

# The Computational Cognitive Neuroscience of Learning and Memory: Principles and Models

Lee I. Newman and Thad A. Polk

University of Michigan, Department of Psychology

## **1. The computational cognitive neuroscience of learning and memory: computational principles and models**

In this chapter, our objective is to provide the reader a brief introduction to learning and memory as viewed from the perspective of neural computation. The computational agenda in cognitive neuroscience focuses on two important tasks, each relevant to the study of learning and memory: to identify and understand how neural computations are carried out in the brain and to use computational simulations to better understand how these neural computations give rise to cognitive function. Our agenda is (i) to present a set of core principles of neural computation that have been derived from empirical work in neuroscience and (ii) to demonstrate how these principles can serve as building blocks for computational models that have been used to explain many aspects of learning and memory.

## **2. Computational learning: three approaches**

Computational cognitive neuroscientists often divide learning into three broad classes: supervised, reinforcement, and unsupervised learning. The three classes share the common objective of forming stored representations to be later recalled and used to guide behavior; they differ in the specific mechanisms by which stored representations are derived from

experience. We will briefly discuss<sup>1</sup> these three types of learning at a behavioral level using a concrete example: the task faced by a young child learning the name of an object visually experienced in the environment.

In a supervised learning task, the learner is provided with explicit feedback on a desired outcome. For example, a child might point to a dog and say “cat” only to be told “dog” by a parent. After a number of such episodes, this corrective feedback ultimately allows the child to form a stored representation associating the visual image of a dog with the utterance “dog.” In reinforcement learning, the learner is not explicitly corrected, but rather receives either immediate or delayed feedback in the form of rewards and/or punishments. Returning to our example, rewarding a child with a snack every time she says “dog” in the presence of a dog might reinforce the correct association and facilitate its acquisition. However, this reinforcement-driven task is typically much more difficult than its supervised counterpart because the nature of the feedback used to drive the learning process is less specific: rather than being told the correct name, the child is simply given a positive or negative reinforcement. This difficulty may be further compounded by the fact that reinforcements may not be immediate and the same reinforcement may be given for different but related behaviors (e.g., the child may receive the same snack for correctly saying “car” in the presence of a car).

Although the supervised and reinforcement-learning tasks differ in the type of feedback received by the learner (i.e., explicit correction vs. less specific rewards), they have in common the fact that experience (i.e., seeing a dog) is always accompanied by some form of behavioral feedback. In contrast, in an unsupervised learning task, the learner must make sense of experience in the absence of any feedback. For example, a child may frequently see a dog and at the same time hear the word “dog,” and ultimately come to associate the word with the object after many instances of such experience.

A great deal of work has attempted to understand how these three types of learning are instantiated in neural computations. One of the earliest proposals was the seminal work by Hebb (1949) in which he offered a neural theory of unsupervised learning, which we will discuss in the following section. Unsupervised learning in the brain has since been widely studied, for example, in the context of research on the perceptual learning in sensory cortices (Fahle & Poggio, 2002; Hinton, G.E. & Sejnowski, 1999; Rolls & Deco, 2002) and the development of topographic cortical maps

---

<sup>1</sup> For more detailed and mathematical treatment of these three classes of learning, see Dayan & Abbott (2001), Hastie, Tibshirani, & Friedman (2001), and Sutton & Barto (1998).

(Obermayer & Sejnowski, 2001). Significant progress has also been made in understanding the neural basis of reinforcement learning in the mesolimbic dopamine system (Berridge & Robinson, 1998; Dayan & Balleine, 2002; O'Doherty, 2004; Schultz, 2006). The neural basis of supervised learning is less well-understood at present, but supervised learning processes can be implemented with some of the same computational mechanisms that support unsupervised learning, and moreover, supervisory feedback can be provided to one population of neurons by the output of others (Dayan & Abbott, 2001). Therefore, it may very well be the case that although supervised and unsupervised learning differ in their informational content at the level of behavioral tasks, at a neural level they may be more similar than different. In this chapter, we have chosen to focus on unsupervised models of learning and memory, in part because their grounding in neurobiology is better understood, and in part because these models are amenable to derivation from established principles of neural computation to which we now turn our attention.

### **3. Principles of neural computation**

Empirical work in neuroscience has led to models of learning and memory at multiple levels ranging from molecular models of synaptic modification (Destexhe, Mainen, & Sejnowski, 1994; Xie, Liaw, Baudry, & Berger, 1997; Zador, Koch, & Brown, 1990), to models of how a population of neurons can encode the shape of an object (Logothetis, Pauls, & Poggio, 1995; Poggio & Edelman, 1990), to models of how multiple brain areas cooperate in support of memory systems (McClelland, McNaughton, & O'Reilly, 1995). Although this broad range of models differs in the detail with which they capture neurobiological constraints, they share a common set of neural building blocks. In this section, we introduce the reader to a core set of these building blocks. For comprehensive and mathematical treatments of the core principles of computational neuroscience, we direct the interested reader to Abbott and Sejnowski (1999), Arbib (2003), Dayan and Abbott (2001), Gerstner and Kistler (2002), Gluck and Myers (2001), Koch and Davis (1994), Hinton and Sejnowski (1999), Obermayer and Sejnowski (2001), and O'Reilly and Munakata (2000).

#### **3.1. PRINCIPLE 1: DISTRIBUTED, OVERLAPPING REPRESENTATIONS**

Throughout cortex, information is represented by patterns of activity across populations of neuronal processing elements as revealed by both electrophysiological recordings (e.g., Georgopoulos, Schwartz, & Kettner, 1986) and functional neuroimaging (e.g., Ishai, Ungerleider,

Martin, & Haxby, 2000). These patterns of activity are known as *distributed* representations because the burden of representing information is shared across many neuronal elements. Distributed representations can be contrasted with *localist* representations in which information is represented in the extreme by a single neuron, often referred to as a grandmother cell (a single cell capable of representing complex information, such as the face of one's grandmother). It is important to note that the degree of distribution of representations and the degree of localization of representations in the brain are not identical concepts. A representation can be distributed over a large number of neurons in a single, small region of cortex (distributed and localized), or more widely distributed over neurons spanning disparate cortical regions (distributed and not localized). Closely related to the distributed nature of neural representations is the fact that distributed cortical representations also tend to be overlapping: two similar patterns of information typically share many neuronal processing elements, while two highly dissimilar patterns of information share few processing elements.

