Cerebral Cortex
Principles of
Operation

Edmund T. Rolls

Oxford Centre for Computational Neuroscience, Oxford, UK

Preface

The overall aim of this book is to provide insight into the principles of operation of the cerebral cortex. These are key to understanding how we, as humans, function.

There have been few previous attempts to set out some of the important principles of operation of the cortex, and this book is pioneering. I have asked some of the leading investigators in neuroscience about their views on this, and most have not had many well formulated answers or hypotheses. As clear hypotheses are needed in this most important area of 21st century science, how our brains work, I have formulated a set of hypotheses to guide thinking and future research. I present evidence for many of the hypotheses, but at the same time we must all recognise that hypotheses and theory in science are there to be tested, and hopefully refined rather than rejected. Nevertheless, such theories and hypotheses are essential to progress, and it is in this frame of reference that I present the theories, hypotheses, and ideas that I have produced and collected together.

This book focusses on the principles of operation of the cerebral cortex, because at this time it is possible to propose and describe many principles, and many are likely to stand the test of time, and provide a foundation I believe for further developments, even if some need to be changed. In this context, I have not attempted to produce an overall theory of operation of the cerebral cortex, because at this stage of our understanding, such a theory would be incorrect or incomplete. I believe though that many of the principles will be important, and that many will provide the foundations for more complete theories of the operation of the cerebral cortex.

Given that many different principles of operation of the cortex are proposed in this book, with often several principles in each Chapter, the reader may find it convenient to take one Chapter at a time, and think about the issues raised in each Chapter, as the overall enterprise is large. The Highlights sections provided at the end of each Chapter may be useful in helping the reader to appreciate the different principles being considered in each Chapter.

To understand how the cortex works, including how it functions in perception, memory, attention, decision-making, and cognitive functions, it is necessary to combine different approaches, including neural computation. Neurophysiology at the single neuron level is needed because this is the level at which information is exchanged between the computing elements of the brain. Evidence from the effects of brain damage, including that available from neuropsychology, is needed to help understand what different parts of the system do, and indeed what each part is necessary for. Neuroimaging is useful to indicate where in the human brain different processes take place, and to show which functions can be dissociated from each other. Knowledge of the biophysical and synaptic properties of neurons is essential to understand how the computing elements of the brain work, and therefore what the building blocks of biologically realistic computational models should be. Knowledge of the anatomical and functional architecture of the cortex is needed to show what types of neuronal network actually perform the computation. And finally the approach of neural computation is needed, as this is required to link together all the empirical evidence to produce an understanding of how the system actually works. This book utilizes evidence from all these disciplines to develop an understanding of how different types of memory, perception, attention, and decision-making are implemented by processing in the cerebral cortex.

I emphasize that to understand how memory, perception, attention, decision-making, cognitive functions, and actions are produced in the cortex, we are dealing with large-scale computational systems with interactions between the parts, and that this understanding requires analysis at the computational and global level of the operation of many neurons to perform together a useful function. Understanding at the molecular level is important for helping to understand how these large-scale computational processes are implemented in the brain, but will not by itself give any account of what computations are performed to implement these cognitive functions. Instead, understanding cognitive functions such as object recognition, memory recall, attention, and decision-making requires single neuron data to be closely linked to computational models of how the interactions between large numbers of neurons and many networks of neurons allow these cognitive problems to be solved. The single neuron level is important in this approach, for the single neurons can be thought of as the computational units of the system, and is the level at which the information is exchanged by the spiking activity between the computational elements of the brain. The single neuron level is therefore, because it is the level at which information is communicated between the computing elements of the brain, the fundamental level of information processing, and the level at which the information can be read out (by recording the spiking activity) in order to understand what information is being represented and processed in each brain area.

With its focus on how the brain and especially how the cortex works at the computational neuroscience level, this book is distinct from the many excellent books on neuroscience that describe much evidence about brain structure and function, but do not aim to provide an understanding of how the brain works at the computational level. This book aims to forge an understanding of how some key brain systems may operate at the computational level, so that we can understand how the cortex actually performs some of its complex and necessarily computational functions in memory, perception, attention, decision-making, cognitive functions, and actions.

A test of whether one's understanding is correct is to simulate the processing on a computer, and to show whether the simulation can perform the tasks of cortical systems, and whether the simulation has similar properties to the real cortex. The approach of neural computation leads to a precise definition of how the computation is performed, and to precise and quantitative tests of the theories produced. How memory systems in the cortex work is a paradigm example of this approach, because memory-like operations which involve altered functionality as a result of synaptic modification are at the heart of how many computations in the cortex are performed. It happens that attention and decision-making can be understood in terms of interactions between and fundamental operations in memory systems in the cortex, and therefore it is natural to treat these areas of cognitive neuroscience in this book. The same fundamental concepts based on the operation of neuronal circuitry can be applied to all these functions, as is shown in this book.

One of the distinctive properties of this book is that it links the neural computation approach not only firmly to neuronal neurophysiology, which provides much of the primary data about how the cortex operates, but also to psychophysical studies (for example of attention); to neuropsychological studies of patients with brain damage; and to functional magnetic resonance imaging (fMRI) (and other neuroimaging) approaches. The empirical evidence that is brought to bear is largely from non-human primates and from humans, because of the considerable similarity of their cortical systems.

In this book, I have not attempted to produce a single computational theory of how the cortex operates. Instead, I have highlighted many different principles of cortical function, most of which are likely to be building blocks of how our cortex operates. The reason for this approach is that many of the principles may well be correct, and useful in understanding how the cortex operates, but some might turn out not to be useful or correct. The aim of this

book is therefore to propose some of the fundamental principles of operation of the cerebral cortex, many or most of which will provide a foundation for understanding the operation of the cortex, rather than to produce a single theory of operation of the cortex, which might be disproved if any one of its elements was found to be weak.

The overall aims of the book are developed further, and the plan of the book is described, in Chapter 1, Section 1.1. Some of the main Principles of Operation of the Cerebral Cortex that I describe can be found in the titles of Chapters 2–22; but in practice, most Chapters include several Principles of Operation, which will appear in the Highlights to each Chapter. Section 26.5 may be useful in addition to the Highlights, for Section 26.5 draws together in a synthesis some of the Principles of Operation of the Cerebral Cortex that are described in the book. Further evidence on how these principles are relevant to the operation of different cortical areas and systems and operate together is provided in Chapters 24–25. In these Chapters, the operation of two major cortical systems, those involved in memory and in visual object recognition, are considered to illustrate how the principles are combined to implement two different key cortical functions. The Appendices provide some of the more formal and quantitative properties of the operation of neuronal systems, and are provided because they provide a route to a deeper understanding on the principles, and to enable the presentation in earlier Chapters to be at a readily approachable level. The Appendices describe many of the building blocks of the neurocomputational approach, and are designed to be useful for teaching. Appendix D describes Matlab software that has been made available with this book to provide simple demonstrations of the operation of some key neuronal networks related to cortical function. The programs are available at http://www.oxcns.org.

Part of the material described in the book reflects work performed in collaboration with many colleagues, whose tremendous contributions are warmly appreciated. The contributions of many will be evident from the references cited in the text. Especial appreciation is due to Gustavo Deco, Simon M. Stringer, and Alessandro Treves who have contributed greatly in an always interesting and fruitful research collaboration on computational aspects of brain function, and to many neurophysiology and functional neuroimaging colleagues who have contributed to the empirical discoveries that provide the foundation to which the computational neuroscience must always be closely linked, and whose names are cited throughout the text. Much of the work described would not have been possible without financial support from a number of sources, particularly the Medical Research Council of the UK, the Human Frontier Science Program, the Wellcome Trust, and the James S. McDonnell Foundation. I am also grateful to many colleagues who I have consulted while writing this book, including Joel Price (Washington University School of Medicine), and Donald Wilson (New York University). Dr Patrick Mills is warmly thanked for his comments on the text. Section 24.3.12 on *ars memoriae* is warmly dedicated to my colleagues at Corpus Christi College, Oxford. The book was typeset by the author using LaTeX and WinEdt.

The cover includes part of the picture *Pandora* painted in 1896 by J. W. Waterhouse. The metaphor is to look inside the system of the mind and the brain, in order to understand how the brain functions, and thereby better to understand and treat its disorders. The cover also includes an image of the dendritic morphology of excitatory neurons in S1 whisker barrel cortex (Fig. 1.14) (adapted from Marcel Oberlaender, Christiaan P.J. de Kock, Randy M. Bruno, Alejandro Ramirez, Hanno S. Meyer, Vincent J. Dercksen, Moritz Helmstaedter and Bert Sakmann, Cell type-specific three-dimensional structure of thalamocortical circuits in a column of rat vibrissal cortex, Cerebral Cortex, 2012, Vol. 22, issue 10, pp. 2375–2391, by permission of Oxford University Press). The cover also includes a diagram of the computational circuitry of the hippocampus by the author (Fig. 24.1). The aim of these second two images is to highlight the importance of moving from the anatomy of the cortex using all the approaches available including neuronal network models that address and

incorporate neurophysiological discoveries to lead to an understanding of how the cortex operates computationally.

Updates to and .pdfs of many of the publications cited in this book are available at http://www.oxcns.org. Updates and corrections to the text and notes are also available at http://www.oxcns.org.

I dedicate this work to the overlapping group: my family, friends, and colleagues – in salutem praesentium, in memoriam absentium.

Contents

25 Invariant visual object recognition learning — 554

Appendix 2 Neuronal network models

B.1 Introduction

Formal models of neural networks are needed in order to provide a basis for understanding the processing and memory functions performed by real neuronal networks in the brain. The formal models included in this Appendix all describe fundamental types of network found in different brain regions, and the computations they perform. Each of the types of network described can be thought of as providing one of the fundamental building blocks that the brain uses. Often these building blocks are combined within a brain area to perform a particular computation.

The aim of this Appendix is to describe a set of fundamental networks used by the brain, including the parts of the brain involved in memory, attention, decision-making, and the building of perceptual representations. As each type of network is introduced, we will point briefly to parts of the brain in which each network is found. Understanding these models provides a basis for understanding the theories of how different types of memory functions are performed. The descriptions of these networks are kept relatively concise in this Appendix. More detailed descriptions of some of the quantitative aspects of storage in pattern associators and autoassociators are provided in the Appendices of Rolls and Treves (1998) *Neural Networks and Brain Function*. Another book that provides a clear and quantitative introduction to some of these networks is Hertz, Krogh and Palmer (1991) *Introduction to the Theory of Neural Computation*, and other useful sources include Dayan and Abbott (2001), Gerstner, Kistler, Naud and Paninski (2014) (who focus on neuronal dynamics), Amit (1989) (for attractor networks), Koch (1999) (for a biophysical approach), Wilson (1999)(on spiking networks), and Rolls (2008d).

Some of the background to the operation of the types of neuronal network described here, including a brief review of the evidence on neuronal structure and function, and on synaptic plasticity and the rules by which synaptic strength is modified, much based on studies with long-term potentiation, is provided in Chapter 1.

The network models on which we focus in this Appendix utilize a local learning rule, that is a rule for synaptic modification, in which the signals needed to alter the synaptic strength are present in the pre- and post-synaptic neurons. We focus on these networks because use of a local learning rule is biologically plausible. We discuss the issue of biological plausibility of the networks described, and show how they differ from less biologically plausible networks such as multilayer backpropagation of error networks, in Section B.12.

B.2 Pattern association memory

A fundamental operation of most nervous systems is to learn to associate a first stimulus with a second that occurs at about the same time, and to retrieve the second stimulus when the first is presented. The first stimulus might be the sight of food, and the second stimulus the taste of food. After the association has been learned, the sight of food would enable its taste to be retrieved. In classical conditioning, the taste of food might elicit an unconditioned response of
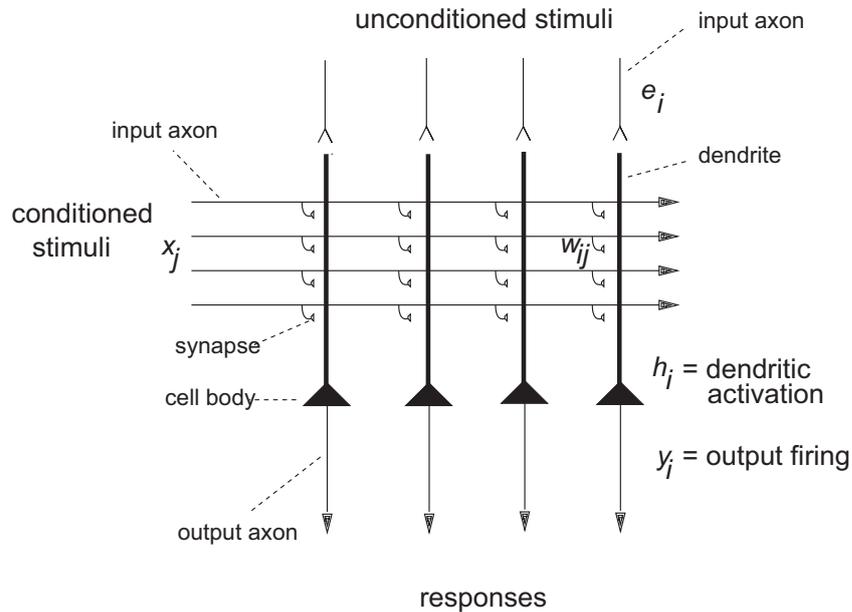
Fig. B.1 A pattern association memory. An unconditioned stimulus has activity or firing rate e_i for the ith neuron, and produces firing y_i of the ith neuron. An unconditioned stimulus may be treated as a vector, across the set of neurons indexed by i, of activity $\mathbf{e}$. The firing rate response can also be thought of as a vector of firing $\mathbf{y}$. The conditioned stimuli have activity or firing rate x_j for the jth axon, which can also be treated as a vector $\mathbf{x}$.

salivation, and if the sight of the food is paired with its taste, then the sight of that food would by learning come to produce salivation. Pattern associators are thus used where the outputs of the visual system interface to learning systems in the orbitofrontal cortex and amygdala that learn associations between the sight of objects and their taste or touch in stimulus–reinforcer association learning (see Chapter 15). Pattern association is also used throughout the cerebral (neo)cortical areas, as it is the architecture that describes the backprojection connections from one cortical area to the preceding cortical area (see Chapters 1, 24 and 6). Pattern association thus contributes to implementing top-down influences in attention, including the effects of attention from higher to lower cortical areas, and thus between the visual object and spatial processing streams (Rolls and Deco 2002) (see Chapter 6); the effects of mood on memory and visual information processing (Rolls and Stringer 2001b); the recall of visual memories; and the operation of short-term memory systems (see Section 4.3.1).

B.2.1 Architecture and operation

The essential elements necessary for pattern association, forming what could be called a prototypical pattern associator network, are shown in Fig. B.1. What we have called the second or unconditioned stimulus pattern is applied through unmodifiable synapses generating an input to each neuron, which, being external with respect to the synaptic matrix we focus on, we can call the external input e_i for the ith neuron. [We can also treat this as a vector, $\mathbf{e}$, as indicated in the legend to Fig. B.1. Vectors and simple operations performed with them are summarized in Appendix A. This unconditioned stimulus is dominant in producing or forcing the firing of the output neurons (y_i for the ith neuron, or the vector $\mathbf{y}$)]. At the same time, the first or conditioned stimulus pattern consisting of the set of firings on the horizontally running

input axons in Fig. B.1 (x_j for the jth axon) (or equivalently the vector **x**) is applied through modifiable synapses w_{ij} to the dendrites of the output neurons. The synapses are modifiable in such a way that if there is presynaptic firing on an input axon x_j paired during learning with postsynaptic activity on neuron i, then the strength or weight w_{ij} between that axon and the dendrite increases. This simple learning rule is often called the Hebb rule, after Donald Hebb who in 1949 formulated the hypothesis that if the firing of one neuron was regularly associated with another, then the strength of the synapse or synapses between the neurons should increase[36]. After learning, presenting the pattern **x** on the input axons will activate the dendrite through the strengthened synapses. If the cue or conditioned stimulus pattern is the same as that learned, the postsynaptic neurons will be activated, even in the absence of the external or unconditioned input, as each of the firing axons produces through a strengthened synapse some activation of the postsynaptic element, the dendrite. The total activation h_i of each postsynaptic neuron i is then the sum of such individual activations. In this way, the 'correct' output neurons, that is those activated during learning, can end up being the ones most strongly activated, and the second or unconditioned stimulus can be effectively recalled. The recall is best when only strong activation of the postsynaptic neuron produces firing, that is if there is a threshold for firing, just like real neurons. The advantages of this are evident when many associations are stored in the memory, as will soon be shown.

Next we introduce a more precise description of the above by writing down explicit mathematical rules for the operation of the simple network model of Fig. B.1, which will help us to understand how pattern association memories in general operate. (In this description we introduce simple vector operations, and, for those who are not familiar with these, refer the reader to Appendix A.) We have denoted above a conditioned stimulus input pattern as **x**. Each of the axons has a firing rate, and if we count or index through the axons using the subscript j, the firing rate of the first axon is x_1, of the second x_2, of the jth x_j, etc. The whole set of axons forms a vector, which is just an ordered (1, 2, 3, etc.) set of elements. The firing rate of each axon x_j is one element of the firing rate vector **x**. Similarly, using i as the index, we can denote the firing rate of any output neuron as y_i, and the firing rate output vector as **y**. With this terminology, we can then identify any synapse onto neuron i from neuron j as w_{ij} (see Fig. B.1). In this book, the first index, i, always refers to the receiving neuron (and thus signifies a dendrite), while the second index, j, refers to the sending neuron (and thus signifies a conditioned stimulus input axon in Fig. B.1). We can now specify the learning and retrieval operations as follows:

B.2.1.1 Learning

The firing rate of every output neuron is forced to a value determined by the unconditioned (or external or forcing stimulus) input e_i. In our simple model this means that for any one neuron i,

$$y_i = \mathrm{f}(e_i) \tag{B.1}$$

which indicates that the firing rate is a function of the dendritic activation, taken in this case to reduce essentially to that resulting from the external forcing input (see Fig. B.1). The function f is called the activation function (see Fig. 1.4), and its precise form is irrelevant, at least during this learning phase. For example, the function at its simplest could be taken to be linear, so that the firing rate would be just proportional to the activation.

---

[36]In fact, the terms in which Hebb put the hypothesis were a little different from an association memory, in that he stated that if one neuron regularly comes to elicit firing in another, then the strength of the synapses should increase. He had in mind the building of what he called cell assemblies. In a pattern associator, the conditioned stimulus need not produce before learning any significant activation of the output neurons. The connection strengths must simply increase if there is associated pre- and postsynaptic firing when, in pattern association, most of the postsynaptic firing is being produced by a different input.

The Hebb rule can then be written as follows:

$$\delta w_{ij} = \alpha y_i x_j \tag{B.2}$$

where δw_{ij} is the change of the synaptic weight w_{ij} that results from the simultaneous (or conjunctive) presence of presynaptic firing x_j and postsynaptic firing or activation y_i, and α is a learning rate constant that specifies how much the synapses alter on any one pairing.

The Hebb rule is expressed in this multiplicative form to reflect the idea that both presynaptic and postsynaptic activity must be present for the synapses to increase in strength. The multiplicative form also reflects the idea that strong pre- and postsynaptic firing will produce a larger change of synaptic weight than smaller firing rates. It is also assumed for now that before any learning takes place, the synaptic strengths are small in relation to the changes that can be produced during Hebbian learning. We will see that this assumption can be relaxed later when a modified Hebb rule is introduced that can lead to a reduction in synaptic strength under some conditions.

B.2.1.2 Recall

When the conditioned stimulus is present on the input axons, the total activation h_i of a neuron i is the sum of all the activations produced through each strengthened synapse w_{ij} by each active neuron x_j. We can express this as

$$h_i = \sum_{j=1}^{C} x_j w_{ij} \tag{B.3}$$

where $\sum_{j=1}^{C}$ indicates that the sum is over the C input axons (or connections) indexed by j to each neuron.

The multiplicative form here indicates that activation should be produced by an axon only if it is firing, and only if it is connected to the dendrite by a strengthened synapse. It also indicates that the strength of the activation reflects how fast the axon x_j is firing, and how strong the synapse w_{ij} is. The sum of all such activations expresses the idea that summation (of synaptic currents in real neurons) occurs along the length of the dendrite, to produce activation at the cell body, where the activation h_i is converted into firing y_i. This conversion can be expressed as

$$y_i = \mathrm{f}(h_i) \tag{B.4}$$

where the function f is again the activation function. The form of the function now becomes more important. Real neurons have thresholds, with firing occurring only if the activation is above the threshold. A threshold linear activation function is shown in Fig. 1.4b on page 7. This has been useful in formal analysis of the properties of neural networks. Neurons also have firing rates that become saturated at a maximum rate, and we could express this as the sigmoid activation function shown in Fig. 1.4c. Yet another simple activation function, used in some models of neural networks, is the binary threshold function (Fig. 1.4d), which indicates that if the activation is below threshold, there is no firing, and that if the activation is above threshold, the neuron fires maximally. Whatever the exact shape of the activation function, some non-linearity is an advantage, for it enables small activations produced by interfering memories to be minimized, and it can enable neurons to perform logical operations, such as to fire or respond only if two or more sets of inputs are present simultaneously.

B.2.2 A simple model

An example of these learning and recall operations is provided in a simple form as follows. The neurons will have simple firing rates, which can be 0 to represent no activity, and 1 to indicate high firing. They are thus binary neurons, which can assume one of two firing rates. If we have a pattern associator with six input axons and four output neurons, we could represent the network before learning, with the same layout as in Fig. B.1, as shown in Fig. B.2:

```
              U  C  S
              1  1  0  0
              ↓  ↓  ↓  ↓
         CS
         1 →   0  0  0  0
         0 →   0  0  0  0
         1 →   0  0  0  0
         0 →   0  0  0  0
         1 →   0  0  0  0
         0 →   0  0  0  0
```

Fig. B.2 Pattern association: before synaptic modification. The unconditioned stimulus (UCS) firing rates are shown as 1 if high and 0 if low as a row vector being applied to force firing of the four output neurons. The six conditioned stimulus (CS) firing rates are shown as a column vector being applied to the vertical dendrites of the output neurons which have initial synaptic weights of 0.

where $\mathbf{x}$ or the conditioned stimulus (CS) is 101010, and $\mathbf{y}$ or the firing produced by the unconditioned stimulus (UCS) is 1100. (The arrows indicate the flow of signals.) The synaptic weights are initially all 0.

After pairing the CS with the UCS during one learning trial, some of the synaptic weights will be incremented according to equation B.2, so that after learning this pair the synaptic weights will become as shown in Fig. B.3:

```
              U  C  S
              1  1  0  0
              ↓  ↓  ↓  ↓
         CS
         1 →   1  1  0  0
         0 →   0  0  0  0
         1 →   1  1  0  0
         0 →   0  0  0  0
         1 →   1  1  0  0
         0 →   0  0  0  0
```

Fig. B.3 Pattern association: after synaptic modification. The synapses where there is conjunctive pre- and post-synaptic activity have been strengthened to value 1.

We can represent what happens during recall, when, for example, we present the CS that has been learned, as shown in Fig. B.4:

```
        CS
        1 →    1  1  0  0
        0 →    0  0  0  0
        1 →    1  1  0  0
        0 →    0  0  0  0
        1 →    1  1  0  0
        0 →    0  0  0  0
               ↓  ↓  ↓  ↓

               3  3  0  0 Activation hᵢ
               1  1  0  0 Firing yᵢ
```

Fig. B.4 Pattern association: recall. The activation h_i of each neuron i is converted with a threshold of 2 to the binary firing rate y_i (1 for high, and 0 for low).

The activation of the four output neurons is 3300, and if we set the threshold of each output neuron to 2, then the output firing is 1100 (where the binary firing rate is 0 if below threshold, and 1 if above). The pattern associator has thus achieved recall of the pattern 1100, which is correct.

We can now illustrate how a number of different associations can be stored in such a pattern associator, and retrieved correctly. Let us associate a new CS pattern 110001 with the UCS 0101 in the same pattern associator. The weights will become as shown next in Fig. B.5 after learning:

```
                U  C  S
                0  1  0  1
                ↓  ↓  ↓  ↓
        CS
        1 →     1  2  0  1
        1 →     0  1  0  1
        0 →     1  1  0  0
        0 →     0  0  0  0
        0 →     1  1  0  0
        1 →     0  1  0  1
```

Fig. B.5 Pattern association: synaptic weights after learning a second pattern association.

If we now present the second CS, the retrieval is as shown in Fig. B.6:

```
CS
1 →    1  2  0  1
1 →    0  1  0  1
0 →    1  1  0  0
0 →    0  0  0  0
0 →    1  1  0  0
1 →    0  1  0  1
       ↓  ↓  ↓  ↓

       1  4  0  3  Activation hᵢ
       0  1  0  1  Firing yᵢ
```

Fig. B.6 Pattern association: recall with the second CS.

The binary output firings were again produced with the threshold set to 2. Recall is perfect.

This illustration shows the value of some threshold non-linearity in the activation function of the neurons. In this case, the activations did reflect some small cross-talk or interference from the previous pattern association of CS1 with UCS1, but this was removed by the threshold operation, to clean up the recall firing. The example also shows that when further associations are learned by a pattern associator trained with the Hebb rule, equation B.2, some synapses will reflect increments above a synaptic strength of 1. It is left as an exercise to the reader to verify that recall is still perfect to CS1, the vector 101010. (The activation vector $\mathbf{h}$ is 3401, and the output firing vector $\mathbf{y}$ with the same threshold of 2 is 1100, which is perfect recall.)

B.2.3 The vector interpretation

The way in which recall is produced, equation B.3, consists for each output neuron i of multiplying each input firing rate x_j by the corresponding synaptic weight w_{ij} and summing the products to obtain the activation h_i. Now we can consider the firing rates x_j where j varies from 1 to N', the number of axons, to be a vector. (A vector is simply an ordered set of numbers – see Appendix A.) Let us call this vector $\mathbf{x}$. Similarly, on a neuron i, the synaptic weights can be treated as a vector, $\mathbf{w}_i$. (The subscript i here indicates that this is the weight vector on the ith neuron.) The operation we have just described to obtain the activation of an output neuron can now be seen to be a simple multiplication operation of two vectors to produce a single output value (called a scalar output). This is the inner product or dot product of two vectors, and can be written

$$h_i = \mathbf{x} \cdot \mathbf{w}_i. \tag{B.5}$$

The inner product of two vectors indicates how similar they are. If two vectors have corresponding elements the same, then the dot product will be maximal. If the two vectors are similar but not identical, then the dot product will be high. If the two vectors are completely different, the dot product will be 0, and the vectors are described as orthogonal. (The term orthogonal means at right angles, and arises from the geometric interpretation of vectors, which is summarized in Appendix A.) Thus the dot product provides a direct measure of how similar two vectors are.

It can now be seen that a fundamental operation many neurons perform is effectively to compute how similar an input pattern vector $\mathbf{x}$ is to their stored weight vector $\mathbf{w}_i$. The

similarity measure they compute, the dot product, is a very good measure of similarity, and indeed, the standard (Pearson product–moment) correlation coefficient used in statistics is the same as a normalized dot product with the mean subtracted from each vector, as shown in Appendix A. (The normalization used in the correlation coefficient results in the coefficient varying always between $+1$ and -1, whereas the actual scalar value of a dot product clearly depends on the length of the vectors from which it is calculated.)

With these concepts, we can now see that during learning, a pattern associator adds to its weight vector a vector $\delta\mathbf{w}_i$ that has the same pattern as the input pattern $\mathbf{x}$, if the postsynaptic neuron i is strongly activated. Indeed, we can express equation B.2 in vector form as

$$\delta\mathbf{w}_i = \alpha y_i \mathbf{x}. \tag{B.6}$$

We can now see that what is recalled by the neuron depends on the similarity of the recall cue vector $\mathbf{x}_r$ to the originally learned vector $\mathbf{x}$. The fact that during recall the output of each neuron reflects the similarity (as measured by the dot product) of the input pattern $\mathbf{x}_r$ to each of the patterns used originally as $\mathbf{x}$ inputs (conditioned stimuli in Fig. B.1) provides a simple way to appreciate many of the interesting and biologically useful properties of pattern associators, as described next.

B.2.4 Properties

B.2.4.1 Generalization

During recall, pattern associators generalize, and produce appropriate outputs if a recall cue vector $\mathbf{x}_r$ is similar to a vector that has been learned already. This occurs because the recall operation involves computing the dot (inner) product of the input pattern vector $\mathbf{x}_r$ with the synaptic weight vector $\mathbf{w}_i$, so that the firing produced, y_i, reflects the similarity of the current input to the previously learned input pattern $\mathbf{x}$. (Generalization will occur to input cue or conditioned stimulus patterns $\mathbf{x}_r$ that are incomplete versions of an original conditioned stimulus $\mathbf{x}$, although the term completion is usually applied to the autoassociation networks described in Section B.3.)

This is an extremely important property of pattern associators, for input stimuli during recall will rarely be absolutely identical to what has been learned previously, and automatic generalization to similar stimuli is extremely useful, and has great adaptive value in biological systems.

Generalization can be illustrated with the simple binary pattern associator considered above. (Those who have appreciated the vector description just given might wish to skip this illustration.) Instead of the second CS, pattern vector 110001, we will use the similar recall cue 110100, as shown in Fig. B.7:

```
        CS
1 →     1   2   0   1
1 →     0   1   0   1
0 →     1   1   0   0
1 →     0   0   0   0
0 →     1   1   0   0
0 →     0   1   0   1
        ↓   ↓   ↓   ↓

        1   3   0   2   Activation $h_i$
        0   1   0   1   Firing $y_i$
```

Fig. B.7 Pattern association: generalization using an input vector similar to the second CS.

It is seen that the output firing rate vector, 0101, is exactly what should be recalled to CS2 (and not to CS1), so correct generalization has occurred. Although this is a small network trained with few examples, the same properties hold for large networks with large numbers of stored patterns, as described more quantitatively in Section B.2.7.1 on capacity below and in Appendix A3 of Rolls and Treves (1998).

B.2.4.2 Graceful degradation or fault tolerance

If the synaptic weight vector w_i (or the weight matrix, which we can call $\mathbf{W}$) has synapses missing (e.g. during development), or loses synapses, then the activation h_i or $\mathbf{h}$ is still reasonable, because h_i is the dot product (correlation) of $\mathbf{x}$ with $\mathbf{w}_i$. The result, especially after passing through the activation function, can frequently be perfect recall. The same property arises if for example one or some of the conditioned stimulus (CS) input axons are lost or damaged. This is a very important property of associative memories, and is not a property of conventional computer memories, which produce incorrect data if even only 1 storage location (for 1 bit or binary digit of data) of their memory is damaged or cannot be accessed. This property of graceful degradation is of great adaptive value for biological systems.

We can illustrate this with a simple example. If we damage two of the synapses in Fig. B.6 to produce the synaptic matrix shown in Fig. B.8 (where x indicates a damaged synapse which has no effect, but was previously 1), and now present the second CS, the retrieval is as follows:

```
        CS
1 →     1   2   0   1
1 →     0   1   0   x
0 →     1   1   0   0
0 →     0   0   0   0
0 →     1   x   0   0
1 →     0   1   0   1
        ↓   ↓   ↓   ↓

        1   4   0   2   Activation $h_i$
        0   1   0   1   Firing $y_i$
```

Fig. B.8 Pattern association: graceful degradation when some synapses are damaged (x).

The binary output firings were again produced with the threshold set to 2. The recalled vector, 0101, is perfect. This illustration again shows the value of some threshold non-linearity in the activation function of the neurons. It is left as an exercise to the reader to verify that recall is still perfect to CS1, the vector 101010. (The output activation vector $\mathbf{h}$ is 3301, and the output firing vector $\mathbf{y}$ with the same threshold of 2 is 1100, which is perfect recall.)

B.2.4.3   The importance of distributed representations for pattern associators

A distributed representation is one in which the firing or activity of all the elements in the vector is used to encode a particular stimulus. For example, in a conditioned stimulus vector CS1 that has the value 101010, we need to know the state of all the elements to know which stimulus is being represented. Another stimulus, CS2, is represented by the vector 110001. We can represent many different events or stimuli with such overlapping sets of elements, and because in general any one element cannot be used to identify the stimulus, but instead the information about which stimulus is present is distributed over the population of elements or neurons, this is called a distributed representation (see Section 8.2). If, for binary neurons, half the neurons are in one state (e.g. 0), and the other half are in the other state (e.g. 1), then the representation is described as fully distributed. The CS representations above are thus fully distributed. If only a smaller proportion of the neurons is active to represent a stimulus, as in the vector 100001, then this is a sparse representation. For binary representations, we can quantify the sparseness by the proportion of neurons in the active (1) state.

In contrast, a local representation is one in which all the information that a particular stimulus or event has occurred is provided by the activity of one of the neurons, or elements in the vector. One stimulus might be represented by the vector 100000, another stimulus by the vector 010000, and a third stimulus by the vector 001000. The activity of neuron or element 1 would indicate that stimulus 1 was present, and of neuron 2, that stimulus 2 was present. The representation is local in that if a particular neuron is active, we know that the stimulus represented by that neuron is present. In neurophysiology, if such cells were present, they might be called 'grandmother cells' (cf. Barlow (1972), (1995); see Chapters 1 and 8 and Appendix C), in that one neuron might represent a stimulus in the environment as complex and specific as one's grandmother. Where the activity of a number of cells must be taken into account in order to represent a stimulus (such as an individual taste), then the representation is sometimes described as using ensemble encoding.

The properties just described for associative memories, generalization, and graceful degradation are only implemented if the representation of the CS or $\mathbf{x}$ vector is distributed. This occurs because the recall operation involves computing the dot (inner) product of the input pattern vector $\mathbf{x}_r$ with the synaptic weight vector $\mathbf{w}_i$. This allows the activation h_i to reflect the similarity of the current input pattern to a previously learned input pattern $\mathbf{x}$ only if several or many elements of the $\mathbf{x}$ and $\mathbf{x}_r$ vectors are in the active state to represent a pattern. If local encoding were used, e.g. 100000, then if the first element of the vector (which might be the firing of axon 1, i.e. x_1, or the strength of synapse $i1$, w_{i1}) is lost, the resulting vector is not similar to any other CS vector, and the activation is 0. In the case of local encoding, the important properties of associative memories, generalization and graceful degradation do not thus emerge. Graceful degradation and generalization are dependent on distributed representations, for then the dot product can reflect similarity even when some elements of the vectors involved are altered. If we think of the correlation between Y and X in a graph, then this correlation is affected only a little if a few X, Y pairs of data are lost (see Appendix A).

B.2.5 Prototype extraction, extraction of central tendency, and noise reduction

If a set of similar conditioned stimulus vectors $\mathbf{x}$ are paired with the same unconditioned stimulus e_i, the weight vector $\mathbf{w}_i$ becomes (or points towards) the sum (or with scaling, the average) of the set of similar vectors $\mathbf{x}$. This follows from the operation of the Hebb rule in equation B.2. When tested at recall, the output of the memory is then best to the average input pattern vector denoted $< \mathbf{x} >$. If the average is thought of as a prototype, then even though the prototype vector $< \mathbf{x} >$ itself may never have been seen, the best output of the neuron or network is to the prototype. This produces 'extraction of the prototype' or 'central tendency'. The same phenomenon is a feature of human memory performance (see McClelland and Rumelhart (1986) Chapter 17), and this simple process with distributed representations in a neural network accounts for the psychological phenomenon.

If the different exemplars of the vector $\mathbf{x}$ are thought of as noisy versions of the true input pattern vector $< \mathbf{x} >$ (with incorrect values for some of the elements), then the pattern associator has performed 'noise reduction', in that the output produced by any one of these vectors will represent the output produced by the true, noiseless, average vector $< \mathbf{x} >$.

B.2.6 Speed

Recall is very fast in a real neuronal network, because the conditioned stimulus input firings x_j ($j = 1, C$ axons) can be applied simultaneously to the synapses w_{ij}, and the activation h_i can be accumulated in one or two time constants of the dendrite (e.g. 10–20 ms). Whenever the threshold of the cell is exceeded, it fires. Thus, in effectively one step, which takes the brain no more than 10–20 ms, all the output neurons of the pattern associator can be firing with rates that reflect the input firing of every axon. This is very different from a conventional digital computer, in which computing h_i in equation B.3 would involve C multiplication and addition operations occurring one after another, or $2C$ time steps.

The brain performs parallel computation in at least two senses in even a pattern associator. One is that for a single neuron, the separate contributions of the firing rate x_j of each axon j multiplied by the synaptic weight w_{ij} are computed in parallel and added in the same timestep. The second is that this can be performed in parallel for all neurons $i = 1, N$ in the network, where there are N output neurons in the network. It is these types of parallel and time-continuous (see Section B.6) processing that enable these classes of neuronal network in the brain to operate so fast, in effectively so few steps.

Learning is also fast ('one-shot') in pattern associators, in that a single pairing of the conditioned stimulus $\mathbf{x}$ and the unconditioned stimulus (UCS) $\mathbf{e}$ which produces the unconditioned output firing $\mathbf{y}$ enables the association to be learned. There is no need to repeat the pairing in order to discover over many trials the appropriate mapping. This is extremely important for biological systems, in which a single co-occurrence of two events may lead to learning that could have life-saving consequences. (For example, the pairing of a visual stimulus with a potentially life-threatening aversive event may enable that event to be avoided in future.) Although repeated pairing with small variations of the vectors is used to obtain the useful properties of prototype extraction, extraction of central tendency, and noise reduction, the essential properties of generalization and graceful degradation are obtained with just one pairing. The actual time scales of the learning in the brain are indicated by studies of associative synaptic modification using long-term potentiation paradigms (LTP, see Section 1.5). Co-occurrence or near simultaneity of the CS and UCS is required for periods of as little as 100 ms, with expression of the synaptic modification being present within typically a few seconds.

(a) Invertebrate   (b) Vertebrate
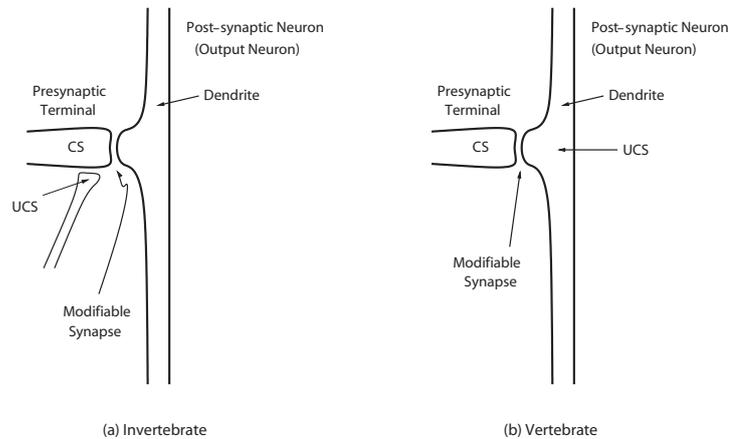
Fig. B.9 (b) In vertebrate pattern association learning, the unconditioned stimulus (UCS) may be made available at all the conditioned stimulus (CS) terminals onto the output neuron because the dendrite of the postsynaptic neuron is electrically short, so that the effect of the UCS spreads for long distances along the dendrite. (a) In contrast, in at least some invertebrate association learning systems, the unconditioned stimulus or teaching input makes a synapse onto the presynaptic terminal carrying the conditioned stimulus.

B.2.7   Local learning rule

The simplest learning rule used in pattern association neural networks, a version of the Hebb rule, is, as shown in equation B.2 above,

$$\delta w_{ij} = \alpha y_i x_j.$$

This is a local learning rule in that the information required to specify the change in synaptic weight is available locally at the synapse, as it is dependent only on the presynaptic firing rate x_j available at the synaptic terminal, and the postsynaptic activation or firing y_i available on the dendrite of the neuron receiving the synapse (see Fig. B.9b). This makes the learning rule biologically plausible, in that the information about how to change the synaptic weight does not have to be carried from a distant source, where it is computed, to every synapse. Such a non-local learning rule would not be biologically plausible, in that there are no appropriate connections known in most parts of the brain to bring in the synaptic training or teacher signal to every synapse.

Evidence that a learning rule with the general form of equation B.2 is implemented in at least some parts of the brain comes from studies of long-term potentiation, described in Section 1.5. Long-term potentiation (LTP) has the synaptic specificity defined by equation B.2, in that only synapses from active afferents, not those from inactive afferents, become strengthened. Synaptic specificity is important for a pattern associator, and most other types of neuronal network, to operate correctly. The number of independently modifiable synapses on each neuron is a primary factor in determining how many different memory patterns can be stored in associative memories (see Sections B.2.7.1 and B.3.3.7).

Another useful property of real neurons in relation to equation B.2 is that the postsynaptic term, y_i, is available on much of the dendrite of a cell, because the electrotonic length of the dendrite is short. In addition, active propagation of spiking activity from the cell body along the dendrite may help to provide a uniform postsynaptic term for the learning. Thus if a neuron is strongly activated with a high value for y_i, then any active synapse onto the cell will be capable of being modified. This enables the cell to learn an association between the pattern

of activity on all its axons and its postsynaptic activation, which is stored as an addition to its weight vector $\mathbf{w}_i$. Then later on, at recall, the output can be produced as a vector dot product operation between the input pattern vector $\mathbf{x}$ and the weight vector $\mathbf{w}_i$, so that the output of the cell can reflect the correlation between the current input vector and what has previously been learned by the cell.

It is interesting that at least many invertebrate neuronal systems may operate very differently from those described here, as described by Rolls and Treves (1998) (see Fig. B.9a). If there were 5,000 conditioned stimulus inputs to a neuron, the implication is that every one would need to have a presynaptic terminal conveying the same UCS to each presynaptic terminal, which is hardly plausible. The implication is that at least some invertebrate neural systems operate very differently to those in vertebrates and, in such systems, the useful properties that arise from using distributed CS representations such as generalization would not arise in the same simple way as a property of the network.

B.2.7.1 Capacity

The question of the storage capacity of a pattern associator is considered in detail in Appendix A3 of Rolls and Treves (1998). It is pointed out there that, for this type of associative network, the number of memories that it can hold simultaneously in storage has to be analysed together with the retrieval quality of each output representation, and then only for a given quality of the representation provided in the input. This is in contrast to autoassociative nets (Section B.3), in which a critical number of stored memories exists (as a function of various parameters of the network), beyond which attempting to store additional memories results in it becoming impossible to retrieve essentially anything. With a pattern associator, instead, one will always retrieve something, but this something will be very small (in information or correlation terms) if too many associations are simultaneously in storage and/or if too little is provided as input.

The conjoint quality–capacity input analysis can be carried out, for any specific instance of a pattern associator, by using formal mathematical models and established analytical procedures (see e.g. Treves (1995), Rolls and Treves (1998), Treves (1990) and Rolls and Treves (1990)). This, however, has to be done case by case. It is anyway useful to develop some intuition for how a pattern associator operates, by considering what its capacity would be in certain well-defined simplified cases.

Linear associative neuronal networks These networks are made up of units with a linear activation function, which appears to make them unsuitable to represent real neurons with their positive-only firing rates. However, even purely linear units have been considered as provisionally relevant models of real neurons, by assuming that the latter operate sometimes in the linear regime of their transfer function. (This implies a high level of spontaneous activity, and may be closer to conditions observed early on in sensory systems rather than in areas more specifically involved in memory.) As usual, the connections are trained by a Hebb (or similar) associative learning rule. The capacity of these networks can be defined as the total number of associations that can be learned independently of each other, given that the linear nature of these systems prevents anything more than a linear transform of the inputs. This implies that if input pattern C can be written as the weighted sum of input patterns A and B, the output to C will be just the same weighted sum of the outputs to A and B. If there are N' input axons, then there can be only at most N' mutually independent input patterns (i.e. none able to be written as a weighted sum of the others), and therefore the capacity of linear networks, defined above, is just N', or equal to the number of inputs to each neuron. In general, a random set of less than N' vectors (the CS input pattern vectors) will tend to be mutually independent but not mutually orthogonal (at $90 \deg$ to each other) (see Appendix A). If they are not orthogonal (the normal situation), then the dot product of them is not 0, and

the output pattern activated by one of the input vectors will be partially activated by other input pattern vectors, in accordance with how similar they are (see equations B.5 and B.6). This amounts to interference, which is therefore the more serious the less orthogonal, on the whole, is the set of input vectors.

Since input patterns are made of elements with positive values, if a simple Hebbian learning rule like the one of equation B.2 is used (in which the input pattern enters directly with no subtraction term), the output resulting from the application of a stored input vector will be the sum of contributions from all other input vectors that have a non-zero dot product with it (see Appendix A), and interference will be disastrous. The only situation in which this would not occur is when different input patterns activate completely different input lines, but this is clearly an uninteresting circumstance for networks operating with distributed representations. A solution to this issue is to use a modified learning rule of the following form:

$$\delta w_{ij} = \alpha y_i (x_j - x) \tag{B.7}$$

where x is a constant, approximately equal to the average value of x_j. This learning rule includes (in proportion to y_i) increasing the synaptic weight if $(x_j - x) > 0$ (long-term potentiation), and decreasing the synaptic weight if $(x_j - x) < 0$ (heterosynaptic long-term depression). It is useful for x to be roughly the average activity of an input axon x_j across patterns, because then the dot product between the various patterns stored on the weights and the input vector will tend to cancel out with the subtractive term, except for the pattern equal to (or correlated with) the input vector itself. Then up to N' input vectors can still be learned by the network, with only minor interference (provided of course that they are mutually independent, as they will in general tend to be).

Table B.1 Effects of pre- and post-synaptic activity on synaptic modification

		Post-synaptic activation	
		0	high
Presynaptic firing	0	No change	Heterosynaptic LTD
	high	Homosynaptic LTD	LTP

This modified learning rule can also be described in terms of a contingency table (Table B.1) showing the synaptic strength modifications produced by different types of learning rule, where LTP indicates an increase in synaptic strength (called long-term potentiation in neurophysiology), and LTD indicates a decrease in synaptic strength (called long-term depression in neurophysiology). Heterosynaptic long-term depression is so-called because it is the decrease in synaptic strength that occurs to a synapse that is other than that through which the postsynaptic cell is being activated. This heterosynaptic long-term depression is the type of change of synaptic strength that is required (in addition to LTP) for effective subtraction of the average presynaptic firing rate, in order, as it were, to make the CS vectors appear more orthogonal to the pattern associator. The rule is sometimes called the Singer–Stent rule, after work by Singer (1987) and Stent (1973), and was discovered in the brain by Levy (Levy 1985, Levy and Desmond 1985) (see also Brown, Kairiss and Keenan (1990b)). Homosynaptic long-term depression is so-called because it is the decrease in synaptic strength that occurs to a synapse which is (the same as that which is) active. For it to occur, the postsynaptic neuron must

simultaneously be inactive, or have only low activity. (This rule is sometimes called the BCM rule after the paper of Bienenstock, Cooper and Munro (1982); see Rolls and Deco (2002), Chapter 7).

Associative neuronal networks with non-linear neurons With non-linear neurons, that is with at least a threshold in the activation function so that the output firing y_i is 0 when the activation h_i is below the threshold, the capacity can be measured in terms of the number of different clusters of output pattern vectors that the network produces. This is because the non-linearities now present (one per output neuron) result in some clustering of the outputs produced by all possible (conditioned stimulus) input patterns $\mathbf{x}$. Input patterns that are similar to a stored input vector can produce, due to the non-linearities, output patterns even closer to the stored output; and vice versa, sufficiently dissimilar inputs can be assigned to different output clusters thereby increasing their mutual dissimilarity. As with the linear counterpart, in order to remove the correlation that would otherwise occur between the patterns because the elements can take only positive values, it is useful to use a modified Hebb rule of the form shown in equation B.7.

With fully distributed output patterns, the number p of associations that leads to different clusters is of order C, the number of input lines (axons) per output neuron (that is, of order N' for a fully connected network), as shown in Appendix A3 of Rolls and Treves (1998). If sparse patterns are used in the output, or alternatively if the learning rule includes a non-linear postsynaptic factor that is effectively equivalent to using sparse output patterns, the coefficient of proportionality between p and C can be much higher than one, that is, many more patterns can be stored than inputs onto each output neuron (see Appendix A3 of Rolls and Treves (1998)). Indeed, the number of different patterns or prototypes p that can be stored can be derived for example in the case of binary units (Gardner 1988) to be

$$p \approx C/[a_o log(1/a_o)] \qquad \text{(B.8)}$$

where a_o is the sparseness of the output firing pattern $\mathbf{y}$ produced by the unconditioned stimulus. p can in this situation be much larger than C (see Appendix A3 of Rolls and Treves (1998), Rolls and Treves (1990) and Treves (1990)). This is an important result for encoding in pattern associators, for it means that provided that the activation functions are non-linear (which is the case with real neurons), there is a very great advantage to using sparse encoding, for then many more than C pattern associations can be stored. Sparse representations may well be present in brain regions involved in associative memory for this reason (see Appendix C).

The non-linearity inherent in the NMDA receptor-based Hebbian plasticity present in the brain may help to make the stored patterns more sparse than the input patterns, and this may be especially beneficial in increasing the storage capacity of associative networks in the brain by allowing participation in the storage of especially those relatively few neurons with high firing rates in the exponential firing rate distributions typical of neurons in sensory systems (see Appendix C).

B.2.7.2 Interference

Interference occurs in linear pattern associators if two vectors are not orthogonal, and is simply dependent on the angle between the originally learned vector and the recall cue or CS vector (see Appendix A), for the activation of the output neuron depends simply on the dot product of the recall vector and the synaptic weight vector (equation B.5). Also in non-linear pattern associators (the interesting case for all practical purposes), interference may occur if two CS patterns are not orthogonal, though the effect can be controlled with sparse encoding of the UCS patterns, effectively by setting high thresholds for the firing of output units. In

Input A	1	0	1
Input B	0	1	1
Required Output	1	1	0

Fig. B.10 A non-linearly separable mapping.

other words, the CS vectors need not be strictly orthogonal, but if they are too similar, some interference will still be likely to occur.

The fact that interference is a property of neural network pattern associator memories is of interest, for interference is a major property of human memory. Indeed, the fact that interference is a property of human memory and of neural network association memories is entirely consistent with the hypothesis that human memory is stored in associative memories of the type described here, or at least that network associative memories of the type described represent a useful exemplar of the class of parallel distributed storage network used in human memory.

It may also be suggested that one reason that interference is tolerated in biological memory is that it is associated with the ability to generalize between stimuli, which is an invaluable feature of biological network associative memories, in that it allows the memory to cope with stimuli that will almost never be identical on different occasions, and in that it allows useful analogies that have survival value to be made.

B.2.7.3 Expansion recoding

If patterns are too similar to be stored in associative memories, then one solution that the brain seems to use repeatedly is to expand the encoding to a form in which the different stimulus patterns are less correlated, that is, more orthogonal, before they are presented as CS stimuli to a pattern associator. The problem can be highlighted by a non-linearly separable mapping (which captures part of the eXclusive OR (XOR) problem), in which the mapping that is desired is as shown in Fig. B.10. The neuron has two inputs, A and B.

This is a mapping of patterns that is impossible for a one-layer network, because the patterns are not linearly separable[37]. A solution is to remap the two input lines A and B to three input lines 1–3, that is to use expansion recoding, as shown in Fig. B.11. This can be performed by a competitive network (see Section B.4). The synaptic weights on the dendrite of the output neuron could then learn the following values using a simple Hebb rule, equation B.2, and the problem could be solved as in Fig. B.12. The whole network would look like that shown in Fig. B.11.

Competitive networks could help with this type of recoding, and could provide very useful preprocessing for a pattern associator in the brain (Rolls and Treves 1998, Rolls 2008d). It is possible that the lateral nucleus of the amygdala performs this function, for it receives inputs from the temporal cortical visual areas, and may preprocess them before they become the inputs to associative networks at the next stage of amygdala processing (Rolls 2008d, Rolls 2014a). The granule cells of the cerebellum may operate similarly (Chapter 23).

---

[37]See Appendix A. There is no set of synaptic weights in a one-layer net that could solve the problem shown in Fig. B.10. Two classes of patterns are not linearly separable if no hyperplane can be positioned in their N-dimensional space so as to separate them (see Appendix A). The XOR problem has the additional constraint that $A = 0, B = 0$ must be mapped to Output = 0.

Competitive Network   Pattern Associator
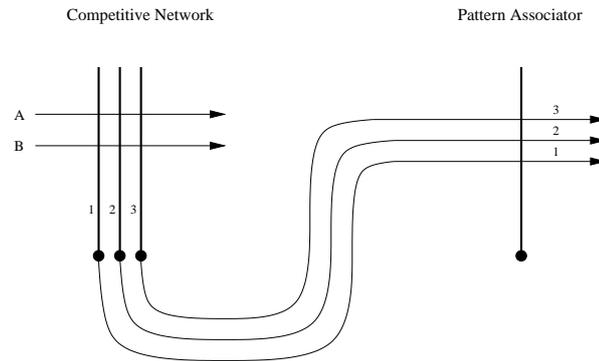
Fig. B.11 Expansion recoding. A competitive network followed by a pattern associator that can enable patterns that are not linearly separable to be learned correctly.

	Synaptic weight
Input 1 (A=1, B=0)	1
Input 2 (A=0, B=1)	1
Input 3 (A=1, B=1)	0

Fig. B.12 Synaptic weights on the dendrite of the output neuron in Fig. B.11.

B.2.8   Implications of different types of coding for storage in pattern associators

Throughout this section, we have made statements about how the properties of pattern associators – such as the number of patterns that can be stored, and whether generalization and graceful degradation occur – depend on the type of encoding of the patterns to be associated. (The types of encoding considered, local, sparse distributed, and fully distributed, are described above.) We draw together these points in Table B.2.

Table B.2 Coding in associative memories*

	Local	**Sparse distributed**	**Fully distributed**
Generalization, completion, graceful degradation	No	Yes	Yes
Number of patterns that can be stored	N (large)	of order $C/[a_o \log(1/a_o)]$ (can be larger)	of order C (usually smaller than N)
Amount of information in each pattern (values if binary)	Minimal ($\log(N)$ bits)	Intermediate ($Na_o \log(1/a_o)$ bits)	Large (N bits)

* N refers here to the number of output units, and C to the average number of inputs to each output unit. a_o is the sparseness of output patterns, or roughly the proportion of output units activated by a UCS pattern. Note: logs are to the base 2.

The amount of information that can be stored in each pattern in a pattern associator is considered in Appendix A3 of Rolls and Treves (1998). That Appendix has been made available at http://www.oxcns.org, and contains a quantitative approach to the capacity of pattern association networks that has not been published elsewhere.

In conclusion, the architecture and properties of pattern association networks make them

very appropriate for stimulus–reinforcer association learning. Their high capacity enables them to learn the reinforcement associations for very large numbers of different stimuli.

B.3 Autoassociation or attractor memory

Autoassociative memories, or attractor neural networks, store memories, each one of which is represented by a pattern of neural activity. The memories are stored in the recurrent synaptic connections between the neurons of the network, for example in the recurrent collateral connections between cortical pyramidal cells. Autoassociative networks can then recall the appropriate memory from the network when provided with a fragment of one of the memories. This is called completion. Many different memories can be stored in the network and retrieved correctly. A feature of this type of memory is that it is content addressable; that is, the information in the memory can be accessed if just the contents of the memory (or a part of the contents of the memory) are used. This is in contrast to a conventional computer, in which the address of what is to be accessed must be supplied, and used to access the contents of the memory. Content addressability is an important simplifying feature of this type of memory, which makes it suitable for use in biological systems. The issue of content addressability will be amplified below.

An autoassociation memory can be used as a short-term memory, in which iterative processing round the recurrent collateral connection loop keeps a representation active by continuing neuronal firing. The short-term memory reflected in continuing neuronal firing for several hundred milliseconds after a visual stimulus is removed which is present in visual cortical areas such as the inferior temporal visual cortex (see Chapter 25) is probably implemented in this way. This short-term memory is one possible mechanism that contributes to the implementation of the trace memory learning rule which can help to implement invariant object recognition as described in Chapter 25. Autoassociation memories also appear to be used in a short-term memory role in the prefrontal cortex. In particular, the temporal visual cortical areas have connections to the ventrolateral prefrontal cortex which help to implement the short-term memory for visual stimuli (in for example delayed match to sample tasks, and visual search tasks, as described in Section 4.3.1). In an analogous way the parietal cortex has connections to the dorsolateral prefrontal cortex for the short-term memory of spatial responses (see Section 4.3.1). These short-term memories provide a mechanism that enables attention to be maintained through backprojections from prefrontal cortex areas to the temporal and parietal areas that send connections to the prefrontal cortex, as described in Chapter 6. Autoassociation networks implemented by the recurrent collateral synapses between cortical pyramidal cells also provide a mechanism for constraint satisfaction and also noise reduction whereby the firing of neighbouring neurons can be taken into account in enabling the network to settle into a state that reflects all the details of the inputs activating the population of connected neurons, as well as the effects of what has been set up during developmental plasticity as well as later experience. Attractor networks are also effectively implemented by virtue of the forward and backward connections between cortical areas (see Sections 11 and 4.3.1). An autoassociation network with rapid synaptic plasticity can learn each memory in one trial. Because of its 'one-shot' rapid learning, and ability to complete, this type of network is well suited for episodic memory storage, in which each past episode must be stored and recalled later from a fragment, and kept separate from other episodic memories (see Chapter 24).
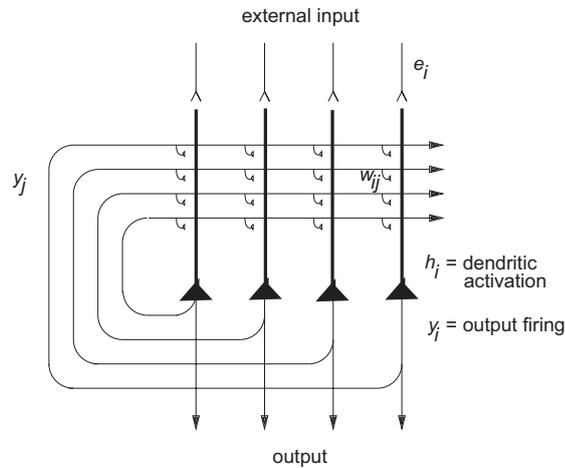
external input

Fig. B.13 The architecture of an autoassociative neural network.

B.3.1 Architecture and operation

The prototypical architecture of an autoassociation memory is shown in Fig. B.13. The external input e_i is applied to each neuron i by unmodifiable synapses. This produces firing y_i of each neuron, or a vector of firing on the output neurons $\mathbf{y}$. Each output neuron i is connected by a recurrent collateral connection to the other neurons in the network, via modifiable connection weights w_{ij}. This architecture effectively enables the output firing vector $\mathbf{y}$ to be associated during learning with itself. Later on, during recall, presentation of part of the external input will force some of the output neurons to fire, but through the recurrent collateral axons and the modified synapses, other neurons in $\mathbf{y}$ can be brought into activity. This process can be repeated a number of times, and recall of a complete pattern may be perfect. Effectively, a pattern can be recalled or recognized because of associations formed between its parts. This of course requires distributed representations.

Next we introduce a more precise and detailed description of the above, and describe the properties of these networks. Ways to analyze formally the operation of these networks are introduced in Appendix A4 of Rolls and Treves (1998) and by Amit (1989).

B.3.1.1 Learning

The firing of every output neuron i is forced to a value y_i determined by the external input e_i. Then a Hebb-like associative local learning rule is applied to the recurrent synapses in the network:

$$\delta w_{ij} = \alpha y_i y_j. \tag{B.9}$$

It is notable that in a fully connected network, this will result in a symmetric matrix of synaptic weights, that is the strength of the connection from neuron 1 to neuron 2 will be the same as the strength of the connection from neuron 2 to neuron 1 (both implemented via recurrent collateral synapses).

It is a factor that is sometimes overlooked that there must be a mechanism for ensuring that during learning y_i does approximate e_i, and must not be influenced much by activity in the recurrent collateral connections, otherwise the new external pattern $\mathbf{e}$ will not be stored in the network, but instead something will be stored that is influenced by the previously stored memories. It is thought that in some parts of the brain, such as the hippocampus, there

are processes that help the external connections to dominate the firing during learning (see Chapter 24, Treves and Rolls (1992b) and Rolls and Treves (1998)).

B.3.1.2 Recall

During recall, the external input e_i is applied, and produces output firing, operating through the non-linear activation function described below. The firing is fed back by the recurrent collateral axons shown in Fig. B.13 to produce activation of each output neuron through the modified synapses on each output neuron. The activation h_i produced by the recurrent collateral effect on the ith neuron is, in the standard way, the sum of the activations produced in proportion to the firing rate of each axon y_j operating through each modified synapse w_{ij}, that is,

$$h_i = \sum_j y_j w_{ij} \qquad (B.10)$$

where $\sum_j$ indicates that the sum is over the C input axons to each neuron, indexed by j.

The output firing y_i is a function of the activation produced by the recurrent collateral effect (internal recall) and by the external input (e_i):

$$y_i = \mathrm{f}(h_i + e_i) \qquad (B.11)$$

The activation function should be non-linear, and may be for example binary threshold, linear threshold, sigmoid, etc. (see Fig. 1.4). The threshold at which the activation function operates is set in part by the effect of the inhibitory neurons in the network (not shown in Fig. B.13). The connectivity is that the pyramidal cells have collateral axons that excite the inhibitory interneurons, which in turn connect back to the population of pyramidal cells to inhibit them by a mixture of shunting (divisive) and subtractive inhibition using GABA (gamma-amino-butyric acid) terminals, as described in Section B.6. There are many fewer inhibitory neurons than excitatory neurons (in the order of 5–10%, see Table 1.1) and of connections to and from inhibitory neurons (see Table 1.1), and partly for this reason the inhibitory neurons are considered to perform generic functions such as threshold setting, rather than to store patterns by modifying their synapses. Similar inhibitory processes are assumed for the other networks described in this Appendix. The non-linear activation function can minimize interference between the pattern being recalled and other patterns stored in the network, and can also be used to ensure that what is a positive feedback system remains stable. The network can be allowed to repeat this recurrent collateral loop a number of times. Each time the loop operates, the output firing becomes more like the originally stored pattern, and this progressive recall is usually complete within 5–15 iterations.

B.3.2 Introduction to the analysis of the operation of autoassociation networks

With complete connectivity in the synaptic matrix, and the use of a Hebb rule, the matrix of synaptic weights formed during learning is symmetric. The learning algorithm is fast, 'one-shot', in that a single presentation of an input pattern is all that is needed to store that pattern.

During recall, a part of one of the originally learned stimuli can be presented as an external input. The resulting firing is allowed to iterate repeatedly round the recurrent collateral system, gradually on each iteration recalling more and more of the originally learned pattern. Completion thus occurs. If a pattern is presented during recall that is similar but not identical to any of the previously learned patterns, then the network settles into a stable recall state in which the firing corresponds to that of the previously learned pattern. The network can

thus generalize in its recall to the most similar previously learned pattern. The activation function of the neurons should be non-linear, since a purely linear system would not produce any categorization of the input patterns it receives, and therefore would not be able to effect anything more than a trivial (i.e. linear) form of completion and generalization.

Recall can be thought of in the following way, relating it to what occurs in pattern associators. The external input $\mathbf{e}$ is applied, produces firing $\mathbf{y}$, which is applied as a recall cue on the recurrent collaterals as $\mathbf{y}^{\mathrm{T}}$. (The notation $\mathbf{y}^{\mathrm{T}}$ signifies the transpose of $\mathbf{y}$, which is implemented by the application of the firing of the neurons $\mathbf{y}$ back via the recurrent collateral axons as the next set of inputs to the neurons.) The activity on the recurrent collaterals is then multiplied by the synaptic weight vector stored during learning on each neuron to produce the new activation h_i which reflects the similarity between $\mathbf{y}^{\mathrm{T}}$ and one of the stored patterns. Partial recall has thus occurred as a result of the recurrent collateral effect. The activations h_i after thresholding (which helps to remove interference from other memories stored in the network, or noise in the recall cue) result in firing y_i, or a vector of all neurons $\mathbf{y}$, which is already more like one of the stored patterns than, at the first iteration, the firing resulting from the recall cue alone, $\mathbf{y} = \mathrm{f}(\mathbf{e})$. This process is repeated a number of times to produce progressive recall of one of the stored patterns.

