective cueing process. Thus, in the SAM model of associative memory, context and items interact in a multiplicative manner, allowing context C to ‘let through’ or select only those items that were associated with the cued context (Raaijmakers and Shiffrin, 1981). Interactive binding can be viewed as a multiplicative interaction, in this case the interaction of two sets of active neurons from which some highly co-activated intersection remains active. Taking the intersection of two sets is often referred to in mathematics as set multiplication. Thus, item-item

interactions and context-item interactions must be highly selective for memory retrieval to operate in accordance with known phenomena in memory psychology.

Recent evidence suggest that entorhinal cortex might carry out a form of interactive binding of context and item (or object) information, with context information being carried by the parahippocampal area and item information by the perirhinal area both of which converge onto the entorhinal cortex, which in turn is the main gateway to and from the hippocampus. A recent implementation indicates that such a model might also explain selective impairments of memory typically found in patients with schizophrenia. In schizophrenia, the parahippocampal area tends to be reduced in volume so that one might hypothesize that context cues fail to select items learned, causing a noticeably reduced free-recall performance. Recognition, however, remains largely intact, as this is much less dependent on the contextual cues. A recent simulation of such a mechanism confirms these intuitions (Talamini et al., in press).

Associative long-term binding

Hebb's learning rule (Hebb, 1949) echoes early ideas about association of 'brain processes' (James, 1890). This rule is often summarized as 'neurons that fire together, wire together'. Formulated in this manner, Hebbian learning is the long-term encoding of representations that are bound co-actively. Connections will grow stronger between co-active neurons, binding them associatively. Also, representations that have been formed through interactive binding process may be associated into long-memory in this manner, both by strengthening the connections between the neurons within a representation (see Figure 4) and the neurons in streams that feed into a representation.

Insert Figure 4 about here

The relationship between associative and synchronous binding is more complicated. If connections exist between coincidentally firing neurons (i.e., neurons that fire at the same time), a rich pattern of correlations may be encoded in the network, and these may be reproduced when the network is stimulated suitably at some later time. Thus, in Figure 2b, connections have developed between neurons 1 and 2, 1 and 5, 2 and 5, creating a small network. A similar network has developed between neurons 3, 4, and 6. In principle, associative binding is thus also suitable for the long-term encoding of synchronously bound representations.

Suppose that a network such as that of Figure 2b is stimulated by a diffuse input that affects all neurons equally and suppose that the neural firing is a chance process, such that the more input they receive the higher the firing probability. Then, firing neurons will tend to contribute to the (simultaneous) firing of other neurons in the network, causing the network as a whole to exhibit a high degree of synchronous firing. If, like in Figure 2b, a network consists of two subnetworks or assemblies, the neurons within such an assembly will fire synchronously, but the correlations in firing between the assemblies will be low. In this way, long-term encoded synchronous firing patterns can be read-out.

If additionally, there are both inhibitory and refractory processes, only a small number of neurons can remain active for only a short time. The inhibition limits the size of an active (synchronously firing) assembly and the refractoriness causes an assembly to be silent for a brief period after firing, perhaps allowing another assembly to become

active. In this way, a set of assemblies encoded in the same neural network may be read-out consecutively. In engineering, the mechanism whereby simultaneously encoded representations are activated sequentially is known as time-multiplexing.

Recently, it has been observed that if presynaptic firing precedes the postsynaptic spike within a small time-window the synapse is potentiated, whereas if it follows the postsynaptic spike, the synapse is depressed (Markram et al., 1997; Bi & Poo, 1998). This phenomenon corresponds to spike timing-dependent Hebbian plasticity (STDP). The temporal window for inducing such synaptic changes is on the order of 10 ms. Therefore, spike timing seems crucial for Hebbian associative binding. A near-synchrony state with a small positive time lag of firing between post and presynaptic neurons seems optimal for long-term synaptic strengthening. A larger time lag between neurons coding for different patterns may be used to cue sequence recall between subsequent assemblies, e.g. in serial recall in working memory. Such a representational differentiation cannot be achieved in terms of pure firing rates or co-activation based binding.

Multiple binding processes in memory

The four neurodynamical binding processes may be associated with different memory processes. Synchronous and co-active binding structure neural representations in working memory. Interactive binding allows highly specific episodic (event) codes to become active during encoding, whereas during retrieval it allows selective activation of memory items on the basis of specific search cues-in-context. Associative binding is responsible for long-term storage. In the following sections, we will review in more detail how these binding processes operate in memory and how their integration is accomplished by large-scale brain architectures.

From perception to working memory

One of the most salient attributes of neural representations in perception is their spatial and temporal discontinuity at different scales. So, an important problem to solve is how coherent perceptual interpretations arise from these discontinuities. The perception of coherent entities (i.e., discretely categorizable real-world objects or sounds) in complex scenes implies the dynamic linking (binding) and separation (segregation) of active neural representations corresponding to these entities. Several solutions to this problem have been suggested. One solution, for example, is to assume higher-order representations (e.g., cardinal cells, Barlow, 1972, 1989) onto which signals from neurons coding for to-be-bound features converge. It is clear that some form of higher-order representation is stored in long-term memory, but given the variability of retinal projections and the multitude of ways in which discrete features can be combined, this solution is often considered implausible since it ultimately would lead to a combinatorial explosion of the number of cells that represent all possible feature combinations.

However, this combinatorial problem may be resolved if the concept of high level representations is combined with the idea that binding can also be accomplished by selective synchronization of time-resolved neuronal responses (von der Malsburg, 1981, 1999; Eckhorn et al, 1988; Engel et al., 1992; Gray, 1999). According to this view, the action potentials of neurons coding features of the same object are synchronized, while being uncorrelated to neurons that code for the features of other objects. Although such synchrony may be supported by the existence of conjunctive codes at a high level, such

neural synchrony may also be decoded by means of categorizing units in a read-out layer without requiring a pre-established mapping onto 'cardinal' representations covering the full combinatorial spectrum.

As an elaboration of the temporal coding hypothesis, Raffone and Van Leeuwen (2002, 2003) suggested that feature binding with multiple active perceptual patterns may be mediated by a dynamic (e.g., chaotic) rather than a static synchronization as is typically observed with coupling of sinusoidal neural oscillators. For instance, given a visual scene with a red circle, a red square and a green square, neurons coding for red would be intermittently synchronized with neurons coding for circle and neurons coding for square, without necessarily implying spurious synchronization of the neurons coding for disjoint features. As was shown in computer simulations, this dynamic and graded synchrony may enable the flexible encoding of multiple active patterns in associative memory networks. Readout of this dynamic synchronization may take place efficiently by means of a fast self-organizing network layer.

The end result of perception is the selection and conscious awareness of a small part of the information that is present in our environment at any moment in time. What is selected and what is in conscious awareness forms the content of working memory. According to Baddeley (1992; Baddeley & Hitch, 1974), the concept of working memory refers to a system for temporarily holding and manipulating information that is required for performing cognitive tasks, such as comprehension, learning, and reasoning. Note that this definition is much broader than is customary in the neuroscience literature, where working memory mainly refers to a maintenance function that is distinguished from other functions like attention and executive control (e.g., Duncan, 2001).

It is generally accepted that the prefrontal cortex (PFC) is of cardinal importance for behavior guided by internal states and intentions, including behavior requiring the selective maintenance of earlier information, the suppression of automatic responses and the establishment of new or rapidly changing mappings between perception and action (e.g., Cohen and Servan-Schreiber, 1992; Duncan, 2001; Miller and Cohen, 2001; Wood and Grafman, 2003). Anatomically the PFC is well positioned to coordinate and control processing in the rest of the brain. It consists of a large number of interconnected areas that collectively have reciprocal connections with almost all other neocortical and subcortical structures. It is also an area that shows late development, both phylogenetically and ontogenetically (Fuster, 2001).

Based on these anatomical considerations it may be argued that executive control functions of working memory have developed as a consequence of an evolutionary development of the anterior parts of the brain that allow neural processes to control primary perception-action relations in the rest of the brain (Phaf and Wolters, 1997). In this view, the role of PFC in controlling behavior is modulatory rather than transmissive (Norman and Shallice, 1996; Miller and Cohen, 2001; O'Reilly, Braver and Cohen, 1999). Whereas simple adaptive behavior rests on an external control loop between perception, action and perception of action results, the extended PFC may have created the possibility of an internalization of this loop. More specifically, it created the possibility to maintain information in an activated state by recurrent connections (loops) between PFC and other areas of the cortex (Phaf and Wolters, 1997).

Apart from maintenance, two other executive functions are top-down control of selective attention and large-scale integration of multi-source information into a unitary

episodic representation (i.e., an episodic buffer, Baddeley, 2000). Large scale integration may be realized as an attractor state in a high level system that receives input from all other subsidiary systems in PFC and from long-term memory. Interactions between this integrative PFC area and modality-specific reverberating loops may modulate the feedback of maintained information to posterior cortical areas, thus allowing top-down control over the selection of actions and over selective attention to sensory input by biasing neural processing in these lower cortical areas. This top-down control may take the form of interactive or ephoric binding.