The distributed and overlapping nature of representations has several important computational implications for learning and memory. First, distributed representations are robust to partial damage because patterns are represented over a large number of neurons. Second, it can be shown mathematically that distributed representations are efficient because they can store a large number of mnemonic patterns relative to localist coding schemes (Rolls & Tovee, 1995; Rolls, Treves, & Tovee, 1997; Rumelhart, McClelland, & PDP ResearchGroup, 1986, Chapter 3). That the brain makes use of this efficiency has been shown empirically (e.g., Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Rolls & Tovee, 1995; Rolls et al., 1997). Third, the tendency of distributed representations to share processing elements provides a natural computational basis for generalization and inference. For example, if a neural representation of the concept *dog* is associated with a representation of the concept *furry*, then there is a neural basis for inferring the association of *furriness* for other objects that are similar to *dog*.

### 3.2. PRINCIPLE 2: RECURRENT CONNECTIVITY

It is well-established that connectivity in cortex is massively recurrent, at multiple levels of processing (Felleman & McClendon, 1991; Fuster, 1995; Kandel, Schwartz, & Jessell, 2000). Within a cortical network, two nearby neurons may be recurrently connected such that they can mutually excite or inhibit each another. However, neurons within a network need not be immediate neighbors to be recurrently connected; activity can spread from one neuron through a chain of many other neurons, ultimately

returning to affect the original neuron. Neurons also need not be in the same network to be recurrently connected. Neural pathways between two networks are typically recurrent, thus allowing entire patterns of activity across networks to mutually affect one another.

There are several important computational implications of recurrent connectivity. First, recurrence can lead to reverberatory activity within a network of neurons, allowing the network to sustain a pattern of activity even when input to the network is no longer present. Second, recurrent connectivity between networks allows the flow of information processing in the brain to proceed not only in a bottom-up manner but also in a top-down manner, thereby enabling higher-order representations (e.g., prior knowledge and expectations) to affect lower-order representations (e.g., visual object identity).

### 3.3. PRINCIPLE 3: HEBBIAN LEARNING

As mentioned previously, Hebb proposed one of the first neural theories of learning (Hebb, 1949). Hebb's idea was that if two connected neurons are frequently active at the same time, some form of physiological change in their connectivity (learning) could render them more likely to be coactive in the future, thus providing a physiological basis for memory. This form of unsupervised learning is frequently captured by the expression "neurons that fire together, wire together." Hebb's principle of learning received empirical support when evidence for synaptic strengthening (called long-term potentiation or LTP) was discovered in neural circuits of the mollusk Aplysia (Castellucci & Kandel, 1976) and in hippocampal neurons of the rabbit (Bliss & Gardner-Medwin, 1973; Bliss & Lomo, 1973). The principle of Hebbian learning operating at the level of two neurons is deceptively simple and yet we cannot overstate its importance to the broader topic of memory formation. By generalizing Hebb's learning rule to networks of neurons, we will show in the next section how these networks are capable of reproducing previously experienced patterns of activity after repeated exposure to the same or similar inputs. The principle of Hebbian learning thus provides an explicit account of how a *pattern of activity* in a network of neurons can be stored in a *pattern of synaptic connections*, thereby serving as a neural substrate of memory.

### 3.4. PRINCIPLE 4: COOPERATIVE AND COMPETITIVE ACTIVATION DYNAMICS

Hebbian learning is a powerful mechanism, but operating in conjunction with recurrent connectivity without other constraints would be problematic for the formation of memory. The problem is that because neurons are highly interconnected, excitatory activity in a few neurons tends to

spread to neighboring neurons, and to their neighbors, and so on. This problem is compounded by the presence of recurrent connectivity, which allows activity to reverberate in the network creating mutually reinforcing activity (cooperation) analogous to the situation when sound from a P.A. system reverberates back to a microphone resulting in the familiar screech of feedback. In the presence of runaway cooperative activity, Hebbian learning would ultimately lead to indiscriminate reinforcement of all synapses in a network (all neurons firing together would get wired together), making it impossible to learn the meaningful patterns necessary for memory formation. Fortunately, competitive processes serve as a balancing force that prevents such a result from occurring (Freund & Buzsaki, 1996; Gibson, Beierlein, & Connors, 1999). Neural competition is implemented via inhibitory synapses that allow active neurons to inhibit less active neurons. As activity spreads through a network, the more active neurons tend to increasingly excite each other (cooperation) and at the same time increasingly inhibit less active neurons (competition). After a period of time, it is only the “winners” of these neural competitions that remain active and that are selectively reinforced by Hebbian learning, making them more likely to win again when the same pattern is again presented to the network. In this way, neurons become specialized: they respond selectively to some patterns of input, but not others.

### 3.5. PRINCIPLE 5: SPATIAL CONSTRAINTS

Studies of the structure of sensory cortices have revealed a columnar organization in which columns of neurons are connected via local excitatory connections and slightly longer-range inhibitory connections (Fisken, Garey, & Powell, 1975; Gilbert, 1992; Hubel & Wiesel, 1962; Mountcastle, 1957, 1997; Szentagothai, 1975). As a result of this differing spatial extent of excitatory and inhibitory connections and the processes of cooperation and competition discussed in the prior section, activity in sensory cortices tends to occur in spatially localized bumps of activity. This fact, combined with the principle of Hebbian learning, implies that learning via synaptic modifications also tends to be spatially localized. This spatial constraint on learning has important implications for the structure of memory, a topic that we address in our discussion of self-organizing maps.

Taken together, the first four principles – distributed and overlapping representation, recurrent connectivity, Hebbian learning, and neural cooperation and competition – provide a foundation for a computational framework of learning and memory. These principles provide an explicit account of how information is represented by patterns of activity in networks of neurons, and how these patterns are transformed into synaptic

connections that serve as a neural substrate of memory. In the next section, we derive from the first four building blocks an important class of computational models known as attractor networks. In the following section, we add the fifth principle and derive a second class of models known as self-organizing maps.

#### 4. Modeling learning and memory with attractor networks

The attractor network model was originally proposed by Hopfield (1982), and a range of variants have subsequently been developed, for example (Ackley, Hinton, & Sejnowski, 1985; Hopfield & Tank, 1986). This type of models is also commonly referred to as autoassociative network or Hopfield network.