Autoassociation networks operate by effectively storing associations between the elements of a pattern. Each element of the pattern vector to be stored is simply the firing of a neuron. What is stored in an autoassociation memory is a set of pattern vectors. The network operates to recall one of the patterns from a fragment of it. Thus, although this network implements recall or recognition of a pattern, it does so by an association learning mechanism, in which associations between the different parts of each pattern are learned. These memories have sometimes been called autocorrelation memories (Kohonen 1977), because they learn correlations between the activity of neurons in the network, in the sense that each pattern learned is defined by a set of simultaneously active neurons. Effectively each pattern is associated by learning with itself. This learning is implemented by an associative (Hebb-like) learning rule.

The system formally resembles spin glass systems of magnets analyzed quantitatively in statistical mechanics. This has led to the analysis of (recurrent) autoassociative networks as dynamical systems made up of many interacting elements, in which the interactions are such as to produce a large variety of basins of attraction of the dynamics. Each basin of attraction corresponds to one of the originally learned patterns, and once the network is within a basin it keeps iterating until a recall state is reached that is the learned pattern itself or a pattern closely similar to it. (Interference effects may prevent an exact identity between the recall state and a learned pattern.) This type of system is contrasted with other, simpler, systems of magnets (e.g. ferromagnets), in which the interactions are such as to produce only a limited number of related basins, since the magnets tend to be, for example, all aligned with each other. The states reached within each basin of attraction are called attractor states, and the analogy between autoassociator neural networks and physical systems with multiple attractors was drawn by Hopfield (1982) in a very influential paper. He was able to show that the recall state can be thought of as the local minimum in an energy landscape, where the energy would be defined as

$$E = -\frac{1}{2} \sum_{i,j} w_{ij}(y_i - <y>)(y_j - <y>). \tag{B.12}$$

This equation can be understood in the following way. If two neurons are both firing above their mean rate (denoted by $<y>$), and are connected by a weight with a positive value, then the firing of these two neurons is consistent with each other, and they mutually support each other, so that they contribute to the system's tendency to remain stable. If across the whole network such mutual support is generally provided, then no further change will take

place, and the system will indeed remain stable. If, on the other hand, either of our pair of neurons was not firing, or if the connecting weight had a negative value, the neurons would not support each other, and indeed the tendency would be for the neurons to try to alter ('flip' in the case of binary units) the state of the other. This would be repeated across the whole network until a situation in which most mutual support, and least 'frustration', was reached. What makes it possible to define an energy function and for these points to hold is that the matrix is symmetric (see Hopfield (1982), Hertz, Krogh and Palmer (1991), Amit (1989)).

Physicists have generally analyzed a system in which the input pattern is presented and then immediately removed, so that the network then 'falls' without further assistance (in what is referred to as the unclamped condition) towards the minimum of its basin of attraction. A more biologically realistic system is one in which the external input is left on contributing to the recall during the fall into the recall state. In this clamped condition, recall is usually faster, and more reliable, so that more memories may be usefully recalled from the network. The approach using methods developed in theoretical physics has led to rapid advances in the understanding of autoassociative networks, and its basic elements are described in Appendix A4 of Rolls and Treves (1998), and by Hertz, Krogh and Palmer (1991) and Amit (1989).

B.3.3   Properties

The internal recall in autoassociation networks involves multiplication of the firing vector of neuronal activity by the vector of synaptic weights on each neuron. This inner product vector multiplication allows the similarity of the firing vector to previously stored firing vectors to be provided by the output (as effectively a correlation), if the patterns learned are distributed. As a result of this type of 'correlation computation' performed if the patterns are distributed, many important properties of these networks arise, including pattern completion (because part of a pattern is correlated with the whole pattern), and graceful degradation (because a damaged synaptic weight vector is still correlated with the original synaptic weight vector). Some of these properties are described next.

B.3.3.1   Completion

Perhaps the most important and useful property of these memories is that they complete an incomplete input vector, allowing recall of a whole memory from a small fraction of it. The memory recalled in response to a fragment is that stored in the memory that is closest in pattern similarity (as measured by the dot product, or correlation). Because the recall is iterative and progressive, the recall can be perfect.

This property and the associative property of pattern associator neural networks are very similar to the properties of human memory. This property may be used when we recall a part of a recent memory of a past episode from a part of that episode. The way in which this could be implemented in the hippocampus is described in Chapter 24.

B.3.3.2   Generalization

The network generalizes in that an input vector similar to one of the stored vectors will lead to recall of the originally stored vector, provided that distributed encoding is used. The principle by which this occurs is similar to that described for a pattern associator.

B.3.3.3   Graceful degradation or fault tolerance

If the synaptic weight vector $\mathbf{w}_i$ on each neuron (or the weight matrix) has synapses missing (e.g. during development), or loses synapses (e.g. with brain damage or ageing), then the activation h_i (or vector of activations $\mathbf{h}$) is still reasonable, because h_i is the dot product (correlation) of $\mathbf{y}^{\mathrm{T}}$ with $\mathbf{w}_i$. The same argument applies if whole input axons are lost. If an

output neuron is lost, then the network cannot itself compensate for this, but the next network in the brain is likely to be able to generalize or complete if its input vector has some elements missing, as would be the case if some output neurons of the autoassociation network were damaged.

B.3.3.4 Prototype extraction, extraction of central tendency, and noise reduction

These arise when a set of similar input pattern vectors $\{\mathbf{e}\}$ (which induce firing of the output neurons $\{\mathbf{y}\}$) are learned by the network. The weight vectors $\mathbf{w}_i$ (or strictly $\mathbf{w}_i^T$) become (or point towards) the average $\{<\mathbf{y}>\}$ of that set of similar vectors. This produces 'extraction of the prototype' or 'extraction of the central tendency', and 'noise reduction'. This process can result in better recognition or recall of the prototype than of any of the exemplars, even though the prototype may never itself have been presented. The general principle by which the effect occurs is similar to that by which it occurs in pattern associators. It of course only occurs if each pattern uses a distributed representation.

Related to outputs of the visual system to long-term memory systems (see Chapter 24), there has been intense debate about whether when human memories are stored, a prototype of what is to be remembered is stored, or whether all the instances or the exemplars are each stored separately so that they can be individually recalled (McClelland and Rumelhart (1986), Chapter 17, p. 172). Evidence favouring the prototype view is that if a number of different examples of an object are shown, then humans may report more confidently that they have seen the prototype before than any of the different exemplars, even though the prototype has never been shown (Posner and Keele 1968, Rosch 1975). Evidence favouring the view that exemplars are stored is that in categorization and perceptual identification tasks the responses made are often sensitive to the congruity between particular training stimuli and particular test stimuli (Brooks 1978, Medin and Schaffer 1978, Jacoby 1983a, Jacoby 1983b, Whittlesea 1983). It is of great interest that both types of phenomena can arise naturally out of distributed information storage in a neuronal network such as an autoassociator. This can be illustrated by the storage in an autoassociation memory of sets of stimuli that are all somewhat different examples of the same pattern. These can be generated, for example, by randomly altering each of the input vectors from the input stimulus. After many such randomly altered exemplars have been learned by the network, recall can be tested, and it is found that the network responds best to the original (prototype) input vector, with which it has never been presented. The reason for this is that the autocorrelation components that build up in the synaptic matrix with repeated presentations of the exemplars represent the average correlation between the different elements of the vector, and this is highest for the prototype. This effect also gives the storage some noise immunity, in that variations in the input that are random noise average out, while the signal that is constant builds up with repeated learning.

B.3.3.5 Speed

The recall operation is fast on each neuron on a single iteration, because the pattern $\mathbf{y}^T$ on the axons can be applied simultaneously to the synapses $\mathbf{w}_i$, and the activation h_i can be accumulated in one or two time constants of the dendrite (e.g. 10–20 ms). If a simple implementation of an autoassociation net such as that described by Hopfield (1982) is simulated on a computer, then 5–15 iterations are typically necessary for completion of an incomplete input cue $\mathbf{e}$. This might be taken to correspond to 50–200 ms in the brain, rather too slow for any one local network in the brain to function. However, it has been shown that if the neurons are treated not as McCulloch–Pitts neurons which are simply 'updated' at each iteration, or cycle of timesteps (and assume the active state if the threshold is exceeded), but instead are analyzed and modelled as 'integrate-and-fire' neurons in real continuous time, then the network can

effectively 'relax' into its recall state very rapidly, in one or two time constants of the synapses (see Section B.6 and Treves (1993), Battaglia and Treves (1998a) and Appendix A5 of Rolls and Treves (1998)). This corresponds to perhaps 20 ms in the brain.

One factor in this rapid dynamics of autoassociative networks with brain-like 'integrate-and-fire' membrane and synaptic properties is that with some spontaneous activity, some of the neurons in the network are close to threshold already before the recall cue is applied, and hence some of the neurons are very quickly pushed by the recall cue into firing, so that information starts to be exchanged very rapidly (within 1–2 ms of brain time) through the modified synapses by the neurons in the network. The progressive exchange of information starting early on within what would otherwise be thought of as an iteration period (of perhaps 20 ms, corresponding to a neuronal firing rate of 50 spikes/s) is the mechanism accounting for rapid recall in an autoassociative neuronal network made biologically realistic in this way. Further analysis of the fast dynamics of these networks if they are implemented in a biologically plausible way with 'integrate-and-fire' neurons is provided in Section B.6, in Appendix A5 of Rolls and Treves (1998), and by Treves (1993). *The general approach applies to other networks with recurrent connections, not just autoassociators, and the fact that such networks can operate much faster than it would seem from simple models that follow discrete time dynamics is probably a major factor in enabling these networks to provide some of the building blocks of brain function.*

Learning is fast, 'one-shot', in that a single presentation of an input pattern $\mathbf{e}$ (producing $\mathbf{y}$) enables the association between the activation of the dendrites (the post-synaptic term h_i) and the firing of the recurrent collateral axons $\mathbf{y}^{\mathrm{T}}$, to be learned. Repeated presentation with small variations of a pattern vector is used to obtain the properties of prototype extraction, extraction of central tendency, and noise reduction, because these arise from the averaging process produced by storing very similar patterns in the network.

B.3.3.6   Local learning rule

The simplest learning used in autoassociation neural networks, a version of the Hebb rule, is (as in equation B.9)

$$\delta w_{ij} = \alpha y_i y_j.$$

The rule is a local learning rule in that the information required to specify the change in synaptic weight is available locally at the synapse, as it is dependent only on the presynaptic firing rate y_j available at the synaptic terminal, and the postsynaptic activation or firing y_i available on the dendrite of the neuron receiving the synapse. This makes the learning rule biologically plausible, in that the information about how to change the synaptic weight does not have to be carried to every synapse from a distant source where it is computed. As with pattern associators, since firing rates are positive quantities, a potentially interfering correlation is induced between different pattern vectors. This can be removed by subtracting the mean of the presynaptic activity from each presynaptic term, using a type of long-term depression. This can be specified as

$$\delta w_{ij} = \alpha y_i (y_j - z) \tag{B.13}$$

where α is a learning rate constant. This learning rule includes (in proportion to y_i) increasing the synaptic weight if $(y_j - z) > 0$ (long-term potentiation), and decreasing the synaptic weight if $(y_j - z) < 0$ (heterosynaptic long-term depression). This procedure works optimally if z is the average activity $< y_j >$ of an axon across patterns.

Evidence that a learning rule with the general form of equation B.9 is implemented in at least some parts of the brain comes from studies of long-term potentiation, described in Section 1.5. One of the important potential functions of heterosynaptic long-term depression

is its ability to allow in effect the average of the presynaptic activity to be subtracted from the presynaptic firing rate (see Appendix A3 of Rolls and Treves (1998), and Rolls and Treves (1990)).

Autoassociation networks can be trained with the error-correction or delta learning rule described in Section B.10. Although a delta rule is less biologically plausible than a Hebb-like rule, a delta rule can help to store separately patterns that are very similar (see McClelland and Rumelhart (1988), Hertz, Krogh and Palmer (1991)).

B.3.3.7 Capacity

One measure of storage capacity is to consider how many orthogonal patterns could be stored, as with pattern associators. If the patterns are orthogonal, there will be no interference between them, and the maximum number p of patterns that can be stored will be the same as the number N of output neurons in a fully connected network. Although in practice the patterns that have to be stored will hardly be orthogonal, this is not a purely academic speculation, since it was shown how one can construct a synaptic matrix that effectively orthogonalizes any set of (linearly independent) patterns (Kohonen 1977, Kohonen 1989, Personnaz, Guyon and Dreyfus 1985, Kanter and Sompolinsky 1987). However, this matrix cannot be learned with a local, one-shot learning rule, and therefore its interest for autoassociators in the brain is limited. The more general case of random non-orthogonal patterns, and of Hebbian learning rules, is considered next. However, it is important to reduce the correlations between patterns to be stored in an autoassociation network to not limit the capacity (Marr 1971, Kohonen 1977, Kohonen 1989, Kohonen et al. 1981, Sompolinsky 1987, Rolls and Treves 1998), and in the brain mechanisms to perform pattern separation are frequently present (Rolls 2016f), including granule cells, as shown in many places in this book.

With non-linear neurons used in the network, the capacity can be measured in terms of the number of input patterns $\mathbf{y}$ (produced by the external input $\mathbf{e}$, see Fig. B.13) that can be stored in the network and recalled later whenever the network settles within each stored pattern's basin of attraction. The first quantitative analysis of storage capacity (Amit, Gutfreund and Sompolinsky 1987) considered a fully connected Hopfield (1982) autoassociator model, in which units are binary elements with an equal probability of being 'on' or 'off' in each pattern, and the number C of inputs per unit is the same as the number N of output units. (Actually it is equal to $N-1$, since a unit is taken not to connect to itself.) Learning is taken to occur by clamping the desired patterns on the network and using a modified Hebb rule, in which the mean of the presynaptic and postsynaptic firings is subtracted from the firing on any one learning trial (this amounts to a covariance learning rule, and is described more fully in Appendix A4 of Rolls and Treves (1998)). With such fully distributed random patterns, the number of patterns that can be learned is (for C large) $p \approx 0.14C = 0.14N$, hence well below what could be achieved with orthogonal patterns or with an 'orthogonalizing' synaptic matrix. Many variations of this 'standard' autoassociator model have been analyzed subsequently.

Treves and Rolls (1991) have extended this analysis to autoassociation networks that are much more biologically relevant in the following ways. First, some or many connections between the recurrent collaterals and the dendrites are missing (this is referred to as diluted connectivity, and results in a non-symmetric synaptic connection matrix in which w_{ij} does not equal w_{ji}, one of the original assumptions made in order to introduce the energy formalism in the Hopfield model). Second, the neurons need not be restricted to binary threshold neurons, but can have a threshold linear activation function (see Fig. 1.4). This enables the neurons to assume real continuously variable firing rates, which are what is found in the brain (Rolls and Tovee 1995b, Treves, Panzeri, Rolls, Booth and Wakeman 1999). Third, the representation need not be fully distributed (with half the neurons 'on', and half 'off'), but instead can have a small proportion of the neurons firing above the spontaneous rate, which is what is found in

parts of the brain such as the hippocampus that are involved in memory (see Treves and Rolls (1994), and Chapter 6 of Rolls and Treves (1998)). Such a representation is defined as being sparse, and the sparseness a of the representation can be measured, by extending the binary notion of the proportion of neurons that are firing, as

$$a = \frac{(\sum\limits_{i=1}^{N} y_i/N)^2}{\sum\limits_{i=1}^{N} y_i^2/N} \tag{B.14}$$

where y_i is the firing rate of the ith neuron in the set of N neurons. Treves and Rolls (1991) have shown that such a network does operate efficiently as an autoassociative network, and can store (and recall correctly) a number of different patterns p as follows

$$p \approx \frac{C^{\mathrm{RC}}}{a\ln(\frac{1}{a})}k \tag{B.15}$$

where C^{RC} is the number of synapses on the dendrites of each neuron devoted to the recurrent collaterals from other neurons in the network, and k is a factor that depends weakly on the detailed structure of the rate distribution, on the connectivity pattern, etc., but is roughly in the order of 0.2–0.3.

The main factors that determine the maximum number of memories that can be stored in an autoassociative network are thus the number of connections on each neuron devoted to the recurrent collaterals, and the sparseness of the representation. For example, for $C^{\mathrm{RC}} = 12,000$ and $a = 0.02$, p is calculated to be approximately $36,000$. This storage capacity can be realized, with little interference between patterns, if the learning rule includes some form of heterosynaptic long-term depression that counterbalances the effects of associative long-term potentiation (Treves and Rolls (1991); see Appendix A4 of Rolls and Treves (1998)). It should be noted that the number of neurons N (which is greater than C^{RC}, the number of recurrent collateral inputs received by any neuron in the network from the other neurons in the network) is not a parameter that influences the number of different memories that can be stored in the network. The implication of this is that increasing the number of neurons (without increasing the number of connections per neuron) does not increase the number of different patterns that can be stored (see Rolls and Treves (1998) Appendix A4), although it may enable simpler encoding of the firing patterns, for example more orthogonal encoding, to be used. This latter point may account in part for why there are generally in the brain more neurons in a recurrent network than there are connections per neuron (see e.g. Chapter 24).

The non-linearity inherent in the NMDA receptor-based Hebbian plasticity present in the brain may help to make the stored patterns more sparse than the input patterns, and this may be especially beneficial in increasing the storage capacity of associative networks in the brain by allowing participation in the storage of especially those relatively few neurons with high firing rates in the exponential firing rate distributions typical of neurons in sensory systems (see Sections B.4.9.3 and C.3.1).

B.3.3.8 Context

The environmental context in which learning occurs can be a very important factor that affects retrieval in humans and other animals. Placing the subject back into the same context in which the original learning occurred can greatly facilitate retrieval.

Context effects arise naturally in association networks if some of the activity in the network reflects the context in which the learning occurs. Retrieval is then better when that context

is present, for the activity contributed by the context becomes part of the retrieval cue for the memory, increasing the correlation of the current state with what was stored. (A strategy for retrieval arises simply from this property. The strategy is to keep trying to recall as many fragments of the original memory situation, including the context, as possible, as this will provide a better cue for complete retrieval of the memory than just a single fragment.)

The effects that mood has on memory including visual memory retrieval may be accounted for by backprojections from brain regions such as the amygdala and orbitofrontal cortex in which the current mood, providing a context, is represented, to brain regions involved in memory such as the perirhinal cortex, and in visual representations such as the inferior temporal visual cortex (see Rolls and Stringer (2001b)). The very well-known effects of context in the human memory literature could arise in the simple way just described. An implication of the explanation is that context effects will be especially important at late stages of memory or information processing systems in the brain, for there information from a wide range of modalities will be mixed, and some of that information could reflect the context in which the learning takes place. One part of the brain where such effects may be strong is the hippocampus, which is implicated in the memory of recent episodes, and which receives inputs derived from most of the cortical information processing streams, including those involved in space (see Chapter 24).

B.3.3.9 Mixture states

If an autoassociation memory is trained on pattern vectors $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A} + \mathbf{B}$ (i.e. $\mathbf{A}$ and $\mathbf{B}$ are both included in the joint vector $\mathbf{A} + \mathbf{B}$; that is if the vectors are not linearly independent), then the autoassociation memory will have difficulty in learning and recalling these three memories as separate, because completion from either $\mathbf{A}$ or $\mathbf{B}$ to $\mathbf{A} + \mathbf{B}$ tends to occur during recall. (The ability to separate such patterns is referred to as configurational learning in the animal learning literature, see e.g. Sutherland and Rudy (1991).) This problem can be minimized by re-representing $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A} + \mathbf{B}$ in such a way that they are different vectors before they are presented to the autoassociation memory. This can be performed by recoding the input vectors to minimize overlap using, for example, a competitive network, and possibly involving expansion recoding, as described for pattern associators (see Section B.2, Fig. B.11). It is suggested that this is a function of the dentate granule cells in the hippocampus, which precede the CA3 recurrent collateral network (Treves and Rolls 1992b, Treves and Rolls 1994) (see Chapter 24).

B.3.3.10 Memory for sequences

One of the first extensions of the standard autoassociator paradigm that has been explored in the literature is the capability to store and retrieve not just individual patterns, but whole sequences of patterns. Hopfield, in the same 1982 paper, suggested that this could be achieved by adding to the standard connection weights, which associate a pattern with itself, a new, asymmetric component, which associates a pattern with the next one in the sequence. In practice this scheme does not work very well, unless the new component is made to operate on a slower time scale than the purely autoassociative component (Kleinfeld 1986, Sompolinsky and Kanter 1986). With two different time scales, the autoassociative component can stabilize a pattern for a while, before the heteroassociative component moves the network, as it were, into the next pattern. The heteroassociative retrieval cue for the next pattern in the sequence is just the previous pattern in the sequence. A particular type of 'slower' operation occurs if the asymmetric component acts after a delay τ. In this case, the network sweeps through the sequence, staying for a time of order τ in each pattern.

One can see how the necessary ingredient for the storage of sequences is only a minor departure from purely Hebbian learning: in fact, the (symmetric) autoassociative component of

the weights can be taken to reflect the Hebbian learning of strictly simultaneous conjunctions of pre- and post-synaptic activity, whereas the (asymmetric) heteroassociative component can be implemented by Hebbian learning of each conjunction of postsynaptic activity with presynaptic activity shifted a time τ in the past. Both components can then be seen as resulting from a generalized Hebbian rule, which increases the weight whenever postsynaptic activity is paired with presynaptic activity occurring within a given time range, which may extend from a few hundred milliseconds in the past up to include strictly simultaneous activity. This is similar to a trace rule (see Chapter 25), which itself matches very well the observed conditions for induction of long-term potentiation, and appears entirely plausible. The learning rule necessary for learning sequences, though, is more complex than a simple trace rule in that the time-shifted conjunctions of activity that are encoded in the weights must in retrieval produce activations that are time-shifted as well (otherwise one falls back into the Hopfield (1982) proposal, which does not quite work). The synaptic weights should therefore keep separate 'traces' of what was simultaneous and what was time-shifted during the original experience, and this is not very plausible.

Levy and colleagues (Levy, Wu and Baxter 1995, Wu, Baxter and Levy 1996) have investigated these issues further, and the temporal asymmetry that may be present in LTP (see Section 1.5) has been suggested as a mechanism that might provide some of the temporal properties that are necessary for the brain to store and recall sequences (Minai and Levy 1993, Abbott and Blum 1996, Markram, Pikus, Gupta and Tsodyks 1998, Abbott and Nelson 2000). A problem with this suggestion is that, given that the temporal dynamics of attractor networks are inherently very fast when the networks have continuous dynamics (see Section B.6), and that the temporal asymmetry in LTP may be in the order of only milliseconds to a few tens of milliseconds (see Section 1.5), the recall of the sequences would be very fast, perhaps 10–20 ms per step of the sequence, with every step of a 10-step sequence effectively retrieved and gone in a quick-fire session of 100–200 ms.

Another way in which a delay could be inserted in a recurrent collateral path in the brain is by inserting another cortical area in the recurrent path. This could fit in with the cortico-cortical backprojection connections described in Chapter 11, which would introduce some conduction delay (see Panzeri, Rolls, Battaglia and Lavis (2001)).

B.3.4 Use of autoassociation networks in the brain

Because of its 'one-shot' rapid learning, and ability to complete, this type of network is well suited for episodic memory storage, in which each episode must be stored and recalled later from a fragment, and kept separate from other episodic memories. It does not take a long time (the 'many epochs' of backpropagation networks) to train this network, because it does not have to 'discover the structure' of a problem. Instead, it stores information in the form in which it is presented to the memory, without altering the representation. An autoassociation network may be used for this function in the CA3 region of the hippocampus (see Chapter 24, and Rolls and Treves (1998) Chapter 6).

An autoassociation memory can also be used as a short-term memory, in which iterative processing round the recurrent collateral loop keeps a representation active until another input cue is received. This may be used to implement many types of short-term memory in the brain (see Section 4.3.1). For example, it may be used in the perirhinal cortex and adjacent temporal lobe cortex to implement short-term visual object memory (Miyashita and Chang 1988, Amit 1995, Hirabayashi, Takeuchi, Tamura and Miyashita 2013); in the dorsolateral prefrontal cortex to implement a short-term memory for spatial responses (Goldman-Rakic 1996); and in the prefrontal cortex to implement a short-term memory for where eye movements should be made in space (see Section 4.3.1 and Rolls (2008d)). Such an autoassociation memory in

the temporal lobe visual cortical areas may be used to implement the firing that continues for often 300 ms after a very brief (16 ms) presentation of a visual stimulus (Rolls and Tovee 1994) (see e.g. Fig. C.17), and may be one way in which a short memory trace is implemented to facilitate invariant learning about visual stimuli (see Chapter 25). In all these cases, the short-term memory may be implemented by the recurrent excitatory collaterals that connect nearby pyramidal cells in the cerebral cortex. The connectivity in this system, that is the probability that a neuron synapses on a nearby neuron, may be in the region of 10% (Braitenberg and Schütz 1991, Abeles 1991) (Chapter 7).

The recurrent connections between nearby neocortical pyramidal cells may also be important in defining the response properties of cortical cells, which may be triggered by external inputs (from for example the thalamus or a preceding cortical area), but may be considerably dependent on the synaptic connections received from nearby cortical pyramidal cells.

The cortico-cortical backprojection connectivity described in Chapters 1, 11, and 24 can be interpreted as a system that allows the forward-projecting neurons in one cortical area to be linked autoassociatively with the backprojecting neurons in the next cortical area (see Chapters 11 and 24). This would be implemented by associative synaptic modification in for example the backprojections. This particular architecture may be especially important in constraint satisfaction (as well as recall), that is it may allow the networks in the two cortical areas to settle into a mutually consistent state. This would effectively enable information in higher cortical areas, which would include information from more divergent sources, to influence the response properties of neurons in earlier cortical processing stages. This is an important function in cortical information processing of interacting associative networks.

B.4 Competitive networks, including self-organizing maps

B.4.1 Function

Competitive neural networks learn to categorize input pattern vectors. Each category of inputs activates a different output neuron (or set of output neurons – see below). The categories formed are based on similarities between the input vectors. Similar, that is correlated, input vectors activate the same output neuron. In that the learning is based on similarities in the input space, and there is no external teacher that forces classification, this is an unsupervised network. The term categorization is used to refer to the process of placing vectors into categories based on their similarity. The term classification is used to refer to the process of placing outputs in particular classes as instructed or taught by a teacher. Examples of classifiers are pattern associators, one-layer delta-rule perceptrons, and multilayer perceptrons taught by error backpropagation (see Sections B.2, B.3, B.10 and B.11). In supervised networks there is usually a teacher for each output neuron.

The categorization produced by competitive nets is of great potential importance in perceptual systems including the whole of the visual cortical processing hierarchies, as described in Chapter 25. Each category formed reflects a set or cluster of active inputs x_j that occur together. This cluster of coactive inputs can be thought of as a feature, and the competitive network can be described as building feature analyzers, where a feature can now be defined as a correlated set of inputs. During learning, a competitive network gradually discovers these features in the input space, and the process of finding these features without a teacher is referred to as self-organization.

Another important use of competitive networks is to remove redundancy from the input space, by allocating output neurons to reflect a set of inputs that co-occur.

input
stimuli x_j

w_{ij}

h_i = dendritic
activation

y_i = output firing

responses
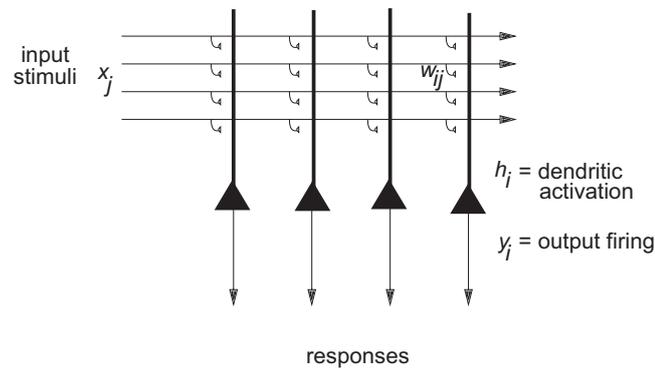
Fig. B.14 The architecture of a competitive network.

Another important aspect of competitive networks is that they separate patterns that are somewhat correlated in the input space, to produce outputs for the different patterns that are less correlated with each other, and may indeed easily be made orthogonal to each other. This has been referred to as orthogonalization.

Another important function of competitive networks is that partly by removing redundancy from the input information space, they can produce sparse output vectors, without losing information. We may refer to this as sparsification.

B.4.2 Architecture and algorithm

B.4.2.1 Architecture

The basic architecture of a competitive network is shown in Fig. B.14. It is a one-layer network with a set of inputs that make modifiable excitatory synapses w_{ij} with the output neurons. The output cells compete with each other (for example by mutual inhibition) in such a way that the most strongly activated neuron or neurons win the competition, and are left firing strongly. The synaptic weights, w_{ij}, are initialized to random values before learning starts. If some of the synapses are missing, that is if there is randomly diluted connectivity, that is not a problem for such networks, and can even help them (see below).

In the brain, the inputs arrive through axons, which make synapses with the dendrites of the output or principal cells of the network. The principal cells are typically pyramidal cells in the cerebral cortex. In the brain, the principal cells are typically excitatory, and mutual inhibition between them is implemented by inhibitory interneurons, which receive excitatory inputs from the principal cells. The inhibitory interneurons then send their axons to make synapses with the pyramidal cells, typically using GABA (gamma-aminobutyric acid) as the inhibitory transmitter.

B.4.2.2 Algorithm

1. Apply an input vector $\mathbf{x}$ and calculate the activation h_i of each neuron

$$h_i = \sum_j x_j w_{ij} \tag{B.16}$$

where the sum is over the C input axons, indexed by j. (It is useful to normalize the length of each input vector $\mathbf{x}$. In the brain, a scaling effect is likely to be achieved both by feedforward inhibition, and by feedback inhibition among the set of input cells (in a preceding network) that give rise to the axons conveying $\mathbf{x}$.)

The output firing y_i^1 is a function of the activation of the neuron

$$y_i^1 = \mathrm{f}(h_i). \tag{B.17}$$

The function f can be linear, sigmoid, monotonically increasing, etc. (see Fig. 1.4).

2. Allow competitive interaction between the output neurons by a mechanism such as lateral or mutual inhibition (possibly with self-excitation), to produce a contrast-enhanced version of the firing rate vector

$$y_i = \mathrm{g}(y_i^1). \tag{B.18}$$

Function g is typically a non-linear operation, and in its most extreme form may be a winner-take-all function, in which after the competition one neuron may be 'on', and the others 'off'. Algorithms that produce softer competition without a single winner to produce a distributed representation are described in Section B.4.9.4 below.

3. Apply an associative Hebb-like learning rule

$$\delta w_{ij} = \alpha y_i x_j. \tag{B.19}$$

4. Normalize the length of the synaptic weight vector on each dendrite to prevent the same few neurons always winning the competition:

$$\sum_j (w_{ij})^2 = 1. \tag{B.20}$$

(A less efficient alternative is to scale the sum of the weights to a constant, e.g. 1.0.)

5. Repeat steps 1–4 for each different input stimulus **x**, in random sequence, a number of times.

B.4.3   Properties

B.4.3.1   Feature discovery by self-organization

Each neuron in a competitive network becomes activated by a set of consistently coactive, that is correlated, input axons, and gradually learns to respond to that cluster of coactive inputs. We can think of competitive networks as discovering features in the input space, where features can now be defined by a set of consistently coactive inputs. Competitive networks thus show how feature analyzers can be built, with no external teacher. The feature analyzers respond to correlations in the input space, and the learning occurs by self-organization in the competitive network. Competitive networks are therefore well suited to the analysis of sensory inputs. Ways in which they may form fundamental building blocks of sensory systems are described in Chapter 25.

The operation of competitive networks can be visualized with the help of Fig. B.15. The input patterns are represented as dots on the surface of a sphere. (The patterns are on the surface of a sphere because the patterns are normalized to the same length.) The directions of the weight vectors of the three neurons are represented by '×'s. The effect of learning is to move the weight vector of each of the neurons to point towards the centre of one of the clusters of inputs. If the neurons are winner-take-all, the result of the learning is that although there are correlations between the input stimuli, the outputs of the three neurons are orthogonal. In this sense, orthogonalization is performed. At the same time, given that each of the patterns within a cluster produces the same output, the correlations between the patterns within a cluster become higher. In a winner-take-all network, the within-pattern correlation becomes 1, and the patterns within a cluster have been placed within the same category.
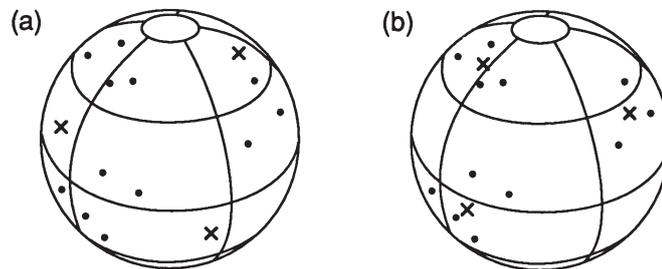
Fig. B.15 Competitive learning. The dots represent the directions of the input vectors, and the '×'s the weights for each of three output neurons. (a) Before learning. (b) After learning. (After Rumelhart and Zipser 1986.)

B.4.3.2 Removal of redundancy

In that competitive networks recode sets of correlated inputs to one or a few output neurons, then redundancy in the input representation is removed. Identifying and removing redundancy in sensory inputs is an important part of processing in sensory systems (cf. Barlow (1989)), in part because a compressed representation is more manageable as an output of sensory systems. The reason for this is that neurons in the receiving systems, for example pattern associators in the orbitofrontal cortex or autoassociation networks in the hippocampus, can then operate with the limited numbers of inputs that each neuron can receive. For example, although the information that a particular face is being viewed is present in the 10^6 fibres in the optic nerve, the information is unusable by associative networks in this form, and is compressed through the visual system until the information about which of many hundreds of faces is present can be represented by less than 100 neurons in the temporal cortical visual areas (Rolls, Treves and Tovee 1997b, Abbott, Rolls and Tovee 1996). (Redundancy can be defined as the difference between the maximum information content of the input data stream (or channel capacity) and its actual content; see Appendix C.)

The recoding of input pattern vectors into a more compressed representation that can be conveyed by a much reduced number of output neurons of a competitive network is referred to in engineering as vector quantization. With a winner-take-all competitive network, each output neuron points to or stands for one of or a cluster of the input vectors, and it is more efficient to transmit the states of the few output neurons than the states of all the input elements. (It is more efficient in the sense that the information transmission rate required, that is the capacity of the channel, can be much smaller.) Vector quantization is of course possible when the input representation contains redundancy.

B.4.3.3 Orthogonalization and categorization

Figure B.15 shows visually how competitive networks reduce the correlation between different clusters of patterns, by allocating them to different output neurons. This is described as orthogonalization. It is a process that is very usefully applied to signals before they are used as inputs to associative networks (pattern associators and autoassociators) trained with Hebbian rules (see Sections B.2 and B.3), because it reduces the interference between patterns stored in these memories. The opposite effect in competitive networks, of bringing closer together very similar input patterns, is referred to as categorization.

These two processes are also illustrated in Fig. B.16, which shows that in a competitive network, very similar input patterns (with correlations higher in this case than approximately 0.8)

a) Before Learning
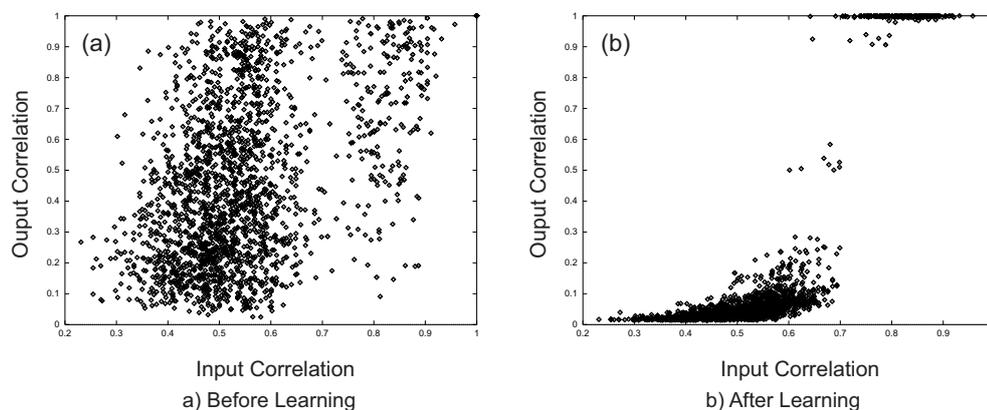
b) After Learning

Fig. B.16 Orthogonalization and categorization in a competitive network: (a) before learning; (b) after learning. The correlations between pairs of output vectors (abscissa) are plotted against the correlations of the corresponding pairs of input vectors that generated the output pair, for all possible pairs in the input set. The competitive net learned for 16 cycles. One cycle consisted of presenting the complete input set of stimuli in a renewing random sequence. The correlation measure shown is the cosine of the angle between two vectors (i.e. the normalized dot product). The network used had 64 input axons to each of 8 output neurons. The net was trained with 64 stimuli, made from 8 initial random binary vectors with each bit having a probability of 0.5 of being 1, from each of which 8 noisy exemplars were made by randomly altering 10 % of the 64 elements. Soft competition was used between the output neurons. (A normalized exponential activation function described in Section B.4.9.4 was used to implement the soft competition.) The sparseness a of the input patterns thus averaged 0.5; and the sparseness a of the output firing vector after learning was close to 0.17 (i.e. after learning, primarily one neuron was active for each input pattern; before learning, the average sparseness of the output patterns produced by each of the inputs was 0.39).

produce more similar outputs (close to 1.0), whereas the correlations between pairs of input patterns that are smaller than approximately 0.7 become much smaller in the output representation. (This simulation used soft competition between neurons with graded firing rates.)

Further analyses of the operation of competitive networks, and how diluted connectivity can help the operation of competitive networks, are provided in Section 7.4 (Rolls 2016f).

B.4.3.4 Sparsification

Competitive networks can produce more sparse representations than those that they receive, depending on the degree of competition. With the greatest competition, winner-take-all, only one output neuron remains active, and the representation is at its most sparse. This effect can be understood further using Figs. B.15 and B.16. This sparsification is useful to apply to representations before input patterns are applied to associative networks, because sparse representations allow many different pattern associations or memories to be stored in these networks (see Sections B.2 and B.3).

B.4.3.5 Capacity

In a competitive net with N output neurons and a simple winner-take-all rule for the competition, it is possible to learn up to N output categories, in that each output neuron may be allocated a category. When the competition acts in a less rudimentary way, the number of categories that can be learned becomes a complex function of various factors, including the number of modifiable connections per cell and the degree of dilution, or incompleteness, of the connections. Such a function has not yet been described analytically in general, but an

upper bound on it can be deduced for the particular case in which the learning is fast, and can be achieved effectively in one shot, or one presentation of each pattern. In that case, the number of categories that can be learned (by the self-organizing process) will at most be equal to the number of associations that can be formed by the corresponding pattern associators, a process that occurs with the additional help of the driving inputs, which effectively determine the categorization in the pattern associator.

Separate constraints on the capacity result if the output vectors are required to be strictly orthogonal. Then, if the output firing rates can assume only positive values, the maximum number p of categories arises, obviously, in the case when only one output neuron is firing for any stimulus, so that up to N categories are formed. If ensemble encoding of output neurons is used (soft competition), again under the orthogonality requirement, then the number of output categories that can be learned will be reduced according to the degree of ensemble encoding. The p categories in the ensemble-encoded case reflect the fact that the between-cluster correlations in the output space are lower than those in the input space. The advantages of ensemble encoding are that dendrites are more evenly allocated to patterns (see Section B.4.9.5), and that correlations between different input stimuli can be reflected in correlations between the corresponding output vectors, so that later networks in the system can generalize usefully. This latter property is of crucial importance, and is utilized for example when an input pattern is presented that has not been learned by the network. The relative similarity of the input pattern to previously learned patterns is indicated by the relative activation of the members of an ensemble of output neurons. This makes the number of different representations that can be reflected in the output of competitive networks with ensemble encoding much higher than with winner-take-all representations, even though with soft competition all these representations cannot strictly be learned.

B.4.3.6 Separation of non-linearly separable patterns

A competitive network can not only separate (e.g. by activating different output neurons) pattern vectors that overlap in almost all elements, but can also help with the separation of vectors that are not linearly separable. An example is that three patterns **A**, **B**, and **A+B** will lead to three different output neurons being activated (see Fig. B.17). For this to occur, the length of the synaptic weight vectors must be normalized (to for example unit length), so that they lie on the surface of a sphere or hypersphere (see Fig. B.15). (If the weight vectors of each neuron are scaled to the same sum, then the weight vectors do not lie on the surface of a hypersphere, and the ability of the network to separate patterns is reduced.)

The property of pattern separation makes a competitive network placed before an auto-association (or pattern association) network very valuable, for it enables the autoassociator to store the three patterns separately, and to recall **A+B** separately from **A** and **B**. This is referred to as the configuration learning problem in animal learning theory (Sutherland and Rudy 1991). Placing a competitive network before a pattern associator will enable a linearly inseparable problem to be solved. For example, three different output neurons of a two-input competitive network could respond to the patterns 01, 10, and 11, and a pattern associator can learn different outputs for neurons 1–3, which are orthogonal to each other (see Fig. B.11). This is an example of expansion recoding (cf. Marr (1969), who used a different algorithm to obtain the expansion). The sparsification that can be produced by the competitive network can also be advantageous in preparing patterns for presentation to a pattern associator or autoassociator, because the sparsification can increase the number of memories that can be associated or stored.
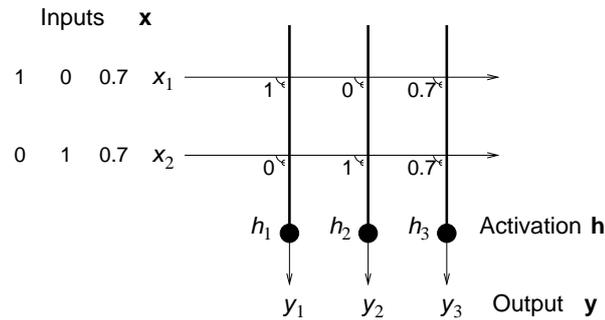
Fig. B.17 Separation of linearly dependent patterns by a competitive network. The network was trained on patterns 10, 01, and 11, applied on the inputs x_1 and x_2. After learning, the network allocated output neuron 1 to pattern 10, neuron 2 to pattern 01, and neuron 3 to pattern 11. The weights in the network produced during the learning are shown. Each input pattern was normalized to unit length, and thus for pattern 11, x_1=0.7 and x_2=0.7, as shown. Because the weight vectors were also normalized to unit length, w_{31}=0.7 and w_{32}=0.7.

B.4.3.7 Stability

These networks are generally stable if the input statistics are stable. If the input statistics keep varying, then the competitive network will keep following the input statistics. If this is a problem, then a critical period in which the input statistics are learned, followed by stabilization, may be useful. This appears to be a solution used in developing sensory systems, which have critical periods beyond which further changes become more difficult. An alternative approach taken by Carpenter and Grossberg in their 'Adaptive Resonance Theory' (Carpenter 1997) is to allow the network to learn only if it does not already have categorizers for a pattern (see Hertz, Krogh and Palmer (1991), p. 228).

Diluted connectivity can help stability, by making neurons tend to find inputs to categorize in only certain parts of the input space, and then making it difficult for the neuron to wander randomly throughout the space later (see further Section 7.4).

B.4.3.8 Frequency of presentation

If some stimuli are presented more frequently than others, then there will be a tendency for the weight vectors to move more rapidly towards frequently presented stimuli, and more neurons may become allocated to the frequently presented stimuli. If winner-take-all competition is used, the result is that the neurons will tend to become allocated during the learning process to the more frequently presented patterns. If soft competition is used, the tendency of neurons to move from patterns that are infrequently or never presented can be reduced by making the competition fairly strong, so that only a few neurons show any learning when each pattern is presented. Provided that the competition is moderately strong (see Section B.4.9.4), the result is that more neurons are allocated to frequently presented patterns, but one or some neurons are allocated to infrequently presented patterns. These points can all be easily demonstrated in simulations.

In an interesting development, it has been shown that if objects consisting of groups of features are presented during training always with another object present, then separate representations of each object can be formed provided that each object is presented many times, but on each occasion is paired with a different object (Stringer and Rolls 2008, Stringer, Rolls and Tromans 2007b). This is related to the fact that in this scenario the frequency of co-occurrence of features within the same object is greater than that of features between different objects (see Section 25.5.6.2).

B.4.3.9   Comparison to principal component analysis (PCA) and cluster analysis

Although competitive networks find clusters of features in the input space, they do not perform hierarchical cluster analysis as typically performed in statistics. In hierarchical cluster analysis, input vectors are joined starting with the most correlated pair, and the level of the joining of vectors is indicated. Competitive nets produce different outputs (i.e. activate different output neurons) for each cluster of vectors (i.e. perform vector quantization), but do not compute the level in the hierarchy, unless the network is redesigned (see Hertz, Krogh and Palmer (1991)).

The feature discovery can also be compared to principal component analysis (PCA). (In PCA, the first principal component of a multidimensional space points in the direction of the vector that accounts for most of the variance, and subsequent principal components account for successively less of the variance, and are mutually orthogonal.) In competitive learning with a winner-take-all algorithm, the outputs are mutually orthogonal, but are not in an ordered series according to the amount of variance accounted for, unless the training algorithm is modified. The modification amounts to allowing each of the neurons in a winner-take-all network to learn one at a time, in sequence. The first neuron learns the first principal component. (Neurons trained with a modified Hebb rule learn to maximize the variance of their outputs – see Hertz, Krogh and Palmer (1991).) The second neuron is then allowed to learn, and because its output is orthogonal to the first, it learns the second principal component. This process is repeated. Details are given by Hertz, Krogh and Palmer (1991), but as this is not a biologically plausible process, it is not considered in detail here.

I note that simple competitive learning is very helpful biologically, because it can separate patterns, but that a full ordered set of principal components as computed by PCA would probably not be very useful in biologically plausible networks. The point here is that biological neuronal networks may operate well if the variance in the input representation is distributed across many input neurons, whereas principal component analysis would tend to result in most of the variance being allocated to a few neurons, and the variance being unevenly distributed across the neurons.

B.4.4   Utility of competitive networks in information processing by the brain

B.4.4.1   Feature analysis and preprocessing

Neurons that respond to correlated combinations of their inputs can be described as feature analyzers. Neurons that act as feature analyzers perform useful preprocessing in many sensory systems (see e.g. Chapter 8 of Rolls and Treves (1998)). The power of competitive networks in multistage hierarchical processing to build combinations of what is found at earlier stages, and thus effectively to build higher-order representations, is also described in Chapter 25 of this book. An interesting development is that competitive networks can learn about individual objects even when multiple objects are presented simultaneously, provided that each object is presented several times more frequently than it is paired with any other individual object (Stringer and Rolls 2008) (see Section 25.5.6.2). This property arises because learning in competitive networks is primarily about forming representations of objects defined by a high correlation of coactive features in the input space (Stringer and Rolls 2008).

B.4.4.2   Removal of redundancy

The removal of redundancy by competition is thought to be a key aspect of how sensory systems, including the ventral cortical visual system, operate. Competitive networks can also be thought of as performing dimension reduction, in that a set of correlated inputs may

be responded to as one category or dimension by a competitive network. The concept of redundancy removal can be linked to the point that individual neurons trained with a modified Hebb rule point their weight vector in the direction of the vector that accounts for most of the variance in the input, that is (acting individually) they find the first principal component of the input space (see Section B.4.3.9 and Hertz, Krogh and Palmer (1991)). Although networks with anti-Hebbian synapses between the principal cells (in which the anti-Hebbian learning forces neurons with initially correlated activity to effectively inhibit each other) (Földiák 1991), and networks that perform Independent Component Analysis (Bell and Sejnowski 1995), could in principle remove redundancy more effectively, it is not clear that they are implemented biologically. In contrast, competitive networks are more biologically plausible, and illustrate redundancy reduction. The more general use of an unsupervised competitive preprocessor is discussed below (see Fig. B.24).

B.4.4.3 Orthogonalization

The orthogonalization performed by competitive networks is very useful for preparing signals for presentation to pattern associators and autoassociators, for this re-representation decreases interference between the patterns stored in such networks. Indeed, this can be essential if patterns are overlapping and not linearly independent, e.g. 01, 10, and 11. If three such binary patterns were presented to an autoassociative network, it would not form separate representations of them, because either of the patterns 01 or 10 would result by completion in recall of the 11 pattern. A competitive network allows a separate neuron to be allocated to each of the three patterns, and this set of orthogonal representations can be learned by associative networks (see Fig. B.17).

B.4.4.4 Sparsification

The sparsification performed by competitive networks is very useful for preparing signals for presentation to pattern associators and autoassociators, for this re-representation increases the number of patterns that can be associated or stored in such networks (see Sections B.2 and B.3).

B.4.4.5 Brain systems in which competitive networks may be used for orthogonalization and sparsification

One system is the hippocampus, in which the dentate granule cells are believed to operate as a competitive network in order to prepare signals for presentation to the CA3 autoassociative network (see Chapter 24). In this case, the operation is enhanced by expansion recoding, in that (in the rat) there are approximately three times as many dentate granule cells as there are cells in the preceding stage, the entorhinal cortex. This expansion recoding will itself tend to reduce correlations between patterns (cf. Marr (1970), and Marr (1969)).

Also in the hippocampus, the CA1 neurons are thought to act as a competitive network that recodes the separate representations of each of the parts of an episode that must be separately represented in CA3, into a form more suitable for the recall using pattern association performed by the backprojections from the hippocampus to the cerebral cortex (see Chapter 24 and Rolls and Treves (1998) Chapter 6).

The granule cells of the cerebellum may perform a similar function, but in this case the principle may be that each of the very large number of granule cells receives a very small random subset of inputs, so that the outputs of the granule cells are decorrelated with respect to the inputs (Marr (1969); see Chapter 23).
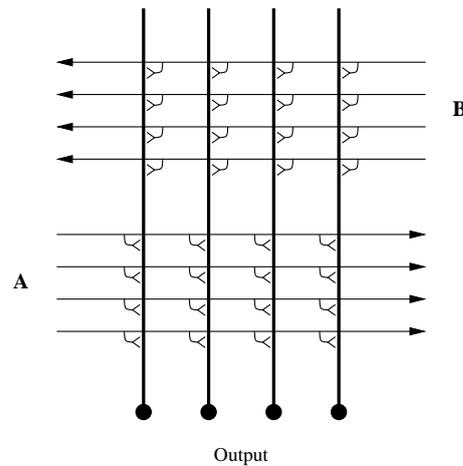
Fig. B.18 Competitive net receiving a normal forward set of inputs **A**, but also another set of inputs **B** that can be used to influence the categories formed in response to **A** inputs. The inputs **B** might be backprojection inputs.

B.4.5 Guidance of competitive learning

Although competitive networks are primarily unsupervised networks, it is possible to influence the categories found by supplying a second input, as follows (Rolls 1989a). Consider a competitive network as shown in Fig. B.18 with the normal set of inputs **A** to be categorized, and with an additional set of inputs **B** from a different source. Both sets of inputs work in the normal way for a competitive network, with random initial weights, competition between the output neurons, and a Hebb-like synaptic modification rule that normalizes the lengths of the synaptic weight vectors onto each neuron. The idea then is to use the **B** inputs to influence the categories formed by the **A** input vectors. The influence of the **B** vectors works best if they are orthogonal to each other. Consider any two **A** vectors. If they occur together with the same **B** vector, then the categories produced by the **A** vectors will be more similar than they would be without the influence of the **B** vectors. The categories will be pulled closer together if soft competition is used, or will be more likely to activate the same neuron if winner-take-all competition is used. Conversely, if any two **A** vectors are paired with two different, preferably orthogonal, **B** vectors, then the categories formed by the **A** vectors will be drawn further apart than they would be without the **B** vectors. The differences in categorization remain present after the learning when just the **A** inputs are used.

This guiding function of one of the inputs is one way in which the consequences of sensory stimuli could be fed back to a sensory system to influence the categories formed when the **A** inputs are presented. This could be one function of backprojections in the cerebral cortex (Rolls 1989c, Rolls 1989a) (Chapter 11). In this case, the **A** inputs of Fig. B.18 would be the forward inputs from a preceding cortical area, and the **B** inputs backprojecting axons from the next cortical area, or from a structure such as the amygdala or hippocampus. If two **A** vectors were both associated with positive reinforcement that was fed back as the same **B** vector from another part of the brain, then the two **A** vectors would be brought closer together in the representational space provided by the output of the neurons. If one of the **A** vectors was associated with positive reinforcement, and the other with negative reinforcement, then the output representations of the two **A** vectors would be further apart. This is one way in which external signals could influence in a mild way the categories formed in sensory systems. Another is that if any **B** vector only occurred for important sensory **A** inputs (as shown by the
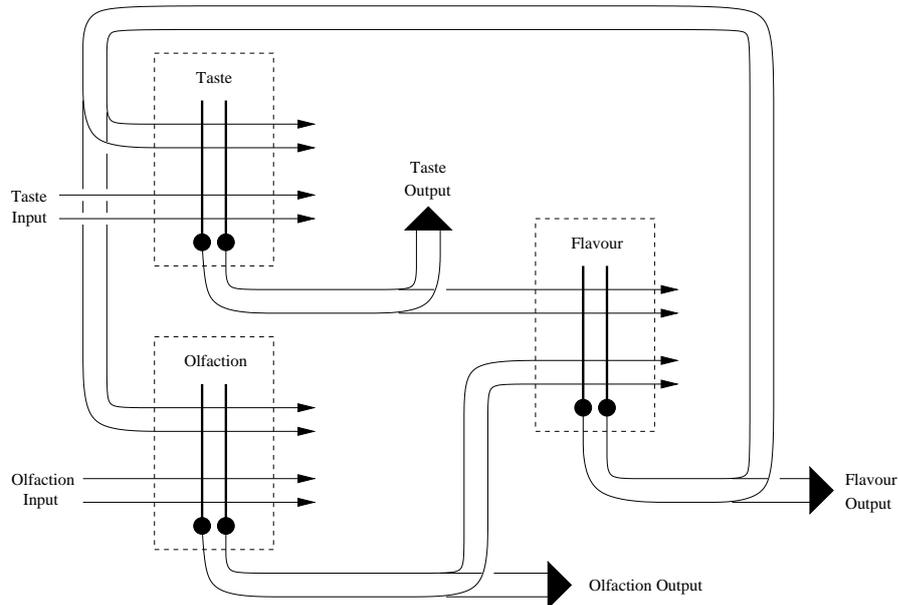
Fig. B.19 A two-layer set of competitive nets in which feedback from layer 2 can influence the categories formed in layer 1. Layer 2 could be a higher cortical visual area with convergence from earlier cortical visual areas (see Chapter 25). In the example, taste and olfactory inputs are received by separate competitive nets in layer 1, and converge into a single competitive net in layer 2. The categories formed in layer 2 (which may be described as representing 'flavour') may be dominated by the relatively orthogonal set of a few tastes that are received by the net. When these layer 2 categories are fed back to layer 1, they may produce in layer 1 categories in, for example, the olfactory network that reflect to some extent the flavour categories of layer 2, and are different from the categories that would otherwise be formed to a large set of rather correlated olfactory inputs. A similar principle may operate in any multilayer hierarchical cortical processing system, such as the ventral visual system, in that the categories that can be formed only at later stages of processing may help earlier stages to form categories relevant to what can be identified at later stages.

immediate consequences of receiving those sensory inputs), then the **A** inputs would simply be more likely to have any representation formed than otherwise, due to strong activation of neurons only when combined **A** and **B** inputs are present.

A similar architecture could be used to provide mild guidance for one sensory system (e.g. olfaction) by another (e.g. taste), as shown in Fig. B.19. (Another example of where this architecture could be used is convergence in the visual system at the next cortical stage of processing, with guiding feedback to influence the categories formed in the different regions of the preceding cortical area, as illustrated in Chapter 11.) The idea is that the taste inputs would be more orthogonal to each other than the olfactory inputs, and that the taste inputs would influence the categories formed in the olfactory input categorizer in layer 1, by feedback from a convergent net in layer 2. The difference from the previous architecture is that we now have a two-layer net, with unimodal or separate networks in layer 1, each feeding forward to a single competitive network in layer 2. The categories formed in layer 2 reflect the co-occurrence of a particular taste with particular odours (which together form flavour in layer 2). Layer 2 then provides feedback connections to both the networks in layer 1. It can be shown in such a network that the categories formed in, for example, the olfactory net in layer 1 are influenced by the tastes with which the odours are paired. The feedback signal is built only in layer 2, after there has been convergence between the different modalities. This architecture

captures some of the properties of sensory systems, in which there are unimodal processing cortical areas followed by multimodal cortical areas. The multimodal cortical areas can build representations that represent the unimodal inputs that tend to co-occur, and the higher level representations may in turn, by the highly developed cortico-cortical backprojections, be able to influence sensory categorization in earlier cortical processing areas (Rolls 1989a). Another such example might be the effect by which the phonemes heard are influenced by the visual inputs produced by seeing mouth movements (cf. McGurk and MacDonald (1976)). This could be implemented by auditory inputs coming together in the cortex in the superior temporal sulcus onto neurons activated by the sight of the lips moving (recorded during experiments of Baylis, Rolls and Leonard (1987), and Hasselmo, Rolls, Baylis and Nalwa (1989b)), using Hebbian learning with co-active inputs. Backprojections from such multimodal areas to the early auditory cortical areas could then influence the responses of auditory cortex neurons to auditory inputs (see Section 4.10 and Fig. 4.5, and cf. Calvert, Bullmore, Brammer, Campbell, Williams, McGuire, Woodruff, Iversen and David (1997)).

A similar principle may operate in any multilayer hierarchical cortical processing system, such as the ventral visual system, in that the categories that can be formed only at later stages of processing may help earlier stages to form categories relevant to what can be identified at the later stages as a result of the operation of backprojections (Rolls 1989a). The idea that the statistical correlation between the inputs received by neighbouring processing streams can be used to guide unsupervised learning within each stream has also been developed by Becker and Hinton (1992) and others (see Phillips, Kay and Smyth (1995)). The networks considered by these authors self-organize under the influence of collateral connections, such as may be implemented by cortico-cortical connections between parallel processing systems in the brain. They use learning rules that, although somewhat complex, are still local in nature, and tend to optimize specific objective functions. The locality of the learning rule, and the simulations performed so far, raise some hope that, once the operation of these types of networks is better understood, they might achieve similar computational capabilities to backpropagation networks (see Section B.11) while retaining biological plausibility.

B.4.6   Topographic map formation

A simple modification to the competitive networks described so far enables them to develop topological maps. In such maps, the closeness in the map reflects the similarity (correlation) between the features in the inputs. The modification that allows such maps to self-organize is to add short-range excitation and long-range inhibition between the neurons. The function to be implemented has a spatial profile that is described as having a Mexican hat shape (see Fig. B.20). The effect of this connectivity between neurons, which need not be modifiable, is to encourage neurons that are close together to respond to similar features in the input space, and to encourage neurons that are far apart to respond to different features in the input space. When these response tendencies are present during learning, the feature analyzers that are built by modifying the synapses from the input onto the activated neurons tend to be similar if they are close together, and different if far apart. This is illustrated in Figs. B.21 and B.22. Feature maps built in this way were described by von der Malsburg (1973) and Willshaw and von der Malsburg (1976). It should be noted that the learning rule needed is simply the modified Hebb rule described above for competitive networks, and is thus local and biologically plausible. (For computational convenience, the algorithm that Kohonen (Kohonen 1982, Kohonen 1989, Kohonen 1995) has mainly used does not use Mexican hat connectivity between the neurons, but instead arranges that when the weights to a winning neuron are updated, so to a smaller extent are those of its neighbours – see further Hertz, Krogh and Palmer (1991).)
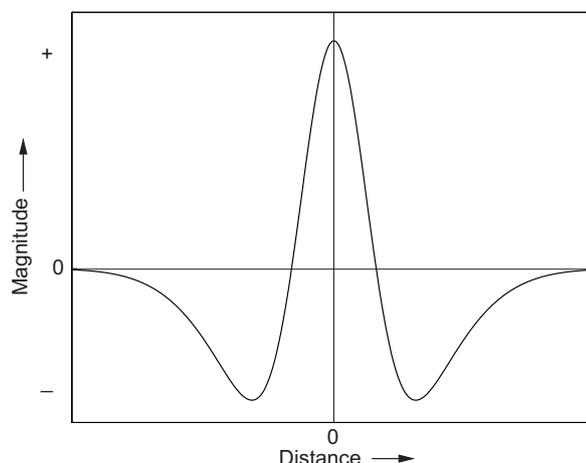
Fig. B.20 Mexican hat lateral spatial interaction profile.

A very common characteristic of connectivity in the brain, found for example throughout the neocortex, consists of short-range excitatory connections between neurons, with inhibition mediated via inhibitory interneurons. The density of the excitatory connectivity even falls gradually as a function of distance from a neuron, extending typically a distance in the order of 1 mm from the neuron (Braitenberg and Schütz 1991), contributing to a spatial function quite like that of a Mexican hat. (Longer-range inhibitory influences would form the negative part of the spatial response profile.) This supports the idea that topological maps, though in some cases probably seeded by chemoaffinity, could develop in the brain with the assistance of the processes just described. It is noted that some cortico-cortical connections even within an area may be longer, skipping past some intermediate neurons, and then making connections after some distance with a further group of neurons. Such longer-range connections are found for example between different columns with similar orientation selectivity in the primary visual cortex. The longer range connections may play a part in stabilizing maps, and again in the exchange of information between neurons performing related computations, in this case about features with the same orientations.

If a low-dimensional space, for example the orientation sensitivity of cortical neurons in the primary visual cortex (which is essentially one-dimensional, the dimension being angle), is mapped to a two-dimensional space such as the surface of the cortex, then the resulting map can have long spatial runs where the value along the dimension (in this case orientation tuning) alters gradually, and continuously. Such self-organization can account for many aspects of the mapping of orientation tuning, and of ocular dominance columns, in V1 (Miller 1994, Harris, Ermentrout and Small 1997). If a high-dimensional information space is mapped to the two-dimensional cortex, then there will be only short runs of groups of neurons with similar feature responsiveness, and then the map must fracture, with a different type of feature mapped for a short distance after the discontinuity. This is exactly what Rolls suggests is the type of topology found in the anterior inferior temporal visual cortex, with the individual groupings representing what can be self-organized by competitive networks combined with a trace rule as described in Section 25.5. Here, visual stimuli are not represented with reference to their position on the retina, because here the neurons are relatively translation invariant. Instead, when recording here, small clumps of neurons with similar responses may be encountered close together, and then one moves into a group of neurons with quite different feature selectivity (personal observations). This topology will arise naturally, given the anatomical
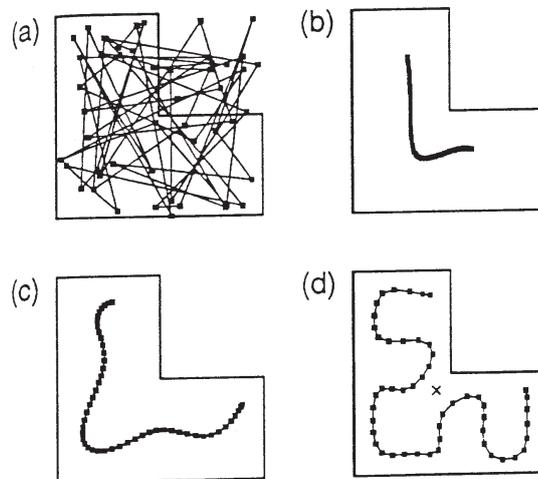
Fig. B.21 Kohonen feature mapping from a two-dimensional L-shaped region to a linear array of 50 units. Each unit has 2 inputs. The input patterns are the X,Y coordinates of points within the L-shape shown. In the diagrams, each point shows the position of a weight vector. Lines connect adjacent units in the 1 D (linear) array of 50 neurons. The weights were initialized to random values within the unit square (a). During feature mapping training, the weights evolved through stages (b) and (c) to (d). By stage (d) the weights have formed so that the positions in the original input space are mapped to a 1 D vector in which adjacent points in the input space activate neighbouring units in the linear array of output units. (Reproduced with permission from Hertz, Krogh and Palmer 1991, Fig. 9.13.)

connectivity of the cortex with its short-range excitatory connections, because there are very many different objects in the world and different types of features that describe objects, with no special continuity between the different combinations of features possible.