Neuroscientific evidence substantiates these theoretical considerations. For example, single cell studies have shown persistent firing during delayed matching tasks both in PFC and in other cortical areas like inferotemporal and parietal cortex (e.g., Fuster, 1995; Goldman-Rakic, 1995; Sakai, Rowe, & Passingham, 2002; Ungerleider, Courtney and Haxby, 1998). Evidence for top-down attentional modulation of neural activity (i.e., relative enhancement of neuronal responses) to task-relevant stimuli and relative suppression of responses to task-irrelevant stimuli has been shown as early as the primary visual cortex (e.g., Chelazzi et al., 1993; Downing, 2000; O'Connor et al., 2002; Reynolds et al., 2000). In addition, task specific activity in PFC may also generate top-down biasing signals involved in long-term memory retrieval (Hasegawa et al., 1998) and storage (Kydd & Bilkey, 2003; Wagner et al., 1998). These findings have led to the biased competition model (Desimone & Duncan, 1995), suggesting that top-down control activates relevant representations, which are then in a better position to compete with irrelevant information for perceptual awareness and motor control. Within PFC, the occurrence of interaction and integration is suggested by findings showing that individual cells in lateral PFC adapt their responsiveness to different combinations of stimulus attributes according to changing task demands (e.g., Asaad et al., 1998; Rushworth and Owen, 1998; Wallis et al., 2001; White and Wise, 1999). A possible location of a large-scale binding system is the anterior prefrontal cortex, which seems to play a specific role in integrating the outcomes of multiple cognitive operations in the pursuit of higher behavioral goals (e.g., Ramnani & Owen, 2004).

A connectionist model of binding in working memory

On the basis of the previous considerations a network model was developed that aimed at demonstrating how core functions of working memory, such as its limited capacity for maintenance, binding (chunking), and modulation of selective attention, can be unitarily accounted for in terms of cortical circuit dynamics involving a recurrent interaction between PFC and other cortical areas (Raffone & Wolters, 2001; also see Raffone, Wolters, & Murre, 2001). Although the model concerned visual working memory, the basic mechanisms of maintenance and binding can be generalized to other working memory systems.

Insert Figure 5 about here

The model (see Figure 5) consisted of a cortical module (located in inferotemporal cortex, IT) and a prefrontal module (e.g., dorsolateral/ventrolateral PFC). Within the IT module, pre-existing cell assemblies were assumed. These were defined by strong intra-assembly connections and weak inter-assembly connections. Assemblies were assumed to code for simple features; multi-feature objects were coded by sets of

coupled (i.e., strongly inter-associated) assemblies. Within the IT module, a global inhibition (competition) mechanism between cell assemblies was implemented. Smaller assemblies, reciprocally connected to corresponding IT assemblies were assumed in the PFC module. The units of the model were spiking neurons modeled after MacGregor and Oliver (1974). Transmission times between connected cells within a module were assumed to be much faster than between cells of different modules. Stimulus input to the IT module was given by stochastic spike trains to specific assemblies coming from lower level visual areas, during a limited onset-offset period.

Figure 6 shows the working of the model when a single cell assembly input was presented. During stimulus input, average activity of the IT assembly (and the matching PFC module after a conduction delay) shows an oscillation with a gamma-band frequency (around 35 Hz). These oscillations are due to the fast synaptic coupling within the assembly. After a certain critical number of its neurons has fired as a consequence of enhanced input, the fast positive activation spreading through the assembly entrains most neurons to fire almost simultaneously. After firing, neurons fall into a collective refractory period. Synchronous firing also occurs in the PFC module. This is generated by the synchronized feedforward input from the IT module. Feedback from the PFC to the IT module suffices to maintain the modules in an oscillatory state after stimulus offset. The matched IT and PFC assemblies reverberate through delayed mutual excitation. In realistic cortical networks, PFC is autonomous and stable against interference in working memory activity, whereas maintenance in posterior cortical areas is PFC-dependent and interference-sensitive (Miller et al., 1996). Raffone and Wolters' (2001) model can accommodate this evidence by the addition of another reverberatory loop within the PFC module, probably mediating a working memory rehearsal process.

In this model, the limited capacity of (visual) working memory emerges due to the interactions via inhibitory interneurons between independent assemblies. Once an assembly becomes oscillatory, the enhancement of its average spike rate will exert a fast, strong and transient inhibitory action on other assemblies. However, after producing a synchronized spike burst this inhibitory effect dissipates quickly, allowing another assembly to gain activation. The capacity of the model was tested by presenting simultaneous input to multiple assemblies. Due to the stochastic nature of the model, the number of stored items varied realistically from trial to trial, but an average capacity to maintain about three to four assemblies after stimulus offset was observed (see Figure 6). This simulated storage capacity is remarkably similar to recent estimations of working memory capacity (see Luck and Vogel, 1997; Cowan, 2000). As can be seen in Figure 6, the inhibitory mechanism caused oscillating assemblies to become spaced optimally in time.

Insert Figure 6 about here

In Figure 6, both in computational and functional terms the limited storage capacity in the model depends on the functional balance of a sufficiently high oscillation frequency (firing rate), enabling a good signal-to-noise ratio, and sufficient phase-segregation between firing of neural assemblies coding for different objects (minimizing interference). The same desynchronizing mechanism that enables phase-segregation between assemblies coding for separate items poses a limit to the number of oscillatory reverberations. In other words, a reduced phase-lag would increase *interference* between

the maintained items and a lower firing rate would reduce the strength or *resolution* of the neural responses coding for a given item. Both interference and reduced resolution would make access or readout in terms of association to response codes or higher-order conjunctive units more difficult.

In a classic paper, Miller (1956) argues that the capacity of working memory has to be expressed in terms of coherent units or ‘chunks’ of information, instead of some formal measure of amount of information. Working memory handles higher-order chunks, where large amounts of organized information can become integrated as one chunk or one single unit of capacity (Ericsson and Kintsch, 1995). Luck and Vogel (1997) showed that the storage capacity of working memory is about four objects and that this capacity limit is indeed independent of the number of features making up the objects. We simulated this chunking capacity by assuming multi-feature objects to be represented as strongly interconnected feature assemblies (assuming mutual synchronizing associations between different assemblies). From these simulations, it appeared that multi-assembly units show the same within-object synchronization and between-object segregation as single assemblies, and that the modeled capacity is largely unaffected by the size or number of features of integrated representations. So, the model can potentially account for chunking based on the strength and stability of composite representations in long-term memory.

The theoretical solution to the problem of limited capacity of short-term memory proposed by Raffone and Wolters (2001) makes computationally-explicit neurophysiological constraints considered in earlier theoretical work (Lisman & Idiart, 1995; Cowan, 2001). The role of oscillations and neural synchrony in working memory suggested by these computational investigations is supported by neurophysiological and electroencephalographic data (Nakamura, Mikami, & Kubota, 1992; Villa and Fuster, 1992; Tallon-Baudry, Bertrand, & Fischer, 2001).

This basic recurrent architecture between PFC and posterior cortex also suggests a possible scheme for implementing a top-down biased competition mechanism for controlling selective attention, via interactive or ephoric binding. As shown by Rainer, Asaad and Miller (1998), prefrontal neurons of monkeys performing a delayed-matching-to-sample task exhibited a higher firing rate when they coded for a target object. Such neurons were suggested to be involved in both the maintenance and the selection of behaviorally relevant information. A top-down selective biasing effect may be assumed to originate from PFC areas presumably involved in supervisory control (e.g., Rushworth & Owen, 1998). It has been shown that both additive and multiplicative modulatory inputs to simulated PFC maintenance modules can be effective in selecting and selectively maintaining biased inputs (Deco and Rolls, 2003; also see Raffone, Murre, & Wolters, 2003).

So far, maintenance, capacity and chunking effects are realized by a binding principle based on synchronous firing mediated by pre-existing representations in long-term memory. Binding within reverberating assemblies may also be possible in the absence of pre-existing associative representations, because correlated spikes from areas at early processing stages may be used to generate synchronous firing in assemblies in higher level areas, as was suggested for example by Roelfsema et al. (1996) and Abeles (1991). Raffone and Wolters showed that even weak connections between IT assemblies can synchronize their reverberatory activities with strongly correlated inputs. It is likely

that early perceptual grouping and long-term memory guidance in terms of pre-existing assemblies with strong connections (i.e., chunks) may cooperate in the formation of integrated entities to be maintained in working memory (see also Jensen and Lisman, 1996).

Thus, with structured input to the system, coherent and synchronized activation patterns representing aspects of the input are constructed and can be maintained over time allowing subsequent operations to be performed. We suggest that one of these subsequent operations is the creation of novel associations between the maintained patterns via the hippocampus (interactive and associative binding), which creates a conjunctive episodic code. Such a code in turn will recursively contribute to further structuring of the elements (chunks) that appear in the assemblies activated in working memory.