##### 4.1. OVERVIEW OF THE MODEL

###### 4.1.1. Architecture

The architecture of an attractor network is shown in Fig. 1a. The network consists of set of neural processing units. The activity of each unit at a given time is specified by a numerical value that serves as an abstract representation of the unit's firing rate. In the attractor net formulation presented here, we consider binary-valued units that are either on (+1) or off (-1). Attractor nets can also be built with continuous-valued units and/or with units that take on only positive values. Units in the network are linked to every other unit via recurrent connections (Principle 2). Associated with each connection is a modifiable and persistent synaptic weight that represents the strength of the connection.

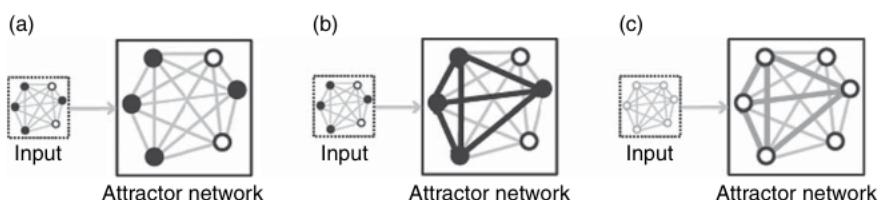


Fig. 1. Attractor network architecture and learning. (a) An attractor net consists of a set of neural units connected by recurrent synaptic connections (gray lines). Presentation of an input pattern forces some units to become active (filled circles) while others remain inactive (open circles). (b) Hebbian learning leads to a strengthening of connections between coactive units (black lines). (c) When input activity is no longer present, the pattern is now stored, and can be later recalled, via the synaptic connections in the network.

#### 4.1.2. Processing

When an external input pattern is presented to the network, each unit in the network is activated based on the level of activation in the corresponding unit in the input pattern. As shown in Fig. 1a, presentation of an input pattern leads to a distributed pattern of activity (Principle 1) in which some units in the network become active (filled circles) while others remain inactive (open circles). In addition to external input, each unit also receives activation from other units in the network via recurrent connections (Principle 2). When an external input is present, it forces a pattern of activity on the network. However, once the external input is removed, the activity of each unit becomes entirely determined by the interaction of the excitatory and inhibitory influence of other units in the network (Principle 4). Units receiving more inhibition than excitation become inactive as neural competition plays out, and units that receive more excitation than inhibition ultimately become the neural “winners” and thus are able to maintain their activity.

#### 4.1.3. Learning

How does an attractor network come to serve as a basis for memory? When a pattern of input activity (+1, -1 values) is imposed on the network, the synaptic connection weights between every pair of neurons are updated based on their coactivity (Principle 3), with the new weight typically being the product of the activities of the two neurons. The connection weights between any two active neurons are strengthened (Fig. 1b) and as a result, in the absence of external input, the input pattern has been stored in the strengthened connection weights of the network (Fig. 1c). Connections between pairs of neurons that are not coactive are weakened (not shown).

## 4.2. PROPERTIES OF THE MODEL

#### 4.2.1. Pattern completion

Just as in biological neural networks, attractor networks store previously experienced patterns of activity in the strength of their synaptic connections. And as in real neural networks, the combination of recurrent connectivity and cooperative and competitive processes allows these simulated networks to recall stored memories based on partial cues. For example, if a noisy or a degraded version of a previously experienced input is presented to a network, it will activate a subset of units in the previously stored pattern (Fig. 2a). With the input removed, the activation of the network will evolve as excitatory and inhibitory activity spreads across the previously strengthened recurrent connections (Fig. 2b). Although

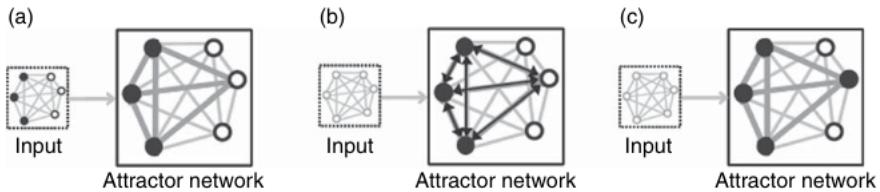


Fig. 2. Pattern completion. (a) A partial version of a previously learned pattern (thick gray lines) is presented to the network and leads activity of a subset of units (black circles). (b) With the input removed, activity spreads across previously strengthened connections (black lines). (c) After activity has settled, the network is able to recall the previously stored pattern (Fig. 1).

beyond the scope of this chapter, it can be shown mathematically that the network will settle into a previously stored pattern of activity that is similar to the novel pattern presented as input (Hopfield, 1982). In this way, complete patterns can be retrieved from memory based only on partial input (Fig. 2c).

#### 4.2.2. Pattern generalization

The ability to generalize information is a critical aspect of cognition, allowing novel objects to be categorized based on their similarity to previously experienced objects. There is a large body of empirical evidence from single-unit recording studies (e.g., Tanaka, Saito, Fukada, & Moriya, 1991), and functional imaging studies (e.g., Ishai et al., 2000) showing that exposure to similar stimuli elicits similar patterns of neural activity in the cortex. Attractor networks exhibit the same kind of generalization as a result of their activation dynamics and the overlap in their stored mnemonic representations. In fact, the generalization property of these networks is directly related to their ability to complete patterns. Consider the network shown in Fig. 3a that receives activation from a novel input. Two patterns have been previously stored in the network. The first pattern is shown by the thick gray connections and the second by the dotted gray connections. When the input is removed, activation spreads across the previously strengthened recurrent connections (Fig. 3b). The two previously stored patterns compete, but because the first input pattern is more similar to the novel input pattern (they share more units), this first pattern receives more cooperative excitation than the second pattern (and is also able to impose greater inhibition on it). The novel pattern is “attracted” to the most similar pattern that has already been stored (thus the name *attractor* network), and in this way, the network is able to generalize on past experience (Fig. 3c).