Rolls' hypothesis contrasts with the view of Tanaka (1996), who has claimed that the inferior temporal cortex provides an alphabet of visual features arranged in discrete modules. The type of mapping found in higher cortical visual areas as proposed by Rolls implies that topological self-organization is an important way in which maps in the brain are formed, for it seems most unlikely that the locations in the map of different types of object seen in an environment could be specified genetically (Rolls and Stringer 2000). Consistent with this, Tsao, Freiwald, Tootell and Livingstone (2006) described with macaque fMRI 'anterior face patches' at A15 to A22. A15 might correspond to where we have analysed face-selective neurons (it might translate to 3 mm posterior to our sphenoid reference, see Section 25.2), and at this level there are separate regions specialized for face identity in areas TEa and TEm on the ventral lip of the superior temporal sulcus and the adjacent gyrus, and for face expression and movement in the cortex deep in the superior temporal sulcus (Hasselmo, Rolls and Baylis 1989a, Baylis, Rolls and Leonard 1987, Rolls 2007e). The 'middle face patch' of Tsao, Freiwald, Tootell and Livingstone (2006) was at A6, which is probably part of the posterior inferior temporal cortex, and, again consistent with self-organizing map principles, has a high concentration of face-selective neurons within the patch.

The biological utility of developing such topology-preserving feature maps may be that if the computation requires neurons with similar types of response to exchange information more than neurons involved in different computations (which is more than reasonable), then the total length of the connections between the neurons is minimized if the neurons that need
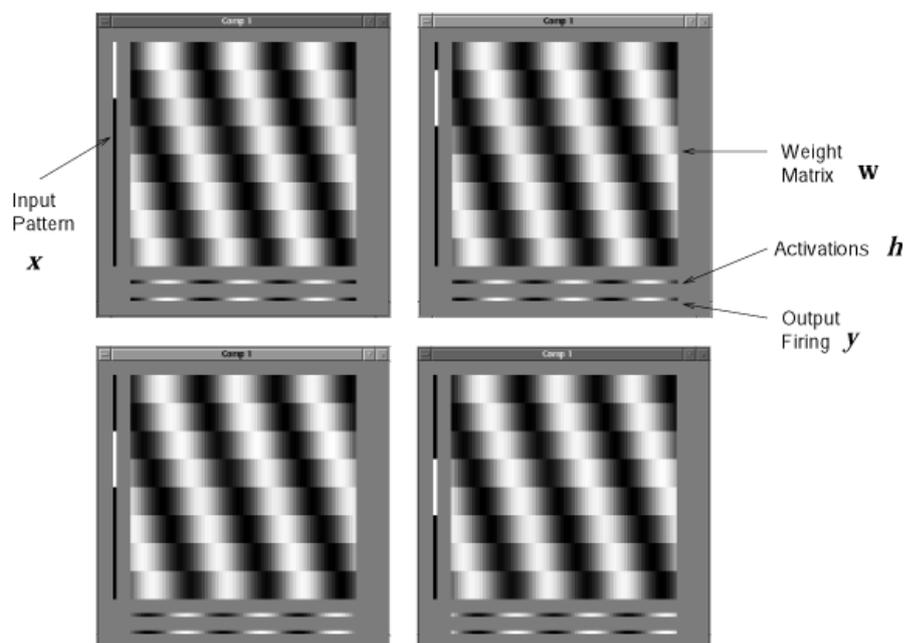
Fig. B.22 Example of a one-dimensional topological map that self-organized from inputs in a low-dimensional space. The network has 64 neurons (vertical elements in the diagram) and 64 inputs per neuron (horizontal elements in the diagram). The four different diagrams represent the net tested with different input patterns. The input patterns x are displayed at the left of each diagram, with white representing firing and black not firing for each of the 64 inputs. The central square of each diagram represents the synaptic weights of the neurons, with white representing a strong weight. The row vector below each weight matrix represents the activations of the 64 output neurons, and the bottom row vector the output firing y. The network was trained with a set of 8 binary input patterns, each of which overlapped in 8 of its 16 'on' elements with the next pattern. The diagram shows that as one moves through correlations in the input space (top left to top right to bottom left to bottom right), so the output neurons activated move steadily across the output array of neurons. Closely correlated inputs are represented close together in the output array of neurons. The way in which this occurs can be seen by inspection of the weight matrix. The network architecture was the same as for a competitive net, except that the activations were converted linearly into output firings, and then each neuron excited its neighbours and inhibited neurons further away. This lateral inhibition was implemented for the simulation by a spatial filter operating on the output firings with the following filter weights (cf. Fig. B.20): 5, 5, 5, 5, 5, 5, 5, 5, 5, 10, 10, 10, 10, 10, 10, 10, 10, 10, 5, 5, 5, 5, 5, 5, 5, 5 which operated on the 64-element firing rate vector.

to exchange information are close together (cf. Cowey (1979), Durbin and Mitchison (1990)). Examples of this include the separation of colour constancy processing in V4 from global motion processing in MT as follows (see Chapter 3 of Rolls and Deco (2002)). In V4, to compute colour constancy, an estimate of the illuminating wavelengths can be obtained by summing the outputs of the pyramidal cells in the inhibitory interneurons over several degrees of visual space, and subtracting this from the excitatory central ON colour-tuned region of the receptive field by (subtractive) feedback inhibition. This enables the cells to discount the illuminating wavelength, and thus compute colour constancy. For this computation, no inputs from motion-selective cells (which in the dorsal stream are colour insensitive) are needed. In MT, to compute global motion (e.g. the motion produced by the average flow of local motion elements, exemplified for example by falling snow), the computation can be performed by averaging in the larger (several degrees) receptive fields of MT the local motion inputs received

by neurons in earlier cortical areas (V1 and V2) with small receptive fields (see Chapter 3 of Rolls and Deco (2002) and Rolls and Stringer (2007)). For this computation, no input from colour cells is useful. Having separate areas (V4 and MT) for these different computations minimizes the wiring lengths, for having intermingled colour and motion cells in a single cortical area would increase the average connection length between the neurons that need to be connected for the computations being performed. Minimizing the total connection length between neurons in the brain is very important in order to keep the size of the brain relatively small.

Placing close to each other neurons that need to exchange information, or that need to receive information from the same source, or that need to project towards the same destination, may also help to minimize the complexity of the rules required to specify cortical (and indeed brain) connectivity (Rolls and Stringer 2000). For example, in the case of V4 and MT, the connectivity rules can be simpler (e.g. connect to neurons in the vicinity, rather than look for colour-, or motion-marked cells, and connect only to the cells with the correct genetically specified label specifying that the cell is either part of motion or of colour processing). Further, the V4 and MT example shows that how the neurons are connected can be specified quite simply, but of course it needs to be specified to be different for different computations. Specifying a general rule for the classes of neurons in a given area also provides a useful simplification to the genetic rules needed to specify the functional architecture of a given cortical area (Rolls and Stringer 2000). In our V4 and MT example, the genetic rules would need to specify the rules separately for different populations of inhibitory interneurons if the computations performed by V4 and MT were performed with intermixed neurons in a single brain area. Together, these two principles, of minimization of wiring length, and allowing simple genetic specification of wiring rules, may underlie the separation of cortical visual information processing into different (e.g. ventral and dorsal) processing streams. The same two principles operating within each brain processing stream may underlie (taken together with the need for hierarchical processing to enable the computations to be biologically plausible in terms of the number of connections per neuron, and the need for local learning rules, see Section B.12) much of the overall architecture of visual cortical processing, and of information processing and its modular architecture throughout the cortex more generally.

The rules of information exchange just described could also tend to produce more gross topography in cortical regions. For example, neurons that respond to animate objects may have certain visual feature requirements in common, and may need to exchange information about these features. Other neurons that respond to inanimate objects might have somewhat different visual feature requirements for their inputs, and might need to exchange information strongly. (For example, selection of whether an object is a chisel or a screwdriver may require competition by mutual (lateral) inhibition to produce the contrast enhancement necessary to result in unambiguous neuronal responses.) The rules just described would account for neurons with responsiveness to inanimate and animate objects tending to be grouped in separate parts of a cortical map or representation, and thus separately susceptible to brain damage (see e.g. Farah (1990), Farah (2000)).

B.4.7 Invariance learning by competitive networks

In conventional competitive learning, the weight vector of a neuron can be thought of as moving towards the centre of a cluster of similar overlapping input stimuli (Rumelhart and Zipser 1985, Hertz, Krogh and Palmer 1991, Rolls and Treves 1998, Rolls and Deco 2002, Perry, Rolls and Stringer 2006, Rolls 2008d). The weight vector points towards the centre of the set of stimuli in the category. The different training stimuli that are placed into the same
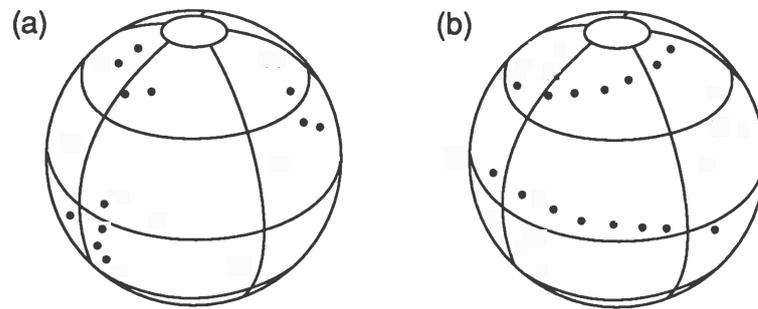
Fig. B.23 (a) Conventional competitive learning. A cluster of overlapping input patterns is categorized as being similar, and this is implemented by a weight vector of an output neuron pointing towards the centre of the cluster. Three clusters are shown, and each cluster might after training have a weight vector pointing towards it. (b) Invariant representation learning. The different transforms of an object may span an elongated region of the space, and the transforms at the far ends of the space may have no overlap (correlation), yet the network must learn to categorize them as similar. The different transforms of two different objects are represented.

category (i.e. activate the same neuron) are typically overlapping in that the pattern vectors are correlated with each other. Figure B.23a illustrates this.

For the formation of invariant representations, there are multiple occurrences of the object at different positions in the space. The object at each position represents a different transform (whether in position, size, view etc.) of the object. The different transforms may be uncorrelated with each other, as would be the case for example with an object translated so far in the space that there would be no active afferents in common between the two transforms. Yet we need these two orthogonal patterns to be mapped to the same output. It may be a very elongated part of the input space that has to be mapped to the same output in invariance learning. These concepts are illustrated in Fig. B.23b.

Objects in the world have temporal and spatial continuity. That is, the statistics of the world are such that we tend to look at one object for some time, during which it may be transforming continuously from one view to another. The temporal continuity property is used in trace rule invariance training, in which a short-term memory in the associative learning rule normally used to train competitive networks is used to help build representations that reflect the continuity in time that characterizes the different transforms of each object, as described in Chapter 25. The transforms of an object also show spatial continuity, and this can also be used in invariance training in what is termed continuous spatial transform learning, described in Section 25.5.11.

In conventional competitive learning the overall weight vector points to the prototypical representation of the object. The only sense in which after normal competitive training (without translations etc) the network generalizes is with respect to the dot product similarity of any input vector compared to the central vector from the training set that the network learns. Continuous spatial transformation learning works by providing a set of training vectors that overlap, and between them cover the whole space over which an invariant transform of the object must be learned. Indeed, it is important for continuous spatial transformation learning that the different exemplars of an object are sufficiently close that the similarity of adjacent training exemplars is sufficient to ensure that the same postsynaptic neuron learns to bridge the continuous space spanned by the whole set of training exemplars of a given object (Stringer,
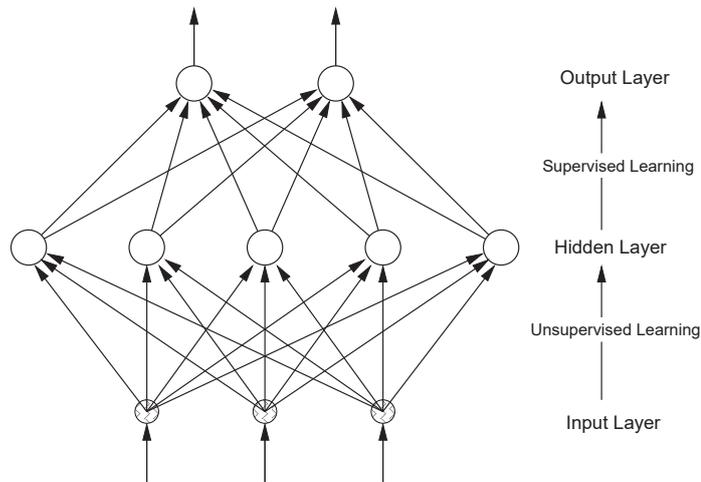
Fig. B.24 A hybrid network, in which for example unsupervised learning rapidly builds relatively orthogonal representations based on input differences, and this is followed by a one-layer supervised network (taught for example by the delta rule) that learns to classify the inputs based on the categorizations formed in the hidden/intermediate layer.

Perry, Rolls and Proske 2006, Perry, Rolls and Stringer 2006, Perry, Rolls and Stringer 2006). This will enable the postsynaptic neuron to span a very elongated space of the different transforms of an object, as described in Section 25.5.11.

B.4.8 Radial Basis Function networks

As noted above, a competitive network can act as a useful preprocessor for other networks. In the neural examples above, competitive networks were useful preprocessors for associative networks. Competitive networks are also used as preprocessors in artificial neural networks, for example in hybrid two-layer networks such as that illustrated in Fig. B.24. The competitive network is advantageous in this hybrid scheme, because as an unsupervised network, it can relatively quickly (with a few presentations of each stimulus) discover the main features in the input space, and code for them. This leaves the second layer of the network to act as a supervised network (taught for example by the delta rule, see Section B.10), which learns to map the features found by the first layer into the output required. This learning scheme is very much faster than that of a (two-layer) backpropagation network, which learns very slowly because it takes it a long time to perform the credit assignment to build useful feature analyzers in layer one (the hidden layer) (see Section B.11).

The general scheme shown in Fig. B.24 is used in radial basis function (RBF) neural networks. The main difference from what has been described is that in an RBF network, the hidden neurons do not use a winner-take-all function (as in some competitive networks), but instead use a normalized activation function in which the measure of distance from a weight vector of the neural input is (instead of the dot product $\mathbf{x} \cdot \mathbf{w}_i$ used for most of the networks described in this book), a Gaussian measure of distance:

$$y_i = \frac{\exp[-(\mathbf{x} - \mathbf{w}_i)^2 / 2\sigma_i^2]}{\sum\limits_k \exp[-(\mathbf{x} - \mathbf{w}_k)^2 / 2\sigma_k^2]}. \tag{B.21}$$

The effect is that the response y_i of neuron i is a maximum if the input stimulus vector $\mathbf{x}$ is centred at $\mathbf{w}_i$, the weight vector of neuron i (this is the upper term in equation B.21). The

magnitude is normalized by dividing by the sum of the activations of all the k neurons in the network. If the input vector $\mathbf{x}$ is not at the centre of the receptive field of the neuron, then the response is decreased according to how far the input vector is from the weight vector $\mathbf{w}_i$ of the neuron, with the weighting decreasing as a Gaussian function with a standard deviation of σ. The idea is like that implemented with soft competition, in that the relative response of different neurons provides an indication of where the input pattern is in relation to the weight vectors of the different neurons. The rapidity with which the response falls off in a Gaussian radial basis function neuron is set by σ_i, which is adjusted so that for any given input pattern vector, a number of RBF neurons are activated. The positions in which the RBF neurons are located (i.e. the directions of their weight vectors, $\mathbf{w}$) are determined usually by unsupervised learning, e.g. the vector quantization that is produced by the normal competitive learning algorithm. The first layer of an RBF network is not different in principle from a network with soft competition, and it is not clear how biologically a Gaussian activation function would be implemented, so the treatment is not developed further here (see Hertz, Krogh and Palmer (1991), Poggio and Girosi (1990a), and Poggio and Girosi (1990b) for further details).

B.4.9 Further details of the algorithms used in competitive networks

B.4.9.1 Normalization of the inputs

Normalization is useful because in step 1 of the training algorithm described in Section B.4.2.2, the neuronal activations, formed by the inner product of the pattern and the normalized weight vector on each neuron, are scaled in such a way that they have a maximum value of 1.0. This helps different input patterns to be equally effective in the learning process. A way in which this normalization could be achieved by a layer of input neurons is given by Grossberg (1976a). In the brain, a number of factors may contribute to normalization of the inputs. One factor is that a set of input axons to a neuron will come from another network in which the firing is controlled by inhibitory feedback, and if the numbers of axons involved is large (hundreds or thousands), then the inputs will be in a reasonable range. Second, there is increasing evidence that the different classes of input to a neuron may activate different types of inhibitory interneuron (e.g. Buhl, Halasy and Somogyi (1994)), which terminate on separate parts of the dendrite, usually close to the site of termination of the corresponding excitatory afferents. This may allow separate feedforward inhibition for the different classes of input. In addition, the feedback inhibitory interneurons also have characteristic termination sites, often on or close to the cell body, where they may be particularly effective in controlling firing of the neuron by shunting (divisive) inhibition, rather than by scaling a class of input (see Section B.6).

B.4.9.2 Normalization of the length of the synaptic weight vector on each dendrite

This is necessary to ensure that one or a few neurons do not always win the competition. (If the weights on one neuron were increased by simple Hebbian learning, and there was no normalization of the weights on the neuron, then it would tend to respond strongly in the future to patterns with some overlap with patterns to which that neuron has previously learned, and gradually that neuron would capture a large number of patterns.) A biologically plausible way to achieve this weight adjustment is to use a modified Hebb rule:

$$\delta w_{ij} = \alpha y_i (x_j - w_{ij}) \tag{B.22}$$

where α is a constant, and x_j and w_{ij} are in appropriate units. In vector notation,

$$\delta \mathbf{w}_i = \alpha y_i (\mathbf{x} - \mathbf{w}_i) \tag{B.23}$$

where $\mathbf{w}_i$ is the synaptic weight vector on neuron i. This implements a Hebb rule that increases synaptic strength according to conjunctive pre- and post-synaptic activity, and also allows the strength of each synapse to decrease in proportion to the firing rate of the postsynaptic neuron (as well as in proportion to the existing synaptic strength). This results in a decrease in synaptic strength for synapses from weakly active presynaptic neurons onto strongly active postsynaptic neurons. Such a modification in synaptic strength is termed heterosynaptic long-term depression in the neurophysiological literature, referring to the fact that the synapses that weaken are other than those that activate the neuron.

This is an important computational use of heterosynaptic long-term depression (LTD). In that the amount of decrease of the synaptic strength depends on how strong the synapse is already, the rule is compatible with what is frequently reported in studies of LTD (see Section 1.5). This rule can maintain the sums of the synaptic weights on each dendrite to be very similar without any need for explicit normalization of the synaptic strengths, and is useful in competitive nets. This rule was used by Willshaw and von der Malsburg (1976). As is made clear with the vector notation above, the modified Hebb rule moves the direction of the weight vector $\mathbf{w}_i$ towards the current input pattern vector $\mathbf{x}$ in proportion to the difference between these two vectors and the firing rate y_i of neuron i.

If explicit weight (vector length) normalization is needed, the appropriate form of the modified Hebb rule is:

$$\delta w_{ij} = \alpha y_i (x_j - y_i w_{ij}). \tag{B.24}$$

This rule, formulated by Oja (1982), makes weight decay proportional to y_i^2, normalizes the synaptic weight vector (see Hertz, Krogh and Palmer (1991)), is still a local learning rule, and is known as the Oja rule.

B.4.9.3 Non-linearity in the learning rule

Non-linearity in the learning rule can assist competition (Rolls 1989b, Rolls 1996c). For example, in the brain, long-term potentiation typically occurs only when strong activation of a neuron has produced sufficient depolarization for the voltage-dependent NMDA receptors to become unblocked, allowing Ca^{2+} to enter the cell (see Section 1.5). This means that synaptic modification occurs only on neurons that are strongly activated, effectively assisting competition to select few winners. The learning rule can be written:

$$\delta w_{ij} = \alpha m_i x_j \tag{B.25}$$

where m_i is a (e.g. threshold) non-linear function of the post-synaptic firing y_i which mimics the operation of the NMDA receptors in learning. (It is noted that in associative networks the same process may result in the stored pattern being more sparse than the input pattern, and that this may be beneficial, especially given the exponential firing rate distribution of neurons, in helping to maximize the number of patterns stored in associative networks (see Sections B.2, B.3, and C.3.1).

B.4.9.4 Competition

In a simulation of a competitive network, a single winner can be selected by searching for the neuron with the maximum activation. If graded competition is required, this can be achieved by an activation function that increases greater than linearly. In some of the networks we have simulated (Rolls 1989b, Rolls 1989a, Wallis and Rolls 1997), raising the activation to a fixed power, typically in the range 2–5, and then rescaling the outputs to a fixed maximum (e.g. 1) is simple to implement. In a real neuronal network, winner-take-all competition can be implemented using mutual (lateral) inhibition between the neurons with non-linear activation

functions, and self-excitation of each neuron (see e.g. Grossberg (1976a), Grossberg (1988), Hertz, Krogh and Palmer (1991)).

Another method to implement soft competition in simulations is to use the normalized exponential or 'softmax' activation function for the neurons (Bridle (1990); see Bishop (1995)):

$$y = \exp(h)/\sum_i \exp(h_i) . \tag{B.26}$$

This function specifies that the firing rate of each neuron is an exponential function of the activation, scaled by the whole vector of activations h_i, $i = 1, N$. The exponential function (in increasing supralinearly) implements soft competition, in that after the competition the faster firing neurons are firing relatively much faster than the slower firing neurons. In fact, the strength of the competition can be adjusted by using a 'temperature' T greater than 0 as follows:

$$y = \exp(h/T)/\sum_i \exp(h_i/T). \tag{B.27}$$

Very low temperatures increase the competition, until with $T \to 0$, the competition becomes 'winner-take-all'. At high temperatures, the competition becomes very soft. (When using the function in simulations, it may be advisable to prescale the firing rates to for example the range 0–1, both to prevent machine overflow, and to set the temperature to operate on a constant range of firing rates, as increasing the range of the inputs has an effect similar to decreasing T.)

The softmax function has the property that activations in the range $-\infty$ to $+\infty$ are mapped into the range 0 to 1.0, and the sum of the firing rates is 1.0. This facilitates interpretation of the firing rates under certain conditions as probabilities, for example that the competitive network firing rate of each neuron reflects the probability that the input vector is within the category or cluster signified by that output neuron (see Bishop (1995)).

B.4.9.5   Soft competition

The use of graded (continuous valued) output neurons in a competitive network, and soft competition rather than winner-take-all competition, has the value that the competitive net generalizes more continuously to an input vector that lies between input vectors that it has learned. Also, with soft competition, neurons with only a small amount of activation by any of the patterns being used will nevertheless learn a little, and move gradually towards the patterns that are being presented. The result is that with soft competition, the output neurons all tend to become allocated to one of the input patterns or one of the clusters of input patterns.

B.4.9.6   Untrained neurons

In competitive networks, especially with winner-take-all or finely tuned neurons, it is possible that some neurons remain unallocated to patterns. This may be useful, in case patterns in the unused part of the space occur in future. Alternatively, unallocated neurons can be made to move towards the parts of the space where patterns are occurring by allowing such losers in the competition to learn a little. Another mechanism is to subtract a bias term μ_i from y_i, and to use a 'conscience' mechanism that raises μ_i if a neuron wins frequently, and lowers μ_i if it wins infrequently (Grossberg 1976b, Bienenstock, Cooper and Munro 1982, De Sieno 1988).

B.4.9.7   Large competitive nets: further aspects

If a large neuronal network is considered, with the number of synapses on each neuron in the region of 10,000, as occurs on large pyramidal cells in some parts of the brain, then there is a potential disadvantage in using neurons with synaptic weights that can take on

only positive values. This difficulty arises in the following way. Consider a set of positive normalized input firing rates and synaptic weight vectors (in which each element of the vector can take on any value between 0.0 and 1.0). Such vectors of random values will on average be more highly aligned with the direction of the central vector (1,1,1,...,1) than with any other vector. An example can be given for the particular case of vectors evenly distributed on the positive 'quadrant' of a high-dimensional hypersphere: the average overlap (i.e. normalized dot product) between two binary random vectors with half the elements on and thus a sparseness of 0.5 (e.g. a random pattern vector and a random dendritic weight vector) will be approximately 0.5, while the average overlap between a random vector and the central vector will be approximately 0.707. A consequence of this will be that if a neuron begins to learn towards several input pattern vectors it will get drawn towards the average of these input patterns which will be closer to the 1,1,1,...,1 direction than to any one of the patterns. As a dendritic weight vector moves towards the central vector, it will become more closely aligned with more and more input patterns so that it is more rapidly drawn towards the central vector. The end result is that in large nets of this type, many of the dendritic weight vectors will point towards the central vector. This effect is not seen so much in small systems, since the fluctuations in the magnitude of the overlaps are sufficiently large that in most cases a dendritic weight vector will have an input pattern very close to it and thus will not learn towards the centre. In large systems, the fluctuations in the overlaps between random vectors become smaller by a factor of $\frac{1}{\sqrt{N}}$ so that the dendrites will not be particularly close to any of the input patterns.

One solution to this problem is to allow the elements of the synaptic weight vectors to take negative as well as positive values. This could be implemented in the brain by feedforward inhibition. A set of vectors taken with random values will then have a reduced mean correlation between any pair, and the competitive net will be able to categorize them effectively. A system with synaptic weights that can be negative as well as positive is not physiologically plausible, but we can instead imagine a system with weights lying on a hypersphere in the positive quadrant of space but with additional inhibition that results in the cumulative effects of some input lines being effectively negative. This can be achieved in a network by using positive input vectors, positive synaptic weight vectors, and thresholding the output neurons at their mean activation. A large competitive network of this general nature does categorize well, and has been described more fully elsewhere (Bennett 1990). In a large network with inhibitory feedback neurons, and principal cells with thresholds, the network could achieve at least in part an approximation to this type of thresholding useful in large competitive networks.

A second way in which nets with positive-only values of the elements could operate is by making the input vectors sparse and initializing the weight vectors to be sparse, or to have a reduced contact probability. (A measure a of neuronal population sparseness is defined (as before) in equation B.28:

$$a = \frac{(\sum\limits_{i=1}^{N} y_i/N)^2}{\sum\limits_{i=1}^{N} y_i^2/N} \tag{B.28}$$

where y_i is the firing rate of the ith neuron in the set of N neurons.) For relatively small net sizes simulated ($N = 100$) with patterns with a sparseness a of, for example, 0.1 or 0.2, learning onto the average vector can be avoided. However, as the net size increases, the sparseness required does become very low. In large nets, a greatly reduced contact probability between neurons (many synapses kept identically zero) would prevent learning of the average vector, thus allowing categorization to occur (see Section 7.4). Reduced contact probability

will, however, prevent complete alignment of synapses with patterns, so that the performance of the network will be affected.

B.5 Continuous attractor networks

B.5.1 Introduction

Single-cell recording studies have shown that some neurons represent the current position along a continuous physical dimension or space even when no inputs are available, for example in darkness (see Chapter 24). Examples include neurons that represent the positions of the eyes (i.e. eye direction with respect to the head), the place where the animal is looking in space, head direction, and the place where the animal is located. In particular, examples of such classes of cells include head direction cells in rats (Ranck 1985, Taube, Muller and Ranck 1990a, Taube, Goodridge, Golob, Dudchenko and Stackman 1996, Muller, Ranck and Taube 1996) and primates (Robertson, Rolls, Georges-François and Panzeri 1999), which respond maximally when the animal's head is facing in a particular preferred direction; place cells in rats (O'Keefe and Dostrovsky 1971, McNaughton, Barnes and O'Keefe 1983, O'Keefe 1984, Muller, Kubie, Bostock, Taube and Quirk 1991, Markus, Qin, Leonard, Skaggs, McNaughton and Barnes 1995) that fire maximally when the animal is in a particular location; and spatial view cells in primates that respond when the monkey is looking towards a particular location in space (Rolls, Robertson and Georges-François 1997a, Georges-François, Rolls and Robertson 1999, Robertson, Rolls and Georges-François 1998). In the parietal cortex there are many spatial representations, in several different coordinate frames (see Chapter 4 of Rolls and Deco (2002) and Andersen, Batista, Snyder, Buneo and Cohen (2000)), and they have some capability to remain active during memory periods when the stimulus is no longer present. Even more than this, the dorsolateral prefrontal cortex networks to which the parietal networks project have the capability to maintain spatial representations active for many seconds or minutes during short-term memory tasks, when the stimulus is no longer present (see Section 4.3.1). In this section, we describe how such networks representing continuous physical space could operate. The locations of such spatial networks in the brain are the parietal areas, the prefrontal areas that implement short-term spatial memory and receive from the parietal cortex (see Section 4.3.1), and the hippocampal system which combines information about objects from the inferior temporal visual cortex with spatial information (see Chapter 24).

A class of network that can maintain the firing of its neurons to represent any location along a continuous physical dimension such as spatial position, head direction, etc. is a 'Continuous Attractor' neural network (CANN). It uses excitatory recurrent collateral connections between the neurons to reflect the distance between the neurons in the state space of the animal (e.g. head direction space). These networks can maintain the bubble of neural activity constant for long periods wherever it is started to represent the current state (head direction, position, etc) of the animal, and are likely to be involved in many aspects of spatial processing and memory, including spatial vision. Global inhibition is used to keep the number of neurons in a bubble or packet of actively firing neurons relatively constant, and to help to ensure that there is only one activity packet. Continuous attractor networks can be thought of as very similar to autoassociation or discrete attractor networks (described in Section B.3), and have the same architecture, as illustrated in Fig. B.25. The main difference is that the patterns stored in a CANN are continuous patterns, with each neuron having broadly tuned firing which decreases with for example a Gaussian function as the distance from the optimal firing location of the cell is varied, and with different neurons having tuning that overlaps throughout the space. Such
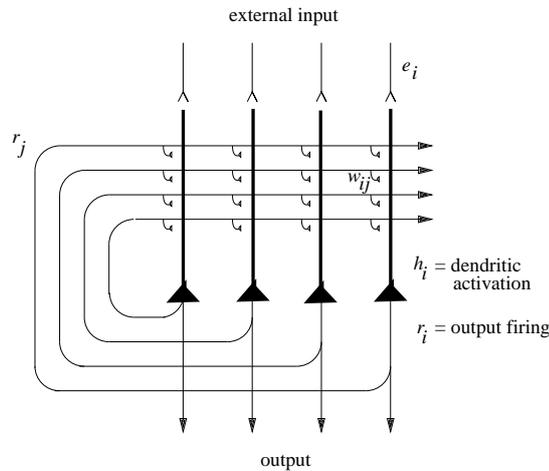
Fig. B.25 The architecture of a continuous attractor neural network (CANN).

tuning is illustrated in Fig. 24.16. For comparison, the autoassociation networks described in Section B.3 have discrete (separate) patterns (each pattern implemented by the firing of a particular subset of the neurons), with no continuous distribution of the patterns throughout the space (see Fig. 24.16). A consequent difference is that the CANN can maintain its firing at any location in the trained continuous space, whereas a discrete attractor or autoassociation network moves its population of active neurons towards one of the previously learned attractor states, and thus implements the recall of a particular previously learned pattern from an incomplete or noisy (distorted) version of one of the previously learned patterns. The energy landscape of a discrete attractor network (see equation B.12) has separate energy minima, each one of which corresponds to a learned pattern, whereas the energy landscape of a continuous attractor network is flat, so that the activity packet remains stable with continuous firing wherever it is started in the state space. (The state space refers to the set of possible spatial states of the animal in its environment, e.g. the set of possible head directions.)

In Section B.5.2, we first describe the operation and properties of continuous attractor networks, which have been studied by for example Amari (1977), Zhang (1996), and Taylor (1999), and then, following Stringer, Trappenberg, Rolls and De Araujo (2002b), address four key issues about the biological application of continuous attractor network models.

One key issue in such continuous attractor neural networks is how the synaptic strengths between the neurons in the continuous attractor network could be learned in biological systems (Section B.5.3).

A second key issue in such continuous attractor neural networks is how the bubble of neuronal firing representing one location in the continuous state space should be updated based on non-visual cues to represent a new location in state space (Section B.5.5). This is essentially the problem of path integration: how a system that represents a memory of where the agent is in physical space could be updated based on idiothetic (self-motion) cues such as vestibular cues (which might represent a head velocity signal), or proprioceptive cues (which might update a representation of place based on movements being made in the space, during for example walking in the dark).

A third key issue is how stability in the bubble of activity representing the current location can be maintained without much drift in darkness, when it is operating as a memory system (Section B.5.6).

A fourth key issue is considered in Section B.5.8 in which we describe networks that store

both continuous patterns and discrete patterns (see Fig. 24.16), which can be used to store for example the location in (continuous, physical) space where an object (a discrete item) is present.

B.5.2 The generic model of a continuous attractor network

The generic model of a continuous attractor is as follows. (The model is described in the context of head direction cells, which represent the head direction of rats (Taube et al. 1996, Muller et al. 1996) and macaques (Robertson, Rolls, Georges-François and Panzeri 1999), and can be reset by visual inputs after gradual drift in darkness.) The model is a recurrent attractor network with global inhibition. It is different from a Hopfield attractor network primarily in that there are no discrete attractors formed by associative learning of discrete patterns. Instead there is a set of neurons that are connected to each other by synaptic weights w_{ij} that are a simple function, for example Gaussian, of the distance between the states of the agent in the physical world (e.g. head directions) represented by the neurons. Neurons that represent similar states (locations in the state space) of the agent in the physical world have strong synaptic connections, which can be set up by an associative learning rule, as described in Section B.5.3. The network updates its firing rates by the following 'leaky-integrator' dynamical equations. The continuously changing activation h_i^{HD} of each head direction cell i is governed by the equation

$$\tau \frac{dh_i^{\text{HD}}(t)}{dt} = -h_i^{\text{HD}}(t) + \frac{\phi_0}{C^{\text{HD}}} \sum_j (w_{ij} - w^{\text{inh}}) r_j^{\text{HD}}(t) + I_i^V, \qquad \text{(B.29)}$$

where r_j^{HD} is the firing rate of head direction cell j, w_{ij} is the excitatory (positive) synaptic weight from head direction cell j to cell i, w^{inh} is a global constant describing the effect of inhibitory interneurons, and τ is the time constant of the system[38]. The term $-h_i^{\text{HD}}(t)$ indicates the amount by which the activation decays (in the leaky integrator neuron) at time t. (The network is updated in a typical simulation at much smaller timesteps than the time constant of the system, τ.) The next term in equation B.29 is the input from other neurons in the network r_j^{HD} weighted by the recurrent collateral synaptic connections w_{ij} (scaled by a constant ϕ_0 and C^{HD} which is the number of synaptic connections received by each head direction cell from other head direction cells in the continuous attractor). The term I_i^V represents a visual input to head direction cell i. Each term I_i^V is set to have a Gaussian response profile in most continuous attractor networks, and this sets the firing of the cells in the continuous attractor to have Gaussian response profiles as a function of where the agent is located in the state space (see e.g. Fig. 24.16 on page 496), but the Gaussian assumption is not crucial. (It is known that the firing rates of head direction cells in both rats (Taube, Goodridge, Golob, Dudchenko and Stackman 1996, Muller, Ranck and Taube 1996) and macaques (Robertson, Rolls, Georges-François and Panzeri 1999) is approximately Gaussian.) When the agent is operating without visual input, in memory mode, then the term I_i^V is set to zero. The firing rate r_i^{HD} of cell i is determined from the activation h_i^{HD} and the sigmoid function

$$r_i^{\text{HD}}(t) = \frac{1}{1 + e^{-2\beta(h_i^{\text{HD}}(t) - \alpha)}}, \qquad \text{(B.30)}$$

where α and β are the sigmoid threshold and slope, respectively.

---

[38]Note that for this section, we use r rather than y to refer to the firing rates of the neurons in the network, remembering that, because this is a recurrently connected network (see Fig. B.13), the output from a neuron y_i might be the input x_j to another neuron.

B.5.3 Learning the synaptic strengths between the neurons that implement a continuous attractor network

So far we have said that the neurons in the continuous attractor network are connected to each other by synaptic weights w_{ij} that are a simple function, for example Gaussian, of the distance between the states of the agent in the physical world (e.g. head directions, spatial views etc) represented by the neurons. In many simulations, the weights are set by formula to have weights with these appropriate Gaussian values. However, Stringer, Trappenberg, Rolls and De Araujo (2002b) showed how the appropriate weights could be set up by learning. They started with the fact that since the neurons have broad tuning that may be Gaussian in shape, nearby neurons in the state space will have overlapping spatial fields, and will thus be co-active to a degree that depends on the distance between them. They postulated that therefore the synaptic weights could be set up by associative learning based on the co-activity of the neurons produced by external stimuli as the animal moved in the state space. For example, head direction cells are forced to fire during learning by visual cues in the environment that produce Gaussian firing as a function of head direction from an optimal head direction for each cell. The learning rule is simply that the weights w_{ij} from head direction cell j with firing rate r_j^{HD} to head direction cell i with firing rate r_i^{HD} are updated according to an associative (Hebb) rule

$$\delta w_{ij} = k r_i^{\mathrm{HD}} \, r_j^{\mathrm{HD}} \tag{B.31}$$

where δw_{ij} is the change of synaptic weight and k is the learning rate constant. During the learning phase, the firing rate r_i^{HD} of each head direction cell i might be the following Gaussian function of the displacement of the head from the optimal firing direction of the cell

$$r_i^{\mathrm{HD}} = e^{-s_{\mathrm{HD}}^2 / 2\sigma_{\mathrm{HD}}^2}, \tag{B.32}$$

where s_{HD} is the difference between the actual head direction x (in degrees) of the agent and the optimal head direction x_i for head direction cell i, and σ_{HD} is the standard deviation.

Stringer, Trappenberg, Rolls and De Araujo (2002b) showed that after training at all head directions, the synaptic connections develop strengths that are an almost Gaussian function of the distance between the cells in head direction space, as shown in Fig. B.26 (left). Interestingly if a non-linearity is introduced into the learning rule that mimics the properties of NMDA receptors by allowing the synapses to modify only after strong postsynaptic firing is present, then the synaptic strengths are still close to a Gaussian function of the distance between the connected cells in head direction space (see Fig. B.26, left). They showed that after training, the continuous attractor network can support stable activity packets in the absence of visual inputs (see Fig. B.26, right) provided that global inhibition is used to prevent all the neurons becoming activated. (The exact stability conditions for such networks have been analyzed by Amari (1977)). Thus Stringer, Trappenberg, Rolls and De Araujo (2002b) demonstrated biologically plausible mechanisms for training the synaptic weights in a continuous attractor using a biologically plausible local learning rule.

Stringer, Trappenberg, Rolls and De Araujo (2002b) went on to show that if there was a short term memory trace built into the operation of the learning rule, then this could help to produce smooth weights in the continuous attractor if only incomplete training was available, that is if the weights were trained at only a few locations. The same rule can take advantage in training the synaptic weights of the temporal probability distributions of firing when they happen to reflect spatial proximity. For example, for head direction cells the agent will necessarily move through similar head directions before reaching quite different head directions, and so the temporal proximity with which the cells fire can be used to set up the appropriate synaptic weights. This new proposal for training continuous
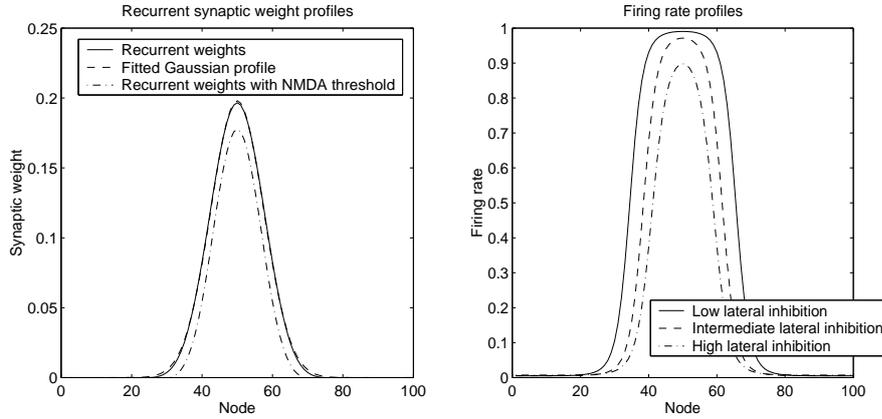
Fig. B.26 Training the weights in a continuous attractor network with an associative rule (equation B.31). Left: the trained recurrent synaptic weights from head direction cell 50 to the other head direction cells in the network arranged in head direction space (solid curve). The dashed line shows a Gaussian curve fitted to the weights shown in the solid curve. The dash-dot curve shows the recurrent synaptic weights trained with rule equation (B.31), but with a non-linearity introduced that mimics the properties of NMDA receptors by allowing the synapses to modify only after strong postsynaptic firing is present. Right: the stable firing rate profiles forming an activity packet in the continuous attractor network during the testing phase when the training (visual) inputs are no longer present. The firing rates are shown after the network has been initially stimulated by visual input to initialize an activity packet, and then allowed to settle to a stable activity profile without visual input. The three graphs show the firing rates for low, intermediate and high values of the lateral inhibition parameter w^{inh}. For both left and right plots, the 100 head direction cells are arranged according to where they fire maximally in the head direction space of the agent when visual cues are available. (After Stringer, Trappenberg, Rolls and de Araujo 2002.)

attractor networks can also help to produce broadly tuned spatial cells even if the driving (e.g. visual) input (I_i^V in equation B.29) during training produces rather narrowly tuned neuronal responses. The learning rule with such temporal properties is a memory trace learning rule that strengthens synaptic connections between neurons, based on the temporal probability distribution of the firing. There are many versions of such rules (Rolls and Milward 2000, Rolls and Stringer 2001a), which are described more fully in Chapter 25, but a simple one that works adequately is

$$\delta w_{ij} = k\bar{r}_i^{\text{HD}}\bar{r}_j^{\text{HD}} \tag{B.33}$$

where δw_{ij} is the change of synaptic weight, and $\bar{r}^{\text{HD}}$ is a local temporal average or trace value of the firing rate of a head direction cell given by

$$\bar{r}^{\text{HD}}(t + \delta t) = (1 - \eta)r^{\text{HD}}(t + \delta t) + \eta\bar{r}^{\text{HD}}(t) \tag{B.34}$$

where η is a parameter set in the interval [0,1] which determines the contribution of the current firing and the previous trace. For $\eta = 0$ the trace rule (B.33) becomes the standard Hebb rule (B.31), while for $\eta > 0$ learning rule (B.33) operates to associate together patterns of activity that occur close together in time. The rule might allow temporal associations to influence the synaptic weights that are learned over times in the order of 1 s. The memory trace required for operation of this rule might be no more complicated than the continuing firing that is an inherent property of attractor networks, but it could also be implemented by a number of biophysical mechanisms, discussed in Chapter 25. Finally, we note that some long-term depression (LTD) in the learning rule could help to maintain the weights of different neurons equally potent (see Section B.4.9.2 and equation B.22), and could compensate for

irregularity during training in which the agent might be trained much more in some than other locations in the space (see Stringer, Trappenberg, Rolls and De Araujo (2002b)).

B.5.4 The capacity of a continuous attractor network: multiple charts and packets

The capacity of a continuous attractor network can be approached on the following bases. First, as there are no discrete attractor states, but instead a continuous physical space is being represented, some concept of spatial resolution must be brought to bear, that is the number of different positions in the space that can be represented. Second, the number of connections per neuron in the continuous attractor will directly influence the number of different spatial positions (locations in the state space) that can be represented. Third, the sparseness of the representation can be thought of as influencing the number of different spatial locations (in the continuous state space) that can be represented, in a way analogous to that described for discrete attractor networks in equation B.14 (Battaglia and Treves 1998b). That is, if the tuning of the neurons is very broad, then fewer locations in the state space may be represented. Fourth, and very interestingly, if representations of different continuous state spaces, for example maps or charts of different environments, are stored in the same network, there may be little cost of adding extra maps or charts. The reason for this is that the large part of the interference between the different memories stored in such a network arises from the correlations between the different positions in any one map, which are typically relatively high because quite broad tuning of individual cells is common. In contrast, there are in general low correlations between the representations of places in different maps or charts, and therefore many different maps can be simultaneously stored in a continuous attractor network (Battaglia and Treves 1998b).

For a similar reason, it is even possible to have the activity packets that operate in different spaces simultaneously active in a single continuous attractor network of neurons, and to move independently of each other in their respective spaces or charts (Stringer, Rolls and Trappenberg 2004).

B.5.5 Continuous attractor models: path integration

So far, we have considered how spatial representations could be stored in continuous attractor networks, and how the activity can be maintained at any location in the state space in a form of short-term memory when the external (e.g. visual) input is removed. However, many networks with spatial representations in the brain can be updated by internal, self-motion (i.e. idiothetic), cues even when there is no external (e.g. visual) input. Examples are head direction cells in the presubiculum of rats and macaques, place cells in the rat hippocampus, and spatial view cells in the primate hippocampus (see Chapter 24). The major question arises about how such idiothetic inputs could drive the activity packet in a continuous attractor network and, in particular, how such a system could be set up biologically by self-organizing learning.

One approach to simulating the movement of an activity packet produced by idiothetic cues (which is a form of path integration whereby the current location is calculated from recent movements) is to employ a look-up table that stores (taking head direction cells as an example), for every possible head direction and head rotational velocity input generated by the vestibular system, the corresponding new head direction (Samsonovich and McNaughton 1997). Another approach involves modulating the strengths of the recurrent synaptic weights in the continuous attractor on one but not the other side of a currently represented position, so that the stable position of the packet of activity, which requires symmetric connections in different directions from each node, is lost, and the packet moves in the direction of the temporarily increased

Visual input I^{V}

r^{ID}

w^{ID}

w^{RC}

r^{HD}
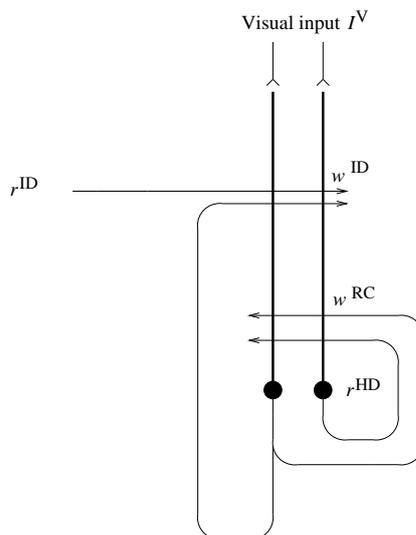
Fig. B.27 General network architecture for a one-dimensional continuous attractor model of head direction cells which can be updated by idiothetic inputs produced by head rotation cell firing r^{ID}. The head direction cell firing is r^{HD}, the continuous attractor synaptic weights are w^{RC}, the idiothetic synaptic weights are w^{ID}, and the external visual input is I^{V}.

weights, although no possible biological implementation was proposed of how the appropriate dynamic synaptic weight changes might be achieved (Zhang 1996). Another mechanism (for head direction cells) (Skaggs, Knierim, Kudrimoti and McNaughton 1995) relies on a set of cells, termed (head) rotation cells, which are co-activated by head direction cells and vestibular cells and drive the activity of the attractor network by anatomically distinct connections for clockwise and counter-clockwise rotation cells, in what is effectively a look-up table. However, no proposal was made about how this could be achieved by a biologically plausible learning process, and this has been the case until recently for most approaches to path integration in continuous attractor networks, which rely heavily on rather artificial pre-set synaptic connectivities.

Stringer, Trappenberg, Rolls and De Araujo (2002b) introduced a proposal with more biological plausibility about how the synaptic connections from idiothetic inputs to a continuous attractor network can be learned by a self-organizing learning process. The essence of the hypothesis is described with Fig. B.27. The continuous attractor synaptic weights w^{RC} are set up under the influence of the external visual inputs I^{V} as described in Section B.5.3. At the same time, the idiothetic synaptic weights w^{ID} (in which the ID refers to the fact that they are in this case produced by idiothetic inputs, produced by cells that fire to represent the velocity of clockwise and anticlockwise head rotation), are set up by associating the change of head direction cell firing that has just occurred (detected by a trace memory mechanism described below) with the current firing of the head rotation cells r^{ID}. For example, when the trace memory mechanism incorporated into the idiothetic synapses w^{ID} detects that the head direction cell firing is at a given location (indicated by the firing r^{HD}) and is moving clockwise (produced by the altering visual inputs I^{V}), and there is simultaneous clockwise head rotation cell firing, the synapses w^{ID} learn the association, so that when that rotation cell firing occurs later without visual input, it takes the current head direction firing in the continuous attractor into account, and moves the location of the head direction attractor in the appropriate direction.

For the learning to operate, the idiothetic synapses onto head direction cell i with firing r_i^{HD} need two inputs: the memory traced term from other head direction cells $\bar{r}_j^{\mathrm{HD}}$ (given by equation B.34), and the head rotation cell input with firing r_k^{ID}; and the learning rule can be written

$$\delta w_{ijk}^{\mathrm{ID}} = \tilde{k}\, r_i^{\mathrm{HD}}\, \bar{r}_j^{\mathrm{HD}}\, r_k^{\mathrm{ID}}, \tag{B.35}$$

where $\tilde{k}$ is the learning rate associated with this type of synaptic connection. The head rotation cell firing (r_k^{ID}) could be as simple as one set of cells that fire for clockwise head rotation (for which k might be 1), and a second set of cells that fire for anticlockwise head rotation (for which k might be 2).

After learning, the firing of the head direction cells would be updated in the dark (when $I_i^V = 0$) by idiothetic head rotation cell firing r_k^{ID} as follows

$$\tau \frac{\mathrm{d}h_i^{\mathrm{HD}}(t)}{\mathrm{d}t} = -h_i^{\mathrm{HD}}(t) + \frac{\phi_0}{C^{\mathrm{HD}}} \sum_j (w_{ij} - w^{inh}) r_j^{\mathrm{HD}}(t) + I_i^V$$

$$+ \phi_1 \Big(\frac{1}{C^{\mathrm{HD \times ID}}} \sum_{j,k} w_{ijk}^{\mathrm{ID}} r_j^{\mathrm{HD}} r_k^{\mathrm{ID}} \Big). \tag{B.36}$$

Equation B.36 is similar to equation B.29, except for the last term, which introduces the effects of the idiothetic synaptic weights w_{ijk}^{ID}, which effectively specify that the current firing of head direction cell i, r_i^{HD}, must be updated by the previously learned combination of the particular head rotation now occurring indicated by r_k^{ID}, and the current head direction indicated by the firings of the other head direction cells r_j^{HD} indexed through j[39]. This makes it clear that the idiothetic synapses operate using combinations of inputs, in this case of two inputs. Neurons that sum the effects of such local products are termed Sigma-Pi neurons (see Section A.2.3). Although such synapses are more complicated than the two-term synapses used throughout the rest of this book, such three-term synapses appear to be useful to solve the computational problem of updating representations based on idiothetic inputs in the way described. Synapses that operate according to Sigma-Pi rules might be implemented in the brain by a number of mechanisms described by Koch (1999) (Section 21.1.1), Jonas and Kaczmarek (1999), and Stringer, Trappenberg, Rolls and De Araujo (2002b), including having two inputs close together on a thin dendrite, so that local synaptic interactions would be emphasized.

Simulations demonstrating the operation of this self-organizing learning to produce movement of the location being represented in a continuous attractor network were described by Stringer, Trappenberg, Rolls and De Araujo (2002b), and one example of the operation is shown in Fig. B.28. They also showed that, after training with just one value of the head rotation cell firing, the network showed the desirable property of moving the head direction being represented in the continuous attractor by an amount that was proportional to the value of the head rotation cell firing. Stringer, Trappenberg, Rolls and De Araujo (2002b) also describe a related model of the idiothetic cell update of the location represented in a continuous attractor, in which the rotation cell firing directly modulates in a multiplicative way the strength of the recurrent connections in the continuous attractor in such a way that clockwise rotation cells modulate the strength of the synaptic connections in the clockwise direction in the continuous attractor, and vice versa.

---

[39]The term $\phi_1/C^{\mathrm{HD \times ID}}$ is a scaling factor that reflects the number $C^{\mathrm{HD \times ID}}$ of inputs to these synapses, and enables the overall magnitude of the idiothetic input to each head direction cell to remain approximately the same as the number of idiothetic connections received by each head direction cell is varied.

Firing rates during rotation
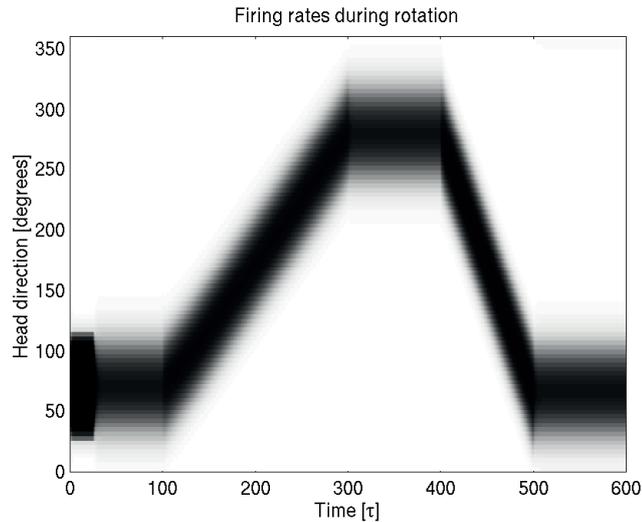
Fig. B.28 Idiothetic update of the location represented in a continuous attractor network. The firing rate of the cells with optima at different head directions (organized according to head direction on the ordinate) is shown by the blackness of the plot, as a function of time. The activity packet was initialized to a head direction of 75 degrees, and the packet was allowed to settle without visual input. For timestep=0 to 100 there was no rotation cell input, and the activity packet in the continuous attractor remained stable at 75 degrees. For timestep=100 to 300 the clockwise rotation cells were active with a firing rate of 0.15 to represent a moderate angular velocity, and the activity packet moved clockwise. For timestep=300 to 400 there was no rotation cell firing, and the activity packet immediately stopped, and remained still. For timestep=400 to 500 the anti-clockwise rotation cells had a high firing rate of 0.3 to represent a high velocity, and the activity packet moved anticlockwise with a greater velocity. For timestep=500 to 600 there was no rotation cell firing, and the activity packet immediately stopped.

It should be emphasized that although the cells are organized in Fig. B.28 according to the spatial position being represented, there is no need for cells in continuous attractors that represent nearby locations in the state space to be close together, as the distance in the state space between any two neurons is represented by the strength of the connection between them, not by where the neurons are physically located. This enables continuous attractor networks to represent spaces with arbitrary topologies, as the topology is represented in the connection strengths (Stringer, Trappenberg, Rolls and De Araujo 2002b, Stringer, Rolls, Trappenberg and De Araujo 2002a, Stringer, Rolls and Trappenberg 2005, Stringer and Rolls 2002). Indeed, it is this that enables many different charts each with its own topology to be represented in a single continuous attractor network (Battaglia and Treves 1998b).

In the network described so far, self-organization occurs, but one set of synapses is Sigma-Pi. We have gone on to show that the Sigma-Pi synapses are not necessary, and can be replaced by a competitive network that learns to respond to combinations of the spatial position and the idiothetic velocity, as illustrated in Fig. B.29 (Stringer and Rolls 2006).

B.5.6 Stabilization of the activity packet within the continuous attractor network when the agent is stationary

With irregular learning conditions (in which identical training with high precision of every node cannot be guaranteed), the recurrent synaptic weights between nodes in the continuous attractor will not be of the perfectly regular and symmetric form normally required in a
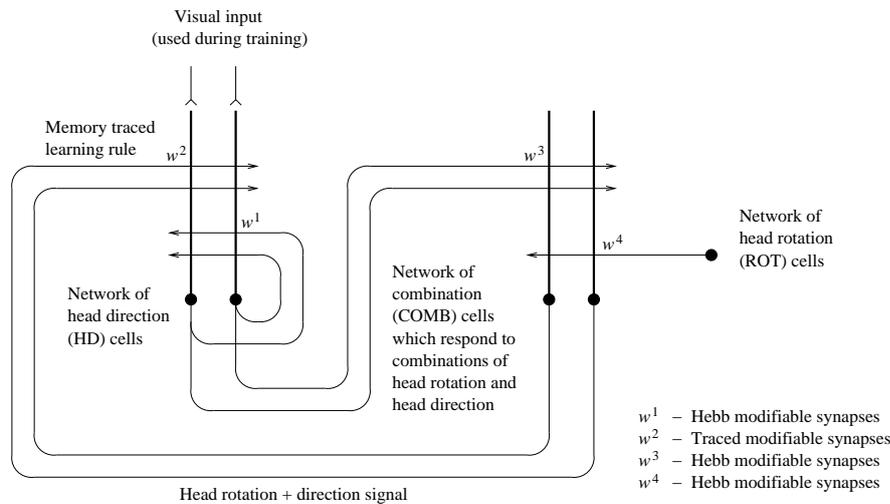
Fig. B.29 Network architecture for a two-layer self-organizing neural network model of the head direction system. The network architecture contains a layer of head direction (HD) cells representing the head direction of the agent, a layer of combination (COMB) cells representing a combination of head direction and rotational velocity, and a layer of rotation (ROT) cells which become active when the agent rotates. There are four types of synaptic connection in the network, which operate as follows. The w_{ij}^1 synapses are Hebb-modifiable recurrent connections between head direction cells. These connections help to support stable packets of activity within this continuous attractor layer of head direction cells in the absence of visual input. The combination cells receive inputs from the head direction cells through the Hebb-modifiable w_{ij}^3 synapses, and inputs from the rotation cells through the Hebb-modifiable w_{ij}^4 synapses. These synaptic inputs encourage combination cells by competitive learning to respond to particular combinations of head direction and rotational velocity. In particular, the combination cells only become significantly active when the agent is rotating. The head direction cells receive inputs from the combination cells through the w_{ij}^2 synapses. The $w_{i,j}^2$ synapses are trained using a 'trace' learning rule, which incorporates a temporal trace of recent combination cell activity. This rule introduces an asymmetry into the w_{ij}^2 weights, which plays an important role in shifting the activity packet through the layer of head direction cells during path integration in the dark. (After Stringer and Rolls 2006.)

continuous attractor neural network. This can lead to drift of the activity packet within the continuous attractor network of e.g. head direction cells when no visual cues are present, even when the agent is not moving. This drift is a common property of the short-term memories of spatial position implemented in the brain, which emphasizes the computational problems that can arise in continuous attractor networks if the weights between nodes are not balanced in different directions in the space being represented. An approach to stabilizing the activity packet when it should not be drifting in real nervous systems, which does help to minimize the drift that can occur, is now described.

The activity packet may be stabilized within a continuous attractor network, when the agent is stationary and the network is operating in memory mode without an external stabilizing input, by enhancing the firing of those cells that are already firing. In biological systems this might be achieved through mechanisms for short-term synaptic enhancement (Koch 1999). Another way is to take advantage of the non-linearity of the activation function of neurons with NMDA receptors, which only contribute to neuronal firing once the neuron is sufficiently depolarized (Wang 1999). The effect is to enhance the firing of neurons that are already reasonably well activated. The effect has been utilized in a model of a network with recurrent excitatory synapses that can maintain active an arbitrary set of neurons that are initially

sufficiently strongly activated by an external stimulus (see Lisman, Fellous and Wang (1998), and, for a discussion on whether these networks could be used to implement short-term memories, see Kesner and Rolls (2001)).

We have incorporated this non-linearity into a model of a head direction continuous attractor network by adjusting the sigmoid threshold α_i (see equation B.30) for each head direction cell i as follows (Stringer, Trappenberg, Rolls and De Araujo 2002b). If the head direction cell firing rate r_i^{HD} is lower than a threshold value, γ, then the sigmoid threshold α_i is set to a relatively high value α^{high}. Otherwise, if the head direction cell firing rate r_i^{HD} is greater than or equal to the threshold value, γ, then the sigmoid threshold α_i is set to a relatively low value α^{low}. It was shown that this procedure has the effect of enhancing the current position of the activity packet within the continuous attractor network, and so prevents the activity packet drifting erratically due to the noise in the recurrent synaptic weights produced for example by irregular learning. An advantage of using the non-linearity in the activation function of a neuron (produced for example by the operation of NMDA receptors) is that this tends to enable packets of activity to be kept active without drift even when the packet is not in one of the energy minima that can result from irregular learning (or from diluted connectivity in the continuous attractor as described below). Thus use of this non-linearity increases the number of locations in the continuous physical state space at which a stable activity packet can be maintained (Stringer, Trappenberg, Rolls and De Araujo 2002b).

The same process might help to stabilize the activity packet against drift caused by the probabilistic spiking of neurons in the network (cf. Chapter 5.6).

B.5.7 Continuous attractor networks in two or more dimensions

Some types of spatial representation used by the brain are of spaces that exist in two or more dimensions. Examples are the two- (or three-) dimensional space representing where one is looking at in a spatial scene. Another is the two- (or three-) dimensional space representing where one is located. It is possible to extend continuous attractor networks to operate in higher dimensional spaces than the one-dimensional spaces considered so far (Taylor 1999, Stringer, Rolls, Trappenberg and De Araujo 2002a). Indeed, it is also possible to extend the analyses of how idiothetic inputs could be used to update two-dimensional state spaces, such as the locations represented by place cells in rats (Stringer, Rolls, Trappenberg and De Araujo 2002a) and the location at which one is looking represented by primate spatial view cells (Stringer, Rolls and Trappenberg 2005, Stringer and Rolls 2002). Interestingly, the number of terms in the synapses implementing idiothetic update do not need to increase beyond three (as in Sigma-Pi synapses) even when higher dimensional state spaces are being considered (Stringer, Rolls, Trappenberg and De Araujo 2002a). Also interestingly, a continuous attractor network can in fact represent the properties of very high dimensional spaces, because the properties of the spaces are captured by the connections between the neurons of the continuous attractor, and these connections are of course, as in the world of discrete attractor networks, capable of representing high dimensional spaces (Stringer, Rolls, Trappenberg and De Araujo 2002a). With these approaches, continuous attractor networks have been developed of the two-dimensional representation of rat hippocampal place cells with idiothetic update by movements in the environment (Stringer, Rolls, Trappenberg and De Araujo 2002a), and of primate hippocampal spatial view cells with idiothetic update by eye and head movements (Stringer, Rolls and Trappenberg 2005, Rolls and Stringer 2005, Stringer and Rolls 2002). Continuous attractor models with some similar properties have also been applied to understanding motor control, for example the generation of a continuous movement in space (Stringer and Rolls 2007, Stringer, Rolls and Taylor 2007a).

B.5.8 Mixed continuous and discrete attractor networks

It has now been shown that attractor networks can store both continuous patterns and discrete patterns, and can thus be used to store for example the location in (continuous, physical) space where an object (a discrete item) is present (see Fig. 24.16 and Rolls, Stringer and Trappenberg (2002)). In this network, when events are stored that have both discrete (object) and continuous (spatial) aspects, then the whole place can be retrieved later by the object, and the object can be retrieved by using the place as a retrieval cue. Such networks are likely to be present in parts of the brain that receive and combine inputs both from systems that contain representations of continuous (physical) space, and from brain systems that contain representations of discrete objects, such as the inferior temporal visual cortex. One such brain system is the hippocampus, which appears to combine and store such representations in a mixed attractor network in the CA3 region, which thus is able to implement episodic memories which typically have a spatial component, for example where an item such as a key is located (see Chapter 24). This network thus shows that in brain regions where the spatial and object processing streams are brought together, then a single network can represent and learn associations between both types of input. Indeed, in brain regions such as the hippocampal system, it is essential that the spatial and object processing streams are brought together in a single network, for it is only when both types of information are in the same network that spatial information can be retrieved from object information, and vice versa, which is a fundamental property of episodic memory (see Chapter 24). It may also be the case that in the prefrontal cortex, attractor networks can store both spatial and discrete (e.g. object-based) types of information in short-term memory (see Section 4.3.1).