Connectionist models of long-term memory and consolidation

Most recent connectionist models of the neural basis of long-term memory emphasize the associative binding role of the hippocampus (Alvarez and Squire, 1994; McClelland et al., 1995; Murre, 1996; Nadel & Moscovitch, 1997; Nadel et al., 2000). These models implement and extend earlier theorizing along similar lines, assuming that long-term memories are initially—and sometimes also thereafter—bound by the hippocampus (Gluck and Meyers, 1993; McNaughton and Nadel, 1990; Marr, 1971; Milner, 1957, 1989; Mishkin, 1982; Nadel et al., 2000; O'Keefe and Nadel, 1978; Paller, 1997; Squire and Alvarez, 1995; Squire, Cohen, and Nadel, 1984; Squire and Zola-Morgan, 1991; Teyler and DiScenna, 1986; Treves and Rolls, 1994; Wickelgren, 1979, 1987).

At least three of the models that have simulated cortico-hippocampal interactions (i.e., Alvarez and Squire, 1994; McClelland et al., 1995; Murre, 1996) share the basic assumption that there is a fast-learning hippocampal memory system and a neocortical memory system in which representations are gradually built up in a consolidation phase, which follows the initial learning episode. This consolidation phase is implemented as a process of reactivation, in which stored patterns are strengthened by rehearsing of the original patterns (McClelland et al., 1995; Robins, 1995) or as 'pseudorehearsal' in which patterns are generated from the network from random cues (Alvarez and Squire, 1994; Meeter and Murre, 2003; Murre, 1996; Robins, 1996). Those patterns are then interleaved with new patterns to protect, repair, or strengthen old ones.

In the TraceLink model of long-term memory and consolidation, for example (Meeter and Murre, 2003; Meeter and Murre, in press a; Murre, 1996), the neocortical memory system is a large layer in which only weak connections are laid down between nodes (artificial neurons) belonging to one pattern. Consolidation is simulated by letting the model—from an initially random state—relax into an attractor (i.e., retrieving an existing memory), and then updating the weights with a Hebbian learning rule (see Figure 7). Eventually, the connections between neocortical nodes built up during consolidation allow the patterns to be retrieved without the support of the hippocampal system. TraceLink and the models by Alvarez and Squire (1994) and McClelland et al. (1995) all assume that neocortical learning is very slow compared to hippocampal learning. It is this assumption that assigns a binding role to the hippocampus in the time period following initial learning of a novel episode.

Insert Figure 7 about here

The consolidation stage that follows this first stage is modeled as the strengthening of connections within a neocortical pattern that is retrieved through the hippocampal system. This implies that there must be consolidation phases in which the hippocampal system reinstates patterns in neocortical memory areas. Such a consolidation mechanism is sensitive to ‘runaway consolidation’, a vicious circle in which one pattern becomes stronger through consolidation, then becomes more likely to be consolidated in the next trial, and ends up monopolizing all consolidation resources while crowding out other memories (Meeter, 2003). In the models runaway consolidation is avoided through the dominance of the hippocampal system, helping to reactivate patterns in the neocortex that have not yet benefited from consolidation (Alvarez and Squire, 1994; McClelland et al., 1995; Meeter and Murre, in press a; Murre, 1996). If during the reactivation learning occurs within the hippocampal system, runaway consolidation immediately rears its head, as now consolidated memories become stronger in both the hippocampus and the neocortex (Meeter and Murre, 2003). Consolidation should therefore take place during a period that the hippocampus is not very plastic.

A different view on the consolidation process was put forward by Nadel and Moscovitch (1997). They dub the approach taken by the above models as the Standard Theory of Consolidation with as a principal characteristic that in the course of the consolidation process, successful retrieval requires less and less hippocampal involvement. In contrast, Nadel and Moscovitch (1997) argue that the hippocampus always remains involved in the retrieval of memories and that consolidation occurs through gradual increase of the strength of the hippocampal trace, rather than the neocortical trace. A connectionist implementation of their model was presented by Nadel et al. (2000).

The models mentioned above differ in details and motivation but they share several crucial assumptions central to the subject matter of this paper.

- They assume a *binding role of the hippocampus*, at least in the initial stages of acquisition.
- They assume that there is a *neural hierarchy* in the sense that the hippocampus receives inputs from a large portion of the cerebral cortex while being able to influence processes, via recurrent connections, in these areas.
- They assume *slower formation of cortico-cortical connections* than (cortico-) hippocampal connections.
- They assume a *long-term consolidation* process in which learned memories continue to evolve well after initial acquisition (also see Meeter & Murre, in press b; Murre, Graham, & Hodges, 2001).

Trace binding role of the hippocampus

If the hippocampus has a binding role, we may ask ourselves: What is being bound? An early model by O'Keefe and Nadel (1978) is based on experiments with rats. With these animals, single-cell recordings reveal many place-sensitive cells in the hippocampus, and the model therefore stresses the role of the hippocampus in navigation (also see O'Keefe, 1990). One might assume that whatever is important for an organism and occurs frequently enough will tend to be represented in the hippocampus. For the rat, quickly

learning new locations seems to be important, for example, to seek shelter in case of danger. To remember places of shelter, it is thus necessary to bind the representation for 'safe spot' to certain locations in the current context. More generally, we might view the binding process as one where one or more objects or attributes (e.g., safe places, boxes with cereal) in a given context (e.g., an experimental maze or my backyard) are bound together.

A functional model by Gluck and Meyers (1993) emphasizes the integration of context and task-related objects by a *compression mechanism* regulated by the hippocampus. An important observation that emerges from this work is that even tasks that can be performed without a hippocampus (e.g., after lesioning) may still show differences when comparing experimental results before and after lesioning. One example of how their model deals with this involvement is the explanation of context-effects in simple conditioning. When a task is moved into a different context, normal animals will show a marked decrease in responding. Animals that undergo lesioning of the hippocampus after having learned the task do not show such a decrease: they respond equally well before and after context shifts (Penick and Solomon, 1991). These and other phenomena are explained by Gluck and Meyers (1993) by assuming that the hippocampus controls a compression mechanism that is automatic and always operative. This mechanism integrates the context and object in a manner akin to the integrative binding mechanism above, in this manner filtering out irrelevant aspects of the context.

After prolonged training the context will be strongly compressed. Stimuli that are central to the task remain prominently represented but those that are coincidental are 'squeezed out'. Thus task aspects such as the lever to be pressed and its consequences will be well-represented (many neurons firing) whereas the color of the floor or any objects that have nothing to do with the task will become to be represented by fewer and fewer neurons. Compression of irrelevant inputs, or attention to relevant aspects of the input, has been part of many theories of cortical functioning. For example, Barlow (1989) has argued that an important function of unsupervised learning in the cortex is to extract redundancy by filtering out irrelevant stimuli. One way in which such extraction can take place is by merging similar stimuli: if stimulus variations are irrelevant it is not necessary to represent them. This leads to categorization or clustering of representations, possibly, but not necessarily, in a topological organization. Many connectionist approaches to unsupervised learning have been proposed, most of these using similar hypothesized processes in the cortex as a justification (e.g., Grossberg, 1976, 1987; Kohonen, 1989; Linsker, 1996 a-c; Murre, 1992; Murre, Phaf, and Wolters, 1992; Rumelhart and Zipser, 1985). These have in common that categories emerge through mutual inhibition of a limited set of categorizing nodes (artificial neurons). This funnels an elaborate representation through just a few nodes. These mechanisms are effective in extracting statistical regularities from the input, thus compressing it. This resembles the mechanism whereby regularities of a computer file (e.g., repeated words or sentence fragments) are utilized by compression algorithms to reduce their size.

McClelland, McNaughton, and O'Reilly (1995) emphasize yet another possible role of the hippocampus. They point out that purely sequential learning may not lead to useful internal representations and that a case can be made for the necessity of a more interleaved mode of learning. In particular, newly learned deviant patterns may disturb already learned representations. There are thus also good behavioral reasons for a slow

(interleaved) learning process, that is, a learning process whereby occasionally an old pattern is given an extra learning trial, amongst newly arrived learning patterns. This mechanism can be interpreted as binding of related patterns that occur at disparate moments in time, a form of interactive binding in time.

We thus see that the hippocampus may take part in several binding processes. The most prevalent one is associative binding, which is very prominent initially, before consolidation to the neocortex has taken place. Additionally, the hippocampus is probably involved in interactive binding of context and task-central items or objects with progressive compression of irrelevant aspects of these representations over time, binding them into a compact representation suitable for efficient long-term storage. It is also likely that the long-term consolidation mechanism, in which the hippocampus is the central area, executes a form of dynamic interactive binding, integrating new representations with similar ones already in long-term memory.