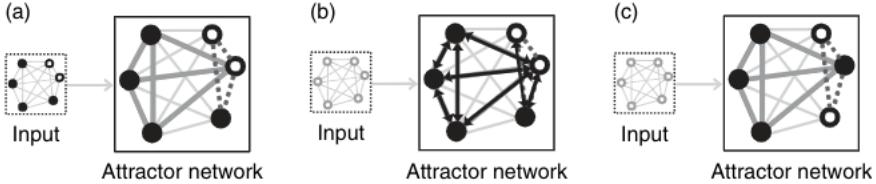


Fig. 3. Pattern generalization. (a) A novel pattern is presented to a network that contains two previously stored patterns (thick gray lines, dotted gray lines). (b) With the input removed, activity spreads across previously strengthened connections, and the two previously stored patterns compete. (c) After activity has settled, the network produces the pattern that is most similar to the novel input.

#### 4.2.3. Similarity-based interference

Figure 4a shows a novel pattern (black units and connections) being learned in a network with a different previously stored pattern (gray connections). The new pattern shares no units with the previously stored pattern and therefore the connections strengthened in learning the new pattern (black connections) will not interfere with the previously stored pattern (gray connections). Now consider (Fig. 4b) what happens when the new pattern shares units with the previously stored pattern such that the two representations are overlapping in addition to being distributed (Principle 1). This representational overlap has important consequences for the network. Unit A is part of both patterns, while unit B is a member of the previously stored pattern and not the input pattern. When unit A becomes active, as a member of the first pattern it should tend to cooperatively

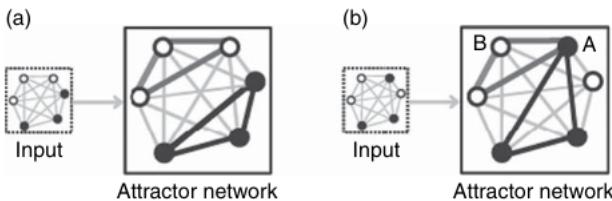


Fig. 4. Similarity-based interference. (a) A novel pattern (black units and connections) is presented to a network with a previously stored pattern (gray connections). These two patterns have no units in common, so new learning does not interfere with previous learning. (b) The novel and previously stored pattern have unit A in common. Unit A will excite unit B based on the previously strengthened excitatory connection, but in learning the pattern, the connection between units A and B should be inhibitory. Learning overlapping patterns thus forces a compromise on the strength of connections.

excite unit B, but as a member of the new pattern to be learned it should result in a connection that competitively inhibits unit B. This poses a problem in that the connection between unit A and B cannot achieve both goals simultaneously. The learning of new patterns thus can interfere with previously stored patterns, resulting in connection weights that must take on values that reflect the relationship between units across all of the input patterns learned by the network. This interference has both desirable and undesirable consequences. On the positive side, a network that learns a set of weights that are a neural compromise based on many experienced patterns is beneficial in that it supports generalization. On the negative side, interference can result in a partial input pattern being recalled as a similar, but incorrect pattern. This type of similarity-based interference is a common behavioral phenomenon for which the attractor net offers a neural explanation (e.g., Chappell & Humphreys, 1994; Jones & Polk, 2002). Lastly, the compromise that must be achieved by the network in learning a set of weights imposes a capacity constraint on the network limiting the number of patterns that can be accurately stored (Abumostafa & St. Jacques, 1985; McEliece, Posner, Rodemich, & Venkatesh, 1987; Treves & Rolls, 1991).

#### *4.2.4. Short-term memory*

As mentioned previously, reverberatory activity over recurrent connections allows an attractor network to maintain a pattern of activity after an input is no longer present. Thus, in addition to serving as a model of long-term memory, attractor networks also provide a computational account of how short-term memory might be instantiated in the brain (Zipser, Kehoe, Littlewort, & Fuster, 1993).

#### *4.2.5. Resistance to damage*

As a result of their distributed representations, recurrent connectivity, and cooperative and competitive activation dynamics, the recall performance of attractor networks is resistant to partial network damage. In a lesioned network, an external input may not result in a complete pattern of activity, but often a partial pattern from one network provides a sufficient input to another network for it to successfully recall the intended pattern (via pattern completion).

### **4.3. APPLICATIONS OF THE MODEL**

The unique properties of attractor networks have been used to simulate a wide range of learning and memory phenomena, for example, working memory (Amit & Brunel, 1997; Compte, Brunel, Goldman-Rakic, & Wang,

2000; Deco & Rolls, 2003; Farrell & Lewandowsky, 2002; Jones & Polk, 2002; Miller, Brody, Romo, & Wang, 2003; Zipser et al., 1993), hippocampally based episodic memory (Byrne, Becker, & Burgess, 2007; Gluck & Myers, 2001; Hasselmo, Wyble, & Wallenstein, 1996; Stringer, Rolls, & Trappenberg, 2004), similarity judgments in semantic memory (Cree, McRae, & McNorgan, 1999; Polk, Behensky, Gonzalez, & Smith, 2002), word reading (Plaut, McClelland, Seidenberg, & Patterson, 1996), and motor skill learning (Newell, Liu, & Mayer-Kress, 2001). Because they can be artificially lesioned, attractor networks have also provided a valuable means for simulating and understanding the patterns of deficits exhibited by patients suffering from different types of brain damage and dysfunction, for example, acquired dyslexia (Hinton, G. E. & Shallice, 1991), schizophrenia (Hoffman & McGlashan, 2001), dysexecutive syndrome (Polk, Simen, Lewis, & Freedman, 2002), and visual agnosia (Brunel, 1993).

## **5. Modeling neural learning and memory with self-organizing maps**

Although attractors have proven useful as tools for understanding the neurocomputational basis of many cognitive phenomena and for generating novel research hypotheses, they are not fully consistent with what is currently known about learning and memory in sensory cortices. Anatomical and physiological studies of sensory cortex have revealed that topography is an important principle of neural organization (Kandel et al., 2000). Topography refers to a form of neural organization in which nearby neurons in the cortex tend to respond to similar inputs. For example, most of the visual cortex is organized retinotopically such that neighboring neurons in cortex respond to stimulation of nearby regions of the retina (e.g., Sereno et al., 1995; Tootell, Switkes, Silverman, & Hamilton, 1988; Tusa, Palmer, & Rosenquist, 1978; Van Essen & Gallant, 1994). Attractor models do not incorporate the spatial constraints outlined in Principle 5, and as a result this class of models cannot explain how topographically organized representations are learned. We, therefore, turn our focus to a second class of computational models called self organizing maps (SOMs). The SOM model was originally proposed by Kohonen (1982a; 1982b) and subsequently many variants have been developed (e.g., Haese & Goodhill, 2001; Heskes, 2001; Kohonen, Kaski, & Lappalainen, 1997; Koikkalainen & Oja, 1990; Luttrell, 1988, 1989; Tereshko & Allinson, 2002) and much theoretical work has been conducted to better understand this class of model analytically (e.g., Cottrell, Fort, & Pages, 1998; Heskes, 2001; Kohonen, 1993).