B.6 Network dynamics: the integrate-and-fire approach

The concept that attractor (autoassociation) networks can operate very rapidly if implemented with neurons that operate dynamically in continuous time was introduced in Section B.3.3.5. The result described was that the principal factor affecting the speed of retrieval is the time constant of the synapses between the neurons that form the attractor ((Treves 1993, Rolls and Treves 1998, Battaglia and Treves 1998a, Panzeri, Rolls, Battaglia and Lavis 2001). This was shown analytically by Treves (1993), and described by Rolls and Treves (1998) Appendix 5. We now describe in more detail the approaches that produce these results, and the actual results found on the speed of processing.

The networks described so far in this chapter, and analyzed in Appendices 3 and 4 of Rolls and Treves (1998), were described in terms of the steady-state activation of networks of neuron-like units. Those may be referred to as 'static' properties, in the sense that they do not involve the time dimension. In order to address 'dynamical' questions, the time dimension has to be reintroduced into the formal models used, and the adequacy of the models themselves has to be reconsidered in view of the specific properties to be discussed.

Consider for example a real network whose operation has been described by an autoassociative formal model that acquires, with learning, a given attractor structure. How does the state of the network approach, in real time during a retrieval operation, one of those attractors? How long does it take? How does the amount of information that can be read off the network's activity evolve with time? Also, which of the potential steady states is indeed a stable state that can be reached asymptotically by the net? How is the stability of different states modulated by external agents? These are examples of dynamical properties, which to be studied require the use of models endowed with some dynamics.

B.6.1 From discrete to continuous time

Already at the level of simple models in which each unit is described by an input–output relation, one may introduce equally simple 'dynamical' rules, in order both to fully specify the model, and to simulate it on computers. These rules are generally formulated in terms of 'updatings': time is considered to be discrete, a succession of time steps, and at each time step the output of one or more of the units is set, or updated, to the value corresponding to its input variable. The input variable may reflect the outputs of other units in the net as updated at the previous time step or, if delays are considered, the outputs as they were at a prescribed number of time steps in the past. If all units in the net are updated together, the dynamics is referred to as parallel; if instead only one unit is updated at each time step, the dynamics is sequential. (One main difference between the Hopfield (1982) model of an autoassociator and a similar model considered earlier by Little (1974) is that the latter was based on parallel rather than sequential dynamics.) Many intermediate possibilities obviously exist, involving the updating of groups of units at a time. The order in which sequential updatings are performed may for instance be chosen at random at the beginning and then left the same in successive cycles across all units in the net; or it may be chosen anew at each cycle; yet a third alternative is to select at each time step a unit, at random, with the possibility that a particular unit may be selected several times before some of the other ones are ever updated. The updating may also be made probabilistic, with the output being set to its new value only with a certain probability, and otherwise remaining at the current value.

Variants of these dynamical rules have been used for decades in the analysis and computer simulation of physical systems in statistical mechanics (and field theory). They can reproduce in simple but effective ways the stochastic nature of transitions among discrete quantum states, and they have been subsequently considered appropriate also in the simulation of neural network models in which units have outputs that take discrete values, implying that a change from one value to another can only occur in a sudden jump. To some extent, different rules are equivalent, in that they lead, in the evolution of the activity of the net along successive steps and cycles, to the same set of possible steady states. For example, it is easy to realize that when no delays are introduced, states that are stable under parallel updating are also stable under sequential updating. The reverse is not necessarily true, but on the other hand states that are stable when updating one unit at a time are stable irrespective of the updating order. Therefore, static properties, which can be deduced from an analysis of stable states, are to some extent robust against differences in the details of the dynamics assigned to the model. (This is a reason for using these dynamical rules in the study of the thermodynamics of physical systems.) Such rules, however, bear no relation to the actual dynamical processes by which the activity of real neurons evolves in time, and are therefore inadequate for the discussion of dynamical issues in neural networks.

A first step towards realism in the dynamics is the substitution of discrete time with continuous time. This somewhat parallels the substitution of the discrete output variables of the most rudimentary models with continuous variables representing firing rates. Although continuous output variables may evolve also in discrete time, and as far as static properties are concerned differences are minimal, with the move from discrete to continuous outputs the main raison d'etre for a dynamics in terms of sudden updatings ceases to exist, since continuous variables can change continuously in continuous time. A paradox arises immediately, however, if a continuous time dynamics is assigned to firing rates. The paradox is that firing rates, although in principle continuous if computed with a generic time-kernel, tend to vary in jumps as new spikes – essentially discrete events – come to be included in the kernel. To avoid this paradox, a continuous time dynamics can be assigned, instead, to instantaneous continuous variables such as membrane potentials. Hopfield (1984), among others, has introduced a

model of an autoassociator in which the output variables represent membrane potentials and evolve continuously in time, and has suggested that under certain conditions the stable states attainable by such a network are essentially the same as for a network of binary units evolving in discrete time. If neurons in the central nervous system communicated with each other via the transmission of graded membrane potentials, as they do in some peripheral and invertebrate neural systems, this model could be an excellent starting point. The fact that, centrally, transmission is primarily via the emission of discrete spikes makes a model based on membrane potentials as output variables inadequate to correctly represent spiking dynamics.

B.6.2 Continuous dynamics with discontinuities

In principle, a solution would be to keep the membrane potential as the basic dynamical variable, evolving in continuous time, and to use as the output variable the spike emission times, as determined by the rapid variation in membrane potential corresponding to each spike. A point-like neuron can generate spikes by altering the membrane potential V according to continuous equations of the Hodgkin–Huxley type:

$$C\frac{\mathrm{d}V}{\mathrm{d}t} = g_0(V_{\text{rest}} - V) + g_{\text{Na}}mh^3(V_{\text{Na}} - V) + g_{\text{K}}n^4(V_{\text{K}} - V) + I \tag{B.37}$$

$$\tau_m\frac{\mathrm{d}m}{\mathrm{d}t} = m_\infty(V) - m \tag{B.38}$$

$$\tau_h\frac{\mathrm{d}h}{\mathrm{d}t} = h_\infty(V) - h \tag{B.39}$$

$$\tau_n\frac{\mathrm{d}n}{\mathrm{d}t} = n_\infty(V) - n. \tag{B.40}$$

The changes in the membrane potential, driven by the input current I, interact with the opening and closing of intrinsic conductances (here a sodium conductance, whose channels are gated by the 'particles' m and h, and a potassium conductance, whose channels are gated by n; Hodgkin and Huxley (1952)). These equations provide an effective description, phenomenological but broadly based on physical principles, of the conductance changes underlying action potentials, and they are treated in any standard neurobiology text.

From the point of view of formal models of neural networks, this level of description is too complicated to be the basis for an analytic understanding of the operation of networks, and it must be simplified. The most widely used simplification is the so-called integrate-and-fire model (see for example MacGregor (1987) and Brunel and Wang (2001)), which is legitimized by the observation that (sodium) action potentials are typically brief and self-similar events. If, in particular, the only relevant variable associated with the spike is its time of emission (at the soma, or axon hillock), which essentially coincides with the time the potential V reaches a certain threshold level V_{thr}, then the conductance changes underlying the rest of the spike can be omitted from the description, and substituted with the ad hoc prescription that (i) a spike is emitted, with its effect on receiving units and on the unit itself; and (ii) after a brief time corresponding to the duration of the spike plus a refractory period, the membrane potential is reset and resumes its integration of the input current I. After a spike the membrane potential is taken to be reset to a value V_{reset}. This type of simplified dynamics of the membrane potentials is thus in continuous time with added discontinuities: continuous in between spikes, with discontinuities occurring at different times for each neuron in a population, every time a neuron emits a spike.

A leaky integrate-and-fire neuron can be modelled as follows. The model describes the depolarization of the membrane potential V (which typically is dynamically changing as a
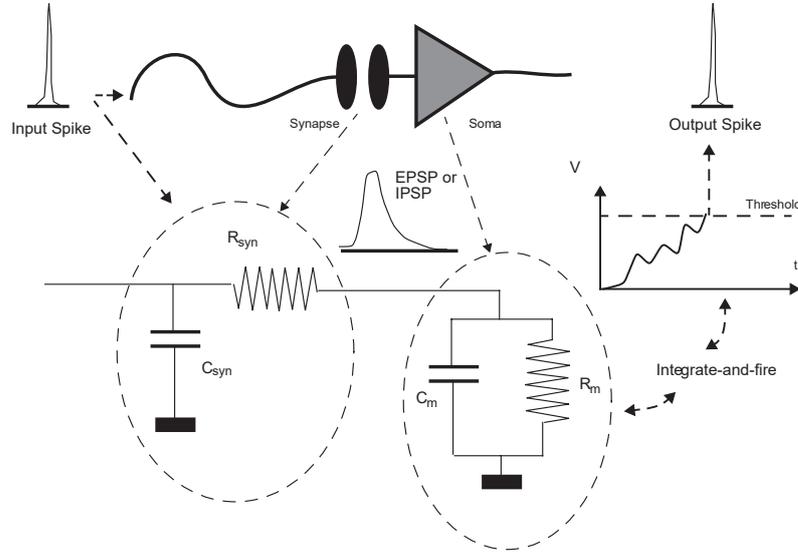
Fig. B.30 Integrate-and-fire neuron. The basic circuit of an integrate-and-fire model consists of the neuron's membrane capacitance C_m in parallel with the membrane's resistance R_m (the reciprocal of the membrane conductance g_m) driven by a synaptic current with a conductance and time constant determined by the synaptic resistance R_{syn} (the reciprocal of the synaptic conductance g_j) and capacitance C_{syn} shown in the Figure. These effects produce excitatory or inhibitory post-synaptic potentials, EPSPs or IPSPs. These potentials are integrated by the cell, and if a threshold V_{thr} is reached a δ-pulse (spike) is fired and transmitted to other neurons, and the membrane potential is reset. (After Deco and Rolls 2003.)

result of the synaptic effects described below between approximately –70 and –50 mV) until threshold V_{thr} (typically –50 mV) is reached when a spike is emitted and the potential is reset to V_{reset} (typically –55 mV). The membrane time constant τ_m is set by the membrane capacitance C_m and the membrane leakage conductance g_m where $\tau_m = C_m/g_m$. V_L denotes the resting potential of the cell, typically –70 mV. Changes in the membrane potential are defined by the following equation

$$C_m \frac{dV(t)}{dt} = g_m\big(V_L - V(t)\big) + \sum_j g_j\big(V_{AMPA} - V(t)\big) + \sum_j g_j\big(V_{GABA} - V(t)\big). \quad \text{(B.41)}$$

The first term on the right of the equation describes how the membrane potential decays back towards the resting potential of the cell depending on how far the cell potential V is from the resting potential V_L, and the membrane leak conductance g_m. The second term on the right represents the excitatory synaptic current that could be injected through AMPA receptors. This is a sum over all AMPA synapses indexed by j on the cell. At each synapse, the current is driven into the cell by the difference between the membrane potential V and the reversal potential V_{AMPA} of the channels opened by AMPA receptors, weighted by the synaptic conductance g_j. This synaptic conductance changes dynamically as a function of the time since a spike reached the synapse, as shown below. Due to their reversal potential V_{AMPA} of typically 0 mV, these currents will tend to depolarize the cell, that is to move the membrane potential towards the firing threshold. The third term on the right represents the inhibitory synaptic current that could be injected through GABA receptors. Due to their reversal potential V_{GABA} of typically –70 mV, these currents will tend to hyperpolarize the cell, that is to move the membrane potential away from the firing threshold. There may be
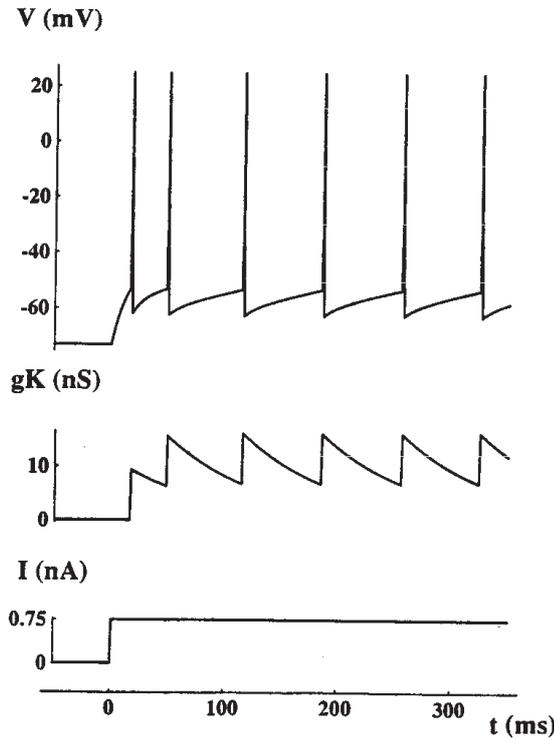
Fig. B.31 Model behaviour of an integrate-and-fire neuron: the membrane potential and adaptation-producing potassium conductance in response to a step of injected current. The spikes were added to the graph by hand, as they do not emerge from the simplified voltage equation. (Reprinted with permission from Network 4, Mean-field analysis of neuronal spike dynamics, Treves, A., 259–284. Copyright ©Taylor and Francis (1993).)

other types of receptor on a cell, for example NMDA receptors, and these operate analogously though with interesting differences, as described in Section B.6.3.

The opening of each synaptic conductance g_j is driven by the arrival of spikes at the presynaptic terminal j, and its closing can often be described as a simple exponential process. A simplified equation for the dynamics of $g_j(t)$ is then

$$\frac{\mathrm{d}g_j(t)}{\mathrm{d}t} = -\frac{g_j(t)}{\tau} + \Delta g_j \sum_l \delta(t - \Delta t - t_l). \tag{B.42}$$

According to the above equation, each synaptic conductance opens instantaneously by a fixed amount Δg_j a time Δt after the emission of the presynaptic spike at t_l. Δt summarizes delays (axonal, synaptic, and dendritic), and each opening superimposes linearly, without saturating, on previous openings. The value of τ in this equation will be different for AMPA receptors (typically 2–10 ms), NMDA receptors (typically 100 ms), and GABA receptors (typically 10 ms).

In order to model the phenomenon of adaptation in the firing rate, prominent especially with pyramidal cells, it is possible to include a time-varying intrinsic (potassium-like) conductance in the cell membrane (Brown, Gähwiler, Griffith and Halliwell 1990a) (shown as g_K in equations B.43 and B.44). This can be done by specifying that this conductance, which if
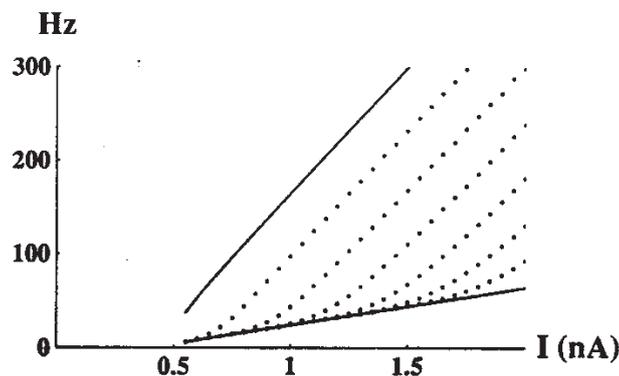
Fig. B.32 Current-to-frequency transduction in a pyramidal cell modelled as an integrate-and-fire neuron. The top solid curve is the firing frequency in the absence of adaptation, $\Delta g_K = 0$. The dotted curves are the instantaneous frequencies computed as the inverse of the ith interspike interval (top to bottom, $i = 1, ..., 6$). The bottom solid curve is the adapted firing curve ($i \to \infty$). With or without adaptation, the input–output transform is close to threshold–linear. (Reprinted with permission from Network 4, Mean-field analysis of neuronal spike dynamics, Treves, A., 259–284. Copyright ©Taylor and Francis (1993).)

open tends to shunt the membrane and thus to prevent firing, opens by a fixed amount with the potential excursion associated with each spike, and then relaxes exponentially to its closed state. In this manner sustained firing driven by a constant input current occurs at lower rates after the first few spikes, in a way similar, if the relevant parameters are set appropriately, to the behaviour observed in vitro of many pyramidal cells (for example, Lanthorn, Storm and Andersen (1984), Mason and Larkman (1990)).

The equations for the dynamics of each neuron with adaptation are then

$$C_m \frac{dV(t)}{dt} = g_m\big(V_L - V(t)\big) + \sum_j g_j\big(V_{AMPA} - V(t)\big) + \sum_j g_j\big(V_{GABA} - V(t)\big) + g_K(t)(V_K - V(t))$$
(B.43)

and

$$\frac{dg_K(t)}{dt} = -\frac{g_K(t)}{\tau_K} + \sum_k \Delta g_K \delta(t - t_k)$$
(B.44)

supplemented by the prescription that when at time $t = t_{k+1}$ the potential reaches the level V_{thr}, a spike is emitted, and hence included also in the sum of equation B.44, and the potential resumes its evolution according to equation B.43 from the reset level V_{reset}. The resulting behaviour is exemplified in Fig. B.31, while Fig. B.32 shows the input–output transform (current to frequency transduction) operated by an integrate-and-fire unit of this type with firing rate adaptation. One should compare this with the transduction operated by real cells, as exemplified for example in Fig. 1.5.

It should be noted that synaptic conductance dynamics is not always included in integrate-and-fire models: sometimes it is substituted with current dynamics, which essentially amounts to neglecting non-linearities due to the appearance of the membrane potential in the driving force for synaptic action (see for example, Amit and Tsodyks (1991), Gerstner (1995)); and sometimes it is simplified altogether by assuming that the membrane potential undergoes small sudden jumps when it receives instantaneous pulses of synaptic current (see the review

in Gerstner (1995)). The latter simplification is quite drastic and changes the character of the dynamics markedly; whereas the former can be a reasonable simplification in some circumstances, but it produces serious distortions in the description of inhibitory GABA$_A$ currents, which, having an equilibrium (Cl$^-$) synaptic potential close to the operating range of the membrane potential, are quite sensitive to the instantaneous value of the membrane potential itself.

B.6.3   An integrate-and-fire implementation

In this subsection the mathematical equations that describe the spiking activity and synapse dynamics in some of the integrate-and-fire simulations performed by Rolls and Deco (Deco and Rolls 2003, Deco, Rolls and Horwitz 2004, Deco and Rolls 2005d, Deco and Rolls 2005c, Deco and Rolls 2005b, Deco and Rolls 2006, Loh, Rolls and Deco 2007a, Rolls, Loh and Deco 2008c, Rolls, Grabenhorst and Deco 2010b, Rolls, Grabenhorst and Deco 2010c, Insabato, Pannunzi, Rolls and Deco 2010, Deco, Rolls and Romo 2010, Rolls and Deco 2011b, Webb, Rolls, Deco and Feng 2011, Rolls and Webb 2012, Rolls, Webb and Deco 2012, Martinez-Garcia, Rolls, Deco and Romo 2011, Rolls, Dempere-Marco and Deco 2013, Rolls and Deco 2015a, Rolls and Deco 2015b, Rolls and Deco 2015b, Rolls and Deco 2016) are set out, in order to show in more detail how an integrate-and-fire simulation is implemented. The equations follow in general the formulation described by Brunel and Wang (2001), though each simulation describes its own architecture to be simulated and neurocomputational questions to be addressed, with additional dynamics introduced where described to implement synaptic facilitation and/or synaptic adaptation (Chapter 10).

Each neuron is described by an integrate-and-fire model. The subthreshold membrane potential $V(t)$ of each neuron evolves according to the following equation:

$$C_{\mathrm{m}}\frac{\mathrm{d}V(t)}{\mathrm{d}t} = -g_{\mathrm{m}}(V(t) - V_{\mathrm{L}}) - I_{\mathrm{syn}}(t) \tag{B.45}$$

where $I_{\mathrm{syn}}(t)$ is the total synaptic current flow into the cell, V_{L} is the resting potential, C_{m} is the membrane capacitance, and g_{m} is the membrane conductance. When the membrane potential $V(t)$ reaches the threshold V_{thr} a spike is generated, and the membrane potential is reset to V_{reset}. The neuron is unable to spike during the first τ_{ref} which is the absolute refractory period.

The total synaptic current is given by the sum of glutamatergic excitatory components (NMDA and AMPA) and inhibitory components (GABA). The external excitatory contributions (ext) from outside the network are produced through AMPA receptors ($I_{\mathrm{AMPA,ext}}$), while the excitatory recurrent synapses (rec) within the network act through AMPA and NMDA receptors ($I_{\mathrm{AMPA,rec}}$ and $I_{\mathrm{NMDA,rec}}$). The total synaptic current is therefore given by:

$$I_{\mathrm{syn}}(t) = I_{\mathrm{AMPA,ext}}(t) + I_{\mathrm{AMPA,rec}}(t) + I_{\mathrm{NMDA,rec}}(t) + I_{\mathrm{GABA}}(t) \tag{B.46}$$

where

$$I_{\mathrm{AMPA,ext}}(t) = g_{\mathrm{AMPA,ext}}(V(t) - V_{\mathrm{E}}) \sum_{j=1}^{N_{\mathrm{ext}}} s_j^{\mathrm{AMPA,ext}}(t) \tag{B.47}$$

$$I_{\mathrm{AMPA,rec}}(t) = g_{\mathrm{AMPA,rec}}(V(t) - V_{\mathrm{E}}) \sum_{j=1}^{N_E} w_j s_j^{\mathrm{AMPA,rec}}(t) \tag{B.48}$$

$$I_{\mathrm{NMDA,rec}}(t) = \frac{g_{\mathrm{NMDA,rec}}(V(t) - V_{\mathrm{E}})}{(1 + [\mathrm{Mg}^{++}] \exp(-0.062V(t))/3.57)} \sum_{j=1}^{N_E} w_j s_j^{\mathrm{NMDA,rec}}(t) \tag{B.49}$$

$$I_{\text{GABA}}(t) = g_{\text{GABA}}(V(t) - V_{\text{I}}) \sum_{j=1}^{N_I} s_j^{\text{GABA}}(t) \tag{B.50}$$

In the preceding equations the reversal potential of the excitatory synaptic currents $V_{\text{E}} = 0$ mV and of the inhibitory synaptic currents $V_{\text{I}} = -70$ mV. The different form for the NMDA receptor-activated channels implements the voltage-dependence of NMDA receptors. This voltage-dependency, and the long time constant of the NMDA receptors, are important in the effects produced through NMDA receptors (Brunel and Wang 2001, Wang 1999). The synaptic strengths w_j are specified in the papers by Rolls and Deco cited above, and depend on the architecture being simulated. The fractions of open channels s are given by:

$$\frac{ds_j^{\text{AMPA,ext}}(t)}{dt} = -\frac{s_j^{\text{AMPA,ext}}(t)}{\tau_{\text{AMPA}}} + \sum_k \delta(t - t_j^k) \tag{B.51}$$

$$\frac{ds_j^{\text{AMPA,rec}}(t)}{dt} = -\frac{s_j^{\text{AMPA,rec}}(t)}{\tau_{\text{AMPA}}} + \sum_k \delta(t - t_j^k) \tag{B.52}$$

$$\frac{ds_j^{\text{NMDA,rec}}(t)}{dt} = -\frac{s_j^{\text{NMDA,rec}}(t)}{\tau_{\text{NMDA,decay}}} + \alpha x_j(t)(1 - s_j^{\text{NMDA,rec}}(t)) \tag{B.53}$$

$$\frac{dx_j(t)}{dt} = -\frac{x_j(t)}{\tau_{\text{NMDA,rise}}} + \sum_k \delta(t - t_j^k) \tag{B.54}$$

$$\frac{ds_j^{\text{GABA}}(t)}{dt} = -\frac{s_j^{\text{GABA}}(t)}{\tau_{\text{GABA}}} + \sum_k \delta(t - t_j^k) \tag{B.55}$$

where the sums over k represent a sum over spikes emitted by presynaptic neuron j at time t_j^k. The value of $\alpha = 0.5$ ms^{-1}.

Typical values of the conductances for pyramidal neurons are: $g_{\text{AMPA,ext}}$=2.08, $g_{\text{AMPA,rec}}$=0.052, $g_{\text{NMDA,rec}}$=0.164, and g_{GABA}=0.67 nS; and for interneurons: $g_{\text{AMPA,ext}}$=1.62, $g_{\text{AMPA,rec}}$=0.0405, $g_{\text{NMDA,rec}}$=0.129 and g_{GABA}=0.49 nS.

A full list of the values of the parameters used is given in Section B.8.3 and in each of the papers cited above in this Section B.6.3.

B.6.4 The speed of processing of one-layer attractor networks with integrate-and-fire neurons

Given that the analytic approach to the rapidity of the dynamics of attractor networks with integrate-and-fire dynamics (Treves 1993, Rolls and Treves 1998) applies mainly when the state is close to the attractor basin, it is of interest to check the performance of such networks by simulation when the completion of partial patterns that may be towards the edge of the attractor basin can be tested. Simmen, Rolls and Treves (1996a) and Treves, Rolls and Simmen (1997) made a start with this, and showed that retrieval could indeed be fast, within 1–2 time constants of the synapses. However, they found that they could not load the systems they simulated with many patterns, and the firing rates during the retrieval process tended to be unstable. The cause of this turned out to be that the inhibition they used to maintain the activity level during retrieval was subtractive, and it turns out that divisive (shunting) inhibition is much more effective in such networks, as described by Rolls and Treves (1998) in Appendix 5. Divisive inhibition is likely to be organized by inhibitory inputs that synapse close to the cell body (where the reversal potential is close to that of the channels opened by GABA receptors),

in contrast to synapses on dendrites, where the different potentials result in opening of the same channels producing hyperpolarization (that is an effectively subtractive influence with respect to the depolarizing currents induced by excitatory (glutamate-releasing) terminals). Battaglia and Treves (1998a) therefore went on to study networks with neurons where the inhibitory neurons could be made to be divisive by having them synapse close to the cell body in neurons modelled with multiple (ten) dendritic compartments. The excitatory inputs terminated on the compartments more distant from the cell body in the model. They found that with this divisive inhibition, the neuronal firing during retrieval was kept under much better control, and the number of patterns that could be successfully stored and retrieved was much higher. Some details of their simulation follow.

Battaglia and Treves (1998a) simulated a network of 800 excitatory and 200 inhibitory cells in its retrieval of one of a number of memory patterns stored in the synaptic weights representing the excitatory-to-excitatory recurrent connections. The memory patterns were assigned at random, drawing the value of each unit in each of the patterns from a binary distribution with sparseness 0.1, that is a probability of 1 in 10 for the unit to be active in the pattern. No baseline excitatory weight was included, but the modifiable weights were instead constrained to remain positive (by clipping at zero synaptic modifications that would make a synaptic weight negative), and a simple exponential decay of the weight with successive modifications was also applied, to prevent runaway synaptic modifications within a rudimentary model of forgetting. Both excitation on inhibitory units and inhibition were mediated by non-modifiable uniform synaptic weights, with values chosen so as to satisfy stability conditions of the type shown in equations A5.13 of Rolls and Treves (1998). Both inhibitory and excitatory neurons were of the general integrate-and-fire type, but excitatory units had in addition an extended dendritic cable, and they received excitatory inputs only at the more distal end of the cable, and inhibitory inputs spread along the cable. In this way, inhibitory inputs reached the soma of excitatory cells with variable delays, and in any case earlier than synchronous excitatory inputs, and at the same time they could shunt the excitatory inputs, resulting in a largely multiplicative form of inhibition (Abbott 1991). The uniform connectivity was not complete, but rather each type of unit could contact units of the other type with a probability of 0.5, and the same was true for inhibitory-to-inhibitory connections.

After 100 (simulated) ms of activity evoked by external inputs uncorrelated with any of the stored patterns, a cue was provided that consisted of the external input becoming correlated with one of the patterns, at various levels of correlation, for 300 ms. After that, external inputs were removed, but when retrieval operated successfully the activity of the units remained strongly correlated with the memory pattern, or even reached a higher level of correlation if a rather corrupted cue had been used, so that, if during the 300 ms the network had stabilized into a state rather distant from the memory pattern, it got much closer to it once the cue was removed. All correlations were quantified using information measures (see Appendix C)), in terms of mutual information between the firing rate pattern across units and the particular memory pattern being retrieved, or in terms of mutual information between the firing rate of one unit and the set of patterns, or, finally, in terms of mutual information between the decoded firing rates of a subpopulation of 10 excitatory cells and the set of memory patterns. The same algorithms were used to extract information measures as were used for example by Rolls, Treves, Tovee and Panzeri (1997d) with real neuronal data from inferior temporal cortex neurons. The firing rates were measured over sliding windows of 30 ms, after checking that shorter windows produced noisier measures. The effect of using a relatively long window, 30 ms, for measuring rates is an apparent linear early rise in information values with time. Nevertheless, in the real system the activity of these cells is 'read' by other cells receiving inputs from them, and that in turn have their own membrane capacitance-determined characteristic time for integrating input activity, a time broadly in the order of 30 ms. Using
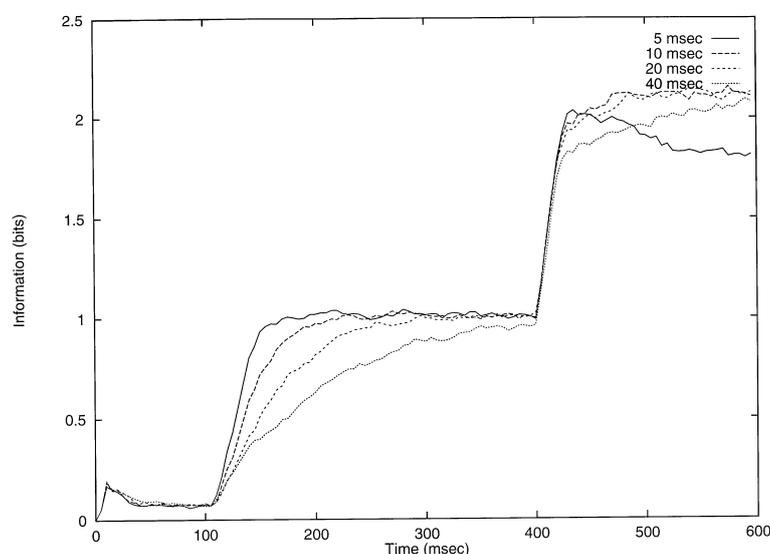
Fig. B.33 Time course of the transinformation about which memory pattern had been selected, as decoded from the firing rates of 10 randomly selected excitatory units. Excitatory conductances closed exponentially with time constants of 5, 10, 20 and 40 ms (curves from top to bottom). A cue of correlation 0.2 with the memory pattern was presented from 100 to 400 ms, uncorrelated external inputs with the same mean strength and sparseness as the cue were applied at earlier times, and no external inputs were applied at later times. (After Battaglia and Treves 1998b.)

such a time window for integrating firing rates did not therefore artificially slow down the read-out process.

It was shown that the time course of different information measures did not depend significantly on the firing rates prevailing during the retrieval state, nor on the resistance–capacitance-determined membrane time constants of the units. Figure B.33 shows that the rise in information after providing the cue at time = 100 ms followed a roughly exponential approach to its steady-state value, which continued until the steady state switched to a new value when the retrieval cue was removed at time = 400 ms. The time constant of the approach to the first steady state was a linear function, as shown in Fig. B.33, of the time constant for excitatory conductances, as predicted by the analysis. (The proportionality factor in the Figure is 2.5, or a collective time constant 2.5 times longer than the synaptic time constant.) The approach to the second steady-state value was more rapid, and the early apparent linear rise prevented the detection of a consistent exponential mode. Therefore, it appears that the cue leads to the basin of attraction of the correct retrieval state by activating transient modes, whose time constant is set by that of excitatory conductances; once the network is in the correct basin, its subsequent reaching the 'very bottom' of the basin after the removal of the cue is not accompanied by any prominent transient mode (see further Battaglia and Treves (1998a)).

Overall, these simulations confirm that recurrent networks, in which excitation is mediated mainly by fast (AMPA, see Section 1.5) channels, can reach asynchronous steady firing states very rapidly, over a few tens of milliseconds, and the rapid approach to steady state is reflected in the relatively rapid rise of information quantities that measure the speed of the operation in functional terms.

An analysis based on integrate-and-fire model units thus indicates that recurrent dynamics

can be so fast as to be practically indistinguishable from purely feedforward dynamics, in contradiction to what simple intuitive arguments would suggest. This makes it hazardous to draw conclusions on the underlying circuitry on the basis of the experimentally observed speed with which selective neuronal responses arise, as attempted by Thorpe and Imbert (1989). The results also show that networks that implement feedback processing can settle into a global retrieval state very rapidly, and that rapid processing is not just a feature of feedforward networks.

We return to the intuitive understanding of this rapid processing. The way in which networks with continuous dynamics (such as networks made of real neurons in the brain, and networks modelled with integrate-and-fire neurons) can be conceptualized as settling so fast into their attractor states is that spontaneous activity in the network ensures that some neurons are close to their firing threshold when the retrieval cue is presented, so that the firing of these neurons is influenced within 1–2 ms by the retrieval cue. These neurons then influence other neurons within milliseconds (given the point that some other neurons will be close to threshold) through the modified recurrent collateral synapses that store the information. In this way, the neurons in networks with continuous dynamics can influence each other within a fraction of the synaptic time constant, and retrieval can be very rapid.

B.6.5 The speed of processing of a four-layer hierarchical network with integrate-and-fire attractor dynamics in each layer

Given that the visual system has a whole series of cortical areas organized predominantly hierarchically (e.g. V1 to V2 to V4 to inferior temporal cortex), the issue arises of whether the rapid information processing that can be performed for object recognition is predominantly feedforward, or whether there is sufficient time for feedback processing within each cortical area implemented by the local recurrent collaterals to contribute to the visual information processing being performed. Some of the constraints are as follows.

An analysis of response latencies indicates that there is sufficient time for only 10–20 ms per processing stage in the visual system. In the primate cortical ventral visual system the response latency difference between neurons in layer $4C\beta$ of V1 and inferior temporal cortical cells is approximately 60 ms (Bullier and Nowak 1995, Nowak and Bullier 1997, Schmolesky, Wang, Hanes, Thompson, Leutgeb, Schall and Leventhal 1998). For example, the latency of the responses of neurons in V1 is approximately 30–40 ms (Celebrini, Thorpe, Trotter and Imbert 1993), and in the temporal cortex visual areas approximately 80–110 ms (Baylis, Rolls and Leonard 1987, Sugase, Yamane, Ueno and Kawano 1999). Given that there are 4–6 stages of processing in the ventral visual system from V1 to the anterior inferior temporal cortex, the difference in latencies between each ventral cortical stage is on this basis approximately 10 ms (Rolls 1992a, Oram and Perrett 1994). Information theoretic analyses of the responses of single visual cortical cells in primates reveal that much of the information that can be extracted from neuronal spike trains is often found to be present in periods as short as 20–30 ms (Tovee, Rolls, Treves and Bellis 1993, Tovee and Rolls 1995, Heller, Hertz, Kjaer and Richmond 1995, Rolls, Tovee and Panzeri 1999b). Backward masking experiments indicate that each cortical area needs to fire for only 20–30 ms to pass information to the next stage (Rolls and Tovee 1994, Rolls, Tovee, Purcell, Stewart and Azzopardi 1994b, Kovacs, Vogels and Orban 1995, Rolls, Tovee and Panzeri 1999b, Rolls 2003) (see Section C.3.4). Rapid serial visual presentation of image sequences shows that cells in the temporal visual cortex are still face selective when faces are presented at the rate of 14 ms/image (Keysers, Xiao, Foldiak and Perrett 2001). Finally, event-related potential studies in humans provide strong evidence that the visual system is able to complete some analyses of complex scenes in less than 150 ms (Thorpe, Fize and Marlot 1996).
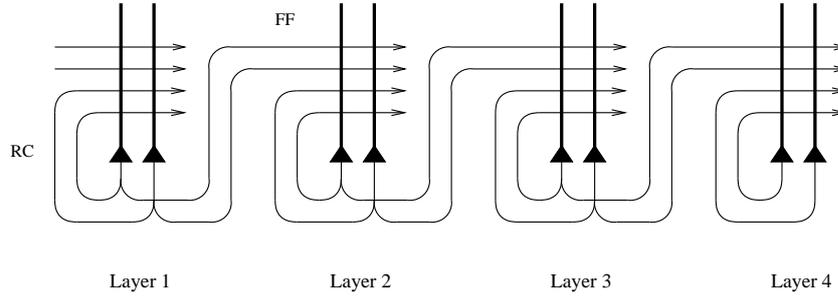
Fig. B.34 The structure of the excitatory connections in the network. There are feedforward (FF) connections between each layer and the next, and excitatory recurrent collaterals (RC) in each layer. Inhibitory connections are also present within each layer, but they are not shown in this Figure. (After Panzeri, Rolls, Battaglia and Lavis 2001.)

To investigate whether feedback processing within each layer could contribute to information processing in such a multilayer system in times as short as 10–20 ms per layer, Panzeri, Rolls, Battaglia and Lavis (2001) simulated a four-layer network with attractor networks in each layer. The network architecture is shown schematically in Fig. B.34. All the neurons realized integrate-and-fire dynamics, and indeed the individual layers and neurons were implemented very similarly to the implementation used by Battaglia and Treves (1998a). In particular, the current flowing from each compartment of the multicompartment neurons to the external medium was expressed as:

$$I(t) = g_{\text{leak}}(V(t) - V^0) + \sum_j g_j(t)(V(t) - V_j), \tag{B.56}$$

where g_{leak} is a constant passive leakage conductance, V^0 the membrane resting potential, $g_j(t)$ the value of the jth synapse conductance at time t, and V_j the reversal potential of the jth synapse. $V(t)$ is the potential in the compartment at time t. The most important parameter in the simulation, the AMPA inactivation time constant, was set to 10 ms. The recurrent collateral (RC) integration time constant of the membrane of excitatory cells was 20 ms long for the simulations presented. The synaptic conductances decayed exponentially in time, obeying the equation

$$\frac{dg_j}{dt} = -\frac{g_j}{\tau_j} + \Delta g_j \sum_k \delta(t - \Delta t - t_k^j), \tag{B.57}$$

where τ_j is the synaptic decay time constant, Δt is a delay term summarizing axonal and synaptic delays, and Δg_j is the amount that the conductance is increased when the presynaptic unit fires a spike. Δg_j thus represents the (unidirectional) coupling strength between the pre-synaptic and the post-synaptic cell. t_k^j is the time at which the pre-synaptic unit fires its kth spike.

An example of the rapid information processing of the system is shown in Fig. B.35, obtained under conditions in which the local recurrent collaterals can contribute to correct performance because the feedforward (FF) inputs from the previous stage are noisy. (The noise implemented in these simulations was some imperfection in the FF signals produced by some alterations to the FF synaptic weights.) Figure B.35 shows that, when the FF carry an incomplete signal, some information is still transmitted successfully in the 'No RC' condition (in which the recurrent collateral connections in each layer are switched off), and with a relatively short latency. However, the noise term in the FF synaptic strengths makes the retrieval fail more and more layer by layer. When in contrast the recurrent collaterals (RC)
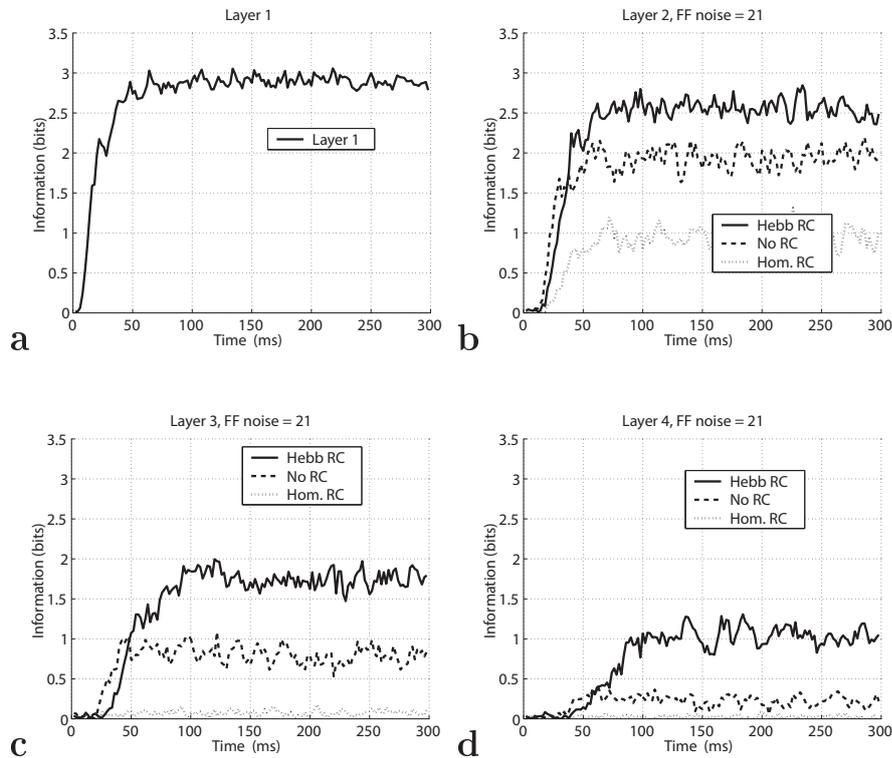
Fig. B.35 The speed of information processing in a 4-layer network with integrate-and-fire neurons. The information time course of the average information carried by the responses of a population of 30 excitatory neurons in each layer. In the simulations considered here, there is noise in the feedforward (FF) synapses. Layer 1 was tested in just one condition. Layers 2–4 are tested in three different conditions: No RC (in which the recurrent collateral synaptic effects do not operate), Hebbian RC (in which the recurrent collaterals have been trained by as associative rule and can help pattern retrieval in each layer), and a control condition named Homogeneous RC (in which the recurrent collaterals could inject current into the neurons, but no useful information was provided by them because they were all set to the same strength). (After Panzeri, Rolls, Battaglia and Lavis 2001.)

are present and operating after Hebbian training, the amount of information retrieved is now much higher, because the RC are able to correct a good part of the erroneous information injected into the neurons by the noisy FF synapses. In Layer 4, 66 ms after cue injection in Layer 1, the information in the Hebbian RC case is 0.2 bits higher than that provided by the FF connections in the 'No RC' condition. This shows that the RC are able to retrieve information in Layer 4 that is not available by any other purely FF mechanism after only roughly 50–55 ms from the time when Layer 1 responds. (This corresponds to 17–18 ms per layer.)

A direct comparison of the latency differences in layers 1–4 of the integrate-and-fire network simulated by Panzeri, Rolls, Battaglia and Lavis (2001) is shown in Fig. B.36. The results are shown for the Hebbian condition illustrated in Fig. B.35, and separate curves are shown for each of the layers 1–4. The Figure shows that, with the time constant of the synapses set to 10 ms, the network can operate with full utilization of and benefit from recurrent processing within each layer in a time which enables the signal to propagate through the 4-layer system with a time course of approximately 17 ms per layer.

The overall results of Panzeri, Rolls, Battaglia and Lavis (2001) were as follows. Through the implementation of continuous dynamics, latency differences were found in information
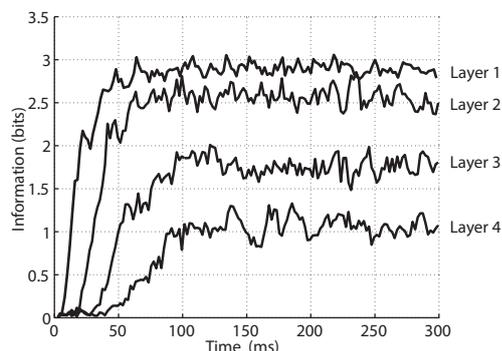
Fig. B.36 The speed of information processing in a 4-layer network with integrate-and-fire neurons. The information time course of the average information carried by the responses of a population of 30 excitatory neurons in each layer. The results are shown for the Hebbian condition illustrated in Fig. B.35, and separate curves are shown for each of the layers 1–4. The Figure shows that, with the time constant of the synapses set to 10 ms, the network can operate with full utilization of and benefit from recurrent processing within each layer in a time in the order of 17 ms per layer. (After Panzeri, Rolls, Battaglia and Lavis 2001.)

retrieval of only 5 ms per layer when local excitation was absent and processing was purely feedforward. However, information latency differences increased significantly when non-associative local excitation (simulating spontaneous firing or unrelated inputs present in the brain) was included. It was also found that local recurrent excitation through associatively modified synapses can contribute significantly to processing in as little as 15 ms per layer, including the feedforward and local feedback processing. Moreover, and in contrast to purely feed-forward processing, the contribution of local recurrent feedback was useful and approximately this rapid even when retrieval was made difficult by noise. These findings provide evidence that cortical information processing can be very fast even when local recurrent circuits are critically involved. The time cost of this recurrent processing is minimal when compared with a feedforward system with spontaneous firing or unrelated inputs already present, and the performance is better than that of a purely feedforward system when noise is present.

It is concluded that local feedback loops within each cortical area can contribute to fast visual processing and cognition.

B.6.6 Spike response model

In this section, we describe another mathematical model that models the activity of single spiking neurons. This model captures the principal effects of real neurons in a realistic way and is simple enough to permit analytical calculations (Gerstner, Ritz and Van Hemmen 1993, Gerstner and Kistler 2002). In contrast to some integrate-and-fire models (Tuckwell 1988), which are essentially given by differential equations, the spike-response model is based on response kernels that describe the integrated effect of spike reception or emission on the membrane potential. In this model, spikes are generated by a threshold process (i.e. the firing time t' is given by the condition that the membrane potential reaches the firing threshold θ; that is, $h(t') = \theta$). Figure B.37 (bottom) shows schematically the spike-generating mechanism.

The membrane potential is given by the integration of the input signal weighted by a kernel defined by the equations

$$h(t') = h^{\text{refr}}(t') + h^{\text{syn}}(t') \tag{B.58}$$

$$h^{\text{refr}}(t') = \int_0^\infty \eta^{\text{refr}}(z)\delta(t' - z - t'_{\text{last}})dz \tag{B.59}$$

$$h^{\text{syn}}(t') = \sum_j J_j \int_0^\infty \Lambda(z', t' - t'_{\text{last}})s(t' - z')dz' \ . \tag{B.60}$$

The kernel $\eta^{\text{refr}}(z)$ is the refractory function. If we consider only absolute refractoriness, $\eta^{\text{refr}}(z)$ is given by:

$$\eta^{\text{refr}}(z) = \begin{cases} -\infty \ \text{for } 0 < z \leq \tau^{\text{refr}} \\ 0 \quad \ \text{for } z \geq \tau^{\text{refr}} \end{cases} \tag{B.61}$$

where τ^{refr} is the absolute refractory time. The time t'_{last} corresponds to the last postsynaptic spike (i.e. the most recent firing of the particular neuron). The second response function is the synaptic kernel $\Lambda(z', t' - t'_{\text{last}})$. It describes the effect of an incoming spike on the membrane potential at the soma of the postsynaptic neuron, and it eventually includes also the dependence on the state of the receiving neuron through the difference $t' - t'_{\text{last}}$ (i.e. through the time that has passed since the last postsynaptic spike). The input spike train yields $s(t' - z') = \sum_i \delta(t' - z' - t_{ij})$, t_{ij} being the ith spike at presynaptic input j. In order to simplify the discussion and without losing generality, let us consider only a single synaptic input, and therefore we can remove the subindex j. In addition, we assume that the synaptic strength J is positive (i.e. excitatory). Integrating equations B.59 and B.60, we obtain

$$h(t') = \eta^{\text{refr}}(t' - t'_{\text{last}}) + J\sum_i \Lambda(t' - t_i, t' - t'_{\text{last}}) \tag{B.62}$$

Synaptic kernels are of the form

$$\Lambda(t' - t_i, t' - t'_{\text{last}}) = \text{H}(t' - t_i)\text{H}(t_i - t'_{\text{last}})\Psi(t' - t_i) \tag{B.63}$$

where $\text{H}(s)$ is the step (or Heaviside) function which vanishes for $s \leq 0$ and takes a value of 1 for $s > 0$. After firing, the membrane potential is reset according to the renewal hypothesis.

This spike-response model is not used in the models described in this book, but is presented to show alternative approaches to modelling the dynamics of network activity.

B.7 Network dynamics: introduction to the mean-field approach

Units whose potential and conductances follow the integrate-and-fire equations in Section B.6 can be assembled together in a model network of any composition and architecture. It is convenient to imagine that units are grouped into classes, such that the parameters quantifying the electrophysiological properties of the units are uniform, or nearly uniform, within each class, while the parameters assigned to synaptic connections are uniform or nearly uniform for all connections from a given presynaptic class to another given postsynaptic class. The parameters that have to be set in a model at this level of description are quite numerous, as listed in Tables B.3 and B.4.

In the limit in which the parameters are constant within each class or pair of classes, a mean-field treatment can be applied to analyze a model network, by summing equations that describe the dynamics of individual units to obtain a more limited number of equations that describe the dynamical behaviour of groups of units (Frolov and Medvedev 1986). The treatment is exact in the further limit in which very many units belong to each class, and is an approximation if each class includes just a few units. Suppose that N_C is the number of

Table B.3 Cellular parameters (chosen according to the class of each unit)

V_{rest}	Resting potential
V_{thr}	Threshold potential
V_{ahp}	Reset potential
V_{K}	Potassium conductance equilibrium potential
C	Membrane capacitance
τ_{K}	Potassium conductance time constant
g_0	Leak conductance
Δg_{K}	Extra potassium conductance following a spike
Δt	Overall transmission delay

Table B.4 Synaptic parameters (chosen according to the classes of presynaptic and postsynaptic units)

V_α	Synapse equilibrium potential
τ_α	Synaptic conductance time constant
Δg_α	Conductance opened by one presynaptic spike
Δt_α	Delay of the connection

classes defined. Summing equations B.43 and B.44 across units of the same class results in N_C functional equations describing the evolution in time of the fraction of cells of a particular class that at a given instant have a given membrane potential. In other words, from a treatment in which the evolution of the variables associated with each unit is followed separately, one moves to a treatment based on density functions, in which the common behaviour of units of the same class is followed together, keeping track solely of the portion of units at any given value of the membrane potential. Summing equation B.42 across connections with the same class of origin and destination results in $N_C \times N_C$ equations describing the dynamics of the overall summed conductance opened on the membrane of a cell of a particular class by all the cells of another given class. A more explicit derivation of mean-field equations is given by Treves (1993) and in Section B.8.

The system of mean-field equations can have many types of asymptotic solutions for long times, including chaotic, periodic, and stationary ones. The stationary solutions are stationary in the sense of the mean fields, but in fact correspond to the units of each class firing tonically at a certain rate. They are of particular interest as the dynamical equivalent of the steady states analyzed by using non-dynamical model networks. In fact, since the neuronal current-to-frequency transfer function resulting from the dynamical equations is rather similar to a threshold linear function (see Fig. B.32), and since each synaptic conductance is constant in time, the stationary solutions are essentially the same as the states described using model networks made up of threshold linear, non-dynamical units. Thus the dynamical formulation reduces to the simpler formulation in terms of steady-state rates when applied to asymptotic stationary solutions; but, among simple rate models, it is equivalent only to those that allow description of the continuous nature of neuronal output, and not to those, for example based on binary units, that do not reproduce this fundamental aspect. The advantages of the dynamical formulation are that (i) it enables one to describe the character and prevalence of other types of asymptotic solutions, and (ii) it enables one to understand how the network reaches, in time, the asymptotic behaviour.

The development of this mean-field approach, and the foundations for its application to models of cortical visual processing and attention, are described in Section B.8.

B.8 Mean-field based neurodynamics

A model of brain functions requires the choice of an appropriate theoretical framework, which permits the investigation and simulation of large-scale biologically realistic neural networks. Starting from the mathematical models of biologically realistic single neurons (i.e. spiking neurons), one can derive models that describe the joint activity of pools of equivalent neurons. This kind of neurodynamical model at the neuronal assembly level is motivated by the experimental observation that cortical neurons of the same type that are near to each other tend to receive similar inputs. As described in the previous section, it is convenient in this simplified approach to neural dynamics to consider all neurons of the same type in a small cortical volume as a computational unit of a neural network. This computational unit is called a neuronal pool or assembly. The mathematical description of the dynamical evolution of neuronal pool activity in multimodular networks, associated with different cortical areas, establishes the roots of the dynamical approach that are used in Chapter 6 and Section 5.6, and in Rolls and Deco (2002) Chapters 9–11. In this Section (B.8), we introduce the mathematical fundamentals utilized for a neurodynamical description of pool activity (see also Section B.8.2). Beginning at the microscopic level and using single spiking neurons to form the pools of a network, we derive the mathematical formulation of the neurodynamics of cell assemblies. Further, we introduce the basic architecture of neuronal pool networks that fulfil the basic mechanisms consistent with the biased competition hypothesis. Each of these networks corresponds to cortical areas that also communicate with each other. We describe therefore the dynamical interaction between different modules or networks, which will be the basis for the implementation of attentional top-down bias.

B.8.1 Population activity

We now introduce thoroughly the concept of a neuronal pool and the differential equations representing the neurodynamics of pool activity.

Starting from individual spiking neurons one can derive a differential equation that describes the dynamical evolution of the averaged activity of a pool of extensively many equivalent neurons. Several areas of the brain contain groups of neurons that are organized in populations of units with (somewhat) similar properties (though in practice the neurons convey independent information, as described in Appendix C). These groups for mean-field modelling purposes are usually called pools of neurons and are constituted by a large and similar population of identical spiking neurons that receive similar external inputs and are mutually coupled by synapses of similar strength. Assemblies of motor neurons (Kandel, Schwartz and Jessell 2000) and the columnar organization in the visual and somatosensory cortex (Hubel and Wiesel 1962, Mountcastle 1957) are examples of these pools. Each single cell in a pool can be described by a spiking model, e.g. the spike response model presented in Section B.6.6. Due to the fact that for large-scale cortical modelling, neuronal pools form a relevant computational unit, we adopt a population code. We take the activity level of each pool of neurons as the relevant dependent variable rather than the spiking activity of individual neurons. We therefore derive a dynamical model for the mean activity of a neural population. In a population of M neurons, the mean activity $A(t)$ is determined by the proportion of active neurons by counting the number of spikes $n_{\text{spikes}}(t, t + \Delta t)$ in a small time interval Δt and dividing by M and by Δt (Gerstner 2000), i.e. formally

$$A(t) = \lim_{\Delta t \to 0} \frac{n_{\text{spikes}}(t, t + \Delta t)}{M \Delta t} \ . \tag{B.64}$$
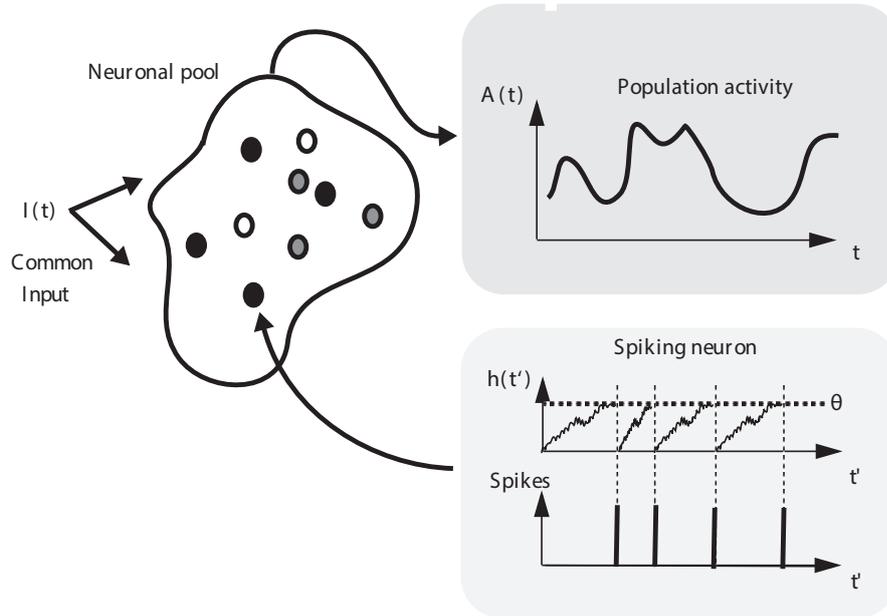
Fig. B.37 Population averaged rate of a neuronal pool of spiking neurons (top) and the action potential generating mechanism of single neurons (bottom). In a neuronal pool, the mean activity $A(t)$ is determined by the proportion of active neurons by counting the number of spikes in a small time interval Δt and dividing by the number of neurons in the pool and by Δt. Spikes are generated by a threshold process. The firing time t' is given by the condition that the membrane potential $h(t')$ reaches the firing threshold θ. The membrane potential $h(t')$ is given by the integration of the input signal weighted by a given kernel (see text for details). (After Rolls and Deco 2002.)

As indicated by Gerstner (2000), and as depicted in Fig. B.37, the concept of pool activity is quite different from the definition of the average firing rate of a single neuron. Contrary to the concept of temporal averaging over many spikes of a single cell, which requires that the input is slowly varying compared with the size of the temporal averaging window, a coding scheme based on pool activity allows rapid adaptation to real-world situations with quickly changing inputs. It is possible to derive dynamical equations for pool activity levels by utilizing the mean-field approximation (Wilson and Cowan 1972, Abbott 1991, Amit and Tsodyks 1991). The mean-field approximation consists of replacing the temporally averaged discharge rate of a cell with an equivalent momentary activity of a neural population (ensemble average) that corresponds to the assumption of ergodicity. According to this approximation, we categorize each cell assembly by means of its activity $A(t)$. A pool of excitatory neurons without external input can be described by the dynamics of the pool activity given by

$$\tau \frac{\partial A(t)}{\partial t} = -A(t) + q\mathrm{F}(A(t)) \tag{B.65}$$

where the first term on the right hand side is a decay term and the second term takes into account the excitatory stimulation between the neurons in the pool. In the previous equation, the non-linearity

$$\mathrm{F}(x) = \frac{1}{T_r - \tau \log(1 - \frac{1}{\tau x})} \tag{B.66}$$

is the response function (transforming current into discharge rate) for a spiking neuron with deterministic input, membrane time constant τ, and absolute refractory time T_r. Equation B.65

was derived by Gerstner (2000) assuming adiabatic conditions. Gerstner (2000) has shown that the population activity in a homogeneous population of neurons can be described by an integral equation. A systematic reduction of the integral equation to a single differential equation of the form B.65 always supposes that the activity changes only slowly compared with the typical interval length. In other words, the mean-field approach described in the above equations and utilized in parts of Chapters 25 and 6, and Section 5.6 generates a dynamics that neglects fast, transient behaviour. This means that we are assuming that rapid oscillations (and synchronization) do not play a computational role at least for the brain functions that we will consider. Rapid oscillations of neural activity could have a relevant functional role, namely of dynamical cooperation between pools in the same or different brain areas. It is well known in the theory of dynamical systems that the synchronization of oscillators is a cooperative phenomenon. Cooperative mechanisms might complement the competitive mechanisms on which our computational cortical model is based.

An example of the application of the mean field approach, used in a model of decision-making (Deco and Rolls 2006), is provided in Section B.8.2, with the parameters used provided in Section B.8.3. The mean-field analysis described is consistent with the integrate-and-fire spiking simulation described in Section B.6.3, that is, the same parameters used in the mean-field analysis can then be used in the integrate-and-fire simulations. Part of the value of the mean-field analysis is that it provides a way of determining the parameters that will lead to the specified steady state behaviour (in the absence of noise), and these parameters can then be used in a well-defined system for the integrate-and-fire simulations to investigate the full dynamics of the system in the presence of the noise generated by the random spike timings of the neurons.

Ways in which noise can be introduced into the mean-field approach are described by Rolls and Deco (2010).

B.8.2 The mean-field approach used in a model of decision-making

The mean-field approximation used by Deco and Rolls (2006) was derived by Brunel and Wang (2001), assuming that the network of integrate-and-fire neurons is in a stationary state. This mean-field formulation includes synaptic dynamics for AMPA, NMDA and GABA activated ion channels (Brunel and Wang 2001). In this formulation the potential of a neuron is calculated as:

$$\tau_x \frac{dV(t)}{dt} = -V(t) + \mu_x + \sigma_x \sqrt{\tau_x} \eta(t) \qquad (B.67)$$

where $V(t)$ is the membrane potential, x labels the populations, τ_x is the effective membrane time constant, μ_x is the mean value the membrane potential would have in the absence of spiking and fluctuations, σ_x measures the magnitude of the fluctuations, and η is a Gaussian process with absolute exponentially decaying correlation function with time constant τ_{AMPA}. The quantities μ_x and σ_x^2 are given by:

$$\mu_x = \frac{(T_{\text{ext}}\nu_{\text{ext}} + T_{\text{AMPA}}n_x^{\text{AMPA}} + \rho_1 n_x^{\text{NMDA}})V_E + \rho_2 n_x^{\text{NMDA}}\langle V \rangle + T_I n_x^{\text{GABA}}V_I + V_L}{S_x} \qquad (B.68)$$

$$\sigma_x^2 = \frac{g_{\text{AMPA,ext}}^2(\langle V \rangle - V_E)^2 N_{\text{ext}}\nu_{\text{ext}}\tau_{\text{AMPA}}^2\tau_x}{g_m^2 \tau_m^2}. \qquad (B.69)$$

where ν_{ext} Hz is the external incoming spiking rate, $\tau_m = C_m/g_m$ with the values for the excitatory or inhibitory neurons depending on the population considered, and the other quantities are given by:

$$S_x = 1 + T_{\text{ext}}\nu_{\text{ext}} + T_{\text{AMPA}}n_x^{\text{AMPA}} + (\rho_1 + \rho_2)n_x^{\text{NMDA}} + T_{\text{I}}n_x^{\text{GABA}} \tag{B.70}$$

$$\tau_x = \frac{C_{\text{m}}}{g_{\text{m}}S_x} \tag{B.71}$$

$$n_x^{\text{AMPA}} = \sum_{j=1}^{p} r_j w_{jx}^{\text{AMPA}}\nu_j \tag{B.72}$$

$$n_x^{\text{NMDA}} = \sum_{j=1}^{p} r_j w_{jx}^{\text{NMDA}}\psi(\nu_j) \tag{B.73}$$

$$n_x^{\text{GABA}} = \sum_{j=1}^{p} r_j w_{jx}^{\text{GABA}}\nu_j \tag{B.74}$$

$$\psi(\nu) = \frac{\nu\tau_{\text{NMDA}}}{1 + \nu\tau_{\text{NMDA}}}\left(1 + \frac{1}{1 + \nu\tau_{\text{NMDA}}}\sum_{n=1}^{\infty}\frac{(-\alpha\tau_{\text{NMDA,rise}})^n T_n(\nu)}{(n+1)!}\right) \tag{B.75}$$

$$T_n(\nu) = \sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{\tau_{\text{NMDA,rise}}(1 + \nu\tau_{\text{NMDA}})}{\tau_{\text{NMDA,rise}}(1 + \nu\tau_{\text{NMDA}}) + k\tau_{\text{NMDA,decay}}} \tag{B.76}$$

$$\tau_{\text{NMDA}} = \alpha\tau_{\text{NMDA,rise}}\tau_{\text{NMDA,decay}} \tag{B.77}$$

$$T_{\text{ext}} = \frac{g_{\text{AMPA,ext}}\tau_{\text{AMPA}}}{g_{\text{m}}} \tag{B.78}$$

$$T_{\text{AMPA}} = \frac{g_{\text{AMPA,rec}}N_{\text{E}}\tau_{\text{AMPA}}}{g_{\text{m}}} \tag{B.79}$$

$$\rho_1 = \frac{g_{\text{NMDA}}N_{\text{E}}}{g_{\text{m}}J} \tag{B.80}$$

$$\rho_2 = \beta\frac{g_{\text{NMDA}}N_{\text{E}}(\langle V_x\rangle - V_{\text{E}})(J-1)}{g_{\text{m}}J^2} \tag{B.81}$$

$$J = 1 + \gamma\exp(-\beta\langle V_x\rangle) \tag{B.82}$$

$$T_{\text{I}} = \frac{g_{\text{GABA}}N_{\text{I}}\tau_{\text{GABA}}}{g_{\text{m}}} \tag{B.83}$$

$$\langle V_x\rangle = \mu_x - (V_{\text{thr}} - V_{\text{reset}})\nu_x\tau_x, \tag{B.84}$$

where p is the number of excitatory populations, r_x is the fraction of neurons in the excitatory population x, w_{jx} the weight of the connections from population x to population j, ν_x is the spiking rate of the population x, $\gamma = [\text{Mg}^{++}]/3.57$, $\beta = 0.062$ and the average membrane potential $\langle V_x\rangle$ has a value between -55 mV and -50 mV.