An integrated architecture for binding in working memory and long term-memory

In the previous sections we have presented four basic neurodynamical binding mechanisms operating with a different spatial and temporal resolution. We have presented essential computational desiderata for flexible representational compositions in perceptual and working memory processing. And we have discussed a model of working memory based on short-term binding for maintenance of integrated information chunks. We have also discussed how transiently active representations in posterior cortex can be quickly bound for intermediate-term storage via the hippocampus with interactive and associative binding processes and how these episodic representations may transfer to posterior cortex in the form of permanent associative binding.

Based on these neurodynamical and structural constraints for memory binding, as an integrative architecture, we suggest a tripartite system with cooperative interactions between posterior cortex and PFC and between posterior cortex and hippocampal or medial temporal lobe area (MTL). Unfortunately, it is still unknown what kind of signaling occurs between PFC and MTL; only recently the anatomical connections between these areas and their possible functional interaction have come under consideration (e.g., Simons and Spiers, 2003). Some of these memory binding interactions have been summarized in the architecture shown in Figure 8.

Insert Figure 8 about here

Of course many interactions between PFC and MTL may be explained by their connections with a common posterior cortex system. For instance, PFC may provide top-down control of encoding processes by selecting, modifying, and elaborating the representations in posterior cortex that are subsequently stored via MTL. Similarly, PFC may control memory search and the selection of retrieval cues or strategies via its interaction with posterior cortex and without direct interaction with the MTL. However, the presence of direct anatomical links between subsystems in MTL and PFC (see, e.g., Simons and Spiers, 2003) suggests direct interactions as well. For example, there are strong reciprocal connections between prefrontal cortex and the perirhinal and entorhinal cortices (Groenewegen and Uylings, 2000). A possible functional interpretation of this

direct link is that it plays a role in the integration and association of spatial and temporal context of events. Recent suggestions on the role of PFC have stressed its ability to integrate the outcomes of separate cognitive operations (e.g., Kroger et al., 2002; Ramnani and Owen, 2004), and to control processing according to spatial or temporal context demands (Koechlin, Ody and Kouneiher, 2003). If it is assumed that the spatial-temporal context of events is integrated in PFC, then direct inputs from PFC to MTL may provide the contextual information that is combined with event specific information to create a spatio-temporal specific episodic trace (see e.g., Kydd & Bilkey, 2003). These suggestions, however, clearly are in need of further substantiation.

The tripartite architecture involves all four forms of neurodynamical binding considered above, with a differential involvement depending on the level and time-scale of neural representation and processing. For instance, attractor dynamics or co-activation based binding may take place in inferotemporal cortex, prefrontal cortex, and the field CA3 of the hippocampus (Rolls & Treves, 1998). These attractor dynamics may be qualitatively different, e.g. due to the differential involvement of fast (AMPA-mediated) and slow (NMDA-mediated) excitatory currents. At a functional level, co-active binding in prefrontal cortex is likely to be associated to executive control processes, probably mediated by NMDA voltage-dependent currents (Wang, 1999; Raffone et al., 2003). Attractor dynamics in posterior cortical areas may be supported by both local intra-regional cooperation and non-local feedback from prefrontal cortex, and it may be sensitive to feedforward input (Miller et al., 1996; Courtney et al., 1997). Multimodal co-active binding in the hippocampus may take place in CA3 due to dense recurrent collateral excitation, in terms of both AMPA- and NMDA-mediated currents, with theta oscillation modulation.

Synchrony-based binding is likely to take place at different representational levels in posterior cortical areas related to perceptual processing (see Singer, 1999). The role of synchrony in higher level attentional and working memory processing has been supported by only a few studies due to technical recording problems (Villa and Fuster, 1992; Fries et al., 2001). Single-cell recording evidence against the role of synchrony in these cognitive functions is not available. Neural synchrony at a system level may also play a crucial role in cooperation between prefrontal cortex, posterior cortical networks and medial temporal lobe neural assemblies, as indicated by a recent MEG study with a complex attentional task (Gross et al., 2004). Finally, temporal coding may play an important role in hippocampal sequence coding in terms of nested theta/gamma oscillations (Jensen & Lisman, 1996; Lisman, 1999).

As considered above, interactive or ecphoric binding of object and multimodal episodic context information may take place in the MTL, specifically in the entorhinal cortex. Ecphoric binding may also take place in prefrontal cortex, where attentional biased competition is primarily mediated (Deco & Rolls, 2003). The convergence of top-down biasing inputs from target-coding prefrontal assemblies and bottom-up input from temporal and parietal networks is likely to give rise to competitive neural dynamics, selecting the neural assemblies with the highest level of activation in PFC. Multiplicative effects in this ecphoric binding may be mediated by NMDA-currents present in prefrontal cortex (Wang, 1999). Interactive binding in terms of voltage-dependent synapses may also take place at lower levels of perceptual processing, to mediate context-effects via

horizontal connections, in the absence of spurious activation spreading that could be caused by co-activation based binding (e.g. Tononi & Edelman, 1992).

As discussed above, associative long-term binding may take place in MTL and neocortical areas, at different stages. However, as suggested by O'Reilly and Rudy (2001), a relative division of labour may exist between neocortical and hippocampal systems, as rats with hippocampal lesions can still learn tasks such as nonlinear discrimination problems. Similar evidence has been gained with amnesic patients. Therefore, there may be a relative independence between cortical and hippocampal binding processes. In our framework, this relative independence in non-episodic learning tasks may be supported by the presence of the four neurodynamical binding mechanisms in both hippocampal and neocortical areas. Differential processing and learning at neocortical and hippocampal networks may depend upon architectural and encoding (mapping) properties emphasizing either idiosyncratic or common features and feature-conjunctions making up entities and events. However, in a proper episodic memory, hippocampal binding is likely to be a prerequisite for a subsequent neocortical binding.

A major unknown aspect concerns the relative role of PFC and posterior cortical areas in episodic information storage and retrieval. It is currently unclear whether PFC stores and retrieve information about object and events or rather operate on posterior and motor cortical areas by means of task-adaptive codes (Duncan, 2001; Wood and Grafman, 2003). In a functional imaging study conducted by Prabhakaran et al. (2000), the activation of prefrontal cortex was greater for maintaining integrated rather than unintegrated representations. These results suggest that prefrontal cortex is involved in integrating verbal and spatial information in working memory, i.e. in mediating the function of the episodic buffer suggested by Baddeley (2000). Therefore, it may be assumed that the transfer of episodic codes from the hippocampus to neocortex during the consolidation process may differentially involve neocortical areas. Associative neocortical binding may be governed by higher order convergence codes in PFC, which can be related to goals, plans and encoding of task setting. PFC may play a role similar to the hippocampus after the consolidation process, in terms of multimodal integration, with access to action-related codes.

To conclude, perception and action do not occur in a vacuum. Perceptual input enters a neural system that is already in an activated state, representing lasting effects of previous inputs and short- and long-term goals and intentions. Which aspects of a momentary input are selectively attended is determined as much (and sometimes even more) by the present activation state of the system as by the actual input itself. Activation states in turn are determined for a large part by the pre-existing neural structures that have been shaped by years of experience making up what is referred to as long-term memory, containing semantic knowledge of concepts, episodic memory of specific events, conditioned stimulus-response associations and perceptual, conceptual and motor skills.

The general scheme we have presented here is that the versatility of the brain in coping with continuously changing environments and demands is due to the capacity to store coherent patterns of input and output in long-term memory and the capacity to control the processing of input by selecting and maintaining task-relevant information in working memory. These processes require the possibility of transiently and permanently binding neural assemblies representing elements of input and output. These binding processes continuously interact. What is transiently bound in working memory governs

what is temporarily and eventually permanently bound in long-term memory. In turn, what is permanently bound affects transient binding in working memory. The interplay of these binding processes determines how the brain develops into a structured system that is cumulatively correlated to its environment, thus implementing a process that is able to lift itself to higher levels of cognitive functioning.

References

- Abeles, M. (1991). *Corticonics: neural circuits of the cerebral cortex*. Cambridge: Cambridge University Press.
- Alvarez, R., & L. R. Squire (1994). Memory consolidation and the medial temporal lobe: a simple network model. *Proceedings of National Academy of Sciences (USA)*, 91, 7041-7045.
- Amit, D. J. (1989). *Modeling brain function*. Cambridge University Press.
- Asaad, W. F., Rainer, G., & Miller, E. K. (2000). Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology*, 84, 451-459.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.
- Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4, 417-423.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation*. London: Academic Press, 47-89.
- Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, 1, 371-394.
- Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, 1, 295-311.
- Bartlett, F. C. (1932). *Remembering*. Cambridge: Cambridge University Press.
- Bi, G. Q., & Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18, 10464-10472.
- Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. A. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, 363, 345-347.
- Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, 99, 45-77.
- Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature*, 386, 608-611.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87-114.
- Damasio, A. R. (1989a). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, 1, 123-132.
- Damasio, A. R. (1989b). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33, 25-62.
- Deco, G., & Rolls, E. T. (2003). Attention and working memory: A dynamical model of neuronal activity in the prefrontal cortex. *European Journal of Neuroscience*, 18, 2374-2390.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neurosciences*, 18, 193-222.
- Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience*, 2, 820-829.
- Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., & Reitboeck, H. J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics*, 60, 121-130.