## 5.1. OVERVIEW OF THE MODEL

### 5.1.1. Architecture

The architecture of an SOM is shown in Fig. 5. The SOM consists of a grid of neural processing units arranged in a two-dimensional map (a  $5 \times 5$  map in Fig. 5) that is a computational abstraction of a locally connected population of neurons in a contiguous region of cortical tissue. Unlike the attractor network in which units are explicitly connected, in the SOM, connectivity is implicit in the computational equations that determine how units become active and how they learn; we will discuss in a later section how this pattern of connectivity can be derived from a subset of our five computational principles. Units within the map are indexed based on their spatial location, and each unit has associated with it a modifiable weight vector that specifies the input pattern for which the unit is best tuned (corresponding to the unit's receptive field). The weight vector for a given unit is a computational abstraction of a set of synaptic connections between the unit and all units (in another map) from which it receives input.

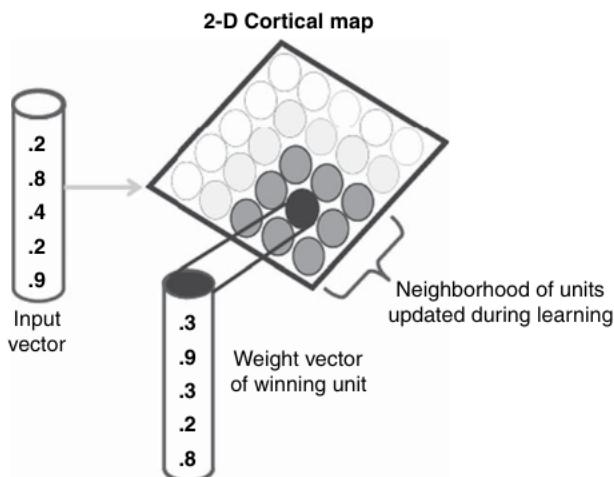


Fig. 5. Self-organizing map (SOM) architecture and learning. The SOM is a two-dimensional map of neural units located on a spatial grid. Each unit is represented by a weight vector representing its preferred features (here five features), also known as its receptive field. When an input vector is presented to the network, similar units become active (black and dark gray units) while dissimilar units remain relatively inactive (light gray and white units). Learning occurs when the weight vectors of the most active cell (black) and its neighbors (dark gray units) are modified so that they become more similar to the input.

### 5.1.2. Processing

When an input pattern is presented to the SOM, each unit in the map has access to the input. The activity of each unit is determined by comparing the similarity between a unit's weight vector and the current input pattern, with similarity typically measured as the inverse of some metric of vector distance such as Euclidean distance. If the input pattern exactly matches a unit's weight vector, this unit will fire at its maximum rate (typically set to 1); poor matches between the input and a unit's weight vector result in little or no activity in that unit. The activity of the map is thus a distributed representation (Principle 1) in which the input is represented as the activity pattern induced in the map as shown in Fig. 5.<sup>2</sup> Furthermore, because units are represented by weight vectors, nearby units can (and typically do) have similarly valued elements (e.g., [0.20, 0.40, 0.60, 0.10, 0.70] and [0.20, 0.46, 0.63, 0.11, 0.70]); SOM representations are therefore also overlapping in addition to being distributed (Principle 1).

### 5.1.3. Learning

Units within the map compete to represent an input pattern. This competitive process is carried out computationally by simply searching the map for the unit with maximum activity; this “winning” unit (black unit in Fig. 5) will be the one whose weight vector is most similar to the input. Learning is accomplished by modifying the weight vector of this winning unit so that its values are more similar to the input pattern, thereby making this unit more likely to win again if the same input is present. Critically, the weight vectors of other units in close spatial proximity to the winning unit (dark gray units in Fig. 5) are also updated (Principle 5), with the magnitude of the updates typically proportional to the distance between each unit and the winning unit. As a result of this spatial constraint on learning, with experience, nearby neurons in the map come to have similar weight vectors and thus come to represent similar input patterns. This spatially constrained learning leads to a process known as self-organization from which topographic representations develop.

Armed with this knowledge of SOM function, we can now explain how Principles 2, 3, and 4 are implicitly present in the model. As mentioned earlier in this chapter, the spatial constraint on learning is based on the

---

<sup>2</sup> In some formulations of the SOM model, the most active unit fully inhibits the activity of all other units thus resulting in what would be considered a purely localist representation rather than a distributed representation. However, in determining which unit is most active, the activity of all units must be compared and therefore in these models, the underlying representation is implicitly distributed.

finding that in sensory cortex, the spatial extent of recurrent excitatory connections is less than the spatial extent of competitive inhibitory connections. As a result, cooperative activity (Principle 4) across recurrent connections (Principle 2) leads to spatially localized “bumps” of activity in cortical tissue. As activity spreads through the network, these bumps of activity compete (Principle 4) via inhibitory connections. As in attractor networks, after the competition has played out, only a subset of units in the network will remain active; however, unlike attractor networks in which winning neurons are typically scattered throughout the network, in a SOM, winning units are confined to a spatially localized region of the cortex—a winning “bump” of activity (black and dark gray units in Fig. 5). Although these activation dynamics are not directly instantiated in the SOM model, they are implicitly instantiated in the spatial constraint imposed on learning. The winning bump of activity corresponds directly to the activities of the winning unit (identified in the learning process) and its neighbors. What about Hebbian learning (Principle 3)? As a reminder, Hebbian learning occurs when the connection between coactive neurons is reinforced. Modifying the weights of the winning unit and its neighbors so that they are more similar to the input corresponds directly to Hebbian reinforcement of the implicit connections between these active units and the input pattern. Therefore, although not immediately obvious in its formulation, the SOM model can, in fact, be derived from our five core principles of neural computation.

## 5.2. PROPERTIES OF THE MODEL

Self-organizing maps share several important properties exhibited by attractor networks and also possess several unique properties.