The spiking rate of a population as a function of the defined quantities is then given by:

$$\nu_x = \phi(\mu_x, \sigma_x), \tag{B.85}$$

where

$$\phi(\mu_x, \sigma_x) = \left(\tau_{\text{rp}} + \tau_x\int_{\beta(\mu_x,\sigma_x)}^{\alpha(\mu_x,\sigma_x)} du\sqrt{\pi}\exp(u^2)[1 + \text{erf}(u)]\right)^{-1} \tag{B.86}$$

$$\alpha(\mu_x, \sigma_x) = \frac{(V_{\text{thr}} - \mu_x)}{\sigma_x}\left(1 + 0.5\frac{\tau_{\text{AMPA}}}{\tau_x}\right) + 1.03\sqrt{\frac{\tau_{\text{AMPA}}}{\tau_x}} - 0.5\frac{\tau_{\text{AMPA}}}{\tau_x} \tag{B.87}$$

$$\beta(\mu_x, \sigma_x) = \frac{(V_{\text{reset}} - \mu_x)}{\sigma_x} \tag{B.88}$$

with erf(u) the error function and τ_{rp} the refractory period which is considered to be 2 ms for excitatory neurons and 1 ms for inhibitory neurons. To solve the equations defined by (B.85) for all xs we integrate numerically (B.84) and the differential equation B.89, which has fixed point solutions corresponding to equation B.85:

$$\tau_x \frac{d\nu_x}{dt} = -\nu_x + \phi(\mu_x, \sigma_x).$$ (B.89)

The values of the parameters used in a mean-field analysis of decision-making are provided in Section B.8.3, with further details of the parameters in Chapter 5.6.

B.8.3 The model parameters used in the mean-field analyses of decision-making

The fixed parameters of the model are shown in Table B.5, and not only provide information about the values of the parameters used in the simulations, but also enable them to be compared to experimentally measured values.

Table B.5 Parameters used in the integrate-and-fire simulations

N_E	800
N_I	200
r	0.1
w_+	2.2
w_I	1.015
N_{ext}	800
ν_{ext}	2.4 kHz
C_m (excitatory)	0.5 nF
C_m (inhibitory)	0.2 nF
g_m (excitatory)	25 nS
g_m (inhibitory)	20 nS
V_L	–70 mV
V_{thr}	–50 mV
V_{reset}	–55 mV
V_E	0 mV
V_I	–70 mV
$g_{AMPA,ext}$ (excitatory)	2.08 nS
$g_{AMPA,rec}$ (excitatory)	0.104 nS
g_{NMDA} (excitatory)	0.327 nS
g_{GABA} (excitatory)	1.25 nS
$g_{AMPA,ext}$ (inhibitory)	1.62 nS
$g_{AMPA,rec}$ (inhibitory)	0.081 nS
g_{NMDA} (inhibitory)	0.258 nS
g_{GABA} (inhibitory)	0.973 nS
$\tau_{NMDA,decay}$	100 ms
$\tau_{NMDA,rise}$	2 ms
τ_{AMPA}	2 ms
τ_{GABA}	10 ms
α	0.5 ms^{-1}
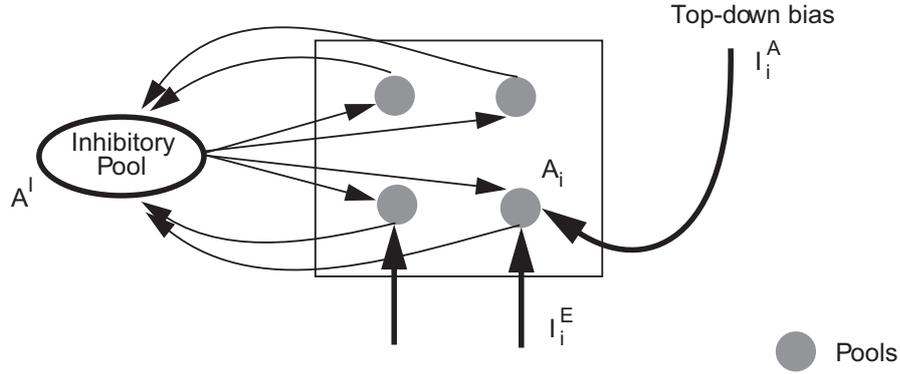
Fig. B.38 Basic computational module for biased competition: a competitive network with external top–down bias. Excitatory pools with activity A_i for the ith pool are connected with a common inhibitory pool with activity A^I in order to implement a competition mechanism. I_i^E is the external sensory input to the cells in pool i, and I_i^A attentional top-down bias, an external input coming from higher modules. The external top-down bias can shift the competition in favour of a specific pool or group of pools. This architecture is similar to that shown in Fig. B.18, but with competition between pools of similar neurons. (After Rolls and Deco 2002.)

B.8.4   A basic computational module based on biased competition

We are interested in the neurodynamics of modules composed of several pools that implement a competitive mechanism[40]. This can be achieved by connecting the pools of a given module with a common inhibitory pool, as is schematically shown in Fig. B.38.

In this way, the more pools of the module that are active, the more active the common inhibitory pool will be, and consequently, the more feedback inhibition will affect the pools in the module, such that only the most excited group of pools will survive the competition. On the other hand, external top-down bias could shift the competition in favour of a specific group of pools. This basic computational module implements therefore the biased competition hypothesis described in Chapter 6. Let us assume that there are m pools in a given module. The system of differential equations describing the dynamics of such a module is given by two differential equations, both of the type of equation B.65. The first differential equation describes the dynamics of the activity level of the excitatory pools (pyramidal neurons) and is mathematically expressed by

$$\tau \frac{\partial A_i(t)}{\partial t} = -A_i(t) + a\mathrm{F}(A_i(t)) - b\mathrm{F}(A^I(t)) +$$
$$I_0 + I_i^E(t) + I_i^A(t) + \nu \quad ; \qquad \text{for } i = 1, ..., m \qquad \text{(B.90)}$$

and the second one describes the dynamics of the activity level of the common inhibitory pool for each feature dimension (stellate neurons)

$$\tau_p \frac{\partial A^I(t)}{\partial t} = -A^I(t) + c\sum_{i=1}^{m} \mathrm{F}(A_i(t)) - d\mathrm{F}(A^I(t)) \qquad \text{(B.91)}$$

where $A_i(t)$ is the activity for pool i, $A^I(t)$ is the activity in the inhibitory pool, I_0 is a diffuse spontaneous background input, $I^E(t)$ is the external sensory input to the cells in pool i, and ν is additive Gaussian noise. The attentional top-down bias $I_i^A(t)$ is defined as an external input coming from higher modules that is not explicitly modelled.

[40]These neurodynamics are used in Chapter 6.
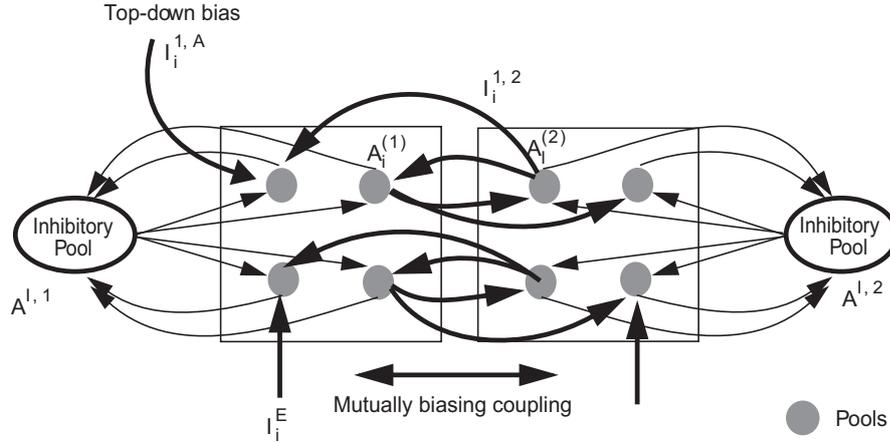
Fig. B.39 Two competitive networks mutually biased through intermodular connections. The activity $A_i^{(1)}$ of the ith excitatory pool in module 1 (on the left) and of the lth excitatory pool in module 2 (on the right) are connected by the mutually biasing coupling $I_i^{1,2}$. The architecture could implement top-down feedback originating from the interaction between brain areas that are explicitly modelled in the system. (Module 2 might be the higher module.) The external top-down bias $I_i^{1,A}$ corresponds to the coupling to pool i of module 1 from brain area A that is not explicitly modelled in the system. (After Rolls and Deco 2002.)

A qualitative description of the main fixed point attractors of the system of differential equations B.90 and B.91 was provided by Usher and Niebur (1996). Basically, we will be interested in the fixed points corresponding to zero activity and larger activation. The parameters will therefore be fixed such that the dynamics evolves to these attractors.

B.8.5 Multimodular neurodynamical architectures

In order to model complex psychophysically and neuropsychologically relevant brain functions such as visual search or object recognition (see e.g. Chapter 6), we must take into account the computational role of individual brain areas and their mutual interaction. The macroscopic phenomenological behaviour will therefore be the result of the mutual interaction of several computational modules.

The dynamical coupling of different basic modules in a multimodular architecture can be described, in our neurodynamical framework, by allowing mutual interaction between pools belonging to different modules. Figure B.39 shows this idea schematically. The system of differential equations describing the global dynamics of such a multimodular system is given by a set of equations of the type of equation B.65.

The excitatory pools belonging to a module obey the following equations:

$$\tau \frac{\partial A_i^{(j)}(t)}{\partial t} = -A_i^{(j)}(t) + a\text{F}(A_i^{(j)}(t)) - b\text{F}(A^{I,j}(t)) +$$
$$I_i^{j,k} + I_0 + I_i^E(t) + I_i^A(t) + \nu \qquad \text{for } i = 1, ..., m \quad \text{(B.92)}$$

and the corresponding inhibitory pools evolve according to

$$\tau_p \frac{\partial A^{I,j}(t)}{\partial t} = -A^{I,j}(t) + c\sum_{i=1}^{m} \text{F}(A_i^{(j)}(t)) - d\text{F}(A^{I,j}(t)). \qquad \text{(B.93)}$$

The mutual coupling $I_i^{j,k}$ between module (j) and (k) is given by

$$I_i^{j,k} = \sum_l W_{il} \mathrm{F}(I_l^{(k)}(t)) \tag{B.94}$$

where W_{il} is the synaptic coupling strength between the pool i of module (j) and pool l of module (k). This mutual coupling term can be interpreted as a top-down bias originating from the interaction between brain areas that is explicitly modelled in our system. On the other hand, the external top-down bias I_i^A corresponds to the coupling with brain areas A that are not explicitly modelled in our system.

Additionally, it is interesting to note that the top-down bias in this kind of architecture modulates the response of the pool activity in a multiplicative manner. Responses of neurons in parietal area 7a are modulated by combined eye and head movement, exhibiting a multiplicative gain modulation that modifies the amplitude of the neural responses to retinal input but does not change the preferred retinal location of a cell, nor in general the width of the receptive field (Brotchie, Andersen, Snyder and Goodman 1995). It has also been suggested that multiplicative gain modulation might play a role in translation invariant object representation (Salinas and Abbott 1997). We will use this multiplicative effect for formulating an architecture for attentional gain modulation, which can contribute to correct translation invariant object recognition in ways to be described in Chapter 6.

We show now that multiplicative-like responses can arise from the top-down biased mutual interaction between pools. Another alternative architecture that can also perform product operations on additive synaptic inputs was proposed by Salinas and Abbott (1996). Our basic architecture for showing this multiplicative effect is presented schematically in Fig. B.40a.

Two pools are mutually connected via fixed weights. The first pool or unit receives a bottom-up visual input I_1^V, modelled by the response of a vertically oriented complex cell. The second pool receives a top-down attentional bias $I_2^A = B$. The two pools are mutually coupled with unity weight. The equations describing the activities of the two pools are given by:

$$\tau \frac{\partial A_1(t)}{\partial t} = -A_1(t) + \alpha \mathrm{F}(A_1(t)) + I_o + I_1^V + \nu \tag{B.95}$$

$$\tau \frac{\partial A_2(t)}{\partial t} = -A_2(t) + \alpha \mathrm{F}(A_2(t)) + I_o + I_2^A + \nu \tag{B.96}$$

where A_1 and A_2 are the activities of pool 1 and pool 2 respectively, $\alpha = 0.95$ is the coefficient of recurrent self-excitation of the pool, ν is the noise input to the pool drawn from a normal distribution $N(\mu = 0, \sigma = 0.02)$, and $I_o = 0.025$ is a direct current biasing input to the pool.

Simulation of the above dynamical equations produces the results shown in Fig. B.40b. The orientation tuning curve of unit (or pool) 1 was modulated by a top-down bias B introduced to unit (pool) 2. The gain modulation was transmitted through the coupling from pool 2 to pool 1 after a few steps of evolution of the dynamical equations. Without the feedback from unit 2, unit 1 exhibits the orientation tuning curve shown as $(B = 0)$. As B increased, the increase in pool 1's response to the vertical bar was significantly greater than the increase in its response to the horizontal bar. Therefore, the attentional gain modulation produced in pool 1 through the mutual coupling was not a simple additive effect, but had a strong multiplicative component. The net effect was that the width of the orientation tuning curve of the cell was roughly preserved under attentional modulation. This was due to the non-linearity in the activation function.

This finding is basically consistent with the effect of attention on the orientation tuning curves of neurons in V4 (McAdams and Maunsell 1999).

(a)

(b)
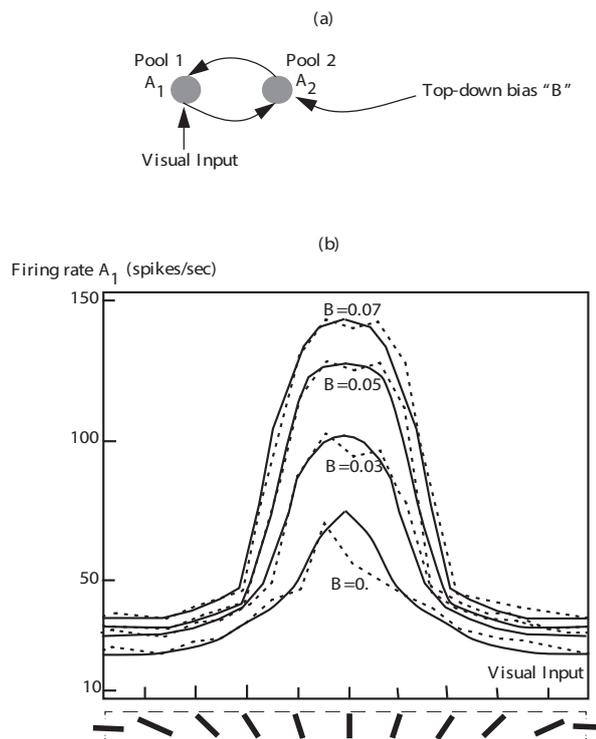
Fig. B.40 (a) The basic building block of the top-down attentional system utilizes non-specific competition between pools within the same module and specific mutual facilitation between pools in different modules. Excitatory neuronal pools within the same module compete with each other through one or more inhibitory neuronal pool(s) I_I with activity A^{I_I}. Excitatory pool 1 (with activity A_1) receives a bottom-up (visual) input I_1^V, and excitatory pool 2 receives a top-down ('attentional') bias input I_2^A. Excitatory neuronal pools in the two different modules can excite each other via mutually biased coupling. (b) The effect of altering the bias input B to pool 2 on the responses or activity of pool 1 to its orientation-tuned visual input (see text). (After Rolls and Deco 2002.)

Summarizing, the neurodynamics of the competitive mechanisms between neuronal pools, and their mutual gain modulation, are the two main ingredients used for proposing a cortical architecture that models attention and different kinds of object search in a visual scene (see Chapter 6 , and Rolls and Deco (2002) Chapters 9–11).

B.9 Sequence memory implemented by adaptation in an attractor network

Sequence memory can be implemented by using synaptic adaptation (Chapter 10) to effectively encode the order of the items in a sequence, as described in Section 10.6 (Deco and Rolls 2005c) and in the model of syntax (Rolls and Deco 2015a) (Chapter 17). Whenever the attractor system is quenched into inactivity, the next member of the sequence emerges out of the spontaneous activity, because the least recently activated member of the sequence has the least synaptic or neuronal adaptation. This mechanism could be implemented in recurrent networks such as the hippocampal CA3, prefrontal cortex, language areas of cortex, etc.
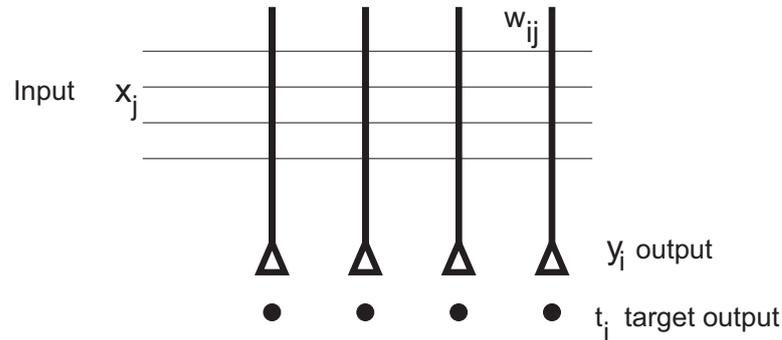
Fig. B.41 One-layer perceptron.

B.10   Perceptrons and one-layer error correction networks

The networks described in the next two Sections (B.10 and B.11) are capable of mapping a set of inputs to a set of required outputs using correction when errors are made. Although some of the networks are very powerful in the types of mapping they can perform, the power is obtained at the cost of learning algorithms that do not use local learning rules. A local learning rule specifies that synaptic strengths should be altered on the basis of information available locally at the synapse, for example the activity of the presynaptic and the post-synaptic neurons. Because the networks described here do not use local learning rules, their biological plausibility remains at present uncertain. One of the aims of future research must be to determine whether comparably difficult problems to those solved by the networks described in Sections B.10 and B.11 can be solved by biologically plausible neuronal networks.

We now describe one-layer networks taught by an error correction algorithm. The term *perceptron* refers strictly to networks with binary threshold activation functions. The outputs might take the values only 1 or 0 for example. The term perceptron arose from networks designed originally to solve perceptual problems (Rosenblatt 1961, Minsky and Papert 1969), and these networks are referred to briefly below. If the output neurons have continuous-valued firing rates, then a more general error-correcting rule called the delta rule is used, and is introduced in this Section (B.10). For such networks, the activation function may be linear, or it may be non-linear but monotonically increasing, without a sharp threshold, as in the sigmoid activation function (see Fig. 1.4).

B.10.1   Architecture and general description

The one-layer error-correcting network has a set of inputs that it is desired to map or classify into a set of outputs (see Fig. B.41). During learning, an input pattern is selected, and produces output firing by activating the output neurons through modifiable synapses, which then fire as a function of their typically non-linear activation function. The output of each neuron is then compared with a target output for that neuron given that input pattern, an error between the actual output and the desired output is determined, and the synaptic weights on that neuron are then adjusted to minimize the error. This process is then repeated for all patterns until the average error across patterns has reached a minimum. A one-layer error-correcting network can thus produce output firing for each pattern in a way that has similarities to a pattern associator. It can perform more powerful mappings than a pattern associator, but requires an error to be computed for each neuron, and for that error to affect the synaptic strength in a way that is not altogether local. A more detailed description follows.

These one-layer networks have a target for each output neuron (for each input pattern). They are thus an example of a supervised network. With the one-layer networks taught with the delta rule or perceptron learning rule described next, there is a separate teacher for each output neuron, as shown in Fig. B.41.

B.10.2 Generic algorithm for a one-layer error correction network

For each input pattern and desired target output:

1. Apply an input pattern to produce input firing $\mathbf{x}$, and obtain the activation of each neuron in the standard way by computing the dot product of the input pattern and the synaptic weight vector. The synaptic weight vector can be initially zero, or have random values.

$$h_i = \sum_j x_j w_{ij} \tag{B.97}$$

where $\sum_j$ indicates that the sum is over the C input axons (or connections) indexed by j to each neuron.

2. Apply an activation function to produce the output firing y_i:

$$y_i = \mathrm{f}(h_i) . \tag{B.98}$$

This activation function f may be sigmoid, linear, binary threshold, linear threshold, etc. If the activation function is non-linear, this helps to classify the inputs into distinct output patterns, but a linear activation function may be used if an optimal linear mapping is desired (see Adaline and Madaline, below).

3. Calculate the difference for each cell i between the target output t_i and the actual output y_i produced by the input, which is the error Δ_i

$$\Delta_i = t_i - y_i. \tag{B.99}$$

4. Apply the following learning rule, which corrects the (continuously variable) weights according to the error and the input firing x_j

$$\delta w_{ij} = k(t_i - y_i)x_j \tag{B.100}$$

where k is a constant that determines the learning rate. This is often called the delta rule, the Widrow–Hoff rule, or the LMS (least mean squares) rule (see below).

5. Repeat steps 1–4 for all input pattern – output target pairs until the root mean square error becomes zero or reaches a minimum.

In general, networks taught by the delta rule may have linear, binary threshold, or non-linear but monotonically increasing (e.g. sigmoid) activation functions, and may be taught with binary or continuous input patterns (see Rolls and Treves (1998), Chapter 5). The properties of these variations are made clear next.

B.10.3 Capability and limitations of single-layer error-correcting networks

Perceptrons perform pattern classification. That is, each neuron classifies the input patterns it receives into classes determined by the teacher. This is thus an example of a supervised

network, with a separate teacher for each output neuron. The classification is most clearly understood if the output neurons are binary, or are strongly non-linear, but the network will still try to obtain an optimal mapping with linear or near-linear output neurons.

When each neuron operates as a binary classifier, we can consider how many input patterns p can be classified by each neuron, and the classes of pattern that can be correctly classified. The result is that the maximum number of patterns that can be correctly classified by a neuron with C inputs is

$$p_{\mathrm{max}} = 2C \tag{B.101}$$

when the inputs have random continuous-valued inputs, but the patterns must be linearly separable (see Section A.2.1, and Hertz, Krogh and Palmer (1991)). For a one-layer network, no set of weights can be found that will perform the XOR (exclusive OR), or any other non-linearly separable function (see Appendix A).

Although the inability of one-layer networks with binary neurons to solve non-linearly separable problems is a limitation, it is not in practice a major limitation on the processing that can be performed in a neural network for a number of reasons. First, if the inputs can take continuous values, then if the patterns are drawn from a random distribution, the one-layer network can map up to $2C$ of them. Second, as described for pattern associators, the perceptron could be preceded by an expansion recoding network such as a competitive network with more output than input neurons. This effectively provides a two-layer network for solving the problem, and multilayer networks are in general capable of solving arbitrary mapping problems. Ways in which such multilayer networks might be trained are discussed later in this Appendix.

We now return to the issue of the capacity of one-layer perceptrons, that is, how many patterns p can be correctly mapped to correct binary outputs if the input patterns are linearly separable.

B.10.3.1   Output neurons with continuous values, and random patterns

Before treating this case, we note that if the inputs are orthogonal, then just as in the pattern associator, C patterns can be correctly classified, where there are C inputs, x_j, $(j = 1, C)$, per neuron. The argument is the same as for a pattern associator.

We consider next the capacity of a one-layer error-correcting network that learns patterns drawn from a random distribution. For neurons with continuous output values, whether the activation function is linear or not, the capacity (for fully distributed inputs) is set by the criterion that the set of input patterns must be linearly independent (see Hertz, Krogh and Palmer (1991)). (Three patterns are linearly independent if any one cannot be formed by addition (with scaling allowed) of the other two patterns – see Appendix A.) Given that there can be a maximum of C linearly independent patterns in a C-dimensional space (see Appendix A), the capacity of the perceptron with such patterns is C patterns. If we choose p random patterns with continuous values, then they will be linearly independent for $p \leq C$ (except for cases with very low probability when the randomly chosen values may not produce linearly independent patterns). (With random continuous values for the input patterns, it is very unlikely that the addition of any two, with scaling allowed, will produce a third pattern in the set.) Thus with continuous valued input patterns,

$$p_{\mathrm{max}} = C. \tag{B.102}$$

If the inputs are not linearly independent, networks trained with the delta rule produce a least mean squares (LMS) error (optimal) solution (see below).

B.10.3.2 Output neurons with binary threshold activation functions

Let us consider here strictly defined perceptrons, that is, networks with (binary) threshold output neurons, and taught by the perceptron learning procedure.

Capacity with fully distributed output patterns

The condition here for correct classification is that described in Appendix A for the AND and XOR functions, that the patterns must be linearly separable. If we consider random continuous-valued inputs, then the capacity is

$$p_{\max} = 2C \tag{B.103}$$

(see Cover (1965), Hertz, Krogh and Palmer (1991); this capacity is the case with C large, and the number of output neurons small). The interesting point to note here is that, even with fully distributed inputs, a perceptron is capable of learning more (fully distributed) patterns than there are inputs per neuron. This formula is in general valid for large C, but happens to hold also for the AND function illustrated in Appendix A.2.1.

Sparse encoding of the patterns

If the output patterns $\mathbf{y}$ are sparse (but still distributed), then just as with the pattern associator, it is possible to map many more than C patterns to correct outputs. Indeed, the number of different patterns or prototypes p that can be stored is

$$p \approx C/a \tag{B.104}$$

where a is the sparseness of the target pattern $\mathbf{t}$. p can in this situation be much larger than C (cf. Rolls and Treves (1990), and Rolls and Treves (1998) Appendix A3).

Perceptron convergence theorem

It can be proved that such networks will learn the desired mapping in a finite number of steps (Block 1962, Minsky and Papert 1969, Hertz, Krogh and Palmer 1991). (This of course depends on there being such a mapping, the condition for this being that the input patterns are linearly separable.) This is important, for it shows that single-layer networks can be proved to be capable of solving certain classes of problem.

As a matter of history, Minsky and Papert (1969) went on to emphasize the point that no one-layer network can correctly classify non-linearly separable patterns. Although it was clear that multilayer networks can solve such mapping problems, Minsky and Papert were pessimistic that an algorithm for training such a multilayer network would be found. Their emphasis that neural networks might not be able to solve general problems in computation, such as computing the XOR, which is a non-linearly separable mapping, resulted in a decline in research activity in neural networks. In retrospect, this was unfortunate, for humans are rather poor at solving parity problems such as the XOR (Thorpe, O'Regan and Pouget 1989), yet can perform many other useful neural network operations very quickly. Algorithms for training multilayer perceptrons were gradually discovered by a number of different investigators, and became widely known after the publication of the algorithm described by Rumelhart, Hinton and Williams (1986b), and Rumelhart, Hinton and Williams (1986a). Even before this, interest in neural network pattern associators, autoassociators and competitive networks was developing (see Hinton and Anderson (1981), Kohonen (1977), Kohonen (1988)), but the acceptance of the algorithm for training multilayer perceptrons led to a great rise in interest in neural networks, partly for use in connectionist models of cognitive function (McClelland and Rumelhart 1986, McLeod, Plunkett and Rolls 1998), and partly for use in applications (see Bishop (1995)).
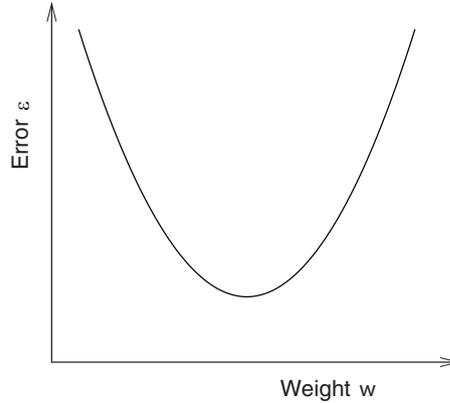
Fig. B.42 The error function ϵ for a neuron in the direction of a particular weight w.

In that perceptrons can correctly classify patterns provided only that they are linearly separable, but pattern associators are more restricted, perceptrons are more powerful learning devices than Hebbian pattern associators.

B.10.3.3 Gradient descent for neurons with continuous-valued outputs

We now consider networks trained by the delta (error correction) rule B.100, and having continuous-valued outputs. The activation function may be linear or non-linear, but provided that it is differentiable (in practice, does not include a sharp threshold), the network can be thought of as gradually decreasing the error on every learning trial, that is as performing some type of gradient descent down a continuous error function. The concept of gradient descent arises from defining an error ϵ for a neuron as

$$\epsilon = \sum_{\mu} (t^{\mu} - y^{\mu})^2 \tag{B.105}$$

where μ indexes the patterns learned by the neuron. The error function for a neuron in the direction of a particular weight would have the form shown in Fig. B.42. The delta rule can be conceptualized as performing gradient descent of this error function, in that for the jth synaptic weight on the neuron

$$\delta w_j = -k \partial \epsilon / \partial w_j \tag{B.106}$$

where $\partial \epsilon / \partial w_j$ is just the slope of the error curve in the direction of w_j in Fig. B.42. This will decrease the weight if the slope is positive and increase the weight if the slope is negative. Given equation B.99, and recalling that $h = \sum_{j} x_j w_j$, equation B.42 becomes

$$\delta w_j = -k \partial / \partial w_j \sum_{\mu} [(t^{\mu} - f(h^{\mu}))^2] \tag{B.107}$$

$$= 2k \sum_{\mu} [(t^{\mu} - y^{\mu})] f'(h) x_j$$

where $f'(h)$ is the derivative of the activation function. Provided that the activation function is monotonically increasing, its derivative will be positive, and the sign of the weight change will only depend on the mean sign of the error. Equation B.108 thus shows one way in which, from a gradient descent conceptualization, equation B.100 can be derived.

   With linear output neurons, this gradient descent is proved to reach the correct mapping (see Hertz, Krogh and Palmer (1991)). (As with all single-layer networks with continuous-valued output neurons, a perfect solution is only found if the input patterns are linearly independent. If they are not, an optimal mapping is achieved, in which the sum of the squares of the errors is a minimum.) With non-linear output neurons (for example with a sigmoid activation function), the error surface may have local minima, and is not guaranteed to reach the optimal solution, although typically a near-optimal solution is achieved. Part of the power of this gradient descent conceptualization is that it can be applied to multilayer networks with neurons with non-linear but differentiable activation functions, for example with sigmoid activation functions (see Hertz, Krogh and Palmer (1991)).

B.10.4   Properties

The properties of single-layer networks trained with a delta rule (and of perceptrons) are similar to those of pattern associators trained with a Hebbian rule in many respects (see Section B.2). In particular, the properties of generalization and graceful degradation are similar, provided that (for both types of network) distributed representations are used. The main differences are in the types of pattern that can be separated correctly, the learning speed (in that delta-rule networks can take advantage of many training trials to learn to separate patterns that could not be learned by Hebbian pattern associators), and in that the delta-rule network needs an error term to be supplied for each neuron, whereas an error term does not have to be supplied for a pattern associator, just an unconditioned or forcing stimulus. Given these overall similarities and differences, the properties of one-layer delta-rule networks are considered here briefly.

B.10.4.1   Generalization

During recall, delta-rule one-layer networks with non-linear output neurons produce appropriate outputs if a recall cue vector $\mathbf{x}_r$ is similar to a vector that has been learned already. This occurs because the recall operation involves computing the dot (inner) product of the input pattern vector $\mathbf{x}_r$ with the synaptic weight vector $\mathbf{w}_i$, so that the firing produced, y_i, reflects the similarity of the current input to the previously learned input pattern $\mathbf{x}$. Distributed representations are needed for this property. If two patterns that a delta-rule network has learned to separate are very similar, then the weights of the network will have been adjusted to force the different outputs to occur correctly. At the same time, this will mean that the way in which the network generalizes in the space between these two vectors will be very sharply defined. (Small changes in the input vector will force it to be classified one way or the other.)

B.10.4.2   Graceful degradation or fault tolerance

One-layer delta-rule networks show graceful degradation provided that the input patterns $\mathbf{x}$ are distributed.

B.10.4.3   Prototype extraction, extraction of central tendency, and noise reduction

These occur as for pattern autoassociators.

B.10.4.4   Speed

Recall is very fast in a one-layer pattern associator or perceptron, because it is a feedforward network (with no recurrent or lateral connections). Recall is also fast if the neuron has cell-like properties, because the stimulus input firings x_j ($j = 1, C$ axons) can be applied simultaneously to the synapses w_{ij}, and the activation h_i can be accumulated in one or two

time constants of the synapses and dendrite (e.g. 10–20 ms) (see Section B.6.5). Whenever the threshold of the cell is exceeded, it fires. Thus, in effectively one time step, which takes the brain no more than 10–20 ms, all the output neurons of the delta-rule network can be firing with rates that reflect the input firing of every axon.

Learning is as fast ('one-shot') in perceptrons as in pattern associators if the input patterns are orthogonal. If the patterns are not orthogonal, so that the error correction rule has to work in order to separate patterns, then the network may take many trials to achieve the best solution (which will be perfect under the conditions described above).

B.10.4.5   Non-local learning rule

The learning rule is not truly local, as it is in pattern associators, autoassociators, and competitive networks, in that with one-layer delta-rule networks, the information required to change each synaptic weight is not available in the presynaptic terminal (reflecting the presynaptic rate) and the postsynaptic activation. Instead, an error for the neuron must be computed, possibly by another neuron, and then this error must be conveyed back to the postsynaptic neuron to provide the postsynaptic error term, which together with the presynaptic rate determines how much the synapse should change, as in equation B.100,

$$\delta w_{ij} = k(t_i - y_i)x_j$$

where $(t_i - y_i)$ is the error.

A rather special architecture would be required if the brain were to utilize delta-rule error-correcting learning. One such architecture might require each output neuron to be supplied with its own error signal by another neuron. The possibility (Albus 1971) that this is implemented in one part of the brain, the cerebellum, is described in Rolls and Treves (1998) Chapter 9. Another functional architecture would require each neuron to compute its own error by subtracting its current activation by its $\mathbf{x}$ inputs from another set of afferents providing the target activation for that neuron. A neurophysiological architecture and mechanism for this is not currently known.

B.10.4.6   Interference

Interference is less of a property of single-layer delta rule networks than of pattern autoassociators and autoassociators, in that delta rule networks can learn to separate patterns even when they are highly correlated. However, if patterns are not linearly independent, then the delta rule will learn a least mean squares solution, and interference can be said to occur.

B.10.4.7   Expansion recoding

As with pattern associators and autoassociators, expansion recoding can separate input patterns into a form that makes them learnable, or that makes learning more rapid with only a few trials needed, by delta rule networks. It has been suggested that this is the role of the granule cells in the cerebellum, which provide for expansion recoding by 1,000:1 of the mossy fibre inputs before they are presented by the parallel fibres to the cerebellar Purkinje cells (Marr 1969, Albus 1971, Rolls and Treves 1998) (Chapter 23).

B.10.4.8   Utility of single-layer error-correcting networks in information processing by the brain

In the cerebellum, each output cell, a Purkinje cell, has its own climbing fibre, that distributes from its inferior olive cell its terminals throughout the dendritic tree of the Purkinje cell. It is this climbing fibre that controls whether learning of the $\mathbf{x}$ inputs supplied by the parallel fibres onto the Purkinje cell occurs, and it has been suggested that the function of this architecture

is for the climbing fibre to bring the error term to every part of the postsynaptic neuron (Chapter 23). This rather special arrangement with each output cell apparently having its own teacher is probably unique in the brain, and shows the lengths to which the brain might need to go to implement a teacher for each output neuron. The requirement for error-correction learning is to have the neuron forced during a learning phase into a state that reflects its error while presynaptic afferents are still active, and rather special arrangements are needed for this.

B.11 Multilayer perceptrons: backpropagation of error networks

B.11.1 Introduction

So far, we have considered how error can be used to train a one-layer network using a delta rule. Minsky and Papert (1969) emphasized the fact that one-layer networks cannot solve certain classes of input–output mapping problems (as described above). It was clear then that these restrictions would not apply to the problems that can be solved by feedforward multilayer networks, if they could be trained. A multilayer feedforward network has two or more connected layers, in which connections allow activity to be projected forward from one layer to the next, and in which there are no lateral connections within a layer. Such a multilayer network has an output layer (which can be trained with a standard delta rule using an error provided for each output neuron), and one or more hidden layers, in which the neurons do not receive separate error signals from an external teacher. (Because they do not provide the outputs of the network directly, and do not directly receive their own teaching error signal, these layers are described as hidden.) To solve an arbitrary mapping problem (in which the inputs are not linearly separable), a multilayer network could have a set of hidden neurons that would remap the inputs in such a way that the output layer can be provided with a linearly separable problem to solve using training of its weights with the delta rule. The problem was: how could the synaptic weights into the hidden neurons be trained in such a way that they would provide an appropriate representation? Minsky and Papert (1969) were pessimistic that such a solution would be found and, partly because of this, interest in computations in neural networks declined for many years. Although some work in neural networks continued in the following years (e.g. (Marr 1969, Marr 1970, Marr 1971, Willshaw and Longuet-Higgins 1969, Willshaw 1981, Malsburg 1973, Grossberg 1976a, Grossberg 1976b, Arbib 1964, Amari 1982, Amari, Yoshida and Kanatani 1977)), widespread interest in neural networks was revived by the type of approach to associative memory and its relation to human memory taken by the work described in the volume edited by Hinton and Anderson (1981), and by Kohonen (Kohonen 1977, Kohonen 1989). Soon after this, a solution to training a multilayer perceptron using backpropagation of error became widely known (Rumelhart, Hinton and Williams 1986b, Rumelhart, Hinton and Williams 1986a) (although earlier solutions had been found), and very great interest in neural networks and also in neural network approaches to cognitive processing (connectionism) developed (Rumelhart and McClelland 1986, McClelland and Rumelhart 1986, McLeod, Plunkett and Rolls 1998).

B.11.2 Architecture and algorithm

An introduction to the way in which a multilayer network can be trained by backpropagation of error is described next. Then we consider whether such a training algorithm is biologically
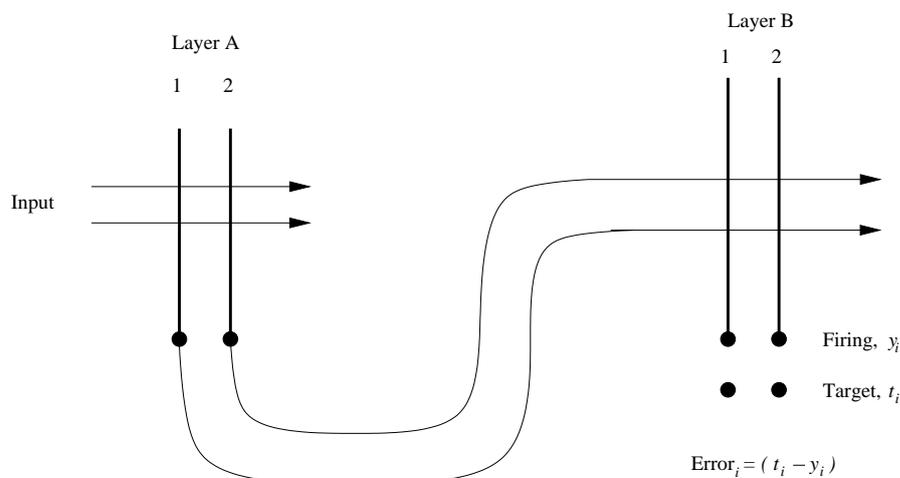
Fig. B.43 A two-layer perceptron. Inputs are applied to layer A through modifiable synapses. The outputs from layer A are applied through modifiable synapses to layer B. Layer B can be trained using a delta rule to produce firing y_i which will approach the target t_i. It is more difficult to modify the weights in layer A, because appropriate error signals must be backpropagated from layer B.

plausible. A more formal account of the training algorithm for multilayer perceptrons (sometimes abbreviated MLP) is given by Rumelhart, Hinton and Williams (1986b), Rumelhart, Hinton and Williams (1986a), and Hertz, Krogh and Palmer (1991), and below.

Consider the two-layer network shown in Fig. B.43. Inputs to the hidden neurons in layer A feed forward activity to the output neurons in layer B. The neurons in the network have a sigmoid activation function. One reason for such an activation function is that it is non-linear, and non-linearity is needed to enable multilayer networks to solve difficult (non-linearly separable) problems. (If the neurons were linear, the multilayer network would be equivalent to a one-layer network, which cannot solve such problems.) Neurons B1 and B2 of the output layer, B, are each trained using a delta rule and an error computed for each output neuron from the target output for that neuron when a given input pattern is being applied to the network. Consider now the error that needs to be used to train neuron A1 by a delta rule. This error clearly influences the error of neuron B1 in a way that depends on the magnitude of the synaptic weight from neuron A1 to B1; and on the error of neuron B2 in a way that depends on the magnitude of the synaptic weight from neuron A1 to B2. In other words, the error for neuron A1 depends on:

the weight from A1 to B1 (w_{11}) · error of neuron B1
+ the weight from A1 to B2 (w_{21}) · error of neuron B2.

In this way, the error calculation can be propagated backwards through the network to any neuron in any hidden layer, so that each neuron in the hidden layer can be trained, once its error is computed, by a delta rule (which uses the computed error for the neuron and the presynaptic firing at the synapse to correct each synaptic weight). For this to work, the way in which each neuron is activated and sends a signal forward must be continuous (not binary), so that the extent to which there is an error in, for example, neuron B1 can be related back in a graded way to provide a continuously variable correction signal to previous stages. This is one of the requirements for enabling the network to descend a continuous error surface. The activation function must be non-linear (e.g. sigmoid) for the network to learn more than

a Input (2)   Hidden (2 sigmoid)   Output (1 sigmoid)

b

$$\Delta z = \frac{\partial z}{\partial y}\,\Delta y$$

$$\Delta y = \frac{\partial y}{\partial x}\,\Delta x$$

$$\Delta z = \frac{\partial z}{\partial y}\cdot\frac{\partial y}{\partial x}\,\Delta x$$

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}\cdot\frac{\partial y}{\partial x}$$

c

Output units

Hidden units H2

Hidden units H1

Input units

$$y_l = f(z_l)$$
$$z_l = \sum_{k\,\varepsilon\,H2} w_{kl}\,y_k$$

$$y_k = f(z_k)$$
$$z_k = \sum_{j\,\varepsilon\,H1} w_{jk}\,y_j$$

$$y_j = f(z_j)$$
$$z_j = \sum_{i\,\varepsilon\,Input} w_{ij}\,x_i$$

d Compare outputs with correct answer to get error derivatives

$$\frac{\partial E}{\partial y_l} = y_l - t_l$$
$$\frac{\partial E}{\partial z_l} = \frac{\partial E}{\partial y_l}\cdot\frac{\partial y_l}{\partial z_l}$$

$$\frac{\partial E}{\partial y_k} = \sum_{l\,\varepsilon\,out} w_{kl}\frac{\partial E}{\partial z_l}$$
$$\frac{\partial E}{\partial z_k} = \frac{\partial E}{\partial y_k}\cdot\frac{\partial y_k}{\partial z_k}$$

$$\frac{\partial E}{\partial y_j} = \sum_{k\,\varepsilon\,H2} w_{jk}\frac{\partial E}{\partial z_k}$$
$$\frac{\partial E}{\partial z_j} = \frac{\partial E}{\partial y_j}\cdot\frac{\partial y_j}{\partial z_j}$$
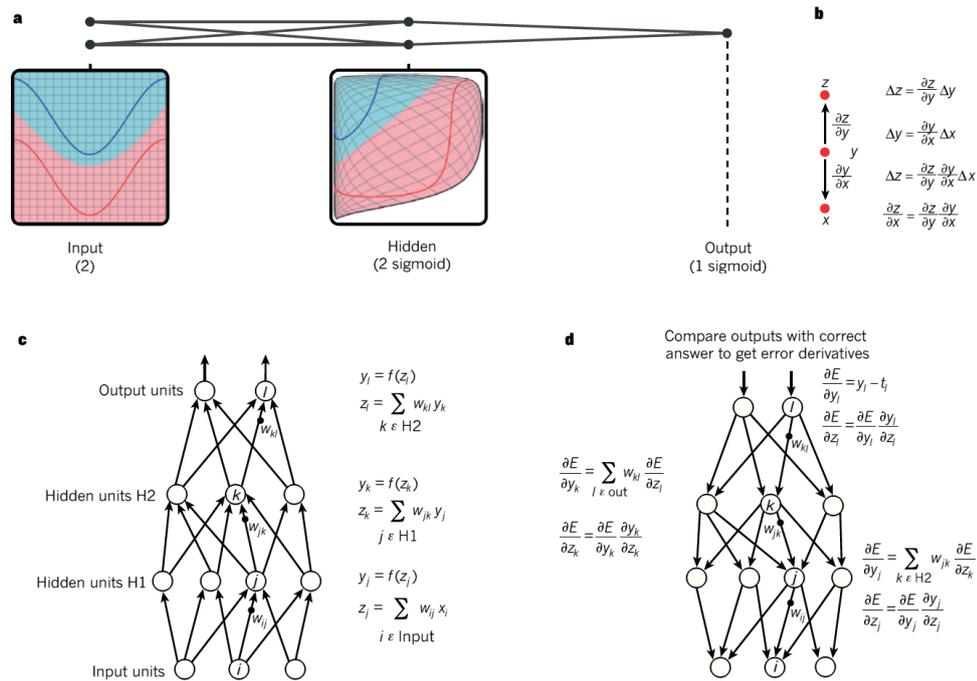
Fig. B.44 Multilayer neural networks and backpropagation (see text). a. How a multilayer neural network (shown by the connected dots) can distort the input space to make the classes of data (examples of which are on the red and blue lines) linearly separable. b. A chain rule of derivatives. c. Equations used for computing the forward pass in a neural net with two hidden layers and one output layer. d. The equations used for computing the backward pass (see text). Note that the indices use different conventions to those used in the remainder of this book. (Reprinted by permission from Macmillan Publishers Ltd: Nature, 521: 436–444, Deep learning, Yann LeCun, Yoshua Bengio, Geoffrey Hinton. Copyright ©2015. a) Reproduced with permission from C. Olah, http://colah.github.io/)

could be learned by a single-layer network. (Remember that a multilayer linear network can always be made equivalent to a single-layer linear network, and that there are some problems that cannot be solved by single-layer networks.) For the way in which the error of each output neuron should be taken into account to be specified in the error correction rule, the position at which the output neuron is operating on its activation function must also be taken into account. For this, the slope of the activation function is needed, and because the slope is needed, the activation function must be differentiable. Although we indicated use of a sigmoid activation function, other activation functions that are non-linear and monotonically increasing (and differentiable) can be used. (For further details, see Rumelhart, Hinton and Williams (1986b), Rumelhart, Hinton and Williams (1986a), and Hertz, Krogh and Palmer (1991)).

More formally, the operation of a backpropagation of error network can be described as follows (Fig. B.44) (LeCun, Bengio and Hinton 2015). Fig. B.44a shows how a multilayer neural network (shown by the connected dots) can distort the input space to make the classes of data (examples of which are on the red and blue lines) linearly separable. Note how a regular grid (shown on the left) in the input space is also transformed (shown in the middle panel) by hidden units. This is an illustrative example with only two input units, two hidden units and one output unit, but the networks may contain hundreds of thousands of units.

Fig. B.44b shows how the chain rule of derivatives tells us how two small effects (that of a small change of x on y, and that of y on z) are composed. A small change Δx in x

gets transformed first into a small change Δx in y by getting multiplied by $\partial y/\partial x$ (that is, the definition of partial derivative). Similarly, the change Δy creates a change Δz in z. Substituting one equation into the other gives the chain rule of derivatives – how Δx gets turned into Δz through multiplication by the product of $\partial y/\partial x$ and $\partial z/\partial x$. It also works when x, y and z are vectors (and the derivatives are Jacobian matrices).

Fig. B.44c shows the equations used for computing the forward pass in a neural net with two hidden layers and one output layer, each constituting a module through which one can backpropagate gradients. At each layer, we first compute the total input z to each unit, which is a weighted sum of the outputs of the units in the layer below. Then a non-linear function $f(.)$ is applied to z to get the output of the unit. For simplicity, bias terms are omitted. The non-linear function used is now typically the rectified linear unit (ReLU) $f(z) = \max(0, z)$ (known in neuroscience as a threshold linear activation function), although sigmoid activation functions have been used.

Fig. B.44d shows the equations used for computing the backward pass. At each hidden layer we compute the error derivative with respect to the output of each unit, which is a weighted sum of the error derivatives with respect to the total inputs to the units in the layer above. We then convert the error derivative with respect to the output into the error derivative with respect to the input by multiplying it by the gradient of $f(z)$. At the output layer, the error derivative with respect to the output of a unit is computed by differentiating the cost function. This gives $y_l - t_l$ if the cost function for unit l is $0.5(y_l - t_l)^2$, where t_l is the target value. Once the $\partial E/\partial z_k$ is known, the error-derivative for the weight w_{jk} on the connection from unit j in the layer below is $y_j \partial E/\partial z_k$.

Once we have this error derivative for a single neuron, we can update the weights into that neuron by performing a gradient descent step in the direction specified by the derivative to minimize the cost function (see LeCun, Bengio and Hinton (2015)).

B.11.3 Properties of multilayer networks trained by error backpropagation

B.11.3.1 Arbitrary mappings

Arbitrary mappings of non-linearly separable patterns can be achieved. For example, such networks can solve the XOR problem, and parity problems in general of which XOR is a special case. (The parity problem is to determine whether the sum of the (binary) bits in a vector is odd or even.) Multilayer feedforward backpropagation of error networks are not guaranteed to converge to the best solution, and may become stuck in local minima in the error surface. However, they generally perform very well.

B.11.3.2 Fast operation

The network operates as a feedforward network, without any recurrent or feedback processing. Thus (once it has learned) the network operates very quickly, with a time proportional to the number of layers.

B.11.3.3 Learning speed

The learning speed can be very slow, taking many thousands of trials. The network learns to gradually approximate the correct input–output mapping required, but the learning is slow because of the credit assignment problem for neurons in the hidden layers. The credit assignment problem refers to the issue of how much to correct the weights of each neuron in the hidden layer. As the example above shows, the error for a hidden neuron could influence the errors of many neurons in the output layers, and the error of each output neuron reflects the error from many hidden neurons. It is thus difficult to assign credit (or blame) on any

single trial to any particular hidden neuron, so an error must be estimated, and the net run until the weights of the crucial hidden neurons have become altered sufficiently to allow good performance of the network. Another factor that can slow learning is that if a neuron operates close to a horizontal part of its activation function, then the output of the neuron will depend rather little on its activation, and correspondingly the error computed to backpropagate will depend rather little on the activation of that neuron, so learning will be slow.

More general approaches to this issue suggest that the number of training trials for such a network will (with a suitable training set) be of the same order of magnitude as the number of synapses in the network (see Cortes, Jaeckel, Solla, Vapnik and Denker (1996)).

B.11.3.4   Number of hidden neurons and generalization

Backpropagation networks are generally intended to discover regular mappings between the input and output, that is mappings in which generalization will occur usefully. If there were one hidden neuron for every combination of inputs that had to be mapped to an output, then this would constitute a look-up table, and no generalization between similar inputs (or inputs not yet received) would occur. The best way to ensure that a backpropagation network learns the structure of the problem space is to set the number of neurons in the hidden layers close to the minimum that will allow the mapping to be implemented. This forces the network not to operate as a look-up table. A problem is that there is no general rule about how many hidden neurons are appropriate, given that this depends on the types of mappings required. In practice, these networks are sometimes trained with different numbers of hidden neurons, until the minimum number required to perform the required mapping has been approximated.

B.12   Biologically plausible networks vs backpropagation

Given that the error for a hidden neuron in an error backpropagation network is calculated by propagating backwards information based on the errors of all the output neurons to which a hidden neuron is connected, and all the relevant synaptic weights, and the activations of the output neurons to define the part of the activation function on which they are operating, it is implausible to suppose that the correct information to provide the appropriate error for each hidden neuron is propagated backwards between real neurons. A hidden neuron would have to 'know', or receive information about, the errors of all the neurons to which it is connected, and its synaptic weights to them, and their current activations. If there were more than one hidden layer, this would be even more difficult.

To expand on the difficulties: first, there would have to be a mechanism in the brain for providing an appropriate error signal to each output neuron in the network. With the possible exception of the cerebellum, an architecture where a separate error signal could be provided for each output neurons is difficult to identify in the brain. Second, any retrograde passing of messages across multiple-layer forward-transmitting pathways in the brain that could be used for backpropagation seems highly implausible, not only because of the difficulty of getting the correct signal to be backpropagated, but also because retrograde signals passed by axonal transport in a multilayer net would take days to arrive, long after the end of any feedback given in the environment indicating a particular error. Third, as noted in Chapter 11, the backprojection pathways that are present in the cortex seem suited to perform recall, and this would make it difficult for them also to have the correct strength to carry the correct error signal.

A problem with the backpropagation of error approach in a biological context is thus that in order to achieve their competence, backpropagation networks use what is almost certainly a learning rule that is much more powerful than those that could be implemented biologically,

and achieve their excellent performance by performing the mapping though a minimal number of hidden neurons. In contrast, real neuronal networks in the brain probably use much less powerful learning rules, in which errors are not propagated backwards, and at the same time have very large numbers of hidden neurons, without the bottleneck that helps to provide backpropagation networks with their good performance. A consequence of these differences between backpropagation and biologically plausible networks may be that the way in which biological networks solve difficult problems may be rather different from the way in which backpropagation networks find mappings. Thus the solutions found by connectionist systems may not always be excellent guides to how biologically plausible networks may perform on similar problems. Part of the challenge for future work is to discover how more biologically plausible networks than backpropagation networks can solve comparably hard problems, and then to examine the properties of these networks, as a perhaps more accurate guide to brain computation.

As stated above, it is a major challenge for brain research to discover whether there are algorithms that will solve comparably difficult problems to backpropagation, but with a local learning rule. Such algorithms may be expected to require many more hidden neurons than backpropagation networks, in that the brain does not appear to use information bottlenecks to help it solve difficult problems. The issue here is that much of the power of backpropagation algorithms arises because there is a minimal number of hidden neurons to perform the required mapping using a final one-layer delta-rule network. Useful generalization arises in such networks because with a minimal number of hidden neurons, the net sets the representation they provide to enable appropriate generalization. The danger with more hidden neurons is that the network becomes a look-up table, with one hidden neuron for every required output, and generalization when the inputs vary becomes poor. The challenge is to find a more biologically plausible type of network that operates with large numbers of neurons, and yet that still provides useful generalization. An example of such an approach is described in Chapter 25.

B.13   Convolution networks

Although not biologically plausible, a class of artificial network that was inspired by neuroscience is the convolution network (LeCun, Kavukcuoglu and Farabet 2010, LeCun, Bengio and Hinton 2015, Bengio, Goodfellow and Courville 2017). This type of network is described here, to compare to biologically plausible networks. A convolution network is a multilayer network that is trained by backpropagation of error and that has convergence from stage to stage. When used, the term 'deep learning' usually refers to convolution learning, and this is an example of a 'deep neural network'.

In more detail, the architecture of a typical convolution network (ConvNet) is structured as a series of stages (LeCun, Bengio and Hinton 2015). The first few stages are composed of two types of layers: convolutional layers and pooling layers.

Units in a convolutional layer are organized in feature maps, within which each unit is connected to local patches in the feature maps of the previous layer through a set of weights called a filter bank. The result of this local weighted sum is then passed through a non-linearity such as a ReLU (a threshold-linear activation function). All units in a feature map share the same filter bank. To be more specific: the local patch of one layer that connects to a neuron in the next layer may be 3x3 units. Whatever weights are learned by backpropagation are the same for all the units across the layer, and this is achieved artificially, effectively by copying the weights laterally. There may be different feature maps in a layer (e.g. for different colours), and each uses a different filter bank. The reason for this architecture is twofold. First, in array

data such as images, local groups of values are often highly correlated, forming distinctive local motifs that are easily detected. Second, the local statistics of images and other signals are invariant to location. In other words, if a motif can appear in one part of the image, it could appear anywhere, hence the idea of units at different locations sharing the same weights and detecting the same pattern in different parts of the array. Mathematically, the filtering operation performed by a feature map is a discrete convolution, hence the name.

Although the role of the convolutional layer is to detect local conjunctions of features from the previous layer, the role of the pooling layer is to merge semantically similar features into one. Because the relative positions of the features forming a motif can vary somewhat, reliably detecting the motif can be done by coarse-graining the position of each feature. A typical pooling unit computes the maximum of a local patch of units in one feature map (or in a few feature maps). To be more specific: a pooling neuron may take the maximum input from a 2 unit x 2 unit patch of units in its input convolution layer. Neighbouring pooling units take input from patches that are shifted by more than one row or column, thereby reducing the dimension of the representation and creating an invariance to small shifts and distortions. The 'stride' may be 2. A convolution net architecture might have these tiny local convolution then pooling of 'max' operator pairs of layers stacked 140 or more on top of one another. Or the architecture might be varied, with several convolution layers before the next pooling layer. At the top of the system, there may be more convolutional and fully-connected pooling layers. To train a ConvNet, stochastic gradient descent is used in conjunction with the gradients computed by backpropagation of error to update the weights to minimize the errors of the output units. The targets for the individual units in the output layer are decided by the experimenter, and might include for one neuron cars, for another neuron bicycles, etc. The whole net is trained with in the order of one million exemplars.

Thus in a convolution network, layers typically alternate, with one layer performing the convolution and filtering followed by a non-linear function such as a threshold linear function (an 'S' layer), and the next layer performing 'pooling' over the neurons in the previous layer (a 'C' layer), to help with transform invariance. The alternation of 'S' and 'C' layers follows the suggestion of Hubel and Wiesel (1962) that complex cells might compute their responses by summing over a set of simple cells. However, there is no evidence for any alternation of S and C layers in most of the primate ventral visual system. The general idea though is that some neurons in the net need to learn to respond to feature combinations, and this is the function of the convolution layer; and the system also needs to do something to cope with transform invariance, and this is done by the max function layers.

The concept of copying weights laterally was used in the Neocognitron (Fukushima 1980) (Section 25.3.5.1), and has the machine advantages that only one small region is learned, and that the 'filters' or neurons are uniformly good (or bad) throughout a layer. The convolution operator in a convolution network has this effect. This lateral copying is of course biologically implausible, and is not used in VisNet (Chapter 25).

With its powerful backpropagation training algorithm, and very large numbers of training trials, a convolution net can learn to activate the correct output neuron in the output layer for a category of object with good accuracy which may be commercially useful (LeCun, Bengio and Hinton 2015). However, unlike VisNet, there is no attempt in general to teach the network transform invariance by presenting images with spatial continuity, and no attempt to take advantage of the statistics of the world to help it learn which transforms are probably of the same object by capitalising on temporal continuity. The network is very biologically implausible, in that each unit (or 'neuron') receives from only typically a 3 x 3 or 2 x 2 unit patch of the preceding area (i.e. the receptive fields are small); by using lateral weight copying within a layer; by having up to 140 or more layers stacked on top of each other

in the hierarchy; by using the non biologically plausible backpropagation of error training algorithm; and by using a teacher for every neuron in the output layer.

B.14 Contrastive Hebbian learning: the Boltzmann machine

In a move towards a learning rule that is more local than in backpropagation networks, yet that can solve similar mapping problems in a multilayer architecture, we describe briefly contrastive Hebbian learning. The multilayer architecture has forward connections through the network to the output layer, and a set of matching backprojections from the output layer through each of the hidden layers to the input layer. The forward connection strength between any pair of neurons has the same value as the backward connection strength between the same two neurons, resulting in a symmetric set of forward and backward connection strengths. An input pattern is applied to the multilayer network, and an output is computed using normal feedforward activation processing with neurons with a sigmoid (non-linear and monotonically increasing) activation function. The output firing then via the backprojections is used to create firing of the input neurons. This process is repeated until the firing rates settle down, in an iterative way (which is similar to the settling of the autoassociative nets described in Section B.3). After settling, the correlations between any two neurons are remembered, for this type of unclamped operation, in which the output neurons fire at the rates that the process just described produces. The correlations reflect the normal presynaptic and postsynaptic terms used in the Hebb rule, e.g. $(x_j y_i)^{\text{uc}}$, where 'uc' refers to the unclamped condition, and as usual x_j is the firing rate of the input neuron, and y_i is the activity of the receiving neuron. The output neurons are then clamped to their target values, and the iterative process just described is repeated, to produce for every pair of synapses in the network $(x_j y_i)^{\text{c}}$, where the c refers now to the clamped condition. An error correction term for each synapse is then computed from the difference between the remembered correlation of the unclamped and the clamped conditions, to produce a synaptic weight correction term as follows:

$$\delta w_{ij} = k[(x_j y_i)^{\text{c}} - (x_j y_i)^{\text{uc}}], \qquad \text{(B.108)}$$

where k is a learning rate constant. This process is then repeated for each input pattern to output pattern to be learned. The whole process is then repeated many times with all patterns until the output neurons fire similarly in the clamped and unclamped conditions, that is until the errors have become small. Further details are provided by Hinton and Sejnowski (1986). The version described above is the mean field (or deterministic) Boltzmann machine (Peterson and Anderson 1987, Hinton 1989). It is sometimes called the '**wake - sleep**' algorithm (Hinton et al. 1995), because when the system is unclamped, with no input, it might be likened to a sleeping or dreaming state. More traditionally, a Boltzmann machine updates one randomly chosen neuron at a time, and each neuron fires with a probability that depends on its activation (Ackley, Hinton and Sejnowski 1985, Hinton and Sejnowski 1986). The latter version makes fewer theoretical assumptions, while the former may operate an order of magnitude faster (Hertz, Krogh and Palmer 1991).

In terms of biological plausibility, it certainly is the case that there are backprojections between adjacent cortical areas (see Chapters 1 and 11). Indeed, there are as many back-projections between adjacent cortical areas as there are forward projections. The backward projections seem to be more diffuse than the forward projections, in that they connect to a wider region of the preceding cortical area than the region that sends the forward projections. If the backward and the forward synapses in such an architecture were Hebb-modifiable, then

there is a possibility that the backward connections would be symmetric with the forward connections. Indeed, such a connection scheme would be useful to implement top-down recall, as summarized in Chapter 24 and described by Rolls and Treves (1998) in their Chapter 6. What seems less biologically plausible is that after an unclamped phase of operation, the correlations between all pairs of neurons would be remembered, there would then be a clamped phase of operation with each output neuron clamped to the required rate for that particular input pattern, and then the synapses would be corrected by an error correction rule that would require a comparison of the correlations between the neuronal firing of every pair of neurons in the unclamped and clamped conditions.

Although this algorithm has the disadvantages that it is not very biologically plausible, and does not operate as well as standard backpropagation, it has been made use of by O'Reilly and Munakata (2000) in approaches to connectionist modelling in cognitive neuroscience.

B.15 Deep Belief Networks

Another artificial network is a Deep Belief Network which is a probabilistic model composed of many Restricted Boltzmann Machines (RBMs) stacked on top of each other with each layer trained on the output of the previous one (Hinton, Osindero and Teh 2006). A restricted Boltzmann Machine (in contrast to a traditional Boltzmann Machine) has no lateral connections within layers, and the connectivity is relaxed to be undirected. This results in the units in a particular layer being conditionally independent given a set of observed activations in the other. This is a key difference that allows for the efficient learning and inference within these models. Each subsequent layer acts like constraints on the activities of the one below. This is a greedy training process and will probably not yield an optimum representation for the entire hierarchy, primarily because the individual layers are trained in isolation. (Once a layer is trained the weights are frozen.) This inefficiency however can be remedied by a relatively small amount of fine tuning (generative or discriminative) after the stack is complete (Hinton and Salakhutdinov 2006). It should be noted that a Deep Belief Network itself is not a deep Restricted Boltzmann Machine. The lower layers do not themselves define an undirected model, and only the top-most two layers keep their undirected nature. However there has been work on modifications that restore the undirected nature of the lower layers allowing inference within the lower layers to make use of top-down connectivity as well as the usual bottom-up connectivity (Salakhutdinov and Larochelle 2010).