- Engel, A. K., König, P., Kreiter, A. K., Schillen, T. B., & Singer, W. (1992). Temporal coding in the visual cortex: New vistas on integration in the nervous system. *Trends in Neuroscience*, 15, 218-226.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211-245.
- Felleman, D. J., & van Essen, D. C. V. (1991). Distributed hierarchical processing in the primate visual cortex. *Cerebral Cortex*, 1, 1-47.
- Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291, 1560-1563.
- Fuster, J. M. (2001). The prefrontal cortex - an update: time is of the essence. *Neuron*, 30, 319-333.
- Gluck, M. A., & Meyers, C. E. (1993). Hippocampal mediation of stimulus representation: a computational theory. *Hippocampus*, 3, 491-516.
- Goldman-Rakic, P. S. (1988). Topography of cognition: parallel distributed networks in primate association cortex. *Annual Review of Neurosciences*, 11, 137-156.
- Gray, C.M. (1999). The temporal correlation hypothesis of visual feature integration: Still alive and well. *Neuron*, 24, 31-47.
- Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338, 334-337.
- Groenewegen, H. J. U., & Uylings, H. B. M. (2000). The prefrontal cortex and the integration of sensory, limbic and autonomic information. *Progress in Brain Research*, 126, 3-28.
- Gross, J., Schmitz, F., Schnitzler, I., Kessler, K., Shapiro, K., Hommel, B., & Schnitzler, A. (2004). Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. *Proceedings of the National Academy of Sciences, USA*, 101, 13050-13055.
- Grossberg, S. (1976). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, 23, 187-202.
- Grossberg, S. (1987). Competitive learning: from interactive activation to adaptive resonance. *Cognitive Science*, 11, 23-63.
- Hasegawa, I., Fukushima, T., Ihara, T., & Miyashita, Y. (1998). Callosal window between prefrontal cortices: Cognitive interaction to retrieve long-term memory. *Science*, 281, 814-818.
- Hebb, D.O. (1949). *The organization of behavior*. New York: Wiley.
- Hinton, G. E., & Anderson, J. A. (Eds.) (1989). *Parallel models of associative memory*. Hillsdale, NJ: Erlbaum.
- 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.
- Horn, D., & Usher, M. (1991). Parallel activation of memories in an oscillatory neural network. *Neural Computation*, 3, 31-43.
- James, W. (1890/1950). *Principles of psychology*. New York: Dover.

- Jensen, O., & Lisman, J. E. (1996). Novel lists of 7 ± 2 known items can be reliably stored in an oscillatory short-term memory network: interaction with long-term memory. *Learning and Memory*, 3, 257-263.
- Koechlin, E., Ody, C., & Kouheiner, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, 302, 1181-1185.
- Kohonen, T. (1989). *Self-organization and associative memory*, 3rd edition. Berlin, Germany: Springer-Verlag.
- Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., & Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning. *Cerebral Cortex*, 12, 477-485.
- Kydd, R. J., & Bilkey, D. K. (2003). Prefrontal cortex lesions modify the spatial properties of hippocampal place cells. *Cerebral Cortex*, 13, 444-451.
- Linsker, R. (1986a). From basic network principles to neural architecture: emergence of spatial-opponent cells. *Proceedings of the National Academy of Sciences, USA*, 83, 7508-7512.
- Linsker, R. (1986b). From basic network principles to neural architecture: emergence of orientation-selective cells. *Proceedings of the National Academy of Sciences, USA*, 83, 8390-8394.
- Linsker, R. (1986c). From basic network principles to neural architecture: emergence of orientation columns. *Proceedings of the National Academy of Sciences, USA*, 83, 8779-8783.
- Lisman, J. E., & Idiart, M. A. P. (1995) Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science*, 267, 1512-1515.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279-281.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419-457.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception. Part I: an account of basic findings. *Psychological Review*, 5, 375-407.
- MacGregor, R. J., & Oliver, R. M. (1974). A model for repetitive firing in neurons. *Kybernetik*, 16, 53-64.
- McNaughton, B. L., & Nadel, L. (1990). Hebb-Marr networks and the neurobiological representations of action in space. In: M. A. Gluck & D. E. Rumelhart (Eds.), *Neuroscience and connectionist theory*. Hillsdale, NJ: Erlbaum, 1-63.
- Markram, H., Lubke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275, 213-215.
- Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical Transactions of the Royal Society B*, 262, 23-81.
- Meeter, M. (2003). Control of consolidation in neural networks: Avoiding runaway effects. *Connection Science*, 15, 45-61.
- Meeter, M., & Murre, J. M. J. (2004). Simulating episodic memory deficits in semantic dementia with the TraceLink model. *Memory*, 12, 272-287.

- Meeter, M., Talamini, L. M., & Murre, J. M. J. (2004). Mode shifting between storage and recall based on novelty detection in oscillating hippocampal circuits. *Hippocampus*, 14, 722 - 741
- Meeter, M., & Murre, J. M. J. (in press a). TraceLink: A model of consolidation and amnesia. *Cognitive Neuropsychology*, in press.
- Meeter, M., & Murre, J. M. J. (in press b). Consolidation of long-term memory: Evidence and alternatives. *Psychological Bulletin*, in press.
- 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.
- Miller, E.K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167-202.
- Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, 16, 5154-5167.
- Milner, P.M. (1957). The cell assembly: Mark II. *Psychological Review*, 64, 242-252.
- Milner, P.M. (1989). A cell assembly theory of hippocampal amnesia. *Neuropsychologia*, 6, 215-234.
- Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transactions of the Royal Society B*, 298, 85-95.
- Murre, J.M.J. (1992). *Categorization and learning in modular neural networks*. Hemel Hempstead: Harvester Wheatsheaf; Hillsdale, NJ: Lawrence Erlbaum.
- Murre, J.M.J., R.H. Phaf, & G. Wolters (1992). CALM: Categorizing And Learning Module. *Neural Networks*, 5, 55-82.
- Murre, J. M. J. (1996). TraceLink: A model of amnesia and consolidation of memory. *Hippocampus*, 6, 675-684.
- Murre, J. M. J., Graham, K. & Hodges, J. (2001). Semantic dementia: new constraints on computational models of long-term memory. *Brain*, 124, 647-675
- Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, 7, 217-227.
- Nadel, L., Samsonovitch, A., Ryan, L., & Moscovitch, M. (2000). Multiple trace theory of human memory: Computational, neuroimaging and neuropsychological results. *Hippocampus*, 10, 352-368.
- Nakamura, K., Mikami, A., & Kubota, K. (1992) Oscillatory neuronal activity related to visual short-term memory in monkey temporal pole. *NeuroReport*, 3, 117-120.
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In: R. J. Davidson, G. E. Schwartz, & D. E. Shapiro (Eds), *Consciousness and self-regulation* (Vol 4). New York: Plenum Press, 1-18.
- O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, 5, 1203-1209.
- O'Keefe, J., & L. Nadel (1978). *The hippocampus as a cognitive map*. Oxford: The Clarendon Press.
- O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically based computational model of working memory. In: A. Miyake & P. Shah (Eds), *Models of working memory*. Cambridge (UK): Cambridge University Press, 375-411.

- Page, M. (2000). Connectionist modelling in psychology: a localist manifesto. *Behavioral and Brain Sciences*, 23, 443-512.
- Penick, S., & Solomon P. (1991). Hippocampus, context and conditioning. *Behavioral Neuroscience*, 105, 611-617.
- Phaf, R.H., & Wolters, G. (1997). A constructivist and connectionist view on conscious and nonconscious processes. *Philosophical Psychology*, 10, 287-307.
- Prabhakaran, V., Narayanan, K., Zhao, Z., & Gabrieli, J. D. E. (2000). Integration of diverse information in working memory within the frontal lobe. *Nature Neuroscience*, 3, 85-90.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 454-457.
- Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience*, 13, 766-785.
- Raffone, A., Wolters, G., & Murre, J. M. J. (2001). A neurophysiological account of working memory limited capacity: Within-chunk integration and between-item segregation. *Behavioral and Brain Sciences*, 24, 139-141.
- Raffone, A., & van Leeuwen, C. (2002). Activation and coherence in memory processes: revisiting the parallel distributed processing approach to retrieval. *Connection Science*, 13, 349-382.
- Raffone A, & van Leeuwen, C. (2003). Dynamic synchronization and chaos in an associative neural network with multiple active memories. *Chaos*, 13, 1090-1104.
- Raffone, A., Murre, J. M. J., & Wolters, G. (2003). NMDA synapses can bias competition between object representations and mediate attentional selection. *Behavioral and Brain Sciences*, 26, 100-101.
- Ramnani, N., & Owen, A.M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience*, 5, 184-194.
- Rainer, G., Asaad, W. F., & Miller, E. K. (1998). Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature*, 393, 577-579.
- Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience*, 5, 184-194.
- Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron*, 26, 703-714.
- Roelfsema, P. R., Engel, A. K., König, P., & Singer, W. (1996). The role of neuronal synchronization in response selection: a biologically plausible theory of structured representations in the visual cortex. *Journal of Cognitive Neuroscience*, 8, 603-625.
- Rolls, E. T., & Treves, A. (1998). *Neural networks and brain function*. Oxford: Oxford University Press.
- Rumelhart, D. E., & D. Zipser (1985). Feature discovery by competitive learning. *Cognitive Science*, 9, 75-112.
- Rushworth, M. F. S., & Owen, A. M. (1998). The functional organization of the lateral frontal cortex: Conjecture or conjuncture in the neurophysiological literature? *Trends in Cognitive Science*, 2, 46-53.
- Sakai, K., Rowe, J. B., & Passingham, R. E. (2002). Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nature Neuroscience*, 5, 479-484.