### 5.2.1. *Pattern generalization*

As in the attractor model, SOMs have an ability to generalize based on prior learning. If presented with a novel input, the unit most similar to this input will become partially active as will its neighboring units. Thus, although a map may never have experienced the input, it is able to produce a best-guess response based on prior experience.

### 5.2.2. *Similarity-based interference*

In the SOM model, learning updates are imposed on a neighborhood of units. If a new input pattern is similar to a previously learned input pattern, it will tend to activate a similar neighborhood of units. As a result, the weight updates made in response to the new pattern will modify the weights of previously updated units. This same similarity-based interference was present in the attractor model, and results in the same set of benefits and problems, which we will not repeat here.

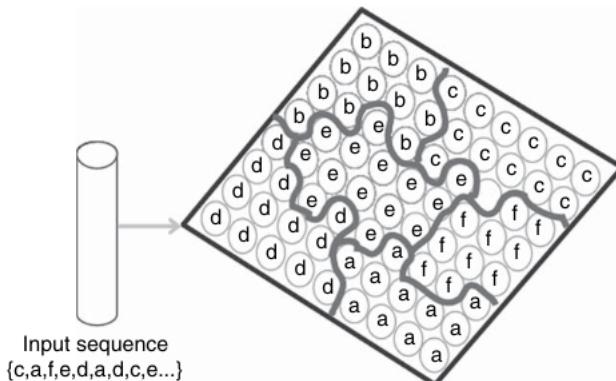


Fig. 6. Self-organized topographic organization. Although experiencing a sequence of different inputs, the SOM identifies statistical clusters based on the features of these inputs. After learning is complete, the map has self-organized such that there are spatial neighborhoods of units (labeled a, b, c, d, e, and f) that have similar weight vectors and therefore will respond to similar types of inputs.

### 5.2.3. Resistance to damage

Although an SOM does not have the ability to complete partial input patterns in the same way as an attractor network, it is nevertheless resistant to damage for a different reason. If a set of units that would normally respond to a particular input is damaged, they are no longer able to respond. However, because nearby neurons in an SOM have similar weight vectors, the winning unit will tend to be an intact unit that is located near the region where the map response should have occurred, and therefore this unit is likely to provide reasonable match to the input.

### 5.2.4. Clustering via emergent topography

As mentioned above, imposing a spatial constraint on learning leads to a self-organizing process in which neurons in nearby regions of a map learn to represent similar types of inputs. As a result, spatially organized representations, or topographies, naturally emerge from the learning process. Self-organizing maps are thus able to find and extract statistical clusters of features from a set of input patterns and to store these clusters in the form of a spatial topography (Fig. 6).

## 5.3. APPLICATIONS OF THE MODEL

Self-organizing maps have served an important role in computational neuroscience by allowing researchers to explicitly test theories of how topographic sensory representations are structured and learned in the brain.

Much of this work has been done in the study of visual cortex where simulations using SOMs have been shown to accurately reproduce the same types of topographies (e.g., orientation, ocular dominance, color blobs, movement direction) found experimentally in the cortex (Barrow, Bray, & Budd, 1996; Carreira-Perpinan, Lister, & Goodhill, 2005; Goodhill, 1993; Goodhill & Willshaw, 1994; Olson & Grossberg, 1998; Sirosh & Miikkulainen, 1997; Sit & Miikkulainen, 2006). At a behavioral level, SOMs have been used to model and help explain a wide range of learning and memory phenomena including categorical speech perception (Guenther & Gjaja, 1996), lexical development (Li, Farkas, & MacWhinney, 2004), and category learning and object recognition (Bradski & Grossberg, 1995; Carpenter, Grossberg, & Rosen, 1991; Newman & Polk, 2007; Polk & Farah, 1998; Ritter & Kohonen, 1989). Self-organizing maps have also been used to help understand learning and memory impairments such as dyslexia (Miikkulainen, 1997; Tuckova & Zetocha, 2006) and noun–verb naming impairments (Vinson & Vigliocco, 2002).

## 6. Conclusions

Experimental work in the neurosciences has produced a wealth of detailed knowledge about neural connectivity, neural processing, and neural representation. In this chapter, we hope to have demonstrated to the reader how knowledge of neurobiology can be directly translated into explicit computational principles; how these principles can serve as building blocks for constructing computational models; and how these computational models can help provide neural explanations of phenomena associated with learning and memory. We have, out of necessity, focused on a small set of computational principles, on only two classes of models, and only a few representative phenomena that these models help to explain. Despite this restricted focus, we hope the reader is left with a compelling sense of how the computational perspective in cognitive neuroscience can inform our understanding of learning and memory.

## References

- Abbott, L., & Sejnowski, T. J. (Eds.). (1999). *Neural codes and distributed representations : foundations of neural computation*. Cambridge, Mass.: MIT Press.
- Abumostafa, Y. S., & Stjacques, J. M. (1985). Information capacity of the Hopfield model. *IEEE Transactions on Information Theory*, 31(4), 461–464.

- Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, 9(1), 147–169.
- Amit, D. J., & Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, 7(3), 237–252.
- Arbib, M. A. (2003). *The Handbook of brain theory and neural networks* (2nd edition). Cambridge, Mass.: MIT Press.
- Barrow, H. G., Bray, A. J., & Budd, J. M. L. (1996). A self-organizing model of “color blob” formation. *Neural Computation*, 8(7), 1427–1448.
- Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, 28(3), 309–369.
- Bliss, T. V. P., & Gardner-Medwin, A. R. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology-London*, 232(2), 357–374.
- Bliss, T. V. P., & Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology-London*, 232(2), 331–356.
- Bradski, G., & Grossberg, S. (1995). Fast-learning VIEWNET architectures for recognizing three-dimensional objects from multiple two-dimensional views. *Neural Networks*, 8(7–8), 1053–1080.
- Brunel, N. (1993). Effect of synapse dilution on the memory retrieval in structured attractor neural networks. *Journal De Physique I*, 3(8), 1693–1715.
- Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychol Rev*, 114(2), 340–375.
- Carpenter, G. A., Grossberg, S., & Rosen, D. B. (1991). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, 4(6), 759–771.
- Carreira-Perpinan, M. A., Lister, R. J., & Goodhill, G. J. (2005). A computational model for the development of multiple maps in primary visual cortex. *Cerebral Cortex*, 15(8), 1222–1233.
- Castellucci, V., & Kandel, E. R. (1976). Presynaptic facilitation as a mechanism for behavioral sensitization in aplysia. *Science*, 194(4270), 1176–1178.
- Chappell, M., & Humphreys, M. S. (1994). An autoassociative neural-network for sparse representations – analysis and application to models of recognition and cued-recall. *Psychological Review*, 101(1), 103–128.
- Compte, A., Brunel, N., Goldman-Rakic, P. S., & Wang, X. J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, 10(9), 910–923.
- Cottrell, M., Fort, J. C., & Pages, G. (1998). Theoretical aspects of the SOM algorithm. *Neurocomputing*, 21(1–3), 119–138.
- Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, 23(3), 371–414.