Deep belief networks do have the advantage, like VisNet, that they are unsupervised, and do not need a teacher for every output neuron to help teach the system how to respond. However, scaling such models to full-sized, high-dimensional images remains a difficult problem. To address this a pooling operation has been tested within a deep belief network, and this can help to form transform-invariant representations (Lee, Grosse, Ranganath and Ng 2011). However, their training algorithm remains very biologically implausible.

B.16 Reinforcement learning

In supervised networks, an error signal is provided for each output neuron in the network, and whenever an input to the network is provided, the error signals specify the magnitude and direction of the error in the output produced by each neuron. These error signals are then used to correct the synaptic weights in the network in such a way that the output errors for each input pattern to be learned gradually diminish over trials (see Sections B.10 and B.11). These networks have an architecture that might be similar to that of the pattern associator shown in

Fig. B.1, except that instead of an unconditioned stimulus, there is an error correction signal provided for each output neuron. Such a network trained by an error correcting (or delta) rule is known as a one-layer perceptron. The architecture is not very plausible for most brain regions, in that it is not clear how an individual error signal could be computed for each of thousands of neurons in a network, and fed into each neuron as its error signal and then used in a delta rule synaptic correction (see Section B.10).

The architecture can be generalized to a multilayer feedforward architecture with many layers between the input and output (Rumelhart, Hinton and Williams 1986a), but the learning is very non-local and rather biologically implausible (see Section B.11), in that an error term (magnitude and direction) for each neuron in the network must be computed from the errors and synaptic weights of all subsequent neurons in the network that any neuron influences, usually on a trial-by-trial basis, by a process known as error backpropagation. Thus although computationally powerful, an issue with perceptrons and multilayer perceptrons that makes them generally biologically implausible for many brain regions is that a separate error signal must be supplied for each output neuron, and that with multilayer perceptrons, computed error backpropagation must occur.

When operating in an environment, usually a simple binary or scalar signal representing success or failure of the whole network or organism is received. This is usually action-dependent feedback that provides a single evaluative measure of the success or failure. Evaluative feedback tells the learner whether or not, and possibly by how much, its behaviour has improved; or it provides a measure of the 'goodness' of the behaviour. Evaluative feedback does not directly tell the learner what it should have done, and although it may provide an index of the degree (i.e. magnitude) of success, it does not include directional information telling the learner how to change its behaviour towards a target, as does error-correction learning (see Barto (1995)). Partly for this reason, there has been some interest in networks that can be taught with such a single reinforcement signal. In this Section (B.16), approaches to such networks are described. It is noted that such networks are classified as reinforcement networks in which there is a single teacher, and that these networks attempt to perform an optimal mapping between an input vector and an output neuron or set of neurons. They thus solve the same class of problems as single layer and multilayer perceptrons. They should be distinguished from pattern-association networks in the brain, which might learn associations between previously neutral stimuli and primary reinforcers such as taste (signals which might be interpreted appropriately by a subsequent part of the brain), but do not attempt to produce arbitrary mappings between an input and an output, using a single reinforcement signal.

A class of problems to which such reinforcement networks might be applied are motor-control problems. It was to such a problem that Barto and Sutton (Barto 1985, Sutton and Barto 1981) applied a reinforcement learning algorithm, the associative reward–penalty algorithm described next. The algorithm can in principle be applied to multilayer networks, and the learning is relatively slow. The algorithm is summarized in Section B.16.1 and by Hertz, Krogh and Palmer (1991). More recent developments in reinforcement learning (see Sections B.16.2 and B.16.3) are described by Sutton and Barto (1998) and reviewed by Dayan and Abbott (2001).

B.16.1   Associative reward–penalty algorithm of Barto and Sutton

The terminology of Barto and Sutton is followed here (see Barto (1985)).
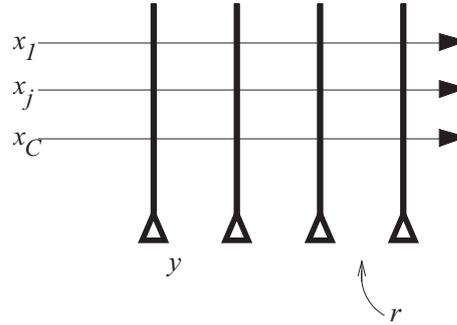
Fig. B.45 A network trained by a single reinforcement input r. The inputs to each neuron are $x_j, j = 1, C$; and y is the output of one of the output neurons.

B.16.1.1 Architecture

The architecture, shown in Fig. B.45, uses a single reinforcement signal, r, = +1 for reward, and –1 for penalty. The inputs x_j take real (continuous) values. The output of a neuron, y, is binary, +1 or –1. The weights on the output neuron are designated w_j.

B.16.1.2 Operation

1. An input vector is applied to the network, and produces activation, h, in the normal way as follows:

$$h = \sum_{j=1}^{C} x_j w_j \tag{B.109}$$

where $\sum_{j=1}^{C}$ indicates that the sum is over the C input axons (or connections) indexed by j to each neuron.

2. The output y is calculated from the activation with a noise term η included. The principle of the network is that if the added noise on a particular trial helps performance, then whatever change it leads to should be incorporated into the synaptic weights, in such a way that the next time that input occurs, the performance is improved.

$$y = \begin{cases} +1 & \text{if } h + \eta \geq 0, \\ -1 & \text{else.} \end{cases} \tag{B.110}$$

where η = the noise added on each trial.

3. Learning rule. The weights are changed as follows:

$$\delta w_j = \begin{cases} \rho(y - \text{E}[y|h])x_j & \text{if } r = +1, \\ \rho\lambda(-y - \text{E}[y|h])x_j & \text{if } r = -1. \end{cases} \tag{B.111}$$

ρ and λ are learning-rate constants. (They are set so that the learning rate is higher when positive reinforcement is received than when negative reinforcement is received.) $\text{E}[y|h]$ is the expectation of y given h (usually a sigmoidal function of h with the range ± 1). $\text{E}[y|h]$ is a (continuously varying) indication of how the neuron usually responds to the current input pattern, i.e. if the actual output y is larger than normally expected, by computing $h = \sum w_j x_j$, because of the noise term, and the reinforcement is +1, increase the weight from x_j; and vice versa. The expectation could be the prediction generated before the noise term is incorporated.

This network combines an associative capacity with its properties of generalization and graceful degradation, with a single 'critic' or error signal for the whole network (Barto 1985). [The term $y - \mathrm{E}[y|h]$ in equation B.111 can be thought of as an error for the output of the neuron: it is the difference between what occurred, and what was expected to occur. The synaptic weight is adjusted according to the sign and magnitude of the error of the postsynaptic firing, multiplied by the presynaptic firing, and depending on the reinforcement r received. The rule is similar to a Hebb synaptic modification rule (equation B.2), except that the postsynaptic term is an error instead of the postsynaptic firing rate, and the learning is modulated by the reinforcement.] The network can solve difficult problems (such as balancing a pole by moving a trolley that supports the pole from side to side, as the pole starts to topple). Although described for single-layer networks, the algorithm can be applied to multilayer networks. The learning rate is very slow, for there is a single reinforcement signal on each trial for the whole network, not a separate error signal for each neuron in the network as is the case in a perceptron trained with an error rule (see Section B.10).

This associative reward–penalty reinforcement-learning algorithm is certainly a move towards biological relevance, in that learning with a single reinforcer can be achieved. That single reinforcer might be broadcast throughout the system by a general projection system. It is not clear yet how a biological system might store the expected output $\mathrm{E}[y|h]$ for comparison with the actual output when noise has been added, and might take into account the sign and magnitude of this difference. Nevertheless, this is an interesting algorithm, which is related to the temporal difference reinforcement learning algorithm described in Section B.16.3.

B.16.2 Reward prediction error or delta rule learning, and classical conditioning

In classical or Pavlovian associative learning, a number of different types of association may be learned (Rolls 2014a, Cardinal, Parkinson, Hall and Everitt 2002). This type of associative learning may be performed by networks with the general architecture and properties of pattern associators (see Section B.2 and Fig. B.1). However, the time course of the acquisition and extinction of these associations can be expressed concisely by a modified type of learning rule in which an error correction term is used (introduced in Section B.16.1), rather than the postsynaptic firing y itself as in equation B.2. Use of this modified, error correction, type of learning also enables some of the properties of classical conditioning to be explained (see Dayan and Abbott (2001) for review), and this type of learning is therefore described briefly here. The rule is known in learning theory as the Rescorla–Wagner rule, after Rescorla and Wagner (1972).

The Rescorla–Wagner rule is a version of error correction or delta-rule learning (see Section B.10), and is based on a simple linear prediction of the expected reward value, denoted by v, associated with a stimulus representation x ($x = 1$ if the stimulus is present, and $x = 0$ if the stimulus is absent). The **expected reward value** v is expressed as the input stimulus variable x multiplied by a weight w

$$v = wx. \tag{B.112}$$

The **reward prediction error** is the difference between the expected reward value v and the actual **reward outcome** r obtained, i.e.

$$\Delta = r - v \tag{B.113}$$

where Δ is the reward prediction error. The value of the weight w is learned by a rule designed to minimize the expected squared error $< (r - v)^2 >$ between the actual

reward outcome r and the predicted reward value v. The angle brackets indicate an average over the presentations of the stimulus and reward. The delta rule will perform the required type of learning:

$$\delta w = k(r - v)x \qquad (B.114)$$

where δw is the change of synaptic weight, k is a constant that determines the learning rate, and the term $(r-v)$ is the reward prediction error Δ (equivalent to the error in the postsynaptic firing, rather than the postsynaptic firing y itself as in equation B.2). Application of this rule during conditioning with the stimulus x presented on every trial results in the weight w approaching the asymptotic limit $w = r$ exponentially over trials as the error Δ becomes zero. In extinction, when $r = 0$, the weight (and thus the output of the system) exponentially decays to $w = 0$. This rule thus helps to capture the time course over trials of the acquisition and extinction of conditioning. The rule also helps to account for a number of properties of classical conditioning, including blocking, inhibitory conditioning, and overshadowing (see Dayan and Abbott (2001)).

How this functionality is implemented in the brain is not yet clear. We consider one suggestion (Schultz et al. 1995, Schultz 2004, Schultz 2006, Schultz 2013) after we introduce a further sophistication of reinforcement learning which allows the time course of events within a trial to be taken into account.

B.16.3   Temporal Difference (TD) learning

An important advance in the area of reinforcement learning was the introduction of algorithms that allow for learning to occur when the reinforcement is delayed or received over a number of time steps, and which allow effects within a trial to be taken into account (Sutton and Barto 1998, Sutton and Barto 1990). A solution to these problems is the addition of an adaptive critic that learns through a time difference (TD) algorithm how to predict the future value of the reinforcer. The time difference algorithm takes into account not only the current reinforcement just received, but also a temporally weighted average of errors in predicting future reinforcements. The temporal difference error is the error by which any two temporally adjacent error predictions are inconsistent (see Barto (1995)). The output of the critic is used as an effective reinforcer instead of the instantaneous reinforcement being received (see Sutton and Barto (1998), Sutton and Barto (1990), and Barto (1995)). This is a solution to the temporal credit assignment problem, and enables future rewards to be predicted. Summaries are provided by Doya (1999), Schultz, Dayan and Montague (1997), and Dayan and Abbott (2001).

In reinforcement learning, a learning agent takes an *action* $\mathbf{u}(t)$ in response to the *state* $\mathbf{x}(t)$ of the environment, which results in the change of the state

$$\mathbf{x}(t + 1) = \mathrm{F}(\mathbf{x}(t), \mathbf{u}(t)), \qquad (B.115)$$

and the delivery of the reinforcement signal, or *reward*

$$r(t + 1) = \mathrm{R}(\mathbf{x}(t), \mathbf{u}(t)). \qquad (B.116)$$

In the above equations, $\mathbf{x}$ is a vector representation of inputs x_j, and equation B.115 indicates that the next state $\mathbf{x}(t + 1)$ at time $(t + 1)$ is a function F of the state at the previous time step of the inputs and actions at that time step in a closed system. In equation B.116 the reward at the next time step is determined by a reward function R which uses the current sensory inputs and action taken. The time t may refer to time within a trial.

The goal is to find a *policy* function G which maps sensory inputs **x** to actions

$$\mathbf{u}(t) = G(\mathbf{x}(t)) \tag{B.117}$$

which maximizes the cumulative sum of the rewards based on the sensory inputs.

The current action $\mathbf{u}(t)$ affects all future states and accordingly all future rewards. The maximization is realized by the use of the *value function* V of the states to predict, given the sensory inputs **x**, the cumulative sum (possibly discounted as a function of time) of all future rewards V(**x**) (possibly within a learning trial) as follows:

$$V(\mathbf{x}) = E[r(t+1) + \gamma r(t+2) + \gamma^2 r(t+3) + ...] \tag{B.118}$$

where $r(t)$ is the reward at time t, and $E[\cdot]$ denotes the expected value of the sum of future rewards up to the end of the trial. $0 \leq \gamma \leq 1$ is a discount factor that makes rewards that arrive sooner more important than rewards that arrive later, according to an exponential decay function. (If $\gamma = 1$ there is no discounting.) It is assumed that the presentation of future cues and rewards depends only on the current sensory cues and not the past sensory cues. The right-hand side of equation B.118 is evaluated for the dynamics in equations B.115–B.117 with the initial condition $\mathbf{x}(t) = \mathbf{x}$. The two basic ingredients in reinforcement learning are the estimation (which we term $\hat{V}$) of the value function V, and then the improvement of the policy or action **u** using the value function (Sutton and Barto 1998).

The basic algorithm for learning the value function is to minimize the *temporal difference* (TD) *error* $\Delta(t)$ for time t within a trial, and this is computed by a 'critic' for the estimated value predictions $\hat{V}(\mathbf{x}(t))$ at successive time steps as

$$\Delta(t) = [r(t) + \gamma\hat{V}(\mathbf{x}(t))] - \hat{V}(\mathbf{x}(t-1)) \tag{B.119}$$

where $\hat{V}(\mathbf{x}(t)) - \hat{V}(\mathbf{x}(t-1))$ is the difference in the reward value prediction at two successive time steps, giving rise to the terminology temporal difference learning. If we introduce the term $\hat{v}$ as the estimate of the cumulated reward by the end of the trial, we can define it as a function $\hat{V}$ of the current sensory input $\mathbf{x}(t)$, i.e. $\hat{v} = \hat{V}(\mathbf{x})$, and we can also write equation B.119 as

$$\Delta(t) = r(t) + \gamma\hat{v}(t) - \hat{v}(t-1) \tag{B.120}$$

which draws out the fact that it is differences at successive timesteps in the reward value predictions $\hat{v}$ that are used to calculate Δ.

$\Delta(t)$ is used to improve the estimates $\hat{v}(t)$ by the 'critic', and can also be used (by an 'actor') to learn appropriate actions.

For example, when the value function is represented (in the critic) as

$$\hat{V}(\mathbf{x}(t)) = \sum_{j=1}^{n} w_j^C x_j(t) \tag{B.121}$$

the learning algorithm for the (value) weight w_j^C in the critic is given by

$$\delta w_j^C = k_c \Delta(t) x_j(t-1) \tag{B.122}$$

where δw_j^C is the change of synaptic weight, k_c is a constant that determines the learning rate for the sensory input x_j, and $\Delta(t)$ is the Temporal Difference error at time t. Under certain conditions this learning rule will cause the estimate $\hat{v}$ to converge to the true value (Dayan and Sejnowski 1994).

A simple way of improving the policy of the actor is to take a stochastic action

$$u_i(t) = g(\sum_{j=1}^{n} w_{ij}^{\mathrm{A}} x_j(t) + \mu_i(t)), \tag{B.123}$$

where g() is a scalar version of the policy function G, w_{ij}^{A} is a weight in the actor, and $\mu_i(t)$ is a noise term. The TD error $\Delta(t)$ as defined in equation B.119 then signals the unexpected delivery of the reward $r(t)$ or the increase in the state value $\hat{V}(\mathbf{x}(t))$ above expectation, possibly due to the previous choice of action $u_i(t-1)$. The learning algorithm for the action weight w_{ij}^{A} in the actor is given by

$$\delta w_{ij}^{\mathrm{A}} = k_a \Delta(t)(u_i(t-1) - <u_i>)x_j(t-1), \tag{B.124}$$

where $<u_i>$ is the average level of the action output, and k_a is a learning rate constant in the actor.

Thus, the TD error $\Delta(t)$, which signals the error in the reward prediction at time t, works as the main teaching signal in both learning the value function (implemented in the critic), and the selection of actions (implemented in the actor). The usefulness of a separate critic is that it enables the TD error to be calculated based on the difference in reward value predictions at two successive time steps as shown in equation B.119.

The algorithm has been applied to modelling the time course of classical conditioning (Sutton and Barto 1990). The algorithm effectively allows the future reinforcement predicted from past history to influence the responses made, and in this sense allows behaviour to be guided not just by immediate reinforcement, but also by 'anticipated' reinforcements. Different types of temporal difference learning are described by Sutton and Barto (1998). An application is to the analysis of decisions when future rewards are discounted with respect to immediate rewards (Dayan and Abbott 2001, Tanaka, Doya, Okada, Ueda, Okamoto and Yamawaki 2004). Another application is to the learning of sequences of actions to take within a trial (Suri and Schultz 1998).

The possibility that dopamine neuron firing may provide an error signal useful in training neuronal systems to predict reward has been discussed in Section 9.6. It has been proposed that the firing of the dopamine neurons can be thought of as an error signal about reward prediction, in that the firing occurs in a task when a reward is given, but then moves forward in time within a trial to the time when a stimulus is presented that can be used to predict when the taste reward will be obtained (Schultz et al. 1995, Schultz 2013) (see Fig. 9.2). The argument is that there is no prediction error when the taste reward is obtained if it has been signalled by a preceding conditioned stimulus, and that is why the dopamine midbrain neurons do not respond at the time of taste reward delivery, but instead, at least during training, to the onset of the conditioned stimulus (Waelti, Dickinson and Schultz 2001). If a different conditioned stimulus is shown that normally predicts that no taste reward will be given, there is no firing of the dopamine neurons to the onset of that conditioned stimulus.

This hypothesis has been built into models of learning in which the error signal is used to train synaptic connections in dopamine pathway recipient regions (such as presumably the striatum and orbitofrontal cortex) (Houk, Adams and Barto 1995, Schultz 2004, Schultz, Dayan and Montague 1997, Waelti, Dickinson and Schultz 2001, Dayan and Abbott 2001, Schultz 2013). Some difficulties with the hypothesis are discussed in Section 9.6 on page 234. The difficulties include the fact that dopamine is released in large quantities by aversive stimuli (see Section 9.6); that error computations for differences between the expected reward and the actual reward received on a trial are computed in the primate orbitofrontal cortex, where expected reward, actual reward, and error neurons are all found, and lesions of which

impair the ability to use changes in reward contingencies to reverse behaviour (see Section 15.14); that the tonic, sustained, firing of the dopamine neurons in the delay period of a task with probabilistic rewards may reflect reward uncertainty, and not the expected reward, nor the magnitude of the prediction error (see Section 9.6 and Shizgal and Arvanitogiannis (2003)); and that reinforcement learning is suited to setting up connections that might be required in fixed tasks such as motor habit or sequence learning, for reinforcement learning algorithms seek to set weights correctly in an 'actor', but are not suited to tasks where rules must be altered flexibly, as in rapid one-trial reversal, for which a very different type of mechanism is described in Section 15.15 (Deco and Rolls 2005d, Rolls and Deco 2016).

The temporal difference approach to reinforcement learning has a weakness that although it can be used to predict internal signals during reinforcement learning in some tasks, it does not directly address learning with respect to actions that are not taken. Q-learning can be considered as an extension of TD learning which adds additional terms to take into account signals from actions that are not taken, for example information gained by observation of others (Montague, King-Casas and Cohen 2006).

Overall, reinforcement learning algorithms are certainly a move towards biological relevance, in that learning with a single reinforcer can be achieved in systems that might learn motor habits or fixed sequences. Whether a single prediction error is broadcast throughout a neural system by a general projection system, such as the dopamine pathways in the brain, which distribute to large parts of the striatum and the prefrontal cortex, remains to be clearly established (see further Chapter 15, Section 9.6.2), though using a modulator of synaptic learning may be computationally useful (Fremaux and Gerstner 2015).

B.17   Highlights

1. The operation and properties of biologically plausible pattern association networks, autoassociation networks, and competitive networks are described.
2. The operation of perceptrons, backpropagation networks, and deep learning networks is described.

Appendix 4 Simulation software for neuronal network models

D.1 Introduction

This Appendix of *Cerebral Cortex: Principles of Operation* (Oxford University Press) (Rolls 2016a) describes the Matlab software that has been made available with *Cerebral Cortex: Principles of Operation* (Rolls 2016a) to provide simple demonstrations of the operation of some key neuronal networks related to cortical function. The software is also provided in connection with *Emotion and Decision-Making Explained* (Oxford University Press) (Rolls 2014a) to illustrate the operation of some neuronal networks involved in emotion and decision-making. The aim of providing the software is to enable those who are learning about these networks to understand how they operate and their properties by running simulations and by providing the programs. The code has been kept simple, but can easily be edited to explore the properties of these networks. The code itself contains comments that help to explain how the code works.

The programs are intended to be used with the descriptions of the networks and their properties provided in Appendix B of *Cerebral Cortex: Principles of Operation* (Rolls 2016a), which is available at http://www.oxcns.org. Most of the papers cited in this Appendix can be found at http://www.oxcns.org. Exercises are suggested below that will provide insight into the properties of these networks, and will make the software suitable for class use. The code is available at http://www.oxcns.org. The programs are written in Matlab^{TM} (Mathworks Inc). The programs should work in any version of Matlab, whether it is running under Windows, Linux, or Apple software. The programs will also work under GNU Octave, which is available for free download.

To get started, copy the Matlab program files PatternAssociationDemo.m, AutoAssociationDemo.m, CompetitiveNetDemo.m and SOMdemo.m, and the function files NormVeclen.m and spars.m, into a directory, and in Matlab cd into that directory.

D.2 Autoassociation or attractor networks

The operation and properties of autoassociation (or attractor) neuronal networks are described in Rolls (2016a) Appendix B.3.

D.2.1 Running the simulation

In the command window of Matlab, after you have performed a 'cd' into the directory where the source files are, type 'AutoAssociationDemo' (with every command followed by 'Enter' or 'Return'). The program will run for a bit, and then pause so that you can inspect what has been produced so far. A paused state is indicated by the words 'Paused: Press any key' in the bottom of the Matlab window. Press 'Enter' to move to the next stage, until the program finishes and the command prompt reappears. You can edit the code to comment out some of the 'pause' statements if you do not want them, or to add a 'pause' statement if you would

like the program to stop at a particular place so that you can inspect the results. To stop the program and exit from it, use 'Ctrl-C'. When at the command prompt, you can access variables by typing the variable name followed by 'Enter'. (Note: if you set '*display = 0*' near the beginning of the program, there will be no pauses, and you will be able to collect data quickly.)

The fully connected autoassociation network has N=100 neurons with the dendrites that receive the synapses shown as vertical columns in the generated figures, and $nSyn = N$ synapses onto each neuron. At the first pause, you will see a figure showing the 10 random binary training patterns with firing rates of 1 or 0, and with a *Sparseness* (referred to as a in the equations below) of 0.5. As these are binary patterns, the *Sparseness* parameter value is the same as the proportion of high firing rates or 1s. At the second pause you will see distorted versions of these patterns to be used as recall cues later. The number of bits that have been flipped $nFlipBits$=14.

Next you will see a figure showing the synaptic matrix being trained with the 10 patterns. Uncomment the pause in the training loop if you wish to see each pattern being presented sequentially. The synaptic weight matrix *SynMat* (the elements of which are referred to as w_{ij} in the equations below) is initialized to zero, and then each pattern is presented as an external input to set the firing rates of the neurons (the postsynaptic term), which because of the recurrent collaterals become also the presynaptic input. Each time a training pattern is presented, the weight change is calculated by a covariance learning rule

$$\delta w_{ij} = \alpha(y_i - a)(y_j - a) \tag{D.1}$$

where α is a learning rate constant, and y_j is the presynaptic firing rate. This learning rule includes (in proportion to y_i the firing rate of the ith postsynaptic neuron) increasing the synaptic weight if $(y_j - a) > 0$ (long-term potentiation), and decreasing the synaptic weight if $(y_j - a) < 0$ (heterosynaptic long-term depression). As these are random binary patterns with sparseness a (the parameter *Sparseness* in AutoassociationDemo.m), a is the average activity $< y_j >$ of an axon across patterns, and a is also the average activity $< y_i >$ of a neuron across patterns. In the exercises, you can try a form of this learning rule that is more biologically plausible, with only heterosynaptic long-term depression. The change of weight is added to the previous synaptic weight. In the figures generated by the code, the maximum weights are shown as white, and the minimum weights are black. Although both these rules lead to some negative synaptic weights, which are not biologically plausible, this limitation can be overcome, as shown in the exercises. In the display of the synaptic weights, remember that each column represents the synaptic weights on a single neuron. The thin row below the synaptic matrix represents the firing rates of the neurons, and the thin column to the left of the synaptic weight matrix the firing rates of the presynaptic input, which are usually the same during training. Rates of 1 are shown as white, and of 0 as black.

Next, testing with the distorted patterns starts, with the output rates allowed to recirculate through the recurrent collateral axons for 9 recall epochs to produce presynaptic inputs that act through the synaptic weight matrix to produce the firing rate for the next recall epoch. The distorted recall cue is presented only on epoch 1 in what is described as the clamped condition, and after that is removed so that the network has a chance settle into a perfect recall state representing the training pattern without being affected by the distorted recall cue. (In the cortex, this may be facilitated by the greater adaptation of the thalamic inputs than of the recurrent collaterals (Rolls 2016a)). (Thus on epoch 1 the distorted recall cue is shown as the PreSynaptic Input column on the left of the display, and the Activation row and the Firing Rate row below the synaptic weight matrix are the Activations and Rates produced by the recall cue. On later epochs, the Presynaptic Input column shows the output firing Rate from the preceding epoch, recirculated by the recurrent collateral connections.)

We are interested in how perfect the recall is, that is how correlated the recalled firing rate state is with the original training pattern, with this correlation having a maximum value of 1. Every time you press 'Enter' a new recall epoch is performed, and you will see that over one to several recall epochs the recall usually will become perfect, that is the correlation, which is provided for you in the command window, will become 1. However, on some trials, the network does not converge to perfect recall, and this occurs if by chance the distorted recall cue happened to be much closer to one of the other stored patterns. Do note that with randomly chosen training patterns in relatively small networks this will sometimes occur, and that this 'statistical fluctuation' in how close some of the training patterns are to each other, and how close the distorted test patterns are to individual training patterns, is to be expected. These effects smooth out as the network become larger. (Remember that cortical neurons have in the order of 10,000 recurrent collateral inputs to each neuron, so these statistical fluctuations will be largely smoothed out. Do note also that the performance of the network will be different each time it is run, because the random number generator is set with a different seed on each run. (Near the beginning of the program, you could uncomment the command that causes the same seed to be used for the random number generator for each run, to help with further program development that you may try.)

After the last test pattern has been presented, the percentage of the patterns that were correctly recalled is shown, using as a criterion that the correlation of the recalled firing rate with the training pattern $r >= 0.98$. For the reasons just described related to the random generation of the training and distorted test patterns, the percentage correct will vary from run to run, so taking the average of several runs is recommended.

Note that the continuing stable firing after the first few recall epochs models the implementation of short-term memory in the cerebral cortex (Rolls 2008d, Rolls 2016a).

Note that the retrieval of the correct pattern from a distorted version that includes missing 1s models completion in the cerebral cortex, which is important in episodic memory as implemented in the hippocampus in which the whole of the episodic memory can be retrieved from any part (Rolls 2008d, Rolls 2016a, Kesner and Rolls 2015).

Note that the retrieval of the correct pattern from a distorted version that includes missing 1s or has 1s instead of 0s models correct memory retrieval in the cerebral cortex even if there is some distortion of the recall cue (Rolls 2008d, Rolls 2016a). This also enables generalization to similar patterns or stimuli that have been encountered previously, which is highly behaviorally adaptive (Rolls 2008d, Rolls 2016a).

D.2.2 Exercises

1. Measure the percent correct as you increase the number of patterns from 10 to 15 with the Sparseness remaining at 0.5. Plot the result. How close is your result to that found by Hopfield (1982), which was $0.14N$ for a sparseness a of 0.5? You could try increasing N by 10 (to 1000) and multiplying the number of training patterns by 10 and increasing $nFlipBits$ to say 100 to see whether your results, with smaller statistical fluctuations, approximate more closely to the theoretical value.

2. Test the effect of reducing the Sparseness (a) to 0.1, which will increase the number of patterns that can be stored and correctly recalled, i.e. the capacity of the network (Treves and Rolls 1991, Rolls 2016a). (Hint: try a number of patterns in the region of 30.) You could also plot the capacity of the network as a function of the Sparseness (a) with values down to say 0.01 in larger networks, with theoretical results provided by Treves and Rolls (1991) and Rolls (2016a).

3. Test the effect of altering the learning rule, from the covariance rule to the rule with heterosynaptic long term depression (LTD), which is as follows (Rolls 2016a):

$$\delta w_{ij} = \alpha(y_i)(y_j - a). \tag{D.2}$$

(Hint: in the training loop in the program, at about line 102, comment in the heterosynaptic LTD rule.)

4. Test whether the network operates with positive-only synaptic weights. (Hint: there are two lines (close to 131) just after the training loop that if uncommented will add a constant to all the numbers in the synaptic weight matrix to make all the numbers positive (with the minimum weight 0).

5. What happens to the recall if you train more patterns than the critical capacity? (Hint: see Rolls (2016a) section B.16.)

6. Does it in practice make much difference if you allow self-connections, especially in large networks? (Hint: in the program in the training loop if you comment out the line 'if syn $\sim=$ neuron' and its corresponding 'end', self-connections will be allowed.)

D.3   Pattern association networks

The operation and properties of pattern association networks are described in Rolls (2016a) Appendix B.2.

D.3.1   Running the simulation

In the command window of Matlab, after you have performed a 'cd' into the directory where the source files are, type 'PatternAssociationDemo'. The program will run for a bit, and then pause so that you can inspect what has been produced so far. A paused state is indicated by the words 'Paused: Press any key'. Press 'Enter' to move to the next stage, until the program finishes and the command prompt reappears. You can edit the code to comment out some of the 'pause' statements if you do not want them, or to add a 'pause' statement if you would like the program to stop at a particular place so that you can inspect the results. To stop the program and exit from it, use 'Ctrl-C'. When at the command prompt, you can access variables by typing the variable name followed by 'Enter'. (Note: if you set 'display = 0' near the beginning of the program, there will be few pauses, and you will be able to collect data quickly.)

The fully connected pattern association network has N=8 output neurons with the dendrites that receive the synapses shown as vertical columns in the generated figures, and $nSyn$=64 synapses onto each neuron. At the first pause, you will see the 8 random binary conditioned stimulus (CS) training patterns with firing rates of 1 or 0, and with a sparseness of 0.25 (*InputSparseness*). As these are binary patterns, the sparseness is the same as the proportion of high firing rates or 1s. At the second pause you will see distorted versions of these patterns to be used as recall cues later. The number of bits that have been flipped $nFlipBits$=8.

Next you will see the synaptic matrix being trained with the 8 pattern associations between each conditioned stimulus (CS) and its unconditioned stimulus (US), the output firing. In this simulation, the default is for each US pattern to have one bit on, so that the sparseness of the US, *OutputSparseness* is $1/N$. Uncomment the pause in the training loop if you wish to see each pattern being presented sequentially. The synaptic weight matrix *SynMat* (elements

of which are referred to as w_{ij} below) is initialized to zero, and then each pattern pair is presented as a CS and US pair, with the CS specifying the presynaptic input, and the US the firing rate in the N neurons present when the CS is presented. Each time a training pattern pair is presented, the weight change is calculated by an associative learning rule that includes heterosynaptic long term depression

$$\delta w_{ij} = \alpha(y_i)(x_j - a). \tag{D.3}$$

where α is a learning rate constant, and x_j is the presynaptic firing rate. This learning rule includes (in proportion to y_i the firing rate of the ith postsynaptic neuron produced by the US) increasing the synaptic weight if $(y_j - a) > 0$ (long-term potentiation), and decreasing the synaptic weight if $(x_j - a) < 0$ (heterosynaptic long-term depression). As the CS patterns are random binary patterns with sparseness a (the parameter *InputSparseness*), a is the average activity $< x_j >$ of an axon across patterns. The reasons for the use of this learning rule are described in Rolls (2016a) section B.3.3.6. In the exercises, you can try a Hebbian associative learning rule that does not include this heterosynaptic long-term depression. The change of weight is added to the previous synaptic weight. In the figures generated by the code, the maximum weights are shown as white, and the minimum weights are black. Although both these rules lead to some negative synaptic weights, which are not biologically plausible, this limitation can be overcome, as shown in the exercises. In the display of the synaptic weights, remember that each column represents the synaptic weights on a single neuron. The top thin row below the synaptic matrix represents the activations of the neurons, the bottom thin row the firing rates of the neurons, and the thin column to the left of the synaptic weight matrix the firing rates of the presynaptic input, i.e. the recall (CS) stimulus. Rates of 1 are shown as white, and of 0 as black.

Next, testing with the distorted CS patterns starts. We are interested in how the pattern association network generalizes with distorted inputs to produce the correct US output that should be produced if there were no distortion. The recall performance is measured by how correlated the recalled firing rate state (the Conditioned Response, CR) is with the output that should be produced to the original training CS pattern, with this correlation having a maximum value of 1. Every time you press 'Enter' a new CS is presented. Do note that with randomly chosen distortions of the CS patterns to produce the recall cue, in relatively small networks the distorted CS will sometimes be close of another CS, and that this 'statistical fluctuation' in how close some of the distorted CSs are to the original CSs is to be expected. These effects smooth out as the network become larger. (Remember that cortical neurons have in the order of 10,000 inputs to each neuron (in addition to the recurrent collaterals, Rolls (2016a)), so these statistical fluctuations will be largely smoothed out. Do note also that the performance of the network will be different each time it is run, because the random number generator is set with a different seed on each run. (Near the beginning of the program, you could uncomment the command that causes the same seed to be used for the random number generator for each run, to help with further program development that you may try.)

After the last distorted testing CS pattern has been presented, the percentage of the patterns that were correctly recalled is shown, using a criterion for the correlation of the recalled firing rate with the training pattern $r >= 0.98$. For the reasons just described related to the random generation of the distorted test patterns, the percentage correct will vary from run to run, so taking the average of several runs is recommended.

Note that the retrieval of the correct output (conditioned response, CR) pattern from a distorted CS recall stimulus models generalization in the cerebral cortex, which is highly behaviorally adaptive (Rolls 2008d, Rolls 2016a).

D.3.2   Exercises

1. Test how well the network generalizes to distorted recall cues, by altering $nFlipBits$ to values between 0 and 14. Remember to take the average of several simulation runs.

2. Test the effect of altering the learning rule, from the rule with heterosynaptic long term depression (LTD) as in Eqn. D.3, to a simple associative synaptic modification rule which is as follows (Rolls 2016a):

$$\delta w_{ij} = \alpha(y_i)(x_j). \tag{D.4}$$

(Hint: in the training loop in the program, at about line 110, comment out the heterosynaptic LTD rule, and uncomment the associative rule.)
   If you see little difference, can you suggest conditions under which there may be a difference, such as higher loading? (Hint: check Rolls (2016a) section B.2.7.)

3. Describe how the threshold non-linearity in the activation function helps to remove interference from other training patterns, and from distortions in the recall cue. (Hint: compare the activations to the firing rates of the output neurons.)

4. In the pattern associator, the conditioned stimulus (CS) retrieves the unconditioned response (UR), that is the effects produced by the unconditioned stimulus (US) with which it has been paired during learning. Can the opposite occur, that is, can the US be used to retrieve the CS? If not, in what type of network is the retrieval symmetrical? (Hint: if you are not sure, read Rolls (2016a) sections B.2 and B.3.)

5. The network simulated uses local representations for the US, with only one neuron on for each US. If you can program in Matlab, can you write code that would generate distributed representations for the US patterns? Then can you produce new CS patterns with random binary representations and sparsenesses in the range 0.1–0.5, like those used for the autoassociator, so that you can then investigate the storage capacity of pattern association networks. (Hint: reference to Rolls (2016a) section B.2.7 and B.2.8 may be useful.)

D.4   Competitive networks and Self-Organizing Maps

The operation and properties of competitive networks are described in Rolls (2016a) Appendix B.4. This is a self-organizing network, with no teacher. The input patterns are presented in random sequence, and the network learns how to categorise the patterns, placing similar patterns into the same category, and different patterns into different categories. The particular neuron or neurons that represent each category depend on the initial random synaptic weight vectors as well as the learning, so the output neurons are different from simulation run to simulation run. Self-organizing maps are investigated in Exercise 4.

D.4.1   Running the simulation

In the command window of Matlab, after you have performed a 'cd' into the directory where the source files are, type 'CompetitiveNetDemo'. The program will run for a bit, and then pause so that you can inspect what has been produced so far. A paused state is indicated by the words 'Paused: Press any key'. Press 'Enter' to move to the next stage, until the program finishes and the command prompt reappears. You can edit the code to comment out some of the 'pause' statements if you do not want them, or to add a 'pause' statement if you would

like the program to stop at a particular place so that you can inspect the results. To stop the program and exit from it, use 'Ctrl-C'. When at the command prompt, you can access variables by typing the variable name followed by 'Enter'. (Note: if you set '*display = 0*' near the beginning of the program, there will be few pauses, and you will be able to collect data quickly.)

The fully connected competitive network has N=100 output neurons with the dendrites that receive the synapses shown as vertical columns, and $nSyn$=100 synapses onto each neuron. With the display on (*display = 1*), at the first pause, you will see the 28 binary input patterns for training with firing rates of 1 or 0. Each pattern is 20 elements long, and each pattern is shifted down 3 elements with respect to the previous pattern, with the parameters supplied. (Each pattern thus overlaps in 17 / 20 locations with its predecessor.)

Next you will see the synaptic matrix being trained with the 28 input patterns, which are presented 20 times each (*nepochs* = 20) in random permuted sequence. Uncomment the pause in the training loop if you wish to see each pattern being presented sequentially. The synaptic weight matrix *SynMat* is initialized so that each neuron has random synaptic weights, with the synaptic weight vector on each neuron normalized to have a length of 1. Normalizing the length of the vector during training is important, for it ensures that no one neuron increases its synaptic strengths to very high values and thus wins for every input pattern. (Biological approximations to this are described by Rolls (2016a) in section B.4.9.2, and involve a type of heterosynaptic long-term depression of synaptic strength that is large when the postsynaptic term is high, and the presynaptic firing is low.) (In addition, for convenience, each input pattern is normalized to a length of 1, so that when the dot product is computed with a synaptic weight vector on a neuron, the maximum activation of a neuron will be 1.)

The firing rate of each neuron is computed using a binary threshold activation function (which can be changed in the code to threshold linear). The sparseness of the firing rate representation can be set in the code using *OutputSparseness*, and as supplied is set to 0.01 for the first demonstration, which results in one of the N=100 output neurons having a high firing rate. Because the (initially random) synaptic weight vectors of each neuron are different, the activations of each of the neurons will be different.

Learning then occurs, using an associative learning rule with synapses increasing between presynaptic neurons with high rates and postsynaptic neurons with high rates. Think of this as moving the synaptic weight vector of a high firing neuron to point closer to the input pattern that is making the neuron fire. At the same time, the other synapses on the same neuron become weaker due to the synaptic weight normalization, so the synaptic weight vector of the neuron moves away from other, different, patterns. In this way, the neurons self-organize so that some neurons point towards some patterns, and other neurons towards other patterns. You can watch these synaptic weight changes during the learning in the figure generated by the code, in which some of the synapses will become stronger (white) on a neuron and correspond to one or several of the patterns, and at the same time the other synapses on the neuron will become weaker (black). In the display of the synaptic weights, remember that each column represents the synaptic weights on a single neuron. The top thin row below the synaptic matrix represents the activations of the neurons, the bottom thin row the firing rates of the neurons, and the thin column to the left of the synaptic weight matrix the firing rates of the presynaptic input, i.e. the pattern currently being presented. Rates of 1 are shown as white, and of 0 as black. Neurons that remain unallocated, and do not respond to any of the input patterns, appear as vertical columns of random synaptic weights onto a neuron. These neurons remain available to potentially categorise different input patterns.

Next, testing with the different patterns starts. We are interested in whether similar input patterns activate the same output neurons; and different input patterns different output neurons. This is described as categorisation (Rolls 2016a). With the display on (*display* = 1), each time

you press 'Enter' the program will step through the patterns in sequence. You will notice in the synaptic matrix that with these initial parameters, only a few neurons have learned. When you step through each testing pattern in sequence, you will observe that one of the neurons responds to several patterns in the sequence, and these patterns are quite closely correlated with each other. Then another neuron becomes active for a further set of close patterns. This is repeated as other patterns are presented, and shows you that similar patterns tend to activate the same output neuron or neurons, and different input patterns tend to activate different output neurons. This illustrates by simulation the categorisation performed by competitive networks.

To illustrate these effects more quantitatively, after the patterns have been presented, you are shown first the correlations between the input patterns, and after that the correlations between the output firing rates produced by each pattern. These results are shown as correlation matrices. You should observe that with the network parameters supplied, the firing rates are grouped into several categories, each corresponding to several input patterns that are highly correlated in the output firing that they produce. The correlation matrix between the output firing rate for the different patterns thus illustrates how a competitive network can categorise similar patterns as similar to each other, and different patterns as different.

D.4.2 Exercises

1. Test the operation of the system with no learning, most easily achieved by setting *nepochs* = 0. This results in the input patterns being mapped to output rates through the random synaptic weight vectors for each neuron. What is your interpretation of what is shown by the correlation matrix between the output firing rates?

(Hint: the randomizing effect of the random synaptic weight vectors on each neuron is to separate out the patterns, with many of the patterns having no correlation with other patterns. In this mode, pattern separation is produced, and not categorisation in that similar inputs are less likely to produce similar outputs, a key property in perception and memory (Rolls 2016a). Pattern separation is important in its own right as a process (Rolls 2016f), and is implemented for example by the random non-associatively modifiable connections of the dentate mossy fibre synapses onto the CA3 neurons (Treves and Rolls 1992b, Rolls 2016a, Rolls 2013b, Kesner and Rolls 2015).)

2. Investigate and describe the effects of altering the similarity between the patterns on the categorization, by for example altering the value of *shift* in the range 2–20.

3. Investigate and describe the effects of making the output firing rate representation (*OutputSparseness*) less sparse. It is suggested that values in the range 0.01–0.1 are investigated.

4. If there is lateral facilitation of nearby neurons, which might be produced by short-range recurrent collaterals in the cortex, then self-organizing maps can be generated, providing a model of topographic maps in the cerebral cortex (Section B.4.6 and Rolls (2016a) Section B.4.6). Investigate this with the modified competitive network 'SOMdemo'. The map is made clear during the testing, in which the patterns are presented in order, and the neurons activated appear in different mapped positions in the firing rate array. Run the program several times to show that the details of the map are different each time, but that the principles are the same, that nearby neurons tend to respond to similar patterns, and that singularities (discontinuities) in the map can occur, just as are found in the primary visual cortex, V1.

The important change of 'SOMdemo' from 'CompetitiveNetDemo' is a local spatial filter for the firing rate array to simulate the effect of facilitation of the firing rates of nearby neurons by the short-range excitatory recurrent collateral connections between nearby cortical

pyramidal cells. This is implemented by the *SpatialFilter* kernel the range of which is set by *FilterRange*. Investigate the effects of altering *FilterRange* from its default value of 11 on the topographic maps that are formed. Other changes to produce 'SOMdemo' include modifying the *OutPutSparseness* of the Firing Rate representation to 0.3. Investigate the effect of modifying this on the maps being formed. Another alteration was to alter the activation function from binary threshold to linear threshold to enable graded firing rate representations to be formed.

What are the advantages of the representations formed when self-organizing maps are present? (Hint: Think of interpolation of untrained input patterns close to those already trained; and check Rolls (2016a) Section B.4.6.)

5. Further work. The code supplied may provide a route for readers to develop their own programs. Further examples of Matlab code used to investigate neural systems are available in Anastasio (2010). Matlab itself has a Neural Network Toolbox, with an introduction to its use in modelling simple networks provided in Wallisch, Lusignan, Benayoun, Baker, Dickey and Hatsoupoulos (2009).

D.5   Highlights

1. Appendix D describes Matlab software that has been made available with *Cerebral Cortex: Principles of Operation* at http://www.oxcns.org to provide simple demonstrations of the operation of some key neuronal networks related to cortical function.
2. The software demonstrates the operation of pattern association networks, autoassociation / attractor networks, competitive networks, and self-organizing maps.

References

Abbott, L. F. (1991). Realistic synaptic inputs for model neural networks, *Network* **2**: 245–258.

Abbott, L. F. and Blum, K. I. (1996). Functional significance of long-term potentiation for sequence learning and prediction, *Cerebral Cortex* **6**: 406–416.

Abbott, L. F. and Nelson, S. B. (2000). Synaptic plasticity: taming the beast, *Nature Neuroscience* **3**: 1178–1183.

Abbott, L. F. and Regehr, W. G. (2004). Synaptic computation, *Nature* **431**: 796–803.

Abbott, L. F., Rolls, E. T. and Tovee, M. J. (1996). Representational capacity of face coding in monkeys, *Cerebral Cortex* **6**: 498–505.

Abbott, L. F., Varela, J. A., Sen, K. and Nelson, S. B. (1997). Synaptic depression and cortical gain control, *Science* **275**: 220–224.

Abeles, M. (1991). *Corticonics: Neural Circuits of the Cerebral Cortex*, Cambridge University Press, Cambridge.

Abramson, N. (1963). *Information Theory and Coding*, McGraw-Hill, New York.

Ackley, D. H. (1987). *A Connectionist Machine for Genetic Hill-Climbing*, Kluwer Academic Publishers, Dordrecht.

Ackley, D. H., Hinton, G. E. and Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines, *Cognitive Science* **9**: 147–169.

Acsady, L., Kamondi, A., Sik, A., Freund, T. and Buzsaki, G. (1998). Gabaergic cells are the major postsynaptic targets of mossy fibers in the rat hippocampus, *Journal of Neuroscience* **18**: 3386–403.

Adrian, E. D. (1928). *The Basis of Sensations*, Christophers, London.

Aertsen, A. M. H. J., Gerstein, G. L., Habib, M. K. and Palm, G. (1989). Dynamics of neuronal firing correlation: modulation of 'effective connectivity', *Journal of Neurophysiology* **61**: 900–917.

Aggelopoulos, N. C. and Rolls, E. T. (2005). Natural scene perception: inferior temporal cortex neurons encode the positions of different objects in the scene, *European Journal of Neuroscience* **22**: 2903–2916.

Aggelopoulos, N. C., Franco, L. and Rolls, E. T. (2005). Object perception in natural scenes: encoding by inferior temporal cortex simultaneously recorded neurons, *Journal of Neurophysiology* **93**: 1342–1357.

Aggleton, J. P. and Nelson, A. J. D. (2015). Why do lesions in the rodent anterior thalamic nuclei cause such severe spatial deficits?, *Neuroscience & Biobehavioral Reviews* **54**: 108–119.

Aigner, T. G., Mitchell, S. J., Aggleton, J. P., DeLong, M. R., Struble, R. G., Price, D. L., Wenk, G. L., Pettigrew, K. D. and Mishkin, M. (1991). Transient impairment of recognition memory following ibotenic acid lesions of the basal forebrain in macaques, *Experimental Brain Research* **86**: 18–26.

Aimone, J. B. and Gage, F. H. (2011). Modeling new neuron function: a history of using computational neuroscience to study adult neurogenesis, *European Journal of Neuroscience* **33**: 1160–9.

Aimone, J. B., Deng, W. and Gage, F. H. (2010). Adult neurogenesis: integrating theories and separating functions, *Trends in Cognitive Sciences* **14**: 325–37.

Akrami, A., Liu, Y., Treves, A. and Jagadeesh, B. (2009). Converging neuronal activity in inferior temporal cortex during the classification of morphed stimuli, *Cerebral Cortex* **19**: 760–776.

Albantakis, L. and Deco, G. (2009). The encoding of alternatives in multiple-choice decision making, *Proceedings of the National Academy of Sciences USA* **106**: 10308–10313.

Alberini, C. M. and Ledoux, J. E. (2013). Memory reconsolidation, *Current Biology* **23**: R746–50.

Albus, J. S. (1971). A theory of cerebellar function, *Mathematical Biosciences* **10**: 25–61.

Aleman, A. and Kahn, R. S. (2005). Strange feelings: do amygdala abnormalities dysregulate the emotional brain in schizophrenia?, *Progress in Neurobiology* **77**(5): 283–298.

Alexander, G. E., Crutcher, M. D. and DeLong, M. R. (1990). Basal ganglia thalamo-cortical circuits: parallel substrates for motor, oculomotor, 'prefrontal' and 'limbic' functions, *Progress in Brain Research* **85**: 119–146.

Alexander, R. D. (1975). The search for a general theory of behavior, *Behavioral Sciences* **20**: 77–100.

Alexander, R. D. (1979). *Darwinism and Human Affairs*, University of Washington Press, Seattle.

Allport, A. (1988). What concept of consciousness?, *in* A. J. Marcel and E. Bisiach (eds), *Consciousness in Contemporary Science*, Oxford University Press, Oxford, pp. 159–182.

Alvarez, P. and Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: a simple network model, *Proceedings of the National Academy of Sciences USA* **91**: 7041–7045.

Amaral, D. G. (1986). Amygdalohippocampal and amygdalocortical projections in the primate brain, *in* R. Schwarcz and Y. Ben-Ari (eds), *Excitatory Amino Acids and Epilepsy*, Plenum Press, New York, pp. 3–18.

Amaral, D. G. (1987). Memory: anatomical organization of candidate brain regions, *in* F. Plum and V. Mountcastle (eds), *Higher Functions of the Brain. Handbook of Physiology, Part I*, American Physiological Society, Washington, DC, pp. 211–294.

Amaral, D. G. (1993). Emerging principles of intrinsic hippocampal organization, *Current Opinion in Neurobiology*

3: 225–229.

Amaral, D. G. and Price, J. L. (1984). Amygdalo-cortical projections in the monkey (Macaca fascicularis), *Journal of Comparative Neurology* **230**: 465–496.

Amaral, D. G. and Witter, M. P. (1989). The three-dimensional organization of the hippocampal formation: a review of anatomical data, *Neuroscience* **31**: 571–591.

Amaral, D. G. and Witter, M. P. (1995). The hippocampal formation, *in* G. Paxinos (ed.), *The Rat Nervous System*, Academic Press, San Diego, pp. 443–493.

Amaral, D. G., Ishizuka, N. and Claiborne, B. (1990). Neurons, numbers and the hippocampal network, *Progress in Brain Research* **83**: 1–11.

Amaral, D. G., Price, J. L., Pitkanen, A. and Carmichael, S. T. (1992). Anatomical organization of the primate amygdaloid complex, *in* J. P. Aggleton (ed.), *The Amygdala*, Wiley-Liss, New York, chapter 1, pp. 1–66.

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields, *Biological Cybernetics* **27**: 77–87.

Amari, S. (1982). Competitive and cooperative aspects in dynamics of neural excitation and self-organization, *in* S. Amari and M. A. Arbib (eds), *Competition and Cooperation in Neural Nets*, Springer, Berlin, chapter 1, pp. 1–28.

Amari, S., Yoshida, K. and Kanatani, K.-I. (1977). A mathematical foundation for statistical neurodynamics, *SIAM Journal of Applied Mathematics* **33**: 95–126.

Amit, D. J. (1989). *Modelling Brain Function*, Cambridge University Press, New York.

Amit, D. J. (1995). The Hebbian paradigm reintegrated: local reverberations as internal representations, *Behavioral and Brain Sciences* **18**: 617–657.

Amit, D. J. and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex, *Cerebral Cortex* **7**: 237–252.

Amit, D. J. and Tsodyks, M. V. (1991). Quantitative study of attractor neural network retrieving at low spike rates. I. Substrate – spikes, rates and neuronal gain, *Network* **2**: 259–273.

Amit, D. J., Gutfreund, H. and Sompolinsky, H. (1987). Statistical mechanics of neural networks near saturation, *Annals of Physics (New York)* **173**: 30–67.

Anastasio, T. J. (2010). *Tutorial on Neural Systems Modelling*, Sinauer, Sunderland, MA.

Andersen, P., Dingledine, R., Gjerstad, L., Langmoen, I. A. and Laursen, A. M. (1980). Two different responses of hippocampal pyramidal cells to application of gamma-aminobutyric acid, *Journal of Physiology* **307**: 279–296.

Andersen, P., Morris, R., Amaral, D., Bliss, T. and O'Keefe, J. (2007). *The Hippocampus Book*, Oxford University Press, London.

Andersen, R. A. (1995). Coordinate transformations and motor planning in the posterior parietal cortex, *in* M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, MIT Press, Cambridge, MA, chapter 33, pp. 519–532.

Andersen, R. A., Batista, A. P., Snyder, L. H., Buneo, C. A. and Cohen, Y. E. (2000). Programming to look and reach in the posterior parietal cortex, *in* M. Gazzaniga (ed.), *The New Cognitive Neurosciences*, 2 edn, MIT Press, Cambridge, MA, chapter 36, pp. 515–524.

Anderson, J. R. (1996). ACT: a simple theory of complex cognition, *American Psychologist* **51**: 355–365.

Arbib, M. A. (1964). *Brains, Machines, and Mathematics*, McGraw-Hill, New York (2nd Edn 1987 Springer).

Arcizet, F., Mirpour, K. and Bisley, J. W. (2011). A pure salience response in posterior parietal cortex, *Cerebral Cortex* **21**: 2498–2506.

Armstrong, D. M. and Malcolm, M. (1984). *Consciousness and Causality*, Blackwell, Oxford.

Arnold, P. D., Rosenberg, D. R., Mundo, E., Tharmalingam, S., Kennedy, J. L. and Richter, M. A. (2004). Association of a glutamate (NMDA) subunit receptor gene (GRIN2B) with obsessive-compulsive disorder: a preliminary study, *Psychopharmacology (Berl)* **174**: 530–538.

Arnsten, A. F. and Li, B. M. (2005). Neurobiology of executive functions: catecholamine influences on prefrontal cortical functions, *Biological Psychiatry* **57**: 1377–1384.

Aron, A. R., Robbins, T. W. and Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on, *Trends in Cognitive Sciences* **18**: 177–85.

Artola, A. and Singer, W. (1993). Long term depression: related mechanisms in cerebellum, neocortex and hippocampus, *in* M. Baudry, R. F. Thompson and J. L. Davis (eds), *Synaptic Plasticity: Molecular, Cellular and Functional Aspects*, MIT Press, Cambridge, MA, chapter 7, pp. 129–146.

Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing?, *Network* **3**: 213–251.

Atick, J. J. and Redlich, A. N. (1990). Towards a theory of early visual processing, *Neural Computation* **2**: 308–320.

Atick, J. J., Griffin, P. A. and Relich, A. N. (1996). The vocabulary of shape: principal shapes for probing perception and neural response, *Network* **7**: 1–5.

Attneave, F. (1954). Some informational aspects of visual perception, *Psychological Review* **61**: 183–193.

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*, Cambridge University Press, New York.

Bacon-Mace, N., Mace, M. J., Fabre-Thorpe, M. and Thorpe, S. J. (2005). The time course of visual processing: backward masking and natural scene categorisation, *Vision Research* **45**: 1459–1469.

Baddeley, R. J., Abbott, L. F., Booth, M. J. A., Sengpiel, F., Freeman, T., Wakeman, E. A. and Rolls, E. T. (1997).

Responses of neurons in primary and inferior temporal visual cortices to natural scenes, *Proceedings of the Royal Society B* **264**: 1775–1783.

Bagal, A. A., Kao, J. P. Y., Tang, C.-M. and Thompson, S. M. (2005). Long-term potentiation of exogenous glutamate responses at single dendritic spines., *Proc Natl Acad Sci U S A* **102**(40): 14434–9.

Balduzzi, D., Vanchinathan, H. and Buhmann, J. (2014). Kickback cuts backprop's red-tape: Biologically plausible credit assignment in neural networks, *arXiv preprint arXiv:1411.6191*.

Ballard, D. H. (1990). Animate vision uses object-centred reference frames, *in* R. Eckmiller (ed.), *Advanced Neural Computers*, North-Holland, Elsevier, Amsterdam, pp. 229–236.

Ballard, D. H. (1993). Subsymbolic modelling of hand-eye co-ordination, *in* D. E. Broadbent (ed.), *The Simulation of Human Intelligence*, Blackwell, Oxford, chapter 3, pp. 71–102.

Balu, D. T. and Coyle, J. T. (2015). The nmda receptor 'glycine modulatory site' in schizophrenia: d-serine, glycine, and beyond, *Current Opinion in Pharmacology* **20C**: 109–115.

Banks, S. J., Sziklas, V., Sodums, D. J. and Jones-Gotman, M. (2012). fMRI of verbal and nonverbal memory processes in healthy and epileptogenic medial temporal lobes, *Epilepsy and Behavior* **25**: 42–49.

Banks, W. P. (1978). Encoding and processing of symbolic information in comparative judgements, *in* G. H. Bower (ed.), *The Psychology of Learning and Motivation: Advances in Theory and Research*, Academic Press, pp. 101–159.

Bannon, S., Gonsalvez, C. J., Croft, R. J. and Boyce, P. M. (2002). Response inhibition deficits in obsessive-compulsive disorder, *Psychiatry Research* **110**: 165–174.

Bannon, S., Gonsalvez, C. J., Croft, R. J. and Boyce, P. M. (2006). Executive functions in obsessive-compulsive disorder: state or trait deficits?, *Aust N Z Journal of Psychiatry* **40**: 1031–1038.

Barbas, H. (1993). Organization of cortical afferent input to the orbitofrontal area in the rhesus monkey, *Neuroscience* **56**: 841–864.

Barkas, L. J., Henderson, J. L., Hamilton, D. A., Redhead, E. S. and Gray, W. P. (2010). Selective temporal resections and spatial memory impairment: cue dependent lateralization effects, *Behavioural Brain Research* **208**: 535–544.

Barlow, H. (1995). The neuron doctrine in perception, *in* M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, MIT Press, Cambridge, MA, chapter 26, pp. 415–435.

Barlow, H. (1997). Single neurons, communal goals, and consciousness, *in* M. Ito, Y. Miyashita and E. T. Rolls (eds), *Cognition, Computation, and Consciousness*, Oxford University Press, Oxford, chapter 7, pp. 121–136.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages, *in* W. Rosenblith (ed.), *Sensory Communication*, MIT Press, Cambridge, MA.

Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology, *Perception* **1**: 371–394.

Barlow, H. B. (1985). Cerebral cortex as model builder, *in* D. Rose and V. G. Dobson (eds), *Models of the Visual Cortex*, Wiley, Chichester, pp. 37–46.

Barlow, H. B. (1989). Unsupervised learning, *Neural Computation* **1**: 295–311.

Barlow, H. B., Kaushal, T. P. and Mitchison, G. J. (1989). Finding minimum entropy codes, *Neural Computation* **1**: 412–423.

Barnes, C. A. (2003). Long-term potentiation and the ageing brain, *Philosophical Transactions of the Royal Society of London B* **358**: 765–772.

Barnes, D. C., Hofacer, R. D., Zaman, A. R., Rennaker, R. L. and Wilson, D. A. (2008). Olfactory perceptual stability and discrimination, *Nature Neuroscience* **11**: 1378–80.

Barrett, H. (2012). A hierarchical model of the evolution of human brain specializations, *Proceedings of the National Academy of Sciences* **109**(Supplement 1): 10733–10740.

Bartlett, M. S. and Sejnowski, T. J. (1997). Viewpoint invariant face recognition using independent component analysis and attractor networks, *in* M. Mozer, M. Jordan and T. Petsche (eds), *Advances in Neural Information Processing Systems 9*, MIT Press, Cambridge, MA, pp. 817–823.

Barto, A. G. (1985). Learning by statistical cooperation of self-interested neuron-like computing elements, *Human Neurobiology* **4**: 229–256.

Barto, A. G. (1995). Adaptive critics and the basal ganglia, *in* J. C. Houk, J. L. Davis and D. G. Beiser (eds), *Models of Information Processing in the Basal Ganglia*, MIT Press, Cambridge, MA, chapter 11, pp. 215–232.

Bartol, T. M., Bromer, C., Kinney, J., Chirillo, M. A., Bourne, J. N., Harris, K. M. and Sejnowski, T. J. (2016). Nanoconnectomic upper bound on the variability of synaptic plasticity, *eLife* **4**: e10778.

Bartus, R. T. (2000). On neurodegenerative diseases, models, and treatment strategies: lessons learned and lessons forgotten a generation following the cholinergic hypothesis, *Experimental Neurology* **163**: 495–529.

Bassett, J. and Taube, J. S. (2005). Head direction signal generation: ascending and descending information streams, *in* S. I. Wiener and J. S. Taube (eds), *Head Direction Cells and the Neural Mechanisms of Spatial Orientation*, MIT Press, Cambridge, MA, chapter 5, pp. 83–109.

Battaglia, F. and Treves, A. (1998a). Stable and rapid recurrent processing in realistic autoassociative memories, *Neural Computation* **10**: 431–450.

Battaglia, F. P. and Treves, A. (1998b). Attractor neural networks storing multiple space representations: a model for hippocampal place fields, *Physical Review E* **58**: 7738–7753.

Baxter, M. G. and Murray, E. A. (2001a). Effects of hippocampal lesions on delayed nonmatching-to-sample in monkeys: a reply to Zola and Squire, *Hippocampus* **11**: 201–203.

Baxter, M. G. and Murray, E. A. (2001b). Opposite relationship of hippocampal and rhinal cortex damage to delayed nonmatching-to-sample deficits in monkeys, *Hippocampus* **11**: 61–71.

Baxter, R. D. and Liddle, P. F. (1998). Neuropsychological deficits associated with schizophrenic syndromes, *Schizophrenia Research* **30**: 239–249.

Baylis, G. C. and Rolls, E. T. (1987). Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks, *Experimental Brain Research* **65**: 614–622.

Baylis, G. C., Rolls, E. T. and Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey, *Brain Research* **342**: 91–102.

Baylis, G. C., Rolls, E. T. and Leonard, C. M. (1987). Functional subdivisions of temporal lobe neocortex, *Journal of Neuroscience* **7**: 330–342.

Baylis, L. L. and Gaffan, D. (1991). Amygdalectomy and ventromedial prefrontal ablation produce similar deficits in food choice and in simple object discrimination learning for an unseen reward, *Experimental Brain Research* **86**: 617–622.

Baylis, L. L. and Rolls, E. T. (1991). Responses of neurons in the primate taste cortex to glutamate, *Physiology and Behavior* **49**: 973–979.

Baylis, L. L., Rolls, E. T. and Baylis, G. C. (1994). Afferent connections of the orbitofrontal cortex taste area of the primate, *Neuroscience* **64**: 801–812.

Bear, M. F. and Singer, W. (1986). Modulation of visual cortical plasticity by acetylcholine and noradrenaline, *Nature* **320**: 172–176.

Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates, *American Journal of Psychiatry* **165**: 969–977.

Beck, D. M. and Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain, *Vision Research* **49**: 1154–1165.

Becker, S. and Hinton, G. E. (1992). Self-organizing neural network that discovers surfaces in random-dot stereograms, *Nature* **355**: 161–163.