- Simons, J. S., & Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Reviews Neuroscience*, 4, 637-648.
- Singer, W. (1999) Neuronal synchrony: a versatile code for the definition of relations? *Neuron*, 24, 49-65.
- Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99, 195-231.
- Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current Opinion in Neurobiology*, 5, 169-177.
- Squire, L. R., Cohen, N. J., & Zola-Morgan, L. (1984). The medial temporal region and memory consolidation: a new hypothesis. In: H. Weingarter & E. Parker (Eds.), *Memory consolidation*. Hillsdale, NJ: Lawrence Erlbaum.
- Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, 253, 1380-1386.
- Talamini, L. M., Meeter, M., Elvevåg, B., Murre, J. M. J., & Goldberg, T. E. (in press). Reduced parahippocampal connectivity produces schizophrenia-like memory deficits in simulated neural circuits. *Archives of General Psychiatry*, in press.
- Tallon-Baudry, C., Bertrand, O., & Fischer, C. (2001). Oscillatory synchrony between human extrastriate areas during visual short-term memory maintenance. *Journal of Neuroscience*, 21, RC177, 1-5.
- Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioral Neuroscience*, 100, 147-154.
- Tononi, G., Sporns, O., & Edelman, G. M. (1992). Reentry and the problem of integration of multiple cortical areas: Simulation of dynamic integration in the visual system. *Cerebral Cortex*, 2, 310-335.
- Tulving, E. (1972). Episodic and Semantic memory. In: E. Tulving & W. Donaldson (Eds.) *Organisation of Memory*, 381-403. New York: Academic Press.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352-373.
- Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, 4, 374-391.
- Ungerleider, L. G., Courtney, S. M., & Haxby, J. V. (1998). A neural system for human visual working memory. *Proceedings of the National Academy of Sciences, USA*, 95, 883-890.
- Villa A. E. P., & Fuster, J. M. (1992). Temporal correlates of information processing during visual short-term memory. *NeuroReport*, 3, 113-116.
- Von der Malsburg, C. (1981). *The correlation theory of brain function*. Internal report 81-2. Göttingen: Max Planck Institute for Biophysical Chemistry.
- Von der Malsburg C. (1999). The what and why of binding: the modeler's perspective. *Neuron*, 24, 95-104.
- Wagner, A. D., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Prefrontal cortex and recognition memory: functional-MRI evidence for context-dependent retrieval processes. *Brain*, 121, 1985-2002.
- Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, 411, 953-956.
- Wang X. J. (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *Journal of Neuroscience*, 19, 9587-9603.

- 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. (1987). Site fragility theory of chunking and consolidation in a distributed associative memory. In: N. W. Milgram, C. M. MacLeod, & T. C. Petit (Eds.), *Neuroplasticity, learning, and memory*. Alan R. Liss, 301-325.
- Willshaw, D.J., O.P. Buneman, & H.C. Longuet-Higgins (1969). Non-holographic associative memory. *Nature*, 222, 960-962.
- Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: Processing and representational perspectives. *Nature Reviews Neuroscience*, 4, 139-147.
- Zipser, D., & R. A. Anderson (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331, 679-684.

Figure Captions

Figure 1. Binding by co-activation, after McClelland and Rumelhart (1981). See text for explanation.

Figure 2. Binding by synchronization. (a) Neurons 1, 2, and 5 fire in synchrony and so do neurons 3, 4, and 6. (b) A connectivity pattern that could underlie the correlations shown in (a).

Figure 3. Interactive-binding. (a) Two streams of activation patterns converge onto a single layer in which three neurons receive inputs from both streams. Through a suitably calibrated inhibitory mechanism these three neurons remain active, suppressing the rest. (b) Illustration of how this mechanism can be viewed as the intersection of two sets of activated neurons.

Figure 4. Associative binding. (a) Bidirectional connections have developed between strongly activated nodes. (b) During read-out (not shown) the connected neurons will exhibit synchronous (or correlated) firing.

Figure 5. Scheme of the working memory model. In the IT module, 20 neural assemblies of 100 neurons (not shown individually) code for 20 hypothetical visual features. Assemblies are activated by input from unspecified lower areas. The figure shows the situation with 5 four-feature objects with strong (synchronizing) connections (the diamond shape configurations). The IT module also comprises small assemblies of inhibitory neurons. Inhibitory input to each object assembly equals the number of firing neurons of the other IT assemblies coding for competing features or objects. The PF module consists of a set of 20 assemblies of 50 neurons “matching” the IT module structure.

Figure 6. (a) Dynamic behavior of one assembly activated by stimulus input to IT and with active feedback from PF. Stimulus onset (at 100 ms) and offset (at 200 ms) times are marked by vertical lines. The panel shows the evolution and continuation of the average spiking activity of one IT assembly (100 interconnected nodes coding a single feature). (b) Phase segregation of IT assemblies coding for independent features. Four out of five assemblies remain active. Due to mutual inhibitory activity, the assemblies become optimally spaced in the oscillatory phase. (c) Phase segregation of objects (chunks) consisting of four interconnected assemblies. Four out of five objects are retained in terms of internally synchronized and mutually desynchronized chunks, whereas all features coding the fifth object are suppressed.

Figure 7. Associative binding in long-term memory as modeled by TraceLink. Stage 1: co-active or synchronous neurons represent the contents of an experience. Stage 2: Through Hebbian learning they have been bound indirectly into a representation via neurons in the link system. Stage 3: Direct cortico-cortical connections have developed. Stage 4: The neurons are now associatively bound through direct interconnections at a cortical level. The representation has become independent of the link system.

Figure 8. Simplified neural hierarchy of some principal areas implicated in short-term and long-term memory.