- Dayan, P., & Abbott, L. F. (2001). Theoretical neuroscience : computational and mathematical modeling of neural systems. Cambridge, Mass.: Massachusetts Institute of Technology Press.
- Dayan, P., & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron*, 36(2), 285–298.
- 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(8), 2374–2390.
- Destexhe, A., Mainen, Z. F., & Sejnowski, T. J. (1994). Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism. *J Comput Neurosci*, 1(3), 195–230.
- Fahle, M., & Poggio, T. (2002). Perceptual learning. Cambridge, Mass.: MIT Press.
- Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review*, 9(1), 59–79.
- Felleman, D. J., & McClendon, E. (1991). Cortical connections of posterior inferotemporal cortex of Macaque monkeys. *Investigative ophthalmology & visual science*, 32(4), 1036–1036.
- Fisken, R. A., Garey, L. J., & Powell, T. P. S. (1975). Intrinsic, association and commissural connections of area 17 of visual-cortex. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 272(919), 487–536.
- Freund, T. F., & Buzsaki, G. (1996). Interneurons of the hippocampus. *Hippocampus*, 6(4), 347–470.
- Fuster, J. M. (1995). Memory in the cerebral cortex : an empirical approach to neural networks in the human and nonhuman primate. Cambridge, Mass.: MIT Press.
- Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233(4771), 1416–1419.
- Gerstner, W., & Kistler, W. M. (2002). Spiking neuron models : single neurons, populations, plasticity. Cambridge, U.K.; New York: Cambridge University Press.
- Gibson, J. R., Beierlein, M., & Connors, B. W. (1999). Two networks of electrically coupled inhibitory neurons in neocortex. *Nature*, 402(6757), 75–79.
- Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. *Neuron*, 9(1), 1–13.
- Gluck, M. A., & Myers, C. (2001). Gateway to memory : an introduction to neural network modeling of the hippocampus and learning. Cambridge, Mass.: MIT Press.
- Goodhill, G. J. (1993). Topography and ocular dominance – a model exploring positive correlations. *Biological Cybernetics*, 69(2), 109–118.
- Goodhill, G. J., & Willshaw, D. J. (1994). Elastic net model of ocular dominance - overall stripe pattern and monocular deprivation. *Neural Computation*, 6(4), 615–621.
- Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, 100(2), 1111–1121.
- Haese, K., & Goodhill, G. J. (2001). Auto-SOM: Recursive parameter estimation for guidance of self-organizing feature maps. *Neural Computation*, 13(3), 595–619.
- Hasselmo, M. E., Wyble, B. P., & Wallenstein, G. V. (1996). Encoding and retrieval of episodic memories: Role of cholinergic and GABAergic modulation in the hippocampus. *Hippocampus*, 6(6), 693–708.

- Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning : data mining, inference, and prediction*. New York: Springer.
- Hebb, D. O. (1949). *The organization of behavior; a neuropsychological theory*. New York,: Wiley.
- Heskes, T. (2001). Self-organizing maps, vector quantization, and mixture modeling. *IEEE Transactions on Neural Networks*, 12(6), 1299–1305.
- Hinton, G. E., & Sejnowski, T. J. (Eds.). (1999). *Unsupervised learning : foundations of neural computation*. Cambridge, Mass.: MIT Press.
- Hinton, G. E., & Shalllice, T. (1991). Lesioning an attractor network – investigations in acquired dyslexia. *Psychological Review*, 98(1), 74–95.
- Hoffman, R. E., & McGlashan, T. H. (2001). Neural network models of schizophrenia. *Neuroscientist*, 7(5), 441–454.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, 79(8), 2554–2558.
- Hopfield, J. J., & Tank, D. W. (1986). Computing with neural circuits – a model. *Science*, 233(4764), 625–633.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in cats visual cortex. *Journal of Physiology-London*, 160(1), 106–154.
- Ishai, A., Ungerleider, L. G., Martin, A., & Haxby, J. V. (2000). The representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience*, 12, 35–51.
- Ishai, A., Ungerleider, L. G., Martin, A., Schouten, H. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 96(16), 9379–9384.
- Jones, M., & Polk, T. A. (2002). An attractor network model of serial recall. *Cognitive Systems Research*, 3, 45–55.
- Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science*. New York: McGraw-Hill, Health Professions Division.
- Koch, C., & Davis, J. L. (Eds.). (1994). *Large-scale neuronal theories of the brain*. Cambridge, Mass.: MIT Press.
- Kohonen, T. (1982a). Analysis of a simple self-organizing process. *Biological Cybernetics*, 44(2), 135–140.
- Kohonen, T. (1982b). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1), 59–69.
- Kohonen, T. (1993). Physiological interpretation of the self-organizing map algorithm. *Neural Networks*, 6(7), 895–905.
- Kohonen, T., Kaski, S., & Lappalainen, H. (1997). Self-organized formation of various invariant-feature filters in the adaptive-subspace SOM. *Neural Computation*, 9(6), 1321–1344.
- Koikkalainen, P., & Oja, E. (1990). Self-organizing hierarchical feature maps. Paper presented at the International Joint Conference on Neural Networks (IJCNN), Washington, D.C.
- Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, 17(8–9), 1345–1362.

- Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current biology : CB*, 5(5), 552–563.
- Luttrell, S. P. (1988). Self-organising multilayer topographic mappings. In *Proceedings of ICNN'88, IEEE International Conference on Neural Networks*, vol. 1, pp. 93–100. IEEE Service Center, Piscataway, NJ.
- Luttrell, S. P. (1989). Hierarchical self-organising networks. In *Proceedings of 1st IEE Conference on Artificial Neural Networks*, pp. 2–6. British Neural Network Society, London, UK.
- 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. *Psychol Rev*, 102(3), 419–457.
- McEliece, R. J., Posner, E. C., Rodemich, E. R., & Venkatesh, S. S. (1987). The capacity of the hopfield associative memory. *IEEE Transactions on Information Theory*, 33(4), 461–482.
- Miikkulainen, R. (1997). Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon. *Brain and Language*, 59(2), 334–366.
- Miller, P., Brody, C. D., Romo, R., & Wang, X. J. (2003). A recurrent network model of somatosensory parametric working memory in the prefrontal cortex. *Cerebral Cortex*, 13(11), 1208–1218.
- Mountcastle, V. B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of Neurophysiology*, 20(4), 408–434.
- Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain*, 120, 701–722.
- Newell, K. M., Liu, Y. T., & Mayer-Kress, G. (2001). Time scales in motor learning and development. *Psychological Review*, 108(1), 57–82.
- Newman, L. I., & Polk, T. A. (2007). The emergence of semantic topography in a neurally-inspired computational model. Paper presented at the Eighth International Conference on Cognitive Modeling.
- O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current opinion in neurobiology*, 14(6), 769–776.
- O'Reilly, R. C., & Munakata, Y. (2000). Computational explorations in cognitive neuroscience understanding the mind by simulating the brain. Cambridge, Mass.: MIT Press.
- Obermayer, K., & Sejnowski, T. J. (Eds.). (2001). *Self-organizing map formation: foundations of neural computation*. Cambridge, Mass.: MIT Press.
- Olson, S. J., & Grossberg, S. (1998). A neural network model for the development of simple and complex cell receptive fields within cortical maps of orientation and ocular dominance. *Neural Networks*, 11(2), 189–208.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56–115.
- Poggio, T., & Edelman, S. (1990). A network that learns to recognize 3-dimensional objects. *Nature*, 343(6255), 263–266.

- Polk, T. A., Behensky, C., Gonzalez, R., & Smith, E. E. (2002). Rating the similarity of simple perceptual stimuli: asymmetries induced by manipulating exposure frequency. *Cognition*, 82(3), B75–B88.
- Polk, T. A., & Farah, M. J. (1998). The neural development and organization of letter recognition: Evidence from functional neuroimaging, computational modeling, and behavioral studies. *Proceedings of the National Academy of Sciences of the United States of America*, 95(3), 847–852.
- Polk, T. A., Simen, P., Lewis, R. L., & Freedman, E. (2002). A computational approach to control in complex cognition. *Cognitive Brain Research*, 15(1), 71–83.
- Ritter, H., & Kohonen, T. (1989). Self-organizing semantic maps. *Biological Cybernetics*, 61(4), 241–254.
- Rolls, E. T., & Deco, G. (2002). Computational neuroscience of vision. Oxford [England] ; New York: Oxford University Press.
- Rolls, E. T., & Tovee, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*, 73(2), 713–726.
- Rolls, E. T., Treves, A., & Tovee, M. J. (1997). The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Experimental Brain Research*, 114(1), 149–162.
- Rumelhart, D. E., McClelland, J. L., & PDPResearchGroup. (1986). Parallel distributed processing: explorations in the microstructure of cognition. Cambridge, Mass.: MIT Press.
- Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annual Review of Psychology*, 57, 87–115.
- Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of Multiple Visual Areas in Humans Revealed by Functional Magnetic-Resonance-Imaging. *Science*, 268(5212), 889–893.
- Sirosh, J., & Miikkulainen, R. (1997). Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Computation*, 9(3), 577–594.
- Sit, Y. F., & Miikkulainen, R. (2006). Self-organization of hierarchical visual maps with feedback connections. *Neurocomputing*, 69(10–12), 1309–1312.
- Stringer, S. M., Rolls, E. T., & Trappenberg, T. P. (2004). Self-organising continuous attractor networks with multiple activity packets, and the representation of space. *Neural Networks*, 17(1), 5–27.
- Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning : an introduction. Cambridge, Mass.: MIT Press.
- Szentagothai, J. (1975). Module-concept in cerebral-cortex architecture. *Brain Research*, 95(2-3), 475–496.
- Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, 66(1), 170–189.
- Tereshko, V., & Allinson, N. M. (2002). Combining lateral and elastic interactions: Topology-preserving elastic nets. *Neural Processing Letters*, 15(3), 213–223.

- Tootell, R. B. H., Switkes, E., Silverman, M. S., & Hamilton, S. L. (1988). Functional-Anatomy of Macaque Striate Cortex .2. Retinotopic Organization. *Journal of Neuroscience*, 8(5), 1531–1568.
- Treves, A., & Rolls, E. T. (1991). What determines the capacity of autoassociative memories in the brain. *Network-Computation in Neural Systems*, 2(4), 371–397.
- Tuckova, J., & Zetocha, P. (2006). Speech analysis of children with developmental dysphasia by Supervised SOM. *Neural Network World*, 16(6), 533–545.
- Tusa, R. J., Palmer, L. A., & Rosenquist, A. C. (1978). Retinotopic organization of area-17 (striate cortex) in cat. *Journal of Comparative Neurology*, 177(2), 213–235.
- Van Essen, D. C., & Gallant, J. L. (1994). Review: Neural Mechanisms of Form and Motion Processing in the Primate Visual-System. *Neuron*, 13(1), 1–10.
- Vinson, D. P., & Vigliocco, G. (2002). A semantic analysis of grammatical class impairments: semantic representations of object nouns, action nouns and action verbs. *Journal of Neurolinguistics*, 15(3–5), 317–351.
- Xie, X. P., Liaw, J. S., Baudry, M., & Berger, T. W. (1997). Novel expression mechanism for synaptic potentiation: Alignment of presynaptic release site and postsynaptic receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 94(13), 6983–6988.
- Zador, A., Koch, C., & Brown, T. H. (1990). Biophysical model of a Hebbian synapse. *Proc Natl Acad Sci U S A*, 87(17), 6718–6722.
- Zipser, D., Kehoe, B., Littlewort, G., & Fuster, J. (1993). A spiking network model of short-term active memory. *Journal of Neuroscience*, 13(8), 3406–3420.