Beckstead, R. M. and Norgren, R. (1979). An autoradiographic examination of the central distribution of the trigeminal, facial, glossopharyngeal, and vagal nerves in the monkey, *Journal of Comparative Neurology* **184**: 455–472.

Beckstead, R. M., Morse, J. R. and Norgren, R. (1980). The nucleus of the solitary tract in the monkey: projections to the thalamus and brainstem nuclei, *Journal of Comparative Neurology* **190**: 259–282.

Bell, A. J. and Sejnowski, T. J. (1995). An information-maximation approach to blind separation and blind deconvolution, *Neural Computation* **7**: 1129–1159.

Benda, J. and Herz, A. V. M. (2003). A universal model for spike-frequency adaptation, *Neural Computation* **15**: 2523–2564.

Bender, W., Albus, M., Moller, H. J. and Tretter, F. (2006). Towards systemic theories in biological psychiatry, *Pharmacopsychiatry* **39 Suppl 1**: S4–S9.

Bengio, Y., Goodfellow, I. J. and Courville, A. (2017). *Deep Learning*, MIT Press, Cambridge, MA.

Bennett, A. (1990). Large competitive networks, *Network* **1**: 449–462.

Berg, E. (1948). A simple objective technique for measuring flexibility in thinking, *Journal of General Psychology* **39**: 15–22.

Berlin, H., Rolls, E. T. and Kischka, U. (2004). Impulsivity, time perception, emotion, and reinforcement sensitivity in patients with orbitofrontal cortex lesions, *Brain* **127**: 1108–1126.

Berlin, H., Rolls, E. T. and Iversen, S. D. (2005). Borderline Personality Disorder, impulsivity, and the orbitofrontal cortex, *American Journal of Psychiatry* **58**: 234–245.

Bermpohl, F., Kahnt, T., Dalanay, U., Hagele, C., Sajonz, B., Wegner, T., Stoy, M., Adli, M., Kruger, S., Wrase, J., Strohle, A., Bauer, M. and Heinz, A. (2010). Altered representation of expected value in the orbitofrontal cortex in mania, *Human Brain Mapping* **31**: 958–969.

Bernard, A., Lubbers, L. S., Tanis, K. Q., Luo, R., Podtelezhnikov, A. A., Finney, E. M., McWhorter, M. M., Serikawa, K., Lemon, T., Morgan, R., Copeland, C., Smith, K., Cullen, V., Davis-Turak, J., Lee, C. K., Sunkin, S. M., Loboda, A. P., Levine, D. M., Stone, D. J., Hawrylycz, M. J., Roberts, C. J., Jones, A. R., Geschwind, D. H. and Lein, E. S. (2012). Transcriptional architecture of the primate neocortex, *Neuron* **73**: 1083–99.

Berridge, K. C., Flynn, F. W., Schulkin, J. and Grill, H. J. (1984). Sodium depletion enhances salt palatability in rats, *Behavioral Neuroscience* **98**: 652–660.

Bhattacharyya, S. and Chakraborty, K. (2007). Glutamatergic dysfunction–newer targets for anti-obsessional drugs, *Recent Patents CNS Drug Discovery* **2**: 47–55.

Bi, G.-Q. and Poo, M.-M. (1998). Activity-induced synaptic modifications in hippocampal culture, dependence on spike timing, synaptic strength and cell type, *Journal of Neuroscience* **18**: 10464–10472.

Bi, G.-Q. and Poo, M.-M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited, *Annual Review of Neuroscience* **24**: 139–166.

Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R. and Warland, D. (1991). Reading a neural code, *Science*

252: 1854–1857.

Biederman, I. (1972). Perceiving real-world scenes, *Science* **177**: 77–80.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding, *Psychological Review* **94**(2): 115–147.

Bienenstock, E. L., Cooper, L. N. and Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex, *Journal of Neuroscience* **2**: 32–48.

Bierer, L. M., Haroutunian, V., Gabriel, S., Knott, P. J., Carlin, L. S., Purohit, D. P., Perl, D. P., Schmeidler, J., Kanof, P. and Davis, K. L. (1995). Neurochemical correlates of dementia severity in Alzheimer's disease: relative importance of the cholinergic deficits, *Journal of Neurochemistry* **64**: 749–760.

Binford, T. O. (1981). Inferring surfaces from images, *Artificial Intelligence* **17**: 205–244.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford.

Bisley, J. W. and Goldberg, M. E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention, *Science* **299**: 81–86.

Bisley, J. W. and Goldberg, M. E. (2006). Neural correlates of attention and distractibility in the lateral intraparietal area, *Journal of Neurophysiology* **95**: 1696–1717.

Bisley, J. W. and Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe, *Annual Review of Neuroscience* **33**: 1–21.

Blair, R. J., Morris, J. S., Frith, C. D., Perrett, D. I. and Dolan, R. J. (1999). Dissociable neural responses to facial expressions of sadness and anger, *Brain* **122**: 883–893.

Blake, R. and Logothetis, N. K. (2002). Visual competition, *Nature Reviews Neuroscience* **3**: 13–21.

Blakemore, C. and van Sluyters, R. C. (1974). Reversal of the physiological effects of monocular deprivation in kittens: Further evidence for a sensitive period, *Journal of Physiology* **237**: 195–216.

Blakemore, C., Garey, L. J. and Vital-Durand, F. (1978). The physiological effects of monocular deprivation and their reversal in the monkeys visual cortex, *Journal of Physiology* **283**: 223–262.

Bliss, T. V. and Collingridge, G. L. (2013). Expression of NMDA receptor-dependent LTP in the hippocampus: bridging the divide, *Molecular Brain* **6**: 5.

Block, H. D. (1962). The perceptron: a model for brain functioning, *Reviews of Modern Physics* **34**: 123–135.

Block, N. (1995a). On a confusion about a function of consciousness, *Behavioral and Brain Sciences* **18**: 22–47.

Block, N. (1995b). Two neural correlates of consciousness, *Trends in Cognitive Sciences* **9**: 46–52.

Bloomfield, S. (1974). Arithmetical operations performed by nerve cells, *Brain Research* **69**: 115–124.

Bolles, R. C. and Cain, R. A. (1982). Recognizing and locating partially visible objects: The local-feature-focus method, *International Journal of Robotics Research* **1**: 57–82.

Bonelli, S. B., Powell, R. H., Yogarajah, M., Samson, R. S., Symms, M. R., Thompson, P. J., Koepp, M. J. and Duncan, J. S. (2010). Imaging memory in temporal lobe epilepsy: predicting the effects of temporal lobe resection, *Brain* **133**: 1186–1199.

Booth, D. A. (1985). Food-conditioned eating preferences and aversions with interoceptive elements: learned appetites and satieties, *Annals of the New York Academy of Sciences* **443**: 22–37.

Booth, M. C. A. and Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex, *Cerebral Cortex* **8**: 510–523.

Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L. and Rauschecker, J. P. (2015). Neurobiological roots of language in primate audition: common computational properties, *Trends in Cognitive Sciences* **19**: 142–150.

Borsini, F. and Rolls, E. T. (1984). Role of noradrenaline and serotonin in the basolateral region of the amygdala in food preferences and learned taste aversions in the rat, *Physiology and Behavior* **33**: 37–43.

Boussaoud, D., Desimone, R. and Ungerleider, L. G. (1991). Visual topography of area TEO in the macaque, *Journal of Computational Neurology* **306**: 554–575.

Bovier, A. and Gayrard, V. (1992). Rigorous bounds on the storage capacity of the dilute Hopfield model, *Journal of Statistical Physics* **69**: 597–627.

Brady, M., Ponce, J., Yuille, A. and Asada, H. (1985). Describing surfaces, *A. I. Memo 882, The Artificial Intelligence* **17**: 285–349.

Braitenberg, V. and Schütz, A. (1991). *Anatomy of the Cortex*, Springer-Verlag, Berlin.

Braitenberg, V. and Schütz, A. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity*, Springer-Verlag, Berlin.

Bressler, S. L. and Seth, A. K. (2011). Wiener-Granger causality: A well established methodology, *Neuroimage* **58**: 323–329.

Bridle, J. S. (1990). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition, *in* F. Fogelman-Soulie and J. Herault (eds), *Neurocomputing: Algorithms, Architectures and Applications*, Springer-Verlag, New York, pp. 227–236.

Brodmann, K. (1925). *Vergleichende Localisationslehre der Grosshirnrinde. Translated into English by L.J.Garey as Localisation in the Cerebral Cortex 1994 London: Smith-Gordon*, 2 edn, Barth, Leipzig.

Bromberg-Martin, E. S., Matsumoto, M. and Hikosaka, O. (2010a). Dopamine in motivational control: rewarding, aversive, and alerting, *Neuron* **68**: 815–834.

Bromberg-Martin, E. S., Matsumoto, M., Hong, S. and Hikosaka, O. (2010b). A pallidus-habenula-dopamine pathway

signals inferred stimulus values, *Journal of Neurophysiology* **104**: 1068–1076.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances, *in* E. Rosch and B. B. Lloyd (eds), *Cognition and Categorization*, Erlbaum, Hillsdale, NJ.

Brotchie, P., Andersen, R., Snyder, L. and Goodman, S. (1995). Head position signals used by parietal neurons to encode locations of visual stimuli, *Nature London* **375**: 232–235.

Brown, D. A., Gähwiler, B. H., Griffith, W. H. and Halliwell, J. V. (1990a). Membrane currents in hippocampal neurons, *Progress in Brain Research* **83**: 141–160.

Brown, J. M. (2009). Visual streams and shifting attention, *Progress in Brain Research* **176**: 47–63.

Brown, M. and Xiang, J. (1998). Recognition memory: neuronal substrates of the judgement of prior occurrence, *Progress in Neurobiology* **55**: 149–189.

Brown, T. H. and Zador, A. (1990). The hippocampus, *in* G. Shepherd (ed.), *The Synaptic Organization of the Brain*, Oxford University Press, New York, pp. 346–388.

Brown, T. H., Kairiss, E. W. and Keenan, C. L. (1990b). Hebbian synapses: biophysical mechanisms and algorithms, *Annual Review of Neuroscience* **13**: 475–511.

Brown, T. H., Ganong, A. H., Kairiss, E. W., Keenan, C. L. and Kelso, S. R. (eds) (1989). *Long-term Potentiation in Two Synaptic Systems of the Hippocampal Brain Slice*, Academic Press, San Diego.

Bruce, V. (1988). *Recognising Faces*, Erlbaum, Hillsdale, NJ.

Brun, V. H., Otnass, M. K., Molden, S., Steffenach, H. A., Witter, M. P., Moser, M. B. and Moser, E. I. (2002). Place cells and place recognition maintained by direct entorhinal–hippocampal circuitry, *Science* **296**: 2243–2246.

Brunel, N. and Wang, X. J. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition, *Journal of Computational Neuroscience* **11**: 63–85.

Buck, L. and Bargmann, C. I. (2013). Smell and taste: the chemical senses, *in* E. Kandel, J. H. Schwartz, T. H. Jessell, S. A. Siegelbaum and A. J. Hudspeth (eds), *Principles of Neural Science*, 5th edn, McGraw-Hill, New York, chapter 32, pp. 712–742.

Buckley, M. J. and Gaffan, D. (2000). The hippocampus, perirhinal cortex, and memory in the monkey, *in* J. J. Bolhuis (ed.), *Brain, Perception, and Memory: Advances in Cognitive Neuroscience*, Oxford University Press, Oxford, pp. 279–298.

Buckley, M. J. and Gaffan, D. (2006). Perirhinal contributions to object perception, *Trends in Cognitive Sciences* **10**: 100–107.

Buckley, M. J., Booth, M. C. A., Rolls, E. T. and Gaffan, D. (2001). Selective perceptual impairments following perirhinal cortex ablation, *Journal of Neuroscience* **21**: 9824–9836.

Buhl, E. H., Halasy, K. and Somogyi, P. (1994). Diverse sources of hippocampal unitary inhibitory postsynaptic potentials and the number of synaptic release sites, *Nature* **368**: 823–828.

Buhmann, J., Lange, J., von der Malsburg, C., Vorbrüggen, J. C. and Würtz, R. P. (1991). Object recognition in the dynamic link architecture: Parallel implementation of a transputer network, *in* B. Kosko (ed.), *Neural Networks for Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, pp. 121–159.

Bullier, J. and Nowak, L. (1995). Parallel versus serial processing: new vistas on the distributed organization of the visual system, *Current Opinion in Neurobiology* **5**: 497–503.

Bunsey, M. and Eichenbaum, H. (1996). Conservation of hippocampal memory function in rats and humans, *Nature* **379**: 255–257.

Buot, A. and Yelnik, J. (2012). Functional anatomy of the basal ganglia: limbic aspects, *Revue Neurologique (Paris)* **168**: 569–575.

Burgess, N. (2008). Spatial cognition and the brain, *Annals of the New York Academy of Sciences* **1124**: 77–97.

Burgess, N., Recce, M. and O'Keefe, J. (1994). A model of hippocampal function, *Neural Networks* **7**: 1065–1081.

Burgess, N., Jackson, A., Hartley, T. and O'Keefe, J. (2000). Predictions derived from modelling the hippocampal role in navigation, *Biological Cybernetics* **83**: 301–312.

Burgess, N., Maguire, E. A. and O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory, *Neuron* **35**: 625–641.

Burgess, P. W. (2000). Strategy application disorder: the role of the frontal lobes in human multitasking, *Psychological Research* **63**: 279–288.

Burke, S. N. and Barnes, C. A. (2006). Neural plasticity in the ageing brain, *Nature Reviews Neuroscience* **7**: 30–40.

Burton, M. J., Rolls, E. T. and Mora, F. (1976). Effects of hunger on the responses of neurones in the lateral hypothalamus to the sight and taste of food, *Experimental Neurology* **51**: 668–677.

Burwell, R. D., Witter, M. P. and Amaral, D. G. (1995). Perirhinal and postrhinal cortices of the rat: a review of the neuroanatomical literature and comparison with findings from the monkey brain, *Hippocampus* **5**: 390–408.

Bussey, T. J. and Saksida, L. M. (2005). Object memory and perception in the medial temporal lobe: an alternative approach, *Current Opinion in Neurobiology* **15**: 730–737.

Bussey, T. J., Saksida, L. M. and Murray, E. A. (2002). Perirhinal cortex resolves feature ambiguity in complex visual discriminations, *European Journal of Neuroscience* **15**: 365–374.

Bussey, T. J., Saksida, L. M. and Murray, E. A. (2003). Impairments in visual discrimination after perirhinal cortex lesions: testing "declarative" versus "perceptual-mnemonic" views of perirhinal cortex function, *European Journal of Neuroscience* **17**: 649–660.

Bussey, T. J., Saksida, L. M. and Murray, E. A. (2005). The perceptual-mnemonic / feature conjunction model of perirhinal cortex function, *Quarterly Journal of Experimental Psychology* **58B**: 269–282.

Butter, C. M. (1969). Perseveration in extinction and in discrimination reversal tasks following selective prefrontal ablations in Macaca mulatta, *Physiology and Behavior* **4**: 163–171.

Butter, C. M., McDonald, J. A. and Snyder, D. R. (1969). Orality, preference behavior, and reinforcement value of non-food objects in monkeys with orbital frontal lesions, *Science* **164**: 1306–1307.

Buxton, R. B. and Frank, L. R. (1997). A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation, *Journal of Cerebral Blood Flow and Metabolism* **17**: 64–72.

Buxton, R. B., Wong, E. C. and Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model, *Magnetic Resonance in Medicine* **39**: 855–864.

Buzsáki, G. (2006). *Rhythms of the Brain*, Oxford University Press, Oxford.

Byrne, R. W. and Whiten, A. (1988). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*, Clarendon Press, Oxford.

Caan, W., Perrett, D. I. and Rolls, E. T. (1984). Responses of striatal neurons in the behaving monkey. 2. Visual processing in the caudal neostriatum, *Brain Research* **290**: 53–65.

Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J. and DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition, *PLoS Computational Biology* **10**: e1003963.

Cahusac, P. M. B., Miyashita, Y. and Rolls, E. T. (1989). Responses of hippocampal formation neurons in the monkey related to delayed spatial response and object-place memory tasks, *Behavioural Brain Research* **33**: 229–240.

Cahusac, P. M. B., Rolls, E. T., Miyashita, Y. and Niki, H. (1993). Modification of the responses of hippocampal neurons in the monkey during the learning of a conditional spatial response task, *Hippocampus* **3**: 29–42.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W. R., Iversen, S. D. and David, A. S. (1997). Activation of auditory cortex during silent lip-reading, *Science* **276**: 593–596.

Camille, N., Tsuchida, A. and Fellows, L. K. (2011). Double dissociation of stimulus-value and action-value learning in humans with orbitofrontal or anterior cingulate cortex damage, *Journal of Neuroscience* **31**: 15048–15052.

Capuano, B., Crosby, I. T. and Lloyd, E. J. (2002). Schizophrenia: genesis, receptorology and current therapeutics, *Current Medicinal Chemistry* **9**: 521–548.

Cardinal, R. N., Parkinson, J. A., Hall, J. and Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex, *Neuroscience and Biobehavioral Reviews* **26**: 321–352.

Carlson, E. T., Rasquinha, R. J., Zhang, K. and Connor, C. E. (2011). A sparse object coding scheme in area v4, *Current Biology* **21**: 288–93.

Carlson, N. R. (2013). *Physiology of Behavior*, 11th edn, Pearson, Boston.

Carlsson, A. (2006). The neurochemical circuitry of schizophrenia, *Pharmacopsychiatry* **39 Suppl 1**: S10–14.

Carmichael, S. T. and Price, J. L. (1994). Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey, *Journal of Comparative Neurology* **346**: 366–402.

Carmichael, S. T. and Price, J. L. (1995a). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys, *Journal of Comparative Neurology* **363**: 615–641.

Carmichael, S. T. and Price, J. L. (1995b). Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys, *Journal of Comparative Neurology* **363**: 642–664.

Carmichael, S. T., Clugnet, M.-C. and Price, J. L. (1994). Central olfactory connections in the macaque monkey, *Journal of Comparative Neurology* **346**: 403–434.

Carpenter, G. A. (1997). Distributed learning, recognition and prediction by ART and ARTMAP neural networks, *Neural Networks* **10**(8): 1473–1494.

Carpenter, R. H. S. and Williams, M. (1995). Neural computation of log likelihood in control of saccadic eye movements, *Nature* **377**: 59–62.

Carruthers, P. (1996). *Language, Thought and Consciousness*, Cambridge University Press, Cambridge.

Carruthers, P. (2000). *Phenomenal Consciousness*, Cambridge University Press, Cambridge.

Carter, C. S., Perlstein, W., Ganguli, R., Brar, J., Mintun, M. and Cohen, J. D. (1998). Functional hypofrontality and working memory dysfunction in schizophrenia, *American Journal of Psychiatry* **155**: 1285–1287.

Caspers, J., Palomero-Gallagher, N., Caspers, S., Schleicher, A., Amunts, K. and Zilles, K. (2015). Receptor architecture of visual areas in the face and word-form recognition region of the posterior fusiform gyrus, *Brain Struct Funct* **220**: 205–19.

Cassaday, H. J. and Rawlins, J. N. (1997). The hippocampus, objects, and their contexts, *Behavioral Neuroscience* **111**: 1228–1244.

Castner, S. A., Williams, G. V. and Goldman-Rakic, P. S. (2000). Reversal of antipsychotic-induced working memory deficits by short-term dopamine D1 receptor stimulation, *Science* **287**: 2020–2022.

Cavanna, A. E. and Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates, *Brain* **129**: 564–583.

Celebrini, S., Thorpe, S., Trotter, Y. and Imbert, M. (1993). Dynamics of orientation coding in area V1 of the awake primate, *Visual Neuroscience* **10**: 811–825.

Cerasti, E. and Treves, A. (2010). How informative are spatial CA3 representations established by the dentate gyrus?, *PLoS Computational Biology* **6**: e1000759.

Cerasti, E. and Treves, A. (2013). The spatial representations acquired in CA3 by self-organizing recurrent connections, *Frontiers in Cellular Neuroscience* **7**: 112.

Cerella, J. (1986). Pigeons and perceptrons, *Pattern Recognition* **19**: 431–438.

Chadwick, M. J., Mullally, S. L. and Maguire, E. A. (2013). The hippocampus extrapolates beyond the view in scenes: an fmri study of boundary extension, *Cortex* **49**: 2067–2079.

Chakrabarty, K., Bhattacharyya, S., Christopher, R. and Khanna, S. (2005). Glutamatergic dysfunction in OCD, *Neuropsychopharmacology* **30**: 1735–1740.

Chakravarty, I. (1979). A generalized line and junction labeling scheme with applications to scene analysis, *IEEE Transactions PAMI* **1**: 202–205.

Chalmers, D. J. (1996). *The Conscious Mind*, Oxford University Press, Oxford.

Chamberlain, S. R., Fineberg, N. A., Blackwell, A. D., Robbins, T. W. and Sahakian, B. J. (2006). Motor inhibition and cognitive flexibility in obsessive-compulsive disorder and trichotillomania, *American Journal of Psychiatry* **163**: 1282–1284.

Chamberlain, S. R., Fineberg, N. A., Menzies, L. A., Blackwell, A. D., Bullmore, E. T., Robbins, T. W. and Sahakian, B. J. (2007). Impaired cognitive flexibility and motor inhibition in unaffected first-degree relatives of patients with obsessive-compulsive disorder, *American Journal of Psychiatry* **164**: 335–338.

Chan, A. M., Dykstra, A. R., Jayaram, V., Leonard, M. K., Travis, K. E., Gygi, B., Baker, J. M., Eskandar, E., Hochberg, L. R., Halgren, E. and Cash, S. S. (2014). Speech-specific tuning of neurons in human superior temporal gyrus, *Cerebral Cortex* **24**: 2679–2693.

Chau, B. K., Sallet, J., Papageorgiou, G. K., Noonan, M. P., Bell, A. H., Walton, M. E. and Rushworth, M. F. (2015). Contrasting roles for orbitofrontal cortex and amygdala in credit assignment and learning in macaques, *Neuron* **87**: 1106–1118.

Chelazzi, L. (1998). Serial attention mechanisms in visual search: a critical look at the evidence, *Psychological Research* **62**: 195–219.

Chelazzi, L., Miller, E., Duncan, J. and Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex, *Nature (London)* **363**: 345–347.

Cheney, D. L. and Seyfarth, R. M. (1990). *How Monkeys See the World*, University of Chicago Press, Chicago.

Cheng, W., Rolls, E. T., Gu, H., Zhang, J. and Feng, J. (2015). Autism: reduced connectivity between cortical areas involved in face expression, theory of mind, and the sense of self, *Brain* **138**: 1382–1398.

Cheng, W., Rolls, E. T., Qiu, J., Liu, W., Tang, Y., Huang, C.-C., Wang, X., Zhang, J., Lin, W., Zheng, L., Pu, J., Tsai, S.-J., Yang, A. C., Lin, C.-P., Wang, F., Xie, P. and Feng, J. (2016). Medial reward and lateral non-reward orbitofrontal cortex circuits change in opposite directions in depression, *Brain* p. in revision.

Chiba, A. A., Kesner, R. P. and Reynolds, A. M. (1994). Memory for spatial location as a function of temporal lag in rats: role of hippocampus and medial prefrontal cortex, *Behavioral and Neural Biology* **61**: 123–131.

Cho, Y. H. and Kesner, R. P. (1995). Relational object association learning in rats with hippocampal lesions, *Behavioural Brain Research* **67**: 91–98.

Cho, Y. K., Li, C. S. and Smith, D. V. (2002). Gustatory projections from the nucleus of the solitary tract to the parabrachial nuclei in the hamster, *Chemical Senses* **27**: 81–90.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*, MIT Press, Cambridge, Massachusetts.

Christie, B. R. (1996). Long-term depression in the hippocampus, *Hippocampus* **6**: 1–2.

Chu, J. and Anderson, S. A. (2015). Development of cortical interneurons, *Neuropsychopharmacology* **40**: 16–23.

Churchland, A. K., Kiani, R. and Shadlen, M. N. (2008). Decision-making with multiple alternatives, *Nature Neuroscience* **11**: 693–702.

Cicero, M. T. (55 BC). *De Oratore II*, Vol. II, Cicero, Rome.

Clelland, C. D., Choi, M., Romberg, C., Clemenson, G. D., J., Fragniere, A., Tyers, P., Jessberger, S., Saksida, L. M., Barker, R. A., Gage, F. H. and Bussey, T. J. (2009). A functional role for adult hippocampal neurogenesis in spatial pattern separation, *Science* **325**: 210–213.

Colby, C. L., Duhamel, J. R. and Goldberg, M. E. (1993). Ventral intraparietal area of the macaque - anatomic location and visual response properties, *Journal of Neurophysiology* **69**: 902–914.

Collingridge, G. L. and Bliss, T. V. P. (1987). NMDA receptors: their role in long-term potentiation, *Trends in Neurosciences* **10**: 288–293.

Compte, A., Brunel, N., Goldman-Rakic, P. and Wang, X. J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model, *Cerebral Cortex* **10**: 910–923.

Connor, C. E., Gallant, J. L., Preddie, D. and Van Essen, D. (1996). Responses in area V4 depend on the spatial relationship between stimulus and attention, *Journal of Neurophysiology* **75**: 1306–1308.

Cooper, J. R., Bloom, F. E. and Roth, R. H. (2003). *The Biochemical Basis of Neuropharmacology*, 8th edn, Oxford University Press, Oxford.

Corchs, S. and Deco, G. (2002). Large-scale neural model for visual attention: integration of experimental single cell and fMRI data, *Cerebral Cortex* **12**: 339–348.

Corr, P. J. and McNaughton, N. (2012). Neuroscience and approach/avoidance personality traits: a two stage

(valuation-motivation) approach, *Neuroscience and Biobehavioural Reviews* **36**: 2339–2254.

Corrado, G. S., Sugrue, L. P., Seung, H. S. and Newsome, W. T. (2005). Linear-nonlinear-Poisson models of primate choice dynamics, *Journal of the Experimental Analysis of Behavior* **84**: 581–617.

Cortes, C., Jaeckel, L. D., Solla, S. A., Vapnik, V. and Denker, J. S. (1996). Learning curves: asymptotic values and rates of convergence, *Neural Information Processing Systems* **6**: 327–334.

Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, *IEEE Transactions on Electronic Computers* **14**: 326–334.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*, Wiley, New York.

Cowell, R. A., Bussey, T. J. and Saksida, L. M. (2006). Why does brain damage impair memory? A connectionist model of object recognition memory in perirhinal cortex, *Journal of Neuroscience* **26**: 12186–12197.

Cowey, A. (1979). Cortical maps and visual perception, *Quarterly Journal of Experimental Psychology* **31**: 1–17.

Coyle, J. T. (2006). Glutamate and schizophrenia: beyond the dopamine hypothesis, *Cellular and Molecular Neurobiology* **26**: 365–384.

Coyle, J. T. (2012). Nmda receptor and schizophrenia: A brief history, *Schizophrenia Bulletin* **38**: 920–6.

Crane, J. and Milner, B. (2005). What went where? Impaired object-location learning in patients with right hippocampal lesions, *Hippocampus* **15**: 216–231.

Creutzfeldt, O. D. (1995). *Cortex Cerebri. Performance, Structural and Functional Organisation of the Cortex*, Oxford University Press, Oxford.

Crick, F. H. C. (1984). Function of the thalamic reticular complex: the searchlight hypothesis, *Proceedings of the National Academy of Sciences USA* **81**: 4586–4590.

Crick, F. H. C. and Koch, C. (1990). Towards a neurobiological theory of consciousness, *Seminars in the Neurosciences* **2**: 263–275.

Crick, F. H. C. and Mitchison, G. (1995). REM sleep and neural nets, *Behavioural Brain Research* **69**: 147–155.

Critchley, H. D. and Harrison, N. A. (2013). Visceral influences on brain and behavior, *Neuron* **77**: 624–638.

Critchley, H. D. and Rolls, E. T. (1996a). Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex, *Journal of Neurophysiology* **75**: 1673–1686.

Critchley, H. D. and Rolls, E. T. (1996b). Olfactory neuronal responses in the primate orbitofrontal cortex: analysis in an olfactory discrimination task, *Journal of Neurophysiology* **75**: 1659–1672.

Critchley, H. D. and Rolls, E. T. (1996c). Responses of primate taste cortex neurons to the astringent tastant tannic acid, *Chemical Senses* **21**: 135–145.

da Costa, N. M. and Martin, K. A. (2010). Whose cortical column would that be?, *Frontiers in Neuroanatomy* **4**: 16.

da Costa, N. M. and Martin, K. A. C. (2013). Sparse reconstruction of brain circuits: or, how to survive without a microscopic connectome, *Neuroimage* **80**: 27–36.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*, Grosset/Putnam, New York.

Damasio, A. R. (2003). *Looking for Spinoza*, Heinemann, London.

Dan, Y. and Poo, M.-M. (2004). Spike-timing dependent plasticity of neural circuits, *Neuron* **44**: 23–30.

Dan, Y. and Poo, M.-M. (2006). Spike-timing dependent plasticity: from synapse to perception, *Physiological Reviews* **86**: 1033–1048.

Dane, C. and Bajcsy, R. (1982). An object-centred three-dimensional model builder, *Proceedings of the 6th International Conference on Pattern Recognition*, Munich, pp. 348–350.

Darwin, C. (1859). *The Origin of Species*, John Murray [reprinted (1982) by Penguin Books Ltd], London.

Daugman, J. (1988). Complete discrete 2D-Gabor transforms by neural networks for image analysis and compression, *IEEE Transactions on Acoustic, Speech, and Signal Processing* **36**: 1169–1179.

Davies, M. (2008). Consciousness and explanation, *in* L. Weiskrantz and M. Davies (eds), *Frontiers of Consciousness*, Oxford University Press, Oxford, chapter 1, pp. 1–54.

Davis, M. (1992). The role of the amygdala in conditioned fear, *in* J. P. Aggleton (ed.), *The Amygdala*, Wiley-Liss, New York, chapter 9, pp. 255–306.

Davis, M. (2000). The role of the amygdala in conditioned and unconditioned fear and anxiety, *in* J. P. Aggleton (ed.), *The Amygdala: a Functional Analysis*, 2nd edn, Oxford University Press, Oxford, chapter 6, pp. 213–287.

Davis, M. (2011). NMDA receptors and fear extinction: implications for cognitive behavioral therapy, *Dialogues in Clinical Neuroscience* **13**: 463–474.

Dawkins, M. S. (1986). *Unravelling Animal Behaviour*, 1st edn, Longman, Harlow.

Dawkins, M. S. (1993). *Through Our Eyes Only? The Search for Animal Consciousness*, Freeman, Oxford.

Dawkins, M. S. (1995). *Unravelling Animal Behaviour*, 2nd edn, Longman, Harlow.

Dawkins, R. (1976). *The Selfish Gene*, Oxford University Press, Oxford.

Dawkins, R. (1989). *The Selfish Gene*, 2nd edn, Oxford University Press, Oxford.

Day, M., Langston, R. and Morris, R. G. (2003). Glutamate-receptor-mediated encoding and retrieval of paired-associate learning, *Nature* **424**: 205–209.

Dayan, P. and Abbott, L. F. (2001). *Theoretical Neuroscience*, MIT Press, Cambridge, MA.

Dayan, P. and Sejnowski, T. J. (1994). TD(λ) converges with probability 1, *Machine Learning* **14**: 295–301.

De Araujo, I. E. T. and Rolls, E. T. (2004). Representation in the human brain of food texture and oral fat, *Journal of Neuroscience* **24**: 3086–3093.

De Araujo, I. E. T., Rolls, E. T. and Stringer, S. M. (2001). A view model which accounts for the response properties of hippocampal primate spatial view cells and rat place cells, *Hippocampus* **11**: 699–706.

De Araujo, I. E. T., Kringelbach, M. L., Rolls, E. T. and Hobden, P. (2003a). Representation of umami taste in the human brain, *Journal of Neurophysiology* **90**: 313–319.

De Araujo, I. E. T., Kringelbach, M. L., Rolls, E. T. and McGlone, F. (2003b). Human cortical responses to water in the mouth, and the effects of thirst, *Journal of Neurophysiology* **90**: 1865–1876.

De Araujo, I. E. T., Rolls, E. T., Kringelbach, M. L., McGlone, F. and Phillips, N. (2003c). Taste-olfactory convergence, and the representation of the pleasantness of flavour in the human brain, *European Journal of Neuroscience* **18**: 2059–2068.

De Araujo, I. E. T., Rolls, E. T., Velazco, M. I., Margot, C. and Cayeux, I. (2005). Cognitive modulation of olfactory processing, *Neuron* **46**: 671–679.

De Gelder, B., Vroomen, J., Pourtois, G. and Weiskrantz, L. (1999). Non-conscious recognition of affect in the absence of striate cortex, *NeuroReport* **10**: 3759–3763.

de Lafuente, V., Jazayeri, M. and Shadlen, M. N. (2015). Representation of accumulating evidence for a decision in two parietal areas, *Journal of Neuroscience* **35**: 4306–4318.

de Ruyter van Steveninck, R. R. and Laughlin, S. B. (1996). The rates of information transfer at graded-potential synapses, *Nature* **379**: 642–645.

De Sieno, D. (1988). Adding a conscience to competitive learning, *IEEE International Conference on Neural Networks (San Diego 1988)*, Vol. 1, IEEE, New York, pp. 117–124.

De Valois, R. L. and De Valois, K. K. (1988). *Spatial Vision*, Oxford University Press, New York.

DeAngelis, G. C., Cumming, B. G. and Newsome, W. T. (2000). A new role for cortical area MT: the perception of stereoscopic depth, *in* M. Gazzaniga (ed.), *The New Cognitive Neurosciences, Second Edition*, MIT Press, Cambridge, MA, chapter 21, pp. 305–314.

Debiec, J., LeDoux, J. E. and Nader, K. (2002). Cellular and systems reconsolidation in the hippocampus, *Neuron* **36**: 527–538.

Debiec, J., Doyere, V., Nader, K. and LeDoux, J. E. (2006). Directly reactivated, but not indirectly reactivated, memories undergo reconsolidation in the amygdala, *Proceedings of the National Academy of Sciences USA* **103**: 3428–3433.

deCharms, R. C. and Merzenich, M. M. (1996). Primary cortical representation of sounds by the coordination of action-potential timing, *Nature* **381**: 610–613.

Deco, G. (2001). Biased competition mechanisms for visual attention, *in* S. Wermter, J. Austin and D. Willshaw (eds), *Emergent Neural Computational Architectures Based on Neuroscience*, Springer, Heidelberg, pp. 114–126.

Deco, G. and Lee, T. S. (2002). A unified model of spatial and object attention based on inter-cortical biased competition, *Neurocomputing* **44–46**: 775–781.

Deco, G. and Lee, T. S. (2004). The role of early visual cortex in visual integration: a neural model of recurrent interaction, *European Journal of Neuroscience* **20**: 1089–1100.

Deco, G. and Rolls, E. T. (2002). Object-based visual neglect: a computational hypothesis, *European Journal of Neuroscience* **16**: 1994–2000.

Deco, G. and 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.

Deco, G. and Rolls, E. T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition, *Vision Research* **44**: 621–644.

Deco, G. and Rolls, E. T. (2005a). Attention, short term memory, and action selection: a unifying theory, *Progress in Neurobiology* **76**: 236–256.

Deco, G. and Rolls, E. T. (2005b). Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons, *Journal of Neurophysiology* **94**: 295–313.

Deco, G. and Rolls, E. T. (2005c). Sequential memory: a putative neural and synaptic dynamical mechanism, *Journal of Cognitive Neuroscience* **17**: 294–307.

Deco, G. and Rolls, E. T. (2005d). Synaptic and spiking dynamics underlying reward reversal in the orbitofrontal cortex, *Cerebral Cortex* **15**: 15–30.

Deco, G. and Rolls, E. T. (2006). A neurophysiological model of decision-making and Weber's law, *European Journal of Neuroscience* **24**: 901–916.

Deco, G. and Zihl, J. (2001a). A neurodynamical model of visual attention: Feedback enhancement of spatial resolution in a hierarchical system, *Journal of Computational Neuroscience* **10**: 231–253.

Deco, G. and Zihl, J. (2001b). Top-down selective visual attention: a neurodynamical approach, *Visual Cognition* **8**: 119–140.

Deco, G. and Zihl, J. (2004). A biased competition based neurodynamical model of visual neglect, *Medical Engineering and Physics* **26**: 733–743.

Deco, G., Pollatos, O. and Zihl, J. (2002). The time course of selective visual attention: theory and experiments, *Vision Research* **42**: 2925–2945.

Deco, G., Rolls, E. T. and Horwitz, B. (2004). 'What' and 'where' in visual working memory: a computational neurodynamical perspective for integrating fMRI and single-neuron data, *Journal of Cognitive Neuroscience*

16: 683–701.

Deco, G., Scarano, L. and Soto-Faraco, S. (2007). Weber's law in decision making: integrating behavioral data in humans with a neurophysiological model, *Journal of Neuroscience* **27**: 11192–11200.

Deco, G., Rolls, E. T. and Romo, R. (2009). Stochastic dynamics as a principle of brain function, *Progress in Neurobiology* **88**: 1–16.

Deco, G., Rolls, E. T. and Romo, R. (2010). Synaptic dynamics and decision-making, *Proceedings of the National Academy of Sciences* **107**: 7545–7549.

Deco, G., Rolls, E. T., Albantakis, L. and Romo, R. (2013). Brain mechanisms for perceptual and reward-related decision-making, *Progress in Neurobiology* **103**: 194–213.

Deen, B., Koldewyn, K., Kanwisher, N. and Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus, *Cerebral Cortex* **25**: 4596–609.

Dehaene, S. and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework, *Cognition* **79**: 1–37.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J. and Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy, *Trends in Cognitive Sciences* **10**: 204–211.

Dehaene, S., Charles, L., King, J. R. and Marti, S. (2014). Toward a computational theory of conscious processing, *Current Opinion in Neurobiology* **25**: 76–84.

Delatour, B. and Witter, M. P. (2002). Projections from the parahippocampal region to the prefrontal cortex in the rat: evidence of multiple pathways, *European Journal of Neuroscience* **15**: 1400–1407.

DeLong, M. and Wichmann, T. (2010). Changing views of basal ganglia circuits and circuit disorders, *Clinical EEG and Neuroscience* **41**: 61–67.

DeLong, M. R., Georgopoulos, A. P., Crutcher, M. D., Mitchell, S. J., Richardson, R. T. and Alexander, G. E. (1984). Functional organization of the basal ganglia: Contributions of single-cell recording studies, *Functions of the Basal Ganglia. CIBA Foundation Symposium*, Pitman, London, pp. 64–78.

Delorme, A. and Thorpe, S. J. (2001). Face identification using one spike per neuron: resistance to image degradations, *Neural Networks* **14**: 795–803.

Delorme, R., Gousse, V., Roy, I., Trandafir, A., Mathieu, F., Mouren-Simeoni, M. C., Betancur, C. and Leboyer, M. (2007). Shared executive dysfunctions in unaffected relatives of patients with autism and obsessive-compulsive disorder, *European Psychiatry* **22**: 32–38.

Deng, W. L., Rolls, E. T., Ji, X., Robbins, T. W., IMAGEN, C. and Feng, J. (2016). Separate neural systems for behavioral change and for emotional responses to failure during behavioral inhibition.

Dennett, D. C. (1987). *The Intentional Stance*, MIT Press, Cambridge, MA.

Dennett, D. C. (1991). *Consciousness Explained*, Penguin, London.

Dere, E., Easton, A., Nadel, L. and Huston, J. P. (eds) (2008). *Handbook of Episodic Memory*, Elsevier, Amsterdam.

Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention, *Annual Review of Neuroscience* **18**: 193–222.

Desimone, R., Albright, T. D., Gross, C. G. and Bruce, C. (1984). Stimulus-selctive responses of inferior temporal neurons in the macaque, *Journal of Neuroscience* **4**: 2051–2062.

DeWeese, M. R. and Meister, M. (1999). How to measure the information gained from one symbol, *Network* **10**: 325–340.

Di Lorenzo, P. M. (1990). Corticofugal influence on taste responses in the parabrachial pons of the rat, *Brain Research* **530**: 73–84.

Diamond, M. E., Huang, W. and Ebner, F. F. (1994). Laminar comparison of somatosensory cortical plasticity, *Science* **265**: 1885–1888.

DiCarlo, J. J. and Maunsell, J. H. R. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position, *Journal of Neurophysiology* **89**: 3264–3278.

DiCarlo, J. J., Zoccolan, D. and Rust, N. C. (2012). How does the brain solve visual object recognition?, *Neuron* **73**: 415–434.

Dillingham, C. M., Frizzati, A., Nelson, A. J. D. and Vann, S. D. (2015). How do mammillary body inputs contribute to anterior thalamic function?, *Neuroscience & Biobehavioral Reviews* **54**: 108–119.

Ding, M., Chen, Y. and Bressler, S. L. (2006). Granger causality: basic theory and application to neuroscience, *in* B. Schelter, M. Winterhalder and J. Timmer (eds), *Handbook of Time Series Analysis*, Wiley, Weinheim, chapter 17, pp. 437–460.

Divac, I. (1975). Magnocellular nuclei of the basal forebrain project to neocortex, brain stem, and olfactory bulb. Review of some functional correlates, *Brain Research* **93**: 385–398.

Dolan, R. J., Fink, G. R., Rolls, E. T., Booth, M., Holmes, A., Frackowiak, R. S. J. and Friston, K. J. (1997). How the brain learns to see objects and faces in an impoverished context, *Nature* **389**: 596–599.

Douglas, R. J. and Martin, K. A. (2004). Neuronal circuits of the neocortex, *Annual Review of Neuroscience* **27**: 419–51.

Douglas, R. J. and Martin, K. A. (2007). Recurrent neuronal circuits in the neocortex, *Current Biology* **17**: R496–500.

Douglas, R. J. and Martin, K. A. C. (1990). Neocortex, *in* G. M. Shepherd (ed.), *The Synaptic Organization of the Brain*, 3rd edn, Oxford University Press, Oxford, chapter 12, pp. 389–438.

Douglas, R. J., Mahowald, M. A. and Martin, K. A. C. (1996). Microarchitecture of cortical columns, *in* A. Aertsen and V. Braitenberg (eds), *Brain Theory: Biological Basis and Computational Theory of Vision*, Elsevier, Amsterdam.

Douglas, R. J., Markram, H. and Martin, K. A. C. (2004). Neocortex, *in* G. M. Shepherd (ed.), *The Synaptic Organization of the Brain*, 5th edn, Oxford University Press, Oxford, chapter 12, pp. 499–558.

Dow, B. W., Snyder, A. Z., Vautin, R. G. and Bauer, R. (1981). Magnification factor and receptive field size in foveal striate cortex of the monkey, *Experimental Brain Research* **44**: 213–218.

Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex?, *Neural Networks* **12**: 961–974.

Doyere, V., Debiec, J., Monfils, M. H., Schafe, G. E. and LeDoux, J. E. (2007). Synapse-specific reconsolidation of distinct fear memories in the lateral amygdala, *Nature Neuroscience* **10**: 414–416.

Drevets, W. C. (2007). Orbitofrontal cortex function and structure in depression, *Annals of the New York Academy of Sciences* **1121**: 499–527.

Dudai, Y. (2012). The restless engram: consolidations never end, *Annual Review of Neuroscience* **35**: 227–247.

Dudchenko, P. A., Wood, E. R. and Eichenbaum, H. (2000). Neurotoxic hippocampal lesions have no effect on odor span and little effect on odor recognition memory but produce significant impairments on spatial span, recognition, and alternation, *Journal of Neuroscience* **20**: 2964–2977.

Duhamel, J. R., Colby, C. L. and Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements, *Science* **255**: 90–92.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli, *Psychological Review* **87**: 272–300.

Duncan, J. (1996). Cooperating brain systems in selective perception and action, *in* T. Inui and J. L. McClelland (eds), *Attention and Performance XVI*, MIT Press, Cambridge, MA, pp. 549–578.

Duncan, J. and Humphreys, G. (1989). Visual search and stimulus similarity, *Psychological Review* **96**: 433–458.

Durbin, R. and Mitchison, G. (1990). A dimension reduction framework for understanding cortical maps, *Nature* **343**: 644–647.

Durstewitz, D. and Seamans, J. K. (2002). The computational role of dopamine D1 receptors in working memory, *Neural Networks* **15**: 561–572.

Durstewitz, D., Kelc, M. and Gunturkun, O. (1999). A neurocomputational theory of the dopaminergic modulation of working memory functions, *Journal of Neuroscience* **19**: 2807–2822.

Durstewitz, D., Seamans, J. K. and Sejnowski, T. J. (2000a). Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex, *Journal of Neurophysiology* **83**: 1733–1750.

Durstewitz, D., Seamans, J. K. and Sejnowski, T. J. (2000b). Neurocomputational models of working memory, *Nature Neuroscience* **3 Suppl**: 1184–1191.

Easton, A. and Gaffan, D. (2000). Amygdala and the memory of reward: the importance of fibres of passage from the basal forebrain, *in* J. P. Aggleton (ed.), *The Amygdala: a Functional Analysis*, 2nd edn, Oxford University Press, Oxford, chapter 17, pp. 569–586.

Easton, A., Ridley, R. M., Baker, H. F. and Gaffan, D. (2002). Unilateral lesions of the cholinergic basal forebrain and fornix in one hemisphere and inferior temporal cortex in the opposite hemisphere produce severe learning impairments in rhesus monkeys, *Cerebral Cortex* **12**: 729–736.

Eccles, J. C. (1984). The cerebral neocortex: a theory of its operation, *in* E. G. Jones and A. Peters (eds), *Cerebral Cortex: Functional Properties of Cortical Cells*, Vol. 2, Plenum, New York, chapter 1, pp. 1–36.

Eccles, J. C., Ito, M. and Szentagothai, J. (1967). *The Cerebellum as a Neuronal Machine*, Springer-Verlag, New York.

Eckhorn, R. and Popel, B. (1974). Rigorous and extended application of information theory to the afferent visual system of the cat. I. Basic concepts, *Kybernetik* **16**: 191–200.

Eckhorn, R. and Popel, B. (1975). Rigorous and extended application of information theory to the afferent visual system of the cat. II. Experimental results, *Kybernetik* **17**: 7–17.

Eckhorn, R., Grusser, O. J., Kroller, J., Pellnitz, K. and Popel, B. (1976). Efficiency of different neural codes: information transfer calculations for three different neuronal systems, *Biological Cybernetics* **22**: 49–60.

Edelman, S. (1999). *Representation and Recognition in Vision*, MIT Press, Cambridge, MA.

Eichenbaum, H. (1997). Declarative memory: insights from cognitive neurobiology, *Annual Review of Psychology* **48**: 547–572.

Eichenbaum, H. (2014). Time cells in the hippocampus: a new dimension for mapping memories, *Nature Reviews Neuroscience* **15**: 732–744.

Eichenbaum, H. and Cohen, N. J. (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain*, Oxford University Press, New York.

Eichenbaum, H., Otto, T. and Cohen, N. J. (1992). The hippocampus - what does it do?, *Behavioural and Neural Biology* **57**: 2–36.

Ekstrom, A. D., Kahana, M. J., Caplan, J. B., Fields, T. A., Isham, E. A., Newman, E. L. and Fried, I. (2003). Cellular networks underlying human spatial navigation, *Nature* **425**: 184–188.

Elliffe, M. C. M., Rolls, E. T., Parga, N. and Renart, A. (2000). A recurrent model of transformation invariance by

association, *Neural Networks* **13**: 225–237.

Elliffe, M. C. M., Rolls, E. T. and Stringer, S. M. (2002). Invariant recognition of feature combinations in the visual system, *Biological Cybernetics* **86**: 59–71.

Elston, G. N., Benavides-Piccione, R., Elston, A., Zietsch, B., Defelipe, J., Manger, P., Casagrande, V. and Kaas, J. H. (2006). Specializations of the granular prefrontal cortex of primates: implications for cognitive processing, *Anatomical Record A Discov Mol Cell Evol Biol* **288**: 26–35.

Engel, A. K., Konig, P., Kreiter, A. K., Schillen, T. B. and Singer, W. (1992). Temporal coding in the visual system: new vistas on integration in the nervous system, *Trends in Neurosciences* **15**: 218–226.

Engel, A. K., Roelfsema, P. R., Fries, P., Brecht, M. and Singer, W. (1997). Role of the temporal domain for response selection and perceptual binding, *Cerebral Cortex* **7**: 571–582.

Enomoto, K., Matsumoto, N., Nakai, S., Satoh, T., Sato, T. K., Ueda, Y., Inokawa, H., Haruno, M. and Kimura, M. (2011). Dopamine neurons learn to encode the long-term value of multiple future rewards, *Proceedings of the National Academy of Sciences U S A* **108**: 15462–15467.

Epstein, J., Stern, E. and Silbersweig, D. (1999). Mesolimbic activity associated with psychosis in schizophrenia. Symptom-specific PET studies, *Annals of the New York Academy of Sciences* **877**: 562–574.

Epstein, R. and Kanwisher, N. (1998). A cortical representation of the local visual environment, *Nature* **392**: 598–601.

Esber, G. R. and Haselgrove, M. (2011). Reconciling the influence of predictiveness and uncertainty on stimulus salience: a model of attention in associative learning, *Proceedings of the Royal Society of London B: Biological Sciences* p. rspb20110836.

Eshel, N. and Roiser, J. P. (2010). Reward and punishment processing in depression, *Biological Psychiatry* **68**: 118–124.

Eskandar, E. N., Richmond, B. J. and Optican, L. M. (1992). Role of inferior temporal neurons in visual memory. I. Temporal encoding of information about visual images, recalled images, and behavioural context, *Journal of Neurophysiology* **68**: 1277–1295.

Espinosa, J. S. and Stryker, M. P. (2012). Development and plasticity of the primary visual cortex, *Neuron* **75**: 230–249.

Estes, W. K. (1986). Memory for temporal information, *in* J. A. Michon and J. L. Jackson (eds), *Time, Mind and Behavior*, Springer-Verlag, New York, pp. 151–168.

Everling, S., Tinsley, C. J., Gaffan, D. and Duncan, J. (2006). Selective representation of task-relevant objects and locations in the monkey prefrontal cortex, *European Journal of Neuroscience* **23**: 2197–2214.

Fahy, F. L., Riches, I. P. and Brown, M. W. (1993). Neuronal activity related to visual recognition memory and the encoding of recency and familiarity information in the primate anterior and medial inferior temporal and rhinal cortex, *Experimental Brain Research* **96**: 457–492.

Faisal, A., Selen, L. and Wolpert, D. (2008). Noise in the nervous system, *Nature Reviews Neuroscience* **9**: 292–303.

Farah, M. J. (1990). *Visual Agnosia*, MIT Press, Cambridge, MA.

Farah, M. J. (2000). *The Cognitive Neuroscience of Vision*, Blackwell, Oxford.

Farah, M. J. (2004). *Visual Agnosia*, 2nd edn, MIT Press, Cambridge, MA.

Farah, M. J., Meyer, M. M. and McMullen, P. A. (1996). The living/nonliving dissociation is not an artifact: giving an *a priori* implausible hypothesis a strong test, *Cognitive Neuropsychology* **13**: 137–154.

Fatt, P. and Katz, B. (1950). Some observations on biological noise, *Nature* **166**: 597–598.

Faugeras (1993). *The Representation, Recognition and Location of 3-D Objects*, MIT Press, Cambridge, MA.

Faugeras, O. D. and Hebert, M. (1986). The representation, recognition and location of 3-D objects, *International Journal of Robotics Research* **5**: 27–52.

Fazeli, M. S. and Collingridge, G. L. (eds) (1996). *Cortical Plasticity: LTP and LTD*, Bios, Oxford.

Feigenbaum, J. D. and Rolls, E. T. (1991). Allocentric and egocentric spatial information processing in the hippocampal formation of the behaving primate, *Psychobiology* **19**: 21–40.

Feldman, D. E. (2009). Synaptic mechanisms for plasticity in neocortex, *Annual Reviews of Neuroscience* **32**: 33–55.

Feldman, D. E. (2012). The spike-timing dependence of plasticity, *Neuron* **75**: 556–571.

Feldman, J. A. (1985). Four frames suffice: a provisional model of vision and space, *Behavioural Brain Sciences* **8**: 265–289.

Felleman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex, *Cerebral Cortex* **1**: 1–47.

Fellows, L. K. (2007). The role of orbitofrontal cortex in decision making: a component process account, *Annalls of the New York Academy of Sciences* **1121**: 421–430.

Fellows, L. K. (2011). Orbitofrontal contributions to value-based decision making: evidence from humans with frontal lobe damage, *Annals of the New York Academy of Sciences* **1239**: 51–58.

Fellows, L. K. and Farah, M. J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm, *Brain* **126**: 1830–1837.

Ferbinteanu, J., Holsinger, R. M. and McDonald, R. J. (1999). Lesions of the medial or lateral perforant path have different effects on hippocampal contributions to place learning and on fear conditioning to context, *Behavioral Brain Research* **101**: 65–84.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells,

Journal of the Optical Society of America, A **4**: 2379–2394.

Field, D. J. (1994). What is the goal of sensory coding?, *Neural Computation* **6**: 559–601.

Finkel, L. H. and Edelman, G. M. (1987). Population rules for synapses in networks, *in* G. M. Edelman, W. E. Gall and W. M. Cowan (eds), *Synaptic Function*, John Wiley & Sons, New York, pp. 711–757.

Fiorillo, C. D., Tobler, P. N. and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons, *Science* **299**: 1898–1902.

Florian, C. and Roullet, P. (2004). Hippocampal CA3-region is crucial for acquisition and memory consolidation in Morris water maze task in mice, *Behavioral Brain Research* **154**: 365–374.

Fodor, J. A. (1994). *The Elm and the Expert: Mentalese and its Semantics*, MIT Press, Cambridge, MA.

Földiák, P. (1991). Learning invariance from transformation sequences, *Neural Computation* **3**: 193–199.

Földiák, P. (1992). Models of sensory coding, *Technical Report CUED/F–INFENG/TR 91*, University of Cambridge, Department of Engineering, Cambridge.

Földiák, P. (2003). Sparse coding in the primate cortex, *in* M. A. Arbib (ed.), *Handbook of Brain Theory and Neural Networks*, 2nd edn, MIT Press, Cambridge, MA, pp. 1064–1608.

Forgeard, M. J., Haigh, E. A., Beck, A. T., Davidson, R. J., Henn, F. A., Maier, S. F., Mayberg, H. S. and Seligman, M. E. (2011). Beyond depression: Towards a process-based approach to research, diagnosis, and treatment, *Clinical Psychology (New York)* **18**: 275–299.

Foster, T. C., Castro, C. A. and McNaughton, B. L. (1989). Spatial selectivity of rat hippocampal neurons: dependence on preparedness for movement, *Science* **244**: 1580–1582.

Francis, S., Rolls, E. T., Bowtell, R., McGlone, F., O'Doherty, J., Browning, A., Clare, S. and Smith, E. (1999). The representation of pleasant touch in the brain and its relationship with taste and olfactory areas, *NeuroReport* **10**: 453–459.

Franco, L., Rolls, E. T., Aggelopoulos, N. C. and Treves, A. (2004). The use of decoding to analyze the contribution to the information of the correlations between the firing of simultaneously recorded neurons, *Experimental Brain Research* **155**: 370–384.

Franco, L., Rolls, E. T., Aggelopoulos, N. C. and Jerez, J. M. (2007). Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex, *Biological Cybernetics* **96**: 547–560.

Frankland, S. M. and Greene, J. D. (2015). An architecture for encoding sentence meaning in left mid-superior temporal cortex, *Proceedings of the National Academy of Sciences of the United States of America* **112**: 11732–7.

Franks, K. M., Stevens, C. F. and Sejnowski, T. J. (2003). Independent sources of quantal variability at single glutamatergic synapses, *Journal of Neuroscience* **23**: 3186–3195.

Franzius, M., Sprekeler, H. and Wiskott, L. (2007). Slowness and sparseness lead to place, head-direction, and spatial-view cells, *PLoS Computational Biology* **3(8)**: e166.

Freedman, D. J. (2015). Learning-dependent plasticity of visual encoding in inferior temporal cortex, *J Vis* **15**: 1420.

Frégnac, Y. (1996). Dynamics of cortical connectivity in visual cortical networks: an overview, *Journal of Physiology, Paris* **90**: 113–139.

Fregnac, Y., Pananceau, M., Rene, A., Huguet, N., Marre, O., Levy, M. and Shulz, D. E. (2010). A re-examination of Hebbian-covariance rules and spike timing-dependent plasticity in cat visual cortex in vivo, *Frontiers in Synaptic Neuroscience* **2**: 147.

Freiwald, W. A., Tsao, D. Y. and Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe, *Nature Neuroscience* **12**: 1187–96.

Fremaux, N. and Gerstner, W. (2015). Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules, *Frontiers in Neural Circuits* **9**: 85.

Freud, S. (1900). *The Interpretation of Dreams*.

Frey, S. and Petrides, M. (2002). Orbitofrontal cortex and memory formation, *Neuron* **36**: 171–176.

Fried, I., MacDonald, K. A. and Wilson, C. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects, *Neuron* **18**: 753–765.

Fried, I., Rutishauser, U., Cerf, M. and Kreiman, G. (2014). *Single Neuron Studies of the Human Brain: Probing Cognition*, MIT Press.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence, *Trends in Cognitive Sciences* **9**: 474–480.

Fries, P. (2009). Neuronal gamma-band synchronization as a fundamental process in cortical computation, *Annual Reviews of Neuroscience* **32**: 209–224.

Fries, P. (2015). Rhythms for cognition: Communication through coherence, *Neuron* **88**: 220–35.

Fries, P., Reynolds, J., Rorie, A. and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention, *Science* **291**: 1560–1563.

Fries, P., Womelsdorf, T., Oostenveld, R. and Desimone, R. (2008). The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area V4, *Journal of Neuroscience* **28**: 4823–4835.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E. T. and Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging, *Neuroimage* **6**: 218–229.

Frolov, A. A. and Medvedev, A. V. (1986). Substantiation of the "point approximation" for describing the total electrical activity of the brain with use of a simulation model, *Biophysics* **31**: 332–337.

Fuhrmann, G., Markram, H. and Tsodyks, M. (2002a). Spike frequency adaptation and neocortical rhythms, *Journal of Neurophysiology* **88**: 761–770.

Fuhrmann, G., Segev, I., Markram, H. and Tsodyks, M. (2002b). Coding of temporal information by activity-dependent synapses, *Journal of Neurophysiology* **87**: 140–148.

Fujimichi, R., Naya, Y., Koyano, K. W., Takeda, M., Takeuchi, D. and Miyashita, Y. (2010). Unitized representation of paired objects in area 35 of the macaque perirhinal cortex, *European Journal of Neuroscience* **32**: 659–667.

Fukushima, K. (1975). Cognitron: a self-organizing neural network, *Biological Cybernetics* **20**: 121–136.

Fukushima, K. (1980). Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, *Biological Cybernetics* **36**: 193–202.

Fukushima, K. (1988). Neocognitron: A hierarchical neural network model capable of visual pattern recognition unaffected by shift in position, *Neural Networks* **1**: 119–130.

Fukushima, K. (1989). Analysis of the process of visual pattern recognition by the neocognitron, *Neural Networks* **2**: 413–420.

Fukushima, K. (1991). Neural networks for visual pattern recognition, *IEEE Transactions E* **74**: 179–190.

Fukushima, K. and Miyake, S. (1982). Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position, *Pattern Recognition* **15**(6): 455–469.

Funahashi, S., Bruce, C. and Goldman-Rakic, P. (1989). Mnemonic coding of visual space in monkey dorsolateral prefrontal cortex, *Journal of Neurophysiology* **61**: 331–349.

Furman, M. and Wang, X.-J. (2008). Similarity effect and optimal control of multiple-choice decision making, *Neuron* **60**: 1153–1168.

Fusi, S. and Abbott, L. F. (2007). Limits on the memory storage capacity of bounded synapses, *Nature Neuroscience* **10**: 485–493.

Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory, *Joural of Neurophysiology* **36**: 61–78.

Fuster, J. M. (1989). *The Prefrontal Cortex*, 2nd edn, Raven Press, New York.

Fuster, J. M. (2000). *Memory Systems in the Brain*, Raven Press, New York.

Fuster, J. M. (2008). *The Prefrontal Cortex*, 4th edn, Academic Press, London.

Fuster, J. M. (2014). The prefrontal cortex makes the brain a preadaptive system, *Proceedings of the IEEE* **102**: 417–426.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I. and Moser, M.-B. (2004). Spatial representation in the entorhinal cortex, *Science* **2004**: 1258–1264.

Gabbott, P. and Rolls, E. T. (2013). Increased neuronal firing in resting and sleep in areas of the macaque medial prefrontal cortex (mPFC) that are part of the default mode network, *European Journal of Neuroscience* **37**: 1737–1746.

Gabbott, P. L., Warner, T. A., Jays, P. R. and Bacon, S. J. (2003). Areal and synaptic interconnectivity of prelimbic (area 32), infralimbic (area 25) and insular cortices in the rat, *Brain Research* **993**: 59–71.

Gaffan, D. (1993). Additive effects of forgetting and fornix transection in the temporal gradient of retrograde amnesia, *Neuropsychologia* **31**: 1055–1066.

Gaffan, D. (1994). Scene-specific memory for objects: a model of episodic memory impairment in monkeys with fornix transection, *Journal of Cognitive Neuroscience* **6**: 305–320.

Gaffan, D. and Harrison, S. (1989a). A comparison of the effects of fornix section and sulcus principalis ablation upon spatial learning by monkeys, *Behavioural Brain Research* **31**: 207–220.

Gaffan, D. and Harrison, S. (1989b). Place memory and scene memory: effects of fornix transection in the monkey, *Experimental Brain Research* **74**: 202–212.

Gaffan, D. and Saunders, R. C. (1985). Running recognition of configural stimuli by fornix transected monkeys, *Quarterly Journal of Experimental Psychology* **37B**: 61–71.

Gaffan, D., Saunders, R. C., Gaffan, E. A., Harrison, S., Shields, C. and Owen, M. J. (1984). Effects of fornix section upon associative memory in monkeys: role of the hippocampus in learned action, *Quarterly Journal of Experimental Psychology* **36B**: 173–221.

Gallant, J. L., Connor, C. E. and Van-Essen, D. C. (1998). Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing, *NeuroReport* **9**: 85–90.

Gardner, E. (1987). Maximum storage capacity in neural networks, *Europhysics Letters* **4**: 481–485.

Gardner, E. (1988). The space of interactions in neural network models, *Journal of Physics A* **21**: 257–270.

Gardner-Medwin, A. R. (1976). The recall of events through the learning of associations between their parts, *Proceedings of the Royal Society of London, Series B* **194**: 375–402.

Gattass, R., Sousa, A. P. B. and Covey, E. (1985). Cortical visual areas of the macaque: possible substrates for pattern recognition mechanisms, *Experimental Brain Research, Supplement 11, Pattern Recognition Mechanisms* **54**: 1–20.

Gawne, T. J. and Richmond, B. J. (1993). How independent are the messages carried by adjacent inferior temporal cortical neurons?, *Journal of Neuroscience* **13**: 2758–2771.

Gaykema, R. P., van der Kuil, J., Hersh, L. B. and Luiten, P. G. (1991). Patterns of direct projections from the hippocampus to the medial septum-diagonal band complex: anterograde tracing with phaseolus vulgaris leucoagglutinin combined with immunohistochemistry of choline acetyltransferase, *Neuroscience* **43**: 349–360.

Gazzaniga, M. S. (1988). Brain modularity: towards a philosophy of conscious experience, *in* A. J. Marcel and E. Bisiach (eds), *Consciousness in Contemporary Science*, Oxford University Press, Oxford, chapter 10, pp. 218–238.

Gazzaniga, M. S. (1995). Consciousness and the cerebral hemispheres, *in* M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, MIT Press, Cambridge, MA, chapter 92, pp. 1392–1400.

Gazzaniga, M. S. and LeDoux, J. (1978). *The Integrated Mind*, Plenum, New York.