Figure 1

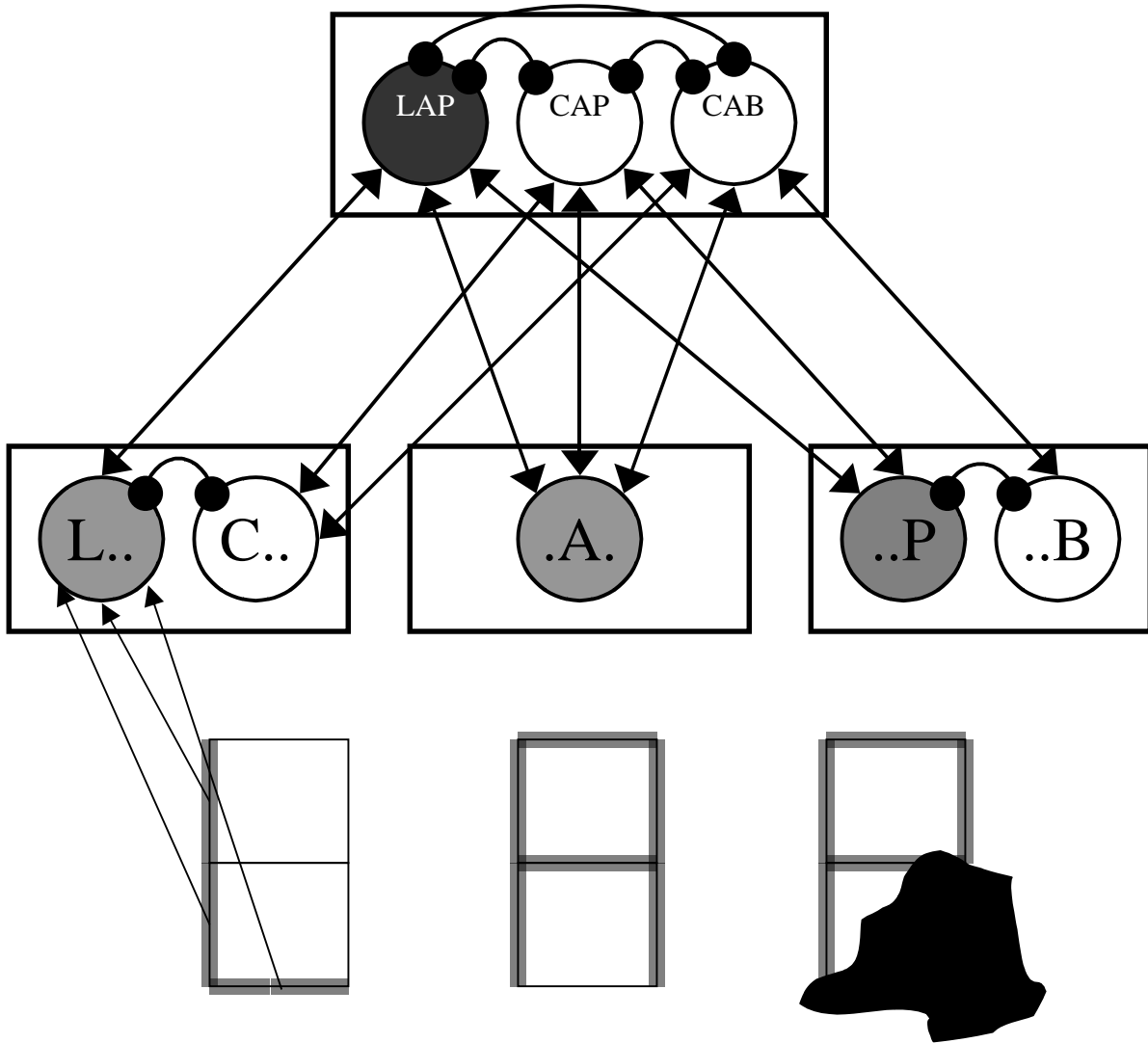


Figure 2

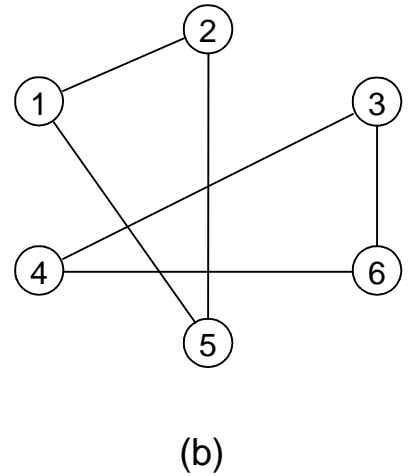
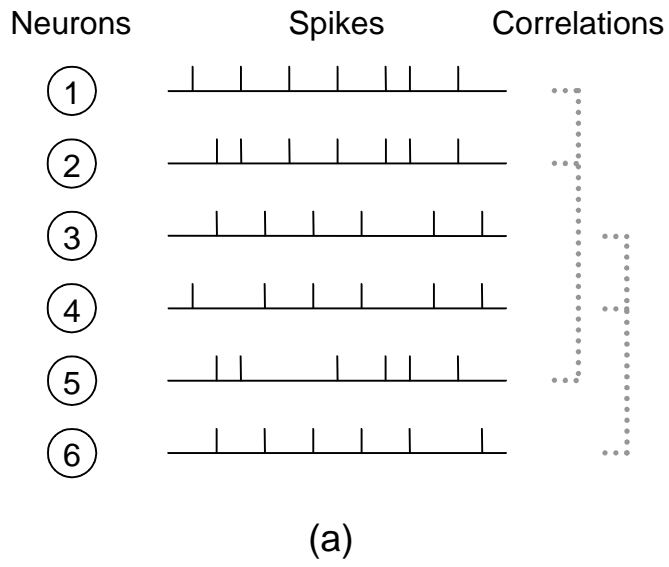
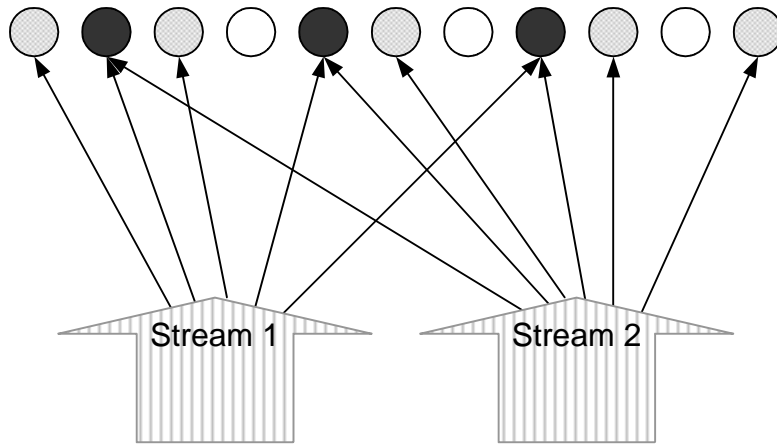
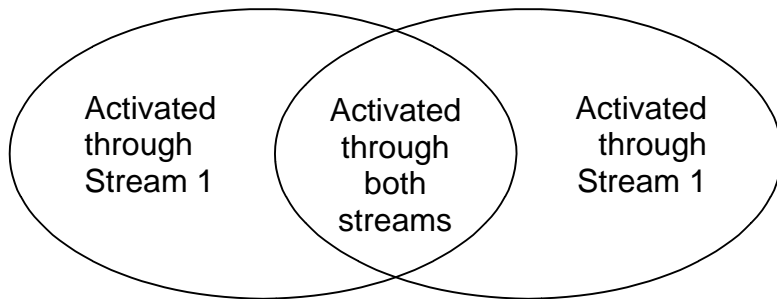


Figure 3



(a)



(b)

Figure 4

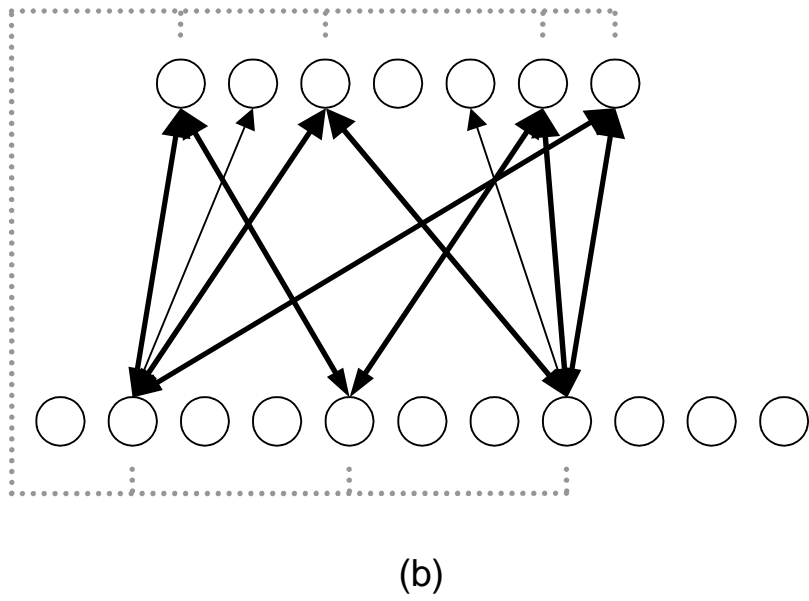
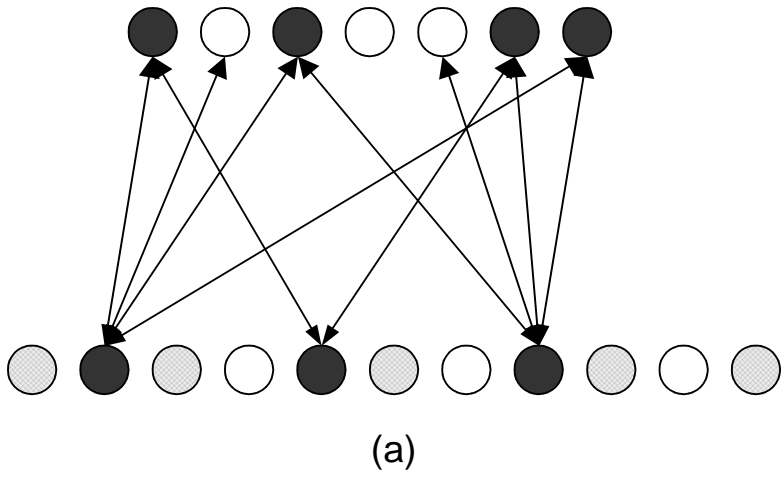


Figure 5

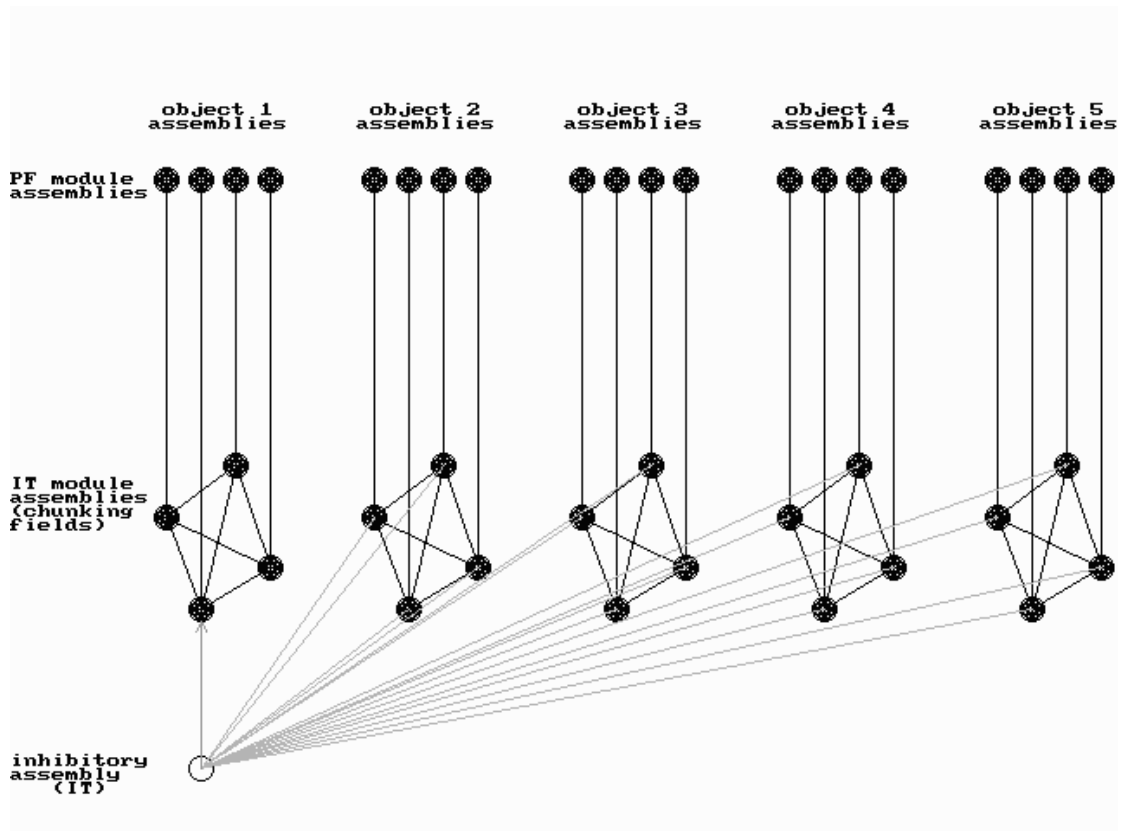


Figure 6

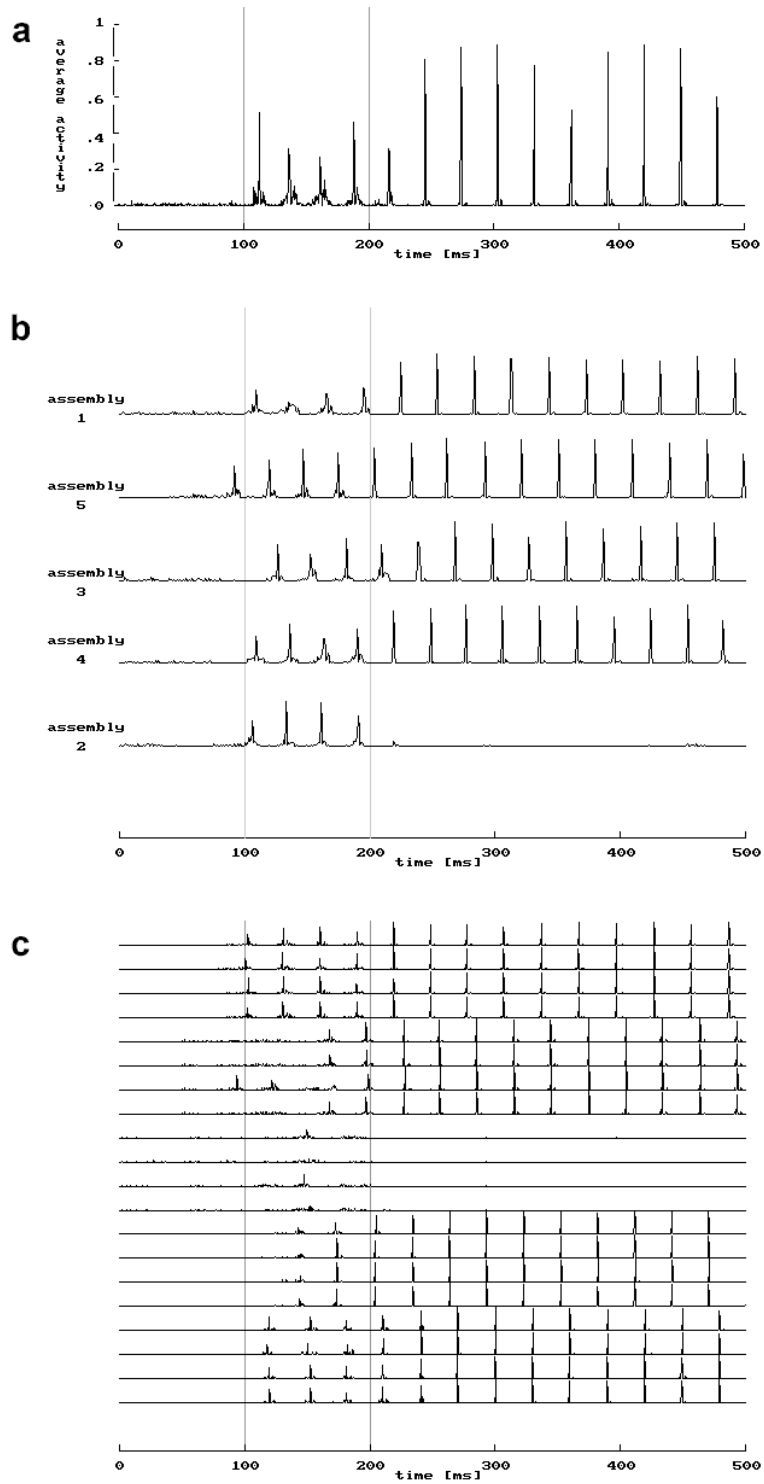


Figure 7

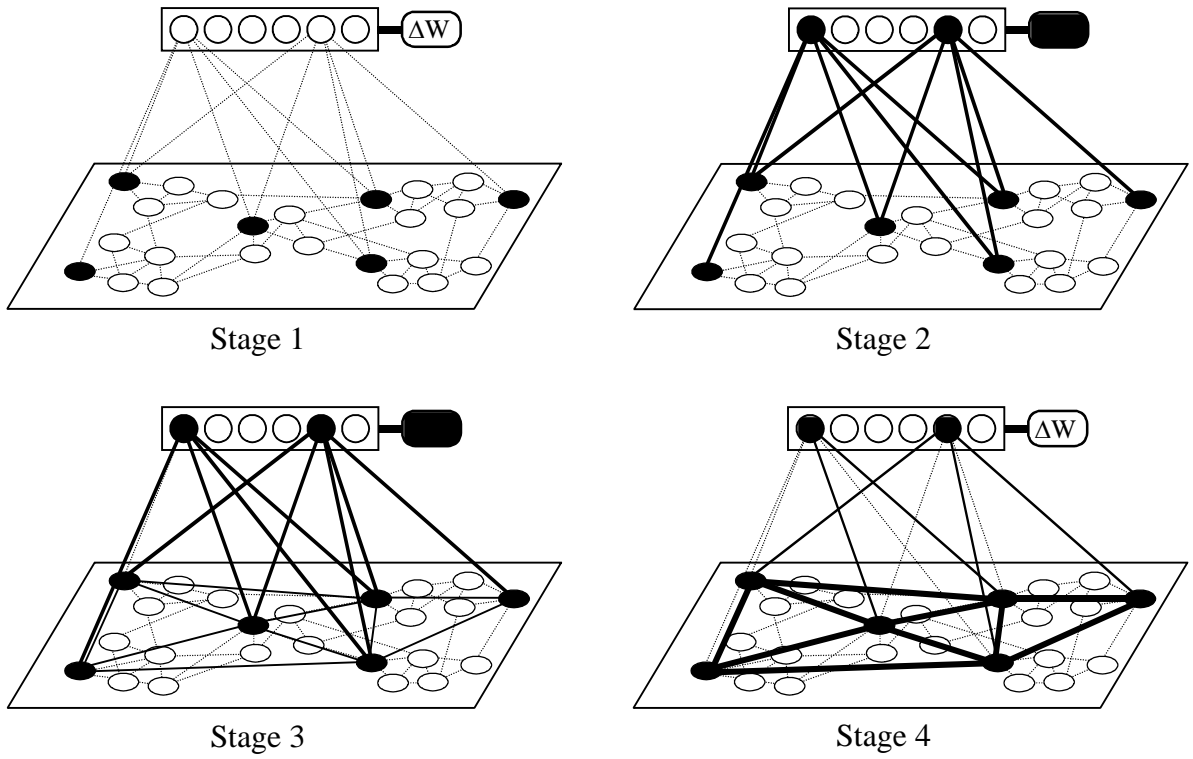


Figure 8