Ge, T., Feng, J., Grabenhorst, F. and Rolls, E. T. (2012). Componential Granger causality, and its application to identifying the source and mechanisms of the top-down biased activation that controls attention to affective vs sensory processing, *Neuroimage* **59**: 1846–1858.

Geesaman, B. J. and Andersen, R. A. (1996). The analysis of complex motion patterns by form/cue invariant MSTd neurons, *Journal of Neuroscience* **16**: 4716–4732.

Gennaro, R. J. (2004). *Higher Order Theories of Consciousness*, John Benjamins, Amsterdam.

Georges-François, P., Rolls, E. T. and Robertson, R. G. (1999). Spatial view cells in the primate hippocampus: allocentric view not head direction or eye position or place, *Cerebral Cortex* **9**: 197–212.

Georgopoulos, A. P. (1995). Motor cortex and cognitive processing, *in* M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, MIT Press, Cambridge, MA, chapter 32, pp. 507–517.

Gerfen, C. R. and Surmeier, D. J. (2011). Modulation of striatal projection systems by dopamine, *Annual Reviews of Neuroscience* **34**: 441–466.

Gerstner, W. (1995). Time structure of the activity in neural network models, *Physical Review E* **51**: 738–758.

Gerstner, W. (2000). Population dynamics of spiking neurons: fast transients, asynchronous states, and locking, *Neural Computation* **12**: 43–89.

Gerstner, W. and Kistler, W. (2002). *Spiking Neuron Models: Single Neurons, Populations and Plasticity*, Cambridge University Press, Cambridge.

Gerstner, W., Ritz, R. and Van Hemmen, L. (1993). A biologically motivated and analytically solvable model of collective oscillations in the cortex, *Biological Cybernetics* **68**: 363–374.

Gerstner, W., Kreiter, A. K., Markram, H. and Herz, A. V. (1997). Neural codes: firing rates and beyond, *Proceedings of the National Academy of Sciences USA* **94**: 12740–12741.

Gerstner, W., Kistler, W., Naud, R. and Paninski, L. (2014). *Neuronal Dynamics – From Single Neurons to Networks and Models of Cognition*, Cambridge University Press, Cambridge.

Gibson, J. J. (1950). *The Perception of the Visual World*, Houghton Mifflin, Boston.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston.

Gil, Z., Connors, B. W. and Amitai, Y. (1997). Differential regulation of neocortical synapses by neuromodulators and activity, *Neuron* **19**: 679–686.

Gilbert, P. E. and Kesner, R. P. (2002). The amygdala but not the hippocampus is involved in pattern separation based on reward value, *Neurobiology of Learning and Memory* **77**: 338–353.

Gilbert, P. E. and Kesner, R. P. (2003). Localization of function within the dorsal hippocampus: the role of the CA3 subregion in paired-associate learning, *Behavioral Neuroscience* **117**: 1385–1394.

Gilbert, P. E. and Kesner, R. P. (2006). The role of dorsal CA3 hippocampal subregion in spatial working memory and pattern separation, *Behavioural Brain Research* **169**: 142–149.

Gilbert, P. E., Kesner, R. P. and DeCoteau, W. E. (1998). Memory for spatial location: role of the hippocampus in mediating spatial pattern separation, *Journal of Neuroscience* **18**: 804–810.

Gilbert, P. E., Kesner, R. P. and Lee, I. (2001). Dissociating hippocampal subregions: double dissociation between dentate gyrus and CA1, *Hippocampus* **11**: 626–636.

Giocomo, L. M. and Hasselmo, M. E. (2007). Neuromodulation by glutamate and acetylcholine can change circuit dynamics by regulating the relative influence of afferent input and excitatory feedback, *Molecular Neurobiology* **36**: 184–200.

Giocomo, L. M., Moser, M. B. and Moser, E. I. (2011). Computational models of grid cells, *Neuron* **71**: 589–603.

Giugliano, M., La Camera, G., Rauch, A., Luescher, H.-R. and Fusi, S. (2002). Non-monotonic current-to-rate response function in a novel integrate-and-fire model neuron, *in* J. Dorronsoro (ed.), *Proceedings of ICANN 2002, LNCS 2415*, Springer, New York, pp. 141–146.

Giza, B. K. and Scott, T. R. (1983). Blood glucose selectively affects taste-evoked activity in rat nucleus tractus solitarius, *Physiology and Behaviour* **31**: 643–650.

Giza, B. K. and Scott, T. R. (1987a). Blood glucose level affects perceived sweetness intensity in rats, *Physiology and Behaviour* **41**: 459–464.

Giza, B. K. and Scott, T. R. (1987b). Intravenous insulin infusions in rats decrease gustatory-evoked responses to sugars, *American Journal of Physiology* **252**: R994–R1002.

Giza, B. K., Scott, T. R. and Vanderweele, D. A. (1992). Administration of satiety factors and gustatory responsiveness in the nucleus tractus solitarius of the rat, *Brain Research Bulletin* **28**: 637–639.

Giza, B. K., Deems, R. O., Vanderweele, D. A. and Scott, T. R. (1993). Pancreatic glucagon suppresses gustatory

responsiveness to glucose, *American Journal of Physiology* **265**: R1231–7.

Glantz, L. A. and Lewis, D. A. (2000). Decreased dendritic spine density on prefrontal cortical pyramidal neurons in schizophrenia, *Archives of General Psychiatry* **57**: 65–73.

Glausier, J. R. and Lewis, D. A. (2013). Dendritic spine pathology in schizophrenia, *Neuroscience* **251**: 90–107.

Gleen, J. F. and Erickson, R. P. (1976). Gastric modulation of gustatory afferent activity, *Physiology and Behaviour* **16**: 561–568.

Glimcher, P. (2003). The neurobiology of visual-saccadic decision making, *Annual Review of Neuroscience* **26**: 133–179.

Glimcher, P. (2004). *Decisions, Uncertainty, and the Brain*, MIT Press, Cambridge, MA.

Glimcher, P. (2005). Indeterminacy in brain and behavior, *Annual Review of Psychology* **56**: 25–56.

Glimcher, P. (2011a). *Foundations of Neuroeconomic Analysis*, Oxford University Press, Oxford.

Glimcher, P. W. (2011b). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis, *Proceedings of the National Academy of Sciences U S A* **108 Suppl 3**: 15647–15654.

Glimcher, P. W. and Fehr, E. (2013). *Neuroeconomics: Decision-Making and the Brain*, 2nd edn, Academic Press, New York.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI, *Neuroimage* **9**: 416–429.

Gnadt, J. W. and Andersen, R. A. (1988). Memory related motor planning activity in posterior parietal cortex of macaque, *Experimental Brain Research* **70**: 216–220.

Gochin, P. M., Colombo, M., Dorfman, G. A., Gerstein, G. L. and Gross, C. G. (1994). Neural ensemble encoding in inferior temporal cortex, *Journal of Neurophysiology* **71**: 2325–2337.

Goff, D. C. and Coyle, J. T. (2001). The emerging role of glutamate in the pathophysiology and treatment of schizophrenia, *American Journal of Psychiatry* **158**: 1367–1377.

Gold, A. E. and Kesner, R. P. (2005). The role of the CA3 subregion of the dorsal hippocampus in spatial pattern completion in the rat, *Hippocampus* **15**: 808–814.

Gold, J. I. and Shadlen, M. N. (2007). The neural basis of decision making, *Annual Review of Neuroscience* **30**: 535–574.

Gold, P. W. (2015). The organization of the stress system and its dysregulation in depressive illness, *Molecular Psychiatry* **20**: 32–47.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Company, Inc.

Goldberg, M. E. and Walker, M. F. (2013). The control of gaze, *in* E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum and A. J. Hudspeth (eds), *Principles of Neural Science*, 5th edn, McGraw-Hill, New York, chapter 40, pp. 894–916.

Goldberg, M. E., Bisley, J. W., Powell, K. D. and Gottlieb, J. (2006). Saccades, salience and attention: the role of the lateral intraparietal area in visual behavior, *Progress in Brain Research* **155**: 157–175.

Goldman-Rakic, P. (1994). Working memory dysfunction in schizophrenia, *Journal of Neuropsychology and Clinical Neuroscience* **6**: 348–357.

Goldman-Rakic, P. S. (1996). The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive, *Philosophical Transactions of the Royal Society B* **351**: 1445–1453.

Goldman-Rakic, P. S. (1999). The physiological approach: functional architecture of working memory and disordered cognition in schizophrenia, *Biological Psychiatry* **46**: 650–661.

Goldman-Rakic, P. S., Castner, S. A., Svensson, T. H., Siever, L. J. and Williams, G. V. (2004). Targeting the dopamine D1 receptor in schizophrenia: insights for cognitive dysfunction, *Psychopharmacology (Berlin)* **174**: 3–16.

Golomb, D., Kleinfeld, D., Reid, R. C., Shapley, R. M. and Shraiman, B. (1994). On temporal codes and the spatiotemporal response of neurons in the lateral geniculate nucleus, *Journal of Neurophysiology* **72**: 2990–3003.

Golomb, D., Hertz, J. A., Panzeri, S., Treves, A. and Richmond, B. J. (1997). How well can we estimate the information carried in neuronal responses from limited samples?, *Neural Computation* **9**: 649–665.

Gonzalez-Burgos, G. and Lewis, D. A. (2012). Nmda receptor hypofunction, parvalbumin-positive neurons, and cortical gamma oscillations in schizophrenia, *Schizophrenia Bulletin* **38**: 950–7.

Gonzalez-Burgos, G., Hashimoto, T. and Lewis, D. A. (2010). Alterations of cortical gaba neurons and network oscillations in schizophrenia, *Curr Psychiatry Rep* **12**: 335–44.

Goodale, M. A. (2004). Perceiving the world and grasping it: dissociations between conscious and unconscious visual processing, *in* M. S. Gazzaniga (ed.), *The Cognitive Neurosciences III*, MIT Press, Cambridge, MA, pp. 1159–1172.

Goodrich-Hunsaker, N. J., Hunsaker, M. R. and Kesner, R. P. (2005). Dissociating the role of the parietal cortex and dorsal hippocampus for spatial information processing, *Behavioral Neuroscience* **119**: 1307–1315.

Goodrich-Hunsaker, N. J., Hunsaker, M. R. and Kesner, R. P. (2008). The interactions and dissociations of the dorsal hippocampus subregions: how the dentate gyrus, ca3, and ca1 process spatial information, *Behavioral Neuroscience* **122**: 16–26.

Gothard, K. M., Battaglia, F. P., Erickson, C. A., Spitler, K. M. and Amaral, D. G. (2007). Neural responses to facial expression and face identity in the monkey amygdala, *Journal of Neurophysiology* **97**: 1671–1683.

Gotlib, I. H. and Hammen, C. L. (2009). *Handbook of Depression*, Guilford Press, New York.

Gottfried, J. A. (2010). Central mechanisms of odour object perception, *Nature Reviews Neuroscience* **11**: 628–641.

Gottfried, J. A. (2015). Structural and functional imaging of the human olfactory system, *in* R. L. Doty (ed.), *Handbook of Olfaction and Gustation*, 3rd edn, Wiley Liss, New York, chapter 13, pp. 279–304.

Gottfried, J. A., Winston, J. S. and Dolan, R. J. (2006). Dissociable codes of odor quality and odorant structure in human piriform cortex, *Neuron* **49**: 467–479.

Grabenhorst, F. and Rolls, E. T. (2008). Selective attention to affective value alters how the brain processes taste stimuli, *European Journal of Neuroscience* **27**: 723–729.

Grabenhorst, F. and Rolls, E. T. (2010). Attentional modulation of affective vs sensory processing: functional connectivity and a top down biased activation theory of selective attention, *Journal of Neurophysiology* **104**: 1649–1660.

Grabenhorst, F. and Rolls, E. T. (2011). Value, pleasure, and choice systems in the ventral prefrontal cortex, *Trends in Cognitive Sciences* **15**: 56–67.

Grabenhorst, F., Rolls, E. T., Margot, C., da Silva, M. and Velazco, M. I. (2007). How pleasant and unpleasant stimuli combine in the brain: odor combinations, *Journal of Neuroscience* **27**: 13532–13540.

Grabenhorst, F., Rolls, E. T. and Bilderbeck, A. (2008a). How cognition modulates affective responses to taste and flavor: top-down influences on the orbitofrontal and pregenual cingulate cortices, *Cerebral Cortex* **18**: 1549–1559.

Grabenhorst, F., Rolls, E. T. and Parris, B. A. (2008b). From affective value to decision-making in the prefrontal cortex, *European Journal of Neuroscience* **28**: 1930–1939.

Grabenhorst, F., D'Souza, A., Parris, B. A., Rolls, E. T. and Passingham, R. E. (2010a). A common neural scale for the subjective value of different primary rewards, *Neuroimage* **51**: 1265–1274.

Grabenhorst, F., Rolls, E. T., Parris, B. A. and D'Souza, A. (2010b). How the brain represents the reward value of fat in the mouth, *Cerebral Cortex* **20**: 1082–1091.

Grady, C. L. (2008). Cognitive neuroscience of aging, *Annals of the New York Academy of Sciences* **1124**: 127–144.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods, *Econometrica* **37**: 414–438.

Gray, C. M. and Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex, *Proceedings of National Academy of Sciences USA* **86**: 1698–1702.

Gray, C. M., Konig, P., Engel, A. K. and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties, *Nature* **338**: 334–337.

Gray, J. A., Young, A. M. J. and Joseph, M. H. (1997). Dopamine's role, *Science* **278**: 1548–1549.

Graybiel, A. M. and Kimura, M. (1995). Adaptive neural networks in the basal ganglia, *in* J. C. Houk, J. L. Davis and D. G. Beiser (eds), *Models of Information Processing in the Basal Ganglia*, MIT Press, Cambridge, MA, chapter 5, pp. 103–116.

Graziano, M. S. A., Andersen, R. A. and Snowden, R. J. (1994). Tuning of MST neurons to spiral motions, *Journal of Neuroscience* **14**: 54–67.

Green, D. and Swets, J. (1966). *Signal Detection Theory and Psychophysics*, Wiley, New York.

Green, M. F. (1996). What are the functional consequences of neurocognitive deficits in schizophrenia?, *American Journal of Psychiatry* **153**: 321–330.

Gregory, R. L. (1970). *The Intelligent Eye*, McGraw-Hill, New York.

Griffin, D. R. (1992). *Animal Minds*, University of Chicago Press, Chicago.

Grill-Spector, K. and Malach, R. (2004). The human visual cortex, *Annual Review of Neuroscience* **27**: 649–677.

Grill-Spector, K., Sayres, R. and Ress, D. (2006). High-resolution imaging reveals highly selective nonface clusters in the fusiform face area, *Nature Neuroscience* **9**: 1177–1185.

Grimson, W. E. L. (1990). *Object Recognition by Computer*, MIT Press, Cambridge, MA.

Griniasty, M., Tsodyks, M. V. and Amit, D. J. (1993). Conversion of temporal correlations between stimuli to spatial correlations between attractors, *Neural Computation* **35**: 1–17.

Gross, C. G. (2002). Genealogy of the "grandmother cell", *Neuroscientist* **8**: 512–518.

Gross, C. G., Rocha-Miranda, C. E. and Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque, *Journal of Neurophysiology* **35**: 96–111.

Gross, C. G., Desimone, R., Albright, T. D. and Schwartz, E. L. (1985). Inferior temporal cortex and pattern recognition, *Experimental Brain Research* **Suppl. 11**: 179–201.

Grossberg, S. (1976a). Adaptive pattern classification and universal recoding i: parallel development and coding of neural feature detectors, *Biological Cybernetics* **23**: 121–134.

Grossberg, S. (1976b). Adaptive pattern classification and universal recoding ii: feedback, expectation, olfaction, illusions, *Biological Cybernetics* **23**: 1187–202.

Grossberg, S. (1988). Non-linear neural networks: principles, mechanisms, and architectures, *Neural Networks* **1**: 17–61.

Groves, P. M. (1983). A theory of the functional organization of the neostriatum and the neostriatal control of voluntary movement, *Brain Research Reviews* **5**: 109–132.

Groves, P. M., Garcia-Munoz, M., Linder, J. C., Manley, M. S., Martone, M. E. and Young, S. J. (1995). Elements of

the intrinsic organization and information processing in the neostriatum, *in* J. C. Houk, J. L. Davis and D. G. Beiser (eds), *Models of Information Processing in the Basal Ganglia*, MIT Press, Cambridge, MA, chapter 4, pp. 51–96.

Grusser, O.-J. and Landis, T. (1991). *Visual Agnosias*, MacMillan, London.

Guest, S., Grabenhorst, F., Essick, G., Chen, Y., Young, M., McGlone, F., de Araujo, I. and Rolls, E. T. (2007). Human cortical representation of oral temperature, *Physiology and Behavior* **92**: 975–984.

Gulyas, A. I., Miles, R., Hajos, N. and Freund, T. F. (1993). Precision and variability in postsynaptic target selection of inhibitory cells in the hippocampal CA3 region, *European Journal of Neuroscience* **5**: 1729–1751.

Gurney, K., Prescott, T. J. and Redgrave, P. (2001a). A computational model of action selection in the basal ganglia I: A new functional anatomy, *Biological Cybernetics* **84**: 401–410.

Gurney, K., Prescott, T. J. and Redgrave, P. (2001b). A computational model of action selection in the basal ganglia II: Analysis and simulation of behaviour, *Biological Cybernetics* **84**: 411–423.

Gutig, R. and Sompolinsky, H. (2006). The tempotron: a neuron that learns spike timing-based decisions, *Nature Neuroscience* **9**: 420–428.

Haber, S. N. and Knutson, B. (2009). The reward circuit: linking primate anatomy and human imaging, *Neuropsychopharmacology* **35**: 4–26.

Haberly, L. B. (2001). Parallel-distributed processing in olfactory cortex: new insights from morphological and physiological analysis of neuronal circuitry, *Chemical Senses* **26**: 551–576.

Hafner, H., Maurer, K., Loffler, W., an der Heiden, W., Hambrecht, M. and Schultze-Lutter, F. (2003). Modeling the early course of schizophrenia, *Schizophrenia Bulletin* **29**: 325–340.

Hafting, T., Fyhn, M., Molden, S., Moser, M. B. and Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex, *Nature* **436**: 801–806.

Hajnal, A., Takenouchi, K. and Norgren, R. (1999). Effect of intraduodenal lipid on parabrachial gustatory coding in awake rats, *Journal of Neuroscience* **19**: 7182–7190.

Hamani, C., Mayberg, H., Snyder, B., Giacobbe, P., Kennedy, S. and Lozano, A. M. (2009). Deep brain stimulation of the subcallosal cingulate gyrus for depression: anatomical location of active contacts in clinical responders and a suggested guideline for targeting, *Journal of Neurosurgery* **111**: 1209–1215.

Hamani, C., Mayberg, H., Stone, S., Laxton, A., Haber, S. and Lozano, A. M. (2011). The subcallosal cingulate gyrus in the context of major depression, *Biological Psychiatry* **69**: 301–308.

Hamilton, J. P., Chen, M. C. and Gotlib, I. H. (2013). Neural systems approaches to understanding major depressive disorder: an intrinsic functional organization perspective, *Neurobiology of Disease* **52**: 4–11.

Hamilton, W. D. (1964). The genetical evolution of social behaviour, *Journal of Theoretical Biology* **7**: 1–52.

Hamilton, W. D. (1996). *Narrow Roads of Gene Land*, W. H. Freeman, New York.

Hamming, R. W. (1990). *Coding and Information Theory*, 2nd edn, Prentice-Hall, Englewood Cliffs, New Jersey.

Hampshire, A. and Owen, A. M. (2006). Fractionating attentional control using event-related fMRI, *Cerebral Cortex* **16**: 1679–1689.

Hampson, R. E., Hedberg, T. and Deadwyler, S. A. (2000). Differential information processing by hippocampal and subicular neurons, *Annals of the New York Academy of Sciences* **911**: 151–165.

Hampton, A. N. and O'Doherty, J. P. (2007). Decoding the neural substrates of reward-related decision making with functional MRI, *Proceedings of the National Academy of Sciences USA* **104**: 1377–1382.

Hampton, R. R. (2001). Rhesus monkeys know when they can remember, *Proceedings of the National Academy of Sciences of the USA* **98**: 5539–5362.

Hampton, R. R. (2005). Monkey perirhinal cortex is critical for visual memory, but not for visual perception: reexamination of the behavioural evidence from monkeys, *Quarterly Journal of Experimental Psychology* **58B**: 283–299.

Handelmann, G. E. and Olton, D. S. (1981). Spatial memory following damage to hippocampal ca3 pyramidal cells with kainic acid: impairment and recovery with preoperative training, *Brain Research* **217**: 41–58.

Hannesson, D. K., Howland, G. J. and Phillips, A. G. (2004). Interaction between perirhinal and medial prefrontal cortex is required for temporal order but not recognition memory for objects in rats, *Journal of Neuroscience* **24**: 4596–4604.

Harel, J., Koch, C. and Perona, P. (2006a). Graph-based visual saliency, *Advances in Neural Information Processing Systems (NIPS)* **19**: 545–552.

Harel, J., Koch, C. and Perona, P. (2006b). A saliency implementation in MATLAB, p. http://www.klab.caltech.edu/ harel/share/gbvs.php.

Hargreaves, E. L., Rao, G., Lee, I. and Knierim, J. J. (2005). Major dissociation between medial and lateral entorhinal input to dorsal hippocampus, *Science* **308**: 1792–1794.

Harmer, C. J. and Cowen, P. J. (2013). 'it's the way that you look at it'–a cognitive neuropsychological account of ssri action in depression, *Philosophical Transactions of the Royal Society of London B Biological Sciences* **368**: 20120407.

Harris, A. E., Ermentrout, G. B. and Small, S. L. (1997). A model of ocular dominance column development by competition for trophic factor, *Proceedings of the National Academy of Sciences USA* **94**: 9944–9949.

Harris, K. D. and Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding, *Nature* **503**: 51–58.

Harris, K. D. and Shepherd, G. M. (2015). The neocortical circuit: themes and variations, *Nature Neuroscience* **18**: 170–181.

Hartley, T., Lever, C., Burgess, N. and O'Keefe, J. (2014). Space in the brain: how the hippocampal formation supports spatial cognition, *Philosophical Transactions of the Royal Society London B Biological Sciences* **369**: 20120510.

Hartston, H. J. and Swerdlow, N. R. (1999). Visuospatial priming and Stroop performance in patients with obsessive compulsive disorder, *Neuropsychology* **13**: 447–457.

Hassabis, D., Chu, C., Rees, G., Weiskopf, N., Molyneux, P. D. and Maguire, E. A. (2009). Decoding neuronal ensembles in the human hippocampus, *Current Biology* **19**: 546–554.

Hasselmo, M. E. and Bower, J. M. (1993). Acetylcholine and memory, *Trends in Neurosciences* **16**: 218–222.

Hasselmo, M. E. and Eichenbaum, H. B. (2005). Hippocampal mechanisms for the context-dependent retrieval of episodes, *Neural Networks* **18**: 1172–1190.

Hasselmo, M. E. and Sarter, M. (2011). Modes and models of forebrain cholinergic neuromodulation of cognition, *Neuropsychopharmacology* **36**: 52–73.

Hasselmo, M. E., Rolls, E. T. and Baylis, G. C. (1989a). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey, *Behavioural Brain Research* **32**: 203–218.

Hasselmo, M. E., Rolls, E. T., Baylis, G. C. and Nalwa, V. (1989b). Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey, *Experimental Brain Research* **75**: 417–429.

Hasselmo, M. E., Schnell, E. and Barkai, E. (1995). Learning and recall at excitatory recurrent synapses and cholinergic modulation in hippocampal region CA3, *Journal of Neuroscience* **15**: 5249–5262.

Hawken, M. J. and Parker, A. J. (1987). Spatial properties of the monkey striate cortex, *Proceedings of the Royal Society, London B* **231**: 251–288.

Haxby, J. V. (2006). Fine structure in representations of faces and objects, *Nature Neuroscience* **9**: 1084–1086.

Haynes, J. D. and Rees, G. (2005a). Predicting the orientation of invisible stimuli from activity in human primary visual cortex, *Nature Neuroscience* **8**: 686–691.

Haynes, J. D. and Rees, G. (2005b). Predicting the stream of consciousness from activity in human visual cortex, *Current Biology* **15**: 1301–1307.

Haynes, J. D. and Rees, G. (2006). Decoding mental states from brain activity in humans, *Nature Reviews Neuroscience* **7**: 523–534.

Haynes, J. D., Sakai, K., Rees, G., Gilbert, S., Frith, C. and Passingham, R. E. (2007). Reading hidden intentions in the human brain, *Current Biology* **17**: 323–328.

He, C., Chen, F., Li, B. and Hu, Z. (2014). Neurophysiology of HCN channels: from cellular functions to multiple regulations, *Progress in Neurobiology* **112**: 1–23.

Hebb, D. O. (1949). *The Organization of Behavior: a Neuropsychological Theory*, Wiley, New York.

Hedges, J. H., Gartshteyn, Y., Kohn, A., Rust, N. C., Shadlen, M. N., Newsome, W. T. and Movshon, J. A. (2011). Dissociation of neuronal and psychophysical responses to local and global motion, *Current Biology* **21**: 2023–2028.

Hegde, J. and Van Essen, D. C. (2000). Selectivity for complex shapes in primate visual area V2, *Journal of Neuroscience* **20**: RC61.

Heinke, D., Deco, G., Zihl, J. and Humphreys, G. (2002). A computational neuroscience account of visual neglect, *Neurocomputing* **44–46**: 811–816.

Heinrichs, D. W. (2005). Antidepressants and the chaotic brain: implications for the respectful treatment of selves, *Philosophy, Psychiatry, and Psychology* **12**: 215–227.

Heller, J., Hertz, J. A., Kjaer, T. W. and Richmond, B. J. (1995). Information flow and temporal coding in primate pattern vision, *Journal of Computational Neuroscience* **2**: 175–193.

Helmholtz, H. v. (1867). *Handbuch der physiologischen Optik*, Voss, Leipzig.

Hempel, C. M., Hartman, K. H., Wang, X. J., Turrigiano, G. G. and Nelson, S. B. (2000). Multiple forms of short-term plasticity at excitatory synapses in rat medial prefrontal cortex, *Journal of Neurophysiology* **83**: 3031–3041.

Henze, D. A., Wittner, L. and Buzsaki, G. (2002). Single granule cells reliably discharge targets in the hippocampal ca3 network in vivo, *Nature Neuroscience* **5**: 790–5.

Herrnstein, R. J. (1984). Objects, categories, and discriminative stimuli, *in* H. L. Roitblat, T. G. Bever and H. S. Terrace (eds), *Animal Cognition*, Lawrence Erlbaum and Associates, Hillsdale, NJ, chapter 14, pp. 233–261.

Hertz, J. A., Krogh, A. and Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*, Addison-Wesley, Wokingham, UK.

Hertz, J. A., Kjaer, T. W., Eskander, E. N. and Richmond, B. J. (1992). Measuring natural neural processing with artificial neural networks, *International Journal of Neural Systems* **3 (Suppl.)**: 91–103.

Hestrin, S., Sah, P. and Nicoll, R. (1990). Mechanisms generating the time course of dual component excitatory synaptic currents recorded in hippocampal slices, *Neuron* **5**: 247–253.

Hevner, R. F. (2016). Evolution of the mammalian dentate gyrus, *Journal of Comparative Neurology* **524**: 578–94.

Heyes, C. (2008). Beast machines? Questions of animal consciousness, *in* L. Weiskrantz and M. Davies (eds), *Frontiers of Consciousness*, Oxford University Press, Oxford, chapter 9, pp. 259–274.

Hikosaka, K. and Watanabe, M. (2000). Delay activity of orbital and lateral prefrontal neurons of the monkey varying

with different rewards, *Cerebral Cortex* **10**: 263–271.

Hill, S. L., Wang, Y., Riachi, I., Schürmann, F. and Markram, H. (2012). Statistical connectivity provides a sufficient foundation for specific functional connectivity in neocortical neural microcircuits, *Proceedings of the National Academy of Sciences* **109**: E2885–E2894.

Hinton, G. E. (1989). Deterministic Boltzmann learning performs steepest descent in weight-space, *Neural Computation* **1**: 143–150.

Hinton, G. E. and Anderson, J. A. (1981). *Parallel Models of Associative Memory*, Erlbaum, Hillsdale, NJ.

Hinton, G. E. and Ghahramani, Z. (1997). Generative models for discovering sparse distributed representations, *Philosophical Transactions of the Royal Society of London, B* **352**: 1177–1190.

Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks, *Science* **313**: 504–507.

Hinton, G. E. and Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines, *in* D. Rumelhart and J. L. McClelland (eds), *Parallel Distributed Processing*, Vol. 1, MIT Press, Cambridge, MA, chapter 7, pp. 282–317.

Hinton, G. E., Dayan, P., Frey, B. J. and Neal, R. M. (1995). The "wake-sleep" algorithm for unsupervised neural networks, *Science* **268**: 1158–1161.

Hinton, G. E., Osindero, S. and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets, *Neural computation* **18**: 1527–1554.

Hirabayashi, T., Takeuchi, D., Tamura, K. and Miyashita, Y. (2013). Functional microcircuit recruited during retrieval of object association memory in monkey perirhinal cortex, *Neuron* **77**: 192–203.

Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve, *Journal of Physiology* **117**: 500–544.

Hoebel, B. G. (1997). Neuroscience, and appetitive behavior research: 25 years, *Appetite* **29**: 119–133.

Hoebel, B. G., Rada, P., Mark, G. P., Parada, M., Puig de Parada, M., Pothos, E. and Hernandez, L. (1996). Hypothalamic control of accumbens dopamine: a system for feeding reinforcement, *in* G. Bray and D. Ryan (eds), *Molecular and Genetic Aspects of Obesity*, Vol. 5, Louisiana State University Press, Baton Rouge, LA, pp. 263–280.

Hoffman, E. A. and Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed neural system for face perception, *Nature Neuroscience* **3**: 80–84.

Hoge, J. and Kesner, R. P. (2007). Role of CA3 and CA1 subregions of the dorsal hippocampus on temporal processing of objects, *Neurobiology of Learning and Memory* **88**: 225–231.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Michigan, USA.

Holland, P. C. and Gallagher, M. (2004). Amygdala-frontal interactions and reward expectancy, *Current Opinion in Neurobiology* **14**: 148–155.

Hölscher, C. and Rolls, E. T. (2002). Perirhinal cortex neuronal activity is actively related to working memory in the macaque, *Neural Plasticity* **9**: 41–51.

Hölscher, C., Rolls, E. T. and Xiang, J. Z. (2003). Perirhinal cortex neuronal activity related to long term familiarity memory in the macaque, *European Journal of Neuroscience* **18**: 2037–2046.

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.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons, *Proceedings of the National Academy of Sciences USA* **81**: 3088–3092.

Hopfield, J. J. and Herz, A. V. (1995). Rapid local synchronization of action potentials: toward computation with coupled integrate-and-fire neurons, *Proceedings of the National Academy of Sciences USA* **92**: 6655–6662.

Hornak, J., Rolls, E. T. and Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage, *Neuropsychologia* **34**: 247–261.

Hornak, J., Bramham, J., Rolls, E. T., Morris, R. G., O'Doherty, J., Bullock, P. R. and Polkey, C. E. (2003). Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices, *Brain* **126**: 1691–1712.

Hornak, J., O'Doherty, J., Bramham, J., Rolls, E. T., Morris, R. G., Bullock, P. R. and Polkey, C. E. (2004). Reward-related reversal learning after surgical excisions in orbitofrontal and dorsolateral prefrontal cortex in humans, *Journal of Cognitive Neuroscience* **16**: 463–478.

Horne, J. (2006). *Sleepfaring: A Journey Through the Science of Sleep*, Oxford University Press, Oxford.

Horvitz, J. C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events, *Neuroscience* **96**: 651–656.

Horwitz, B. and Tagamets, M.-A. (1999). Predicting human functional maps with neural net modeling, *Human Brain Mapping* **8**: 137–142.

Horwitz, B., Tagamets, M.-A. and McIntosh, A. R. (1999). Neural modeling, functional brain imaging, and cognition, *Trends in Cognitive Sciences* **3**: 85–122.

Houk, J. C., Adams, J. L. and Barto, A. C. (1995). A model of how the basal ganglia generates and uses neural signals that predict reinforcement, *in* J. C. Houk, J. L. Davies and D. G. Beiser (eds), *Models of Information Processing in the Basal Ganglia*, MIT Press, Cambridge, MA, chapter 13, pp. 249–270.

Howard, M. W. and Eichenbaum, H. (2015). Time and space in the hippocampus, *Brain Research* **1621**: 345–354.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex, *Journal of Physiology* **160**: 106–154.

Hubel, D. H. and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex, *Journal of Physiology, London* **195**: 215–243.

Hubel, D. H. and Wiesel, T. N. (1977). Functional architecture of the macaque monkey visual cortex, *Proceedings of the Royal Society, London [B]* **198**: 1–59.

Huerta, P. T., Sun, L. D., Wilson, M. A. and Tonegawa, S. (2000). Formation of temporal memory requires NMDA receptors within CA1 pyramidal neurons, *Neuron* **25**: 473–480.

Hughlings Jackson, J. (1878). *Selected writings of John Hughlings Jackson 2, Edited by J. Taylor, 1932*, Hodder and Staughton, London.

Human Genome Sequencing, C. I. (2004). Finishing the euchromatic sequence of the human genome, *Nature* **431**: 931–945.

Hummel, J. E. and Biederman, I. (1992). Dynamic binding in a neural network for shape recognition, *Psychological Review* **99**: 480–517.

Humphrey, N. K. (1980). Nature's psychologists, *in* B. D. Josephson and V. S. Ramachandran (eds), *Consciousness and the Physical World*, Pergamon, Oxford, pp. 57–80.

Humphrey, N. K. (1986). *The Inner Eye*, Faber, London.

Hunsaker, M. R., Thorup, J. A., Welch, T. and Kesner, R. P. (2006). The role of ca3 and ca1 in the acquisition of an object-trace-place paired-associate task, *Behavioral Neuroscience* **120**: 1252–6.

Hunsaker, M. R., Rogers, J. L. and Kesner, R. P. (2007). Behavioral characterization of a transection of dorsal ca3 subcortical efferents: comparison with scopolamine and physostigmine infusions into dorsal ca3, *Neurobiology of Learning and Memory* **88**: 127–36.

Hunsaker, M. R., Lee, B. and Kesner, R. P. (2008). Evaluating the temporal context of episodic memory: the role of ca3 and ca1, *Behavioural Brain Research* **188**: 310–5.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E. and Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex, *Nature* **532**: 453–458.

Huttenlocher, D. P. and Ullman, S. (1990). Recognizing solid objects by alignment with an image, *International Journal of Computer Vision* **5**: 195–212.

Iadarola, N. D., Niciu, M. J., Richards, E. M., Vande Voort, J. L., Ballard, E. D., Lundin, N. B., Nugent, A. C., Machado-Vieira, R. and Zarate, C. A., J. (2015). Ketamine and other n-methyl-d-aspartate receptor antagonists in the treatment of depression: a perspective review, *Therapeutic Advances in Chronic Disease* **6**: 97–114.

Ingvar, D. H. and Franzen, G. (1974). Abnormalities of cerebral blood flow distribution in patients with chronic schizophrenia, *Acta Psychiatrica Scandinavica* **50**: 425–462.

Insabato, A., Pannunzi, M., Rolls, E. T. and Deco, G. (2010). Confidence-related decision-making, *Journal of Neurophysiology* **104**: 539–547.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L. and Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway, *Proceedings of the National Academy of Sciences USA* **96**: 9379–9384.

Ishizuka, N., Weber, J. and Amaral, D. G. (1990). Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat, *Journal of Comparative Neurology* **295**: 580–623.

Ison, M. J., Quian Quiroga, R. and Fried, I. (2015). Rapid encoding of new memories by individual neurons in the human brain, *Neuron* **87**: 220–230.

Issa, E. B. and DiCarlo, J. J. (2012). Precedence of the eye region in neural processing of faces, *The Journal of Neuroscience* **32**: 16666–16682.

Ito, M. (1984). *The Cerebellum and Neural Control*, Raven Press, New York.

Ito, M. (1989). Long-term depression, *Annual Review of Neuroscience* **12**: 85–102.

Ito, M. (1993a). Cerebellar mechanisms of long-term depression, *in* M. Baudry, R. F. Thompson and J. L. Davis (eds), *Synaptic Plasticity: Molecular, Cellular and Functional Aspects*, MIT Press, Cambridge, MA, chapter 6, pp. 117–128.

Ito, M. (1993b). Synaptic plasticity in the cerebellar cortex and its role in motor learning, *Canadian Journal of Neurological Science* **Suppl. 3**: S70–S74.

Ito, M. (2006). Cerebellar circuitry as a neuronal machine, *Progress in Neurobiology* **78**: 272–303.

Ito, M. (2010). Cerebellar cortex, *in* G. M. Shepherd and S. Grillner (eds), *Handbook of Brain Microcircuits*, Oxford University Press, Oxford, chapter 28, pp. 293–300.

Ito, M. (2013). Error detection and representation in the olivo-cerebellar system, *Frontiers in Neural Circuits* **7**: 1.

Ito, M. and Komatsu, H. (2004). Representation of angles embedded within contour stimuli in area V2 of macaque monkeys, *Journal of Neuroscience* **24**: 3313–3324.

Ito, M., Yamaguchi, K., Nagao, S. and Yamazaki, T. (2014). Long-term depression as a model of cerebellar plasticity, *Progress in Brain Research* **210**: 1–30.

Itskov, P. M., Vinnik, E. and Diamond, M. E. (2011). Hippocampal representation of touch-guided behavior in rats: persistent and independent traces of stimulus and reward location, *PLoS One* **6**: e16462.

Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention, *Vision Research* **40**: 1489–1506.

Itti, L. and Koch, C. (2001). Computational modelling of visual attention, *Nature Reviews Neuroscience* **2**: 194–203.

Iversen, S. D. and Mishkin, M. (1970). Perseverative interference in monkey following selective lesions of the inferior prefrontal convexity, *Experimental Brain Research* **11**: 376–386.

Jackendoff, R. (2002). *Foundations of Language*, Oxford University Press, Oxford.

Jackson, B. S. (2004). Including long-range dependence in integrate-and-fire models of the high interspike-interval variability of cortical neurons, *Neural Computation* **16**: 2125–2195.

Jackson, M. B. (2013). Recall of spatial patterns stored in a hippocampal slice by long-term potentiation, *Journal of Neurophysiology* **110**: 2511–2519.

Jacobs, J., Weidemann, C. T., Miller, J. F., Solway, A., Burke, J. F., Wei, X. X., Suthana, N., Sperling, M. R., Sharan, A. D., Fried, I. and Kahana, M. J. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation, *Nature Neuroscience* **16**: 1188–1190.

Jacoby, L. L. (1983a). Perceptual enhancement: persistent effects of an experience, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **9**: 21–38.

Jacoby, L. L. (1983b). Remembering the data: analyzing interaction processes in reading, *Journal of Verbal Learning and Verbal Behavior* **22**: 485–508.

Jahr, C. and Stevens, C. (1990). Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics, *Journal of Neuroscience* **10**: 3178–3182.

Jakab, R. L. and Leranth, C. (1995). Septum, *in* G. Paxinos (ed.), *The Rat Nervous System*, Academic Press, San Diego.

James, W. (1884). What is an emotion?, *Mind* **9**: 188–205.

Janssen, P., Vogels, R. and Orban, G. A. (1999). Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes, *Proceedings of the National Academy of Sciences* **96**: 8217–8222.

Janssen, P., Vogels, R. and Orban, G. A. (2000). Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex, *Science* **288**: 2054–2056.

Janssen, P., Vogels, R., Liu, Y. and Orban, G. A. (2003). At least at the level of inferior temporal cortex, the stereo correspondence problem is solved, *Neuron* **37**: 693–701.

Jarrard, L. E. (1993). On the role of the hippocampus in learning and memory in the rat, *Behavioral and Neural Biology* **60**: 9–26.

Jarrard, L. E. and Davidson, T. L. (1990). Acquisition of concurrent conditional discriminations in rats with ibotenate lesions of hippocampus and subiculum, *Psychobiology* **18**: 68–73.

Jarrett, K., Kavukcuoglu, K., Ranzato, M. and LeCun, Y. (2009). What is the best multi-stage architecture for object recognition?, *2009 IEEE 12th International Conference on Computer Vision (ICCV)* pp. 2146–2153.

Jay, T. M. and Witter, M. P. (1991). Distribution of hippocampal CA1 and subicular efferents in the prefrontal cortex of the rat studied by means of anterograde transport of Phaseolus vulgaris–leucoagglutinin, *Journal of Comparative Neurology* **313**: 574–586.

Jeffery, K. J. and Hayman, R. (2004). Plasticity of the hippocampal place cell representation, *Reviews in the Neurosciences* **15**: 309–331.

Jeffery, K. J., Anderson, M. I., Hayman, R. and Chakraborty, S. (2004). A proposed architecture for the neural representation of spatial context, *Neuroscience and Biobehavioral Reviews* **28**: 201–218.

Jensen, O. and Lisman, J. E. (1996). Theta/gamma networks with slow NMDA channels learn sequences and encode episodic memory: role of NMDA channels in recall, *Learning and Memory* **3**: 264–278.

Jensen, O. and Lisman, J. E. (2005). Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer, *Trends in the Neurosciences* **28**: 67–72.

Jensen, O., Idiart, M. A. and Lisman, J. E. (1996). Physiologically realistic formation of autoassociative memory in networks with theta/gamma oscillations: role of fast NMDA channels, *Learning and Memory* **3**: 243–256.

Jerman, T., Kesner, R. P. and Hunsaker, M. R. (2006). Disconnection analysis of CA3 and DG in mediating encoding but not retrieval in a spatial maze learning task, *Learning and Memory* **13**: 458–464.

Jezek, K., Henriksen, E. J., Treves, A., Moser, E. I. and Moser, M.-B. (2011). Theta-paced flickering between place-cell maps in the hippocampus, *Nature* **278**: 246–249.

Jiang, X., Shen, S., Cadwell, C. R., Berens, P., Sinz, F., Ecker, A. S., Patel, S. and Tolias, A. S. (2015). Principles of connectivity among morphologically defined cell types in adult neocortex, *Science* **350**: aac9462.

Johansen-Berg, H. and Lloyd, D. M. (2000). The physiology and psychology of selective attention to touch, *Frontiers in Bioscience* **5**: D894–904.

Johansen-Berg, H., Gutman, D. A., Behrens, T. E., Matthews, P. M., Rushworth, M. F., Katz, E., Lozano, A. M. and Mayberg, H. S. (2008). Anatomical connectivity of the subgenual cingulate region targeted with deep brain stimulation for treatment-resistant depression, *Cerebral Cortex* **18**: 1374–1383.

Johnson-Laird, P. N. (1988). *The Computer and the Mind: An Introduction to Cognitive Science*, Harvard University Press, Cambridge, MA.

Johnston, D. and Amaral, D. (2004). Hippocampus, *in* G. M. Shepherd (ed.), *The Synaptic Organization of the Brain*, 5th edn, Oxford University Press, Oxford, chapter 11, pp. 455–498.

Johnston, S. T., Shtrahman, M., Parylak, S., Goncalves, J. T. and Gage, F. H. (2015). Paradox of pattern separation and adult neurogenesis: A dual role for new neurons balancing memory resolution and robustness, *Neurobiology*

of Learning and Memory.

Jonas, E. A. and Kaczmarek, L. K. (1999). The inside story: subcellular mechanisms of neuromodulation, *in* P. S. Katz (ed.), *Beyond Neurotransmission*, Oxford University Press, New York, chapter 3, pp. 83–120.

Jones, B. and Mishkin, M. (1972). Limbic lesions and the problem of stimulus–reinforcement associations, *Experimental Neurology* **36**: 362–377.

Jones, E. G. and Peters, A. (eds) (1984). *Cerebral Cortex, Functional Properties of Cortical Cells*, Vol. 2, Plenum, New York.

Jones, E. G. and Powell, T. P. S. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey, *Brain* **93**: 793–820.

Jung, M. W. and McNaughton, B. L. (1993). Spatial selectivity of unit activity in the hippocampal granular layer, *Hippocampus* **3**: 165–182.

Kaas, J. H. and Preuss, T. M. (2014). Human brain evolution, *in* L. Squire, D. Berg, S. du Lac, A. Ghosh and N. C. Spitzer (eds), *Fundamental Neuroscience*, 4th edn, Academic Press, London, chapter 42.

Kacelnik, A. and Brito e Abreu, F. (1998). Risky choice and Weber's Law, *Journal of Theoretical Biology* **194**: 289–298.

Kadohisa, M. and Wilson, D. A. (2006). Separate encoding of identity and similarity of complex familiar odors in piriform cortex, *Proceedings of the National Academy of Sciences U S A* **103**: 15206–15211.

Kadohisa, M., Rolls, E. T. and Verhagen, J. V. (2004). Orbitofrontal cortex neuronal representation of temperature and capsaicin in the mouth, *Neuroscience* **127**: 207–221.

Kadohisa, M., Rolls, E. T. and Verhagen, J. V. (2005a). Neuronal representations of stimuli in the mouth: the primate insular taste cortex, orbitofrontal cortex, and amygdala, *Chemical Senses* **30**: 401–419.

Kadohisa, M., Rolls, E. T. and Verhagen, J. V. (2005b). The primate amygdala: neuronal representations of the viscosity, fat texture, grittiness and taste of foods, *Neuroscience* **132**: 33–48.

Kagel, J. H., Battalio, R. C. and Green, L. (1995). *Economic Choice Theory: An Experimental Analysis of Animal Behaviour*, Cambridge University Press, Cambridge.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning, *Punishment and Aversive Behavior* pp. 279–296.

Kandel, E. R., Schwartz, J. H. and Jessell, T. H. (eds) (2000). *Principles of Neural Science*, 4th edn, Elsevier, Amsterdam.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A. and Hudspeth, A. J. (eds) (2013). *Principles of Neural Science*, 5th edn, McGraw-Hill, New York.

Kanter, I. and Sompolinsky, H. (1987). Associative recall of memories without errors, *Physical Review A* **35**: 380–392.

Kanwisher, N., McDermott, J. and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception, *Journal of Neuroscience* **17**: 4301–4311.

Karno, M., Golding, J. M., Sorenson, S. B. and Burnam, M. A. (1988). The epidemiology of obsessive-compulsive disorder in five US communities, *Archives of General Psychiatry* **45**: 1094–1099.

Kastner, S., De Weerd, P., Desimone, R. and Ungerleider, L. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI, *Science* **282**: 108–111.

Kastner, S., Pinsk, M., De Weerd, P., Desimone, R. and Ungerleider, L. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation, *Neuron* **22**: 751–761.

Kelly, K. M., Nadon, N. L., Morrison, J. H., Thibault, O., Barnes, C. A. and Blalock, E. M. (2006). The neurobiology of aging, *Epilepsy Research* **68, Supplement 1**: S5–S20.

Kelly, R. M. and Strick, P. L. (2004). Macro-architecture of basal ganglia loops with the cerebral cortex: use of rabies virus to reveal multisynaptic circuits, *Progress in Brain Research* **143**: 449–459.

Kemp, J. M. and Powell, T. P. S. (1970). The cortico-striate projections in the monkey, *Brain* **93**: 525–546.

Kennard, C. and Swash, M. (eds) (1989). *Hierarchies in Neurology*, Springer, London.

Kepecs, A., Uchida, N., Zariwala, H. A. and Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence, *Nature* **455**: 227–231.

Kesner, R. P. (1998). Neural mediation of memory for time: role of hippocampus and medial prefrontal cortex, *Psychological Bulletin Reviews* **5**: 585–596.

Kesner, R. P. and Gilbert, P. E. (2006). The role of the medial caudate nucleus, but not the hippocampus, in a delayed-matching-to sample task for motor response, *European Journal of Neuroscience* **23**: 1888–1894.

Kesner, R. P. and Rolls, E. T. (2001). Role of long term synaptic modification in short term memory, *Hippocampus* **11**: 240–250.

Kesner, R. P. and Rolls, E. T. (2015). A computational theory of hippocampal function, and tests of the theory: New developments, *Neuroscience and Biobehavioral Reviews* **48**: 92–147.

Kesner, R. P. and Warthen, D. K. (2010). Implications of ca3 nmda and opiate receptors for spatial pattern completion in rats, *Hippocampus* **20**: 550–7.

Kesner, R. P., Gilbert, P. E. and Barua, L. A. (2002). The role of the hippocampus in memory for the temporal order of a sequence of odors, *Behavioral Neuroscience* **116**: 286–290.

Kesner, R. P., Lee, I. and Gilbert, P. (2004). A behavioral assessment of hippocampal function based on a subregional analysis, *Reviews in the Neurosciences* **15**: 333–351.

Kesner, R. P., Hunsaker, M. R. and Gilbert, P. E. (2005). The role of CA1 in the acquisition of an object-trace-odor paired associate task, *Behavioral Neuroscience* **119**: 781–786.

Kesner, R. P., Hunsaker, M. R. and Warthen, M. W. (2008). The ca3 subregion of the hippocampus is critical for episodic memory processing by means of relational encoding in rats, *Behavioral Neuroscience* **122**: 1217–1225.

Kesner, R. P., Hunsaker, M. R. and Ziegler, W. (2011). The role of the dorsal and ventral hippocampus in olfactory working memory, *Neurobiology of Learning and Memory* **96**: 361–6.

Kesner, R. P., Hui, X., Sommer, T., Wright, C., Barrera, V. R. and Fanselow, M. S. (2014). The role of postnatal neurogenesis in supporting remote memory and spatial metric processing, *Hippocampus* **24**: 1663–71.

Keysers, C. and Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition, *Trends in Cognitive Sciences* **6**: 120–125.

Keysers, C., Xiao, D., Foldiak, P. and Perrett, D. (2001). The speed of sight, *Journal of Cognitive Neuroscience* **13**: 90–101.

Khaligh-Razavi, S.-M. and Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain it cortical representation, *PLoS Computational Biology* **10**(11): e1003915.

Khalil, H. (1996). *Nonlinear Systems*, Prentice Hall, Upper Saddle River, NJ.

Kievit, J. and Kuypers, H. G. J. M. (1975). Subcortical afferents to the frontal lobe in the rhesus monkey studied by means of retrograde horseradish peroxidase transport, *Brain Research* **85**: 261–266.

Killian, N. J., Jutras, M. J. and Buffalo, E. A. (2012). A map of visual space in the primate entorhinal cortex, *Nature* **491**: 761–764.

Kim, J. G. and Biederman, I. (2012). Greater sensitivity to nonaccidental than metric changes in the relations between simple shapes in the lateral occipital cortex, *Neuroimage* **63**: 1818–26.

Kircher, T. T. and Thienel, R. (2005). Functional brain imaging of symptoms and cognition in schizophrenia, *Progress in Brain Research* **150**: 299–308.

Kirkwood, A., Dudek, S. M., Gold, J. T., Aizenman, C. D. and Bear, M. F. (1993). Common forms of synaptic plasticity in the hippocampus and neocortex "in vitro", *Science* **260**: 1518–1521.

Kjaer, T. W., Hertz, J. A. and Richmond, B. J. (1994). Decoding cortical neuronal signals: network models, information estimation and spatial tuning, *Journal of Computational Neuroscience* **1**: 109–139.

Kleinfeld, D. (1986). Sequential state generation by model neural networks, *Proceedings of the National Academy of Sciences of the USA* **83**: 9469–9473.

Knierim, J. J. and Neunuebel, J. P. (2016). Tracking the flow of hippocampal computation: Pattern separation, pattern completion, and attractor dynamics, *Neurobiology of Learning and Memory* **129**: 38–49.

Knierim, J. J., Neunuebel, J. P. and Deshmukh, S. S. (2014). Functional correlates of the lateral and medial entorhinal cortex: objects, path integration and local-global reference frames, *Philosophical Transactions of the Royal Society London B Biological Sciences* **369**: 20130369.

Knudsen, E. I. (2011). Control from below: the role of a midbrain network in spatial attention, *European Journal of Neuroscience* **33**: 1961–1972.

Kobayashi, S. (2012). Organization of neural systems for aversive information processing: pain, error, and punishment, *Frontiers in Neuroscience* **6**: 136.

Koch, C. (1999). *Biophysics of Computation*, Oxford University Press, Oxford.

Koch, C. (2004). *The Quest for Consciousness*, Roberts, Englewood, CO.

Koch, K. W. and Fuster, J. M. (1989). Unit activity in monkey parietal cortex related to haptic perception and temporary memory, *Experimental Brain Research* **76**: 292–306.

Koenderink, J. J. (1990). *Solid Shape*, MIT Press, Cambridge, MA.

Koenderink, J. J. and Van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision, *Biological Cybernetics* **32**: 211–217.

Koenderink, J. J. and van Doorn, A. J. (1991). Affine structure from motion, *Journal of the Optical Society of America, A* **8**: 377–385.

Kohonen, T. (1977). *Associative Memory: A System Theoretical Approach*, Springer, New York.

Kohonen, T. (1982). Clustering, taxonomy, and topological maps of patterns, *in* M. Lang (ed.), *Proceedings of the Sixth International Conference on Pattern Recognition*, IEEE Computer Society Press, Silver Spring, MD, pp. 114–125.

Kohonen, T. (1988). *Self-Organization and Associative Memory*, 2nd edn, Springer-Verlag, New York.

Kohonen, T. (1989). *Self-Organization and Associative Memory*, 3rd (1984, 1st edn; 1988, 2nd edn) edn, Springer-Verlag, Berlin.

Kohonen, T. (1995). *Self-Organizing Maps*, Springer-Verlag, Berlin.

Kohonen, T., Oja, E. and Lehtio, P. (1981). Storage and processing of information in distributed memory systems, *in* G. E. Hinton and J. A. Anderson (eds), *Parallel Models of Associative Memory*, Erlbaum, Hillsdale, NJ, chapter 4, pp. 105–143.

Kolb, B. and Whishaw, I. Q. (2015). *Fundamentals of Human Neuropsychology*, 7th edn, Macmillan, New York.

Kondo, H., Lavenex, P. and Amaral, D. G. (2009). Intrinsic connections of the macaque monkey hippocampal formation: II. CA3 connections, *Journal of Comparative Neurology* **515**: 349–377.

Konopaske, G. T., Lange, N., Coyle, J. T. and Benes, F. M. (2014). Prefrontal cortical dendritic spine pathology in

schizophrenia and bipolar disorder, *JAMA Psychiatry* **71**: 1323–31.

Kosar, E., Grill, H. J. and Norgren, R. (1986). Gustatory cortex in the rat. II. Thalamocortical projections, *Brain Research* **379**: 342–352.

Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate*, MIT Press, Cambridge, MA.

Kourtzi, Z. and Connor, C. E. (2011). Neural representations for object perception: structure, category, and adaptive coding, *Annual Review of Neuroscience* **34**: 45–67.

Kovacs, G., Vogels, R. and Orban, G. A. (1995). Cortical correlate of pattern backward masking, *Proceedings of the National Academy of Sciences USA* **92**: 5587–5591.

Kraus, B. J., Brandon, M. P., Robinson, R. J., n., Connerney, M. A., Hasselmo, M. E. and Eichenbaum, H. (2015). During running in place, grid cells integrate elapsed time and distance run, *Neuron* **88**: 578–589.

Krauzlis, R. J., Lovejoy, L. P. and Zenon, A. (2013). Superior colliculus and visual spatial attention, *Annual Review of Neuroscience* **36**: 165–182.

Krebs, J. R. and Davies, N. B. (1991). *Behavioural Ecology*, 3rd edn, Blackwell, Oxford.

Krebs, J. R. and Kacelnik, A. (1991). Decision making, *in* J. R. Krebs and N. B. Davies (eds), *Behavioural Ecology*, 3rd edn, Blackwell, Oxford, chapter 4, pp. 105–136.

Kreiman, G., Koch, C. and Freid, I. (2000). Category-specific visual responses of single neurons in the human temporal lobe, *Nature Neuroscience* **3**: 946–953.

Kringelbach, M. L. and Rolls, E. T. (2003). Neural correlates of rapid reversal learning in a simple model of human social interaction, *Neuroimage* **20**: 1371–1383.

Kringelbach, M. L. and Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology, *Progress in Neurobiology* **72**: 341–372.

Kringelbach, M. L., O'Doherty, J., Rolls, E. T. and Andrews, C. (2003). Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness, *Cerebral Cortex* **13**: 1064–1071.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, *in* F. Pereira, C. Burges, L. Bottou and K. Weinberger (eds), *Advances in Neural Information Processing Systems 25*, Curran Associates, Inc., pp. 1097–1105.

Kropff, E. and Treves, A. (2005). The storage capacity of Potts models for semantic memory retrieval, *Journal of Statistical Mechanics: Theory and Experiment* **2005**: P08010.

Kropff, E. and Treves, A. (2008). The emergence of grid cells: Intelligent design or just adaptation?, *Hippocampus* **18**: 1256–1269.

Kropff, E., Carmichael, J. E., Moser, M. B. and Moser, E. I. (2015). Speed cells in the medial entorhinal cortex, *Nature* **523**: 419–424.

Kubota, Y. (2014). Untangling gabaergic wiring in the cortical microcircuit, *Current Opinion in Neurobiology* **26**: 7–14.

Lally, N., Nugent, A. C., Luckenbaugh, D. A., Niciu, M. J., Roiser, J. P. and Zarate, C. A., J. (2015). Neural correlates of change in major depressive disorder anhedonia following open-label ketamine, *Journal of Psychopharmacology* **29**: 596–607.

Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex, *Journal of Neuroscience* **15**: 1605–1615.

Land, M. F. (1999). Motion and vision: why animals move their eyes, *Journal of Comparative Physiology A* **185**: 341–352.

Land, M. F. and Collett, T. S. (1997). A survey of active vision in invertebrates, *in* M. V. Srinivasan and S. Venkatesh (eds), *From Living Eyes to Seeing Machines*, Oxford University Press, Oxford, pp. 16–36.

Lange, C. (1885). The emotions, *in* E. Dunlap (ed.), *The Emotions*, 1922 edn, Williams and Wilkins, Baltimore.

Lanthorn, T., Storm, J. and Andersen, P. (1984). Current-to-frequency transduction in CA1 hippocampal pyramidal cells: slow prepotentials dominate the primary range firing, *Experimental Brain Research* **53**: 431–443.

Larochelle, H. and Hinton, G. E. (2010). Learning to combine foveal glimpses with a third-order Boltzmann machine, *Advances in Neural Information Processing Systems (NIPS)* **1**: 1243–1251.

Lassalle, J. M., Bataille, T. and Halley, H. (2000). Reversible inactivation of the hippocampal mossy fiber synapses in mice impairs spatial learning, but neither consolidation nor memory retrieval, in the Morris navigation task, *Neurobiology of Learning and Memory* **73**: 243–257.

Lau, H. and Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness, *Trends in Cognitive Sciences* **15**: 365–373.

Lau, H. C., Rogers, R. D. and Passingham, R. E. (2006). On measuring the perceived onsets of spontaneous actions, *Journal of Neuroscience* **26**: 7265–7271.

Lavenex, P. and Amaral, D. G. (2000). Hippocampal-neocortical interaction: a hierarchy of associativity, *Hippocampus* **10**: 420–430.

Lavenex, P., Suzuki, W. A. and Amaral, D. G. (2004). Perirhinal and parahippocampal cortices of the macaque monkey: Intrinsic projections and interconnections, *Journal of Comparative Neurology* **472**: 371–394.

Laxton, A. W., Neimat, J. S., Davis, K. D., Womelsdorf, T., Hutchison, W. D., Dostrovsky, J. O., Hamani, C., Mayberg, H. S. and Lozano, A. M. (2013). Neuronal coding of implicit emotion categories in the subcallosal cortex in patients with depression, *Biological Psychiatry* **74**: 714–719.

Leak, G. K. and Christopher, S. B. (1982). Freudian psychoanalysis and sociobiology: a synthesis, *American Psychologist* **37**: 313–322.

LeCun, Y., Kavukcuoglu, K. and Farabet, C. (2010). Convolutional networks and applications in vision, *2010 IEEE International Symposium on Circuits and Systems* pp. 253–256.

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning, *Nature* **521**: 436–444.

LeDoux, J. E. (1992). Emotion and the amygdala, *in* J. P. Aggleton (ed.), *The Amygdala*, Wiley-Liss, New York, chapter 12, pp. 339–351.

LeDoux, J. E. (1995). Emotion: clues from the brain, *Annual Review of Psychology* **46**: 209–235.

LeDoux, J. E. (1996). *The Emotional Brain*, Simon and Schuster, New York.

LeDoux, J. E. (2008). Emotional coloration of consciousness: how feelings come about, *in* L. Weiskrantz and M. Davies (eds), *Frontiers of Consciousness*, Oxford University Press, Oxford, pp. 69–130.

LeDoux, J. E. (2011). Rethinking the emotional brain, *Neuron* **73**: 653–676.

Lee, H., Grosse, R., Ranganath, R. and Ng, A. Y. (2011). Unsupervised learning of hierarchical representations with convolutional deep belief networks, *Communications of the ACM* **54**: 95–103.

Lee, I. and Kesner, R. P. (2002). Differential contribution of NMDA receptors in hippocampal subregions to spatial working memory, *Nature Neuroscience* **5**: 162–168.

Lee, I. and Kesner, R. P. (2003a). Differential roles of dorsal hippocampal subregions in spatial working memory with short versus intermediate delay, *Behavioral Neuroscience* **117**: 1044–1053.

Lee, I. and Kesner, R. P. (2003b). Time-dependent relationship between the dorsal hippocampus and the prefrontal cortex in spatial memory, *Journal of Neurosciennce* **23**: 1517–1523.

Lee, I. and Kesner, R. P. (2004a). Differential contributions of dorsal hippocampal subregions to memory acquisition and retrieval in contextual fear-conditioning, *Hippocampus* **14**: 301–310.

Lee, I. and Kesner, R. P. (2004b). Encoding versus retrieval of spatial memory: double dissociation between the dentate gyrus and the perforant path inputs into CA3 in the dorsal hippocampus, *Hippocampus* **14**: 66–76.

Lee, I., Rao, G. and Knierim, J. J. (2004). A double dissociation between hippocampal subfields: differential time course of CA3 and CA1 place cells for processing changed environments, *Neuron* **42**: 803–815.

Lee, I., Jerman, T. S. and Kesner, R. P. (2005). Disruption of delayed memory for a sequence of spatial locations following CA1 or CA3 lesions of the dorsal hippocampus, *Neurobiology of Learning and Memory* **84**: 138–147.

Lee, T. S. (1996). Image representation using 2D Gabor wavelets, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18,10**: 959–971.

Lee, T. S., Mumford, D., Romero, R. D. and Lamme, V. A. F. (1998). The role of primary visual cortex in higher level vision, *Vision Research* **38**: 2429–2454.

Lee, T. S., Yang, C. Y., Romero, R. D. and Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher order perceptual saliency, *Nature Neuroscience* **5**: 589–597.

Lehky, S. R., Sejnowski, T. J. and Desimone, R. (2005). Selectivity and sparseness in the responses of striate complex cells, *Vision Research* **45**: 57–73.

Lehn, H., Steffenach, H. A., van Strien, N. M., Veltman, D. J., Witter, M. P. and Haberg, A. K. (2009). A specific role of the human hippocampus in recall of temporal sequences, *Journal of Neuroscience* **29**: 3475–3484.

Leibowitz, S. F. and Hoebel, B. G. (1998). Behavioral neuroscience and obesity, *in* G. A. Bray, C. Bouchard and P. T. James (eds), *The Handbook of Obesity*, Dekker, New York, pp. 313–358.

Leonard, C. M., Rolls, E. T., Wilson, F. A. W. and Baylis, G. C. (1985). Neurons in the amygdala of the monkey with responses selective for faces, *Behavioural Brain Research* **15**: 159–176.

Leuner, K. and Muller, W. E. (2006). The complexity of the dopaminergic synapses and their modulation by antipsychotics, *Pharmacopsychiatry* **39 Suppl 1**: S15–20.

Leutgeb, J. K., Leutgeb, S., Moser, M. B. and Moser, E. I. (2007). Pattern separation in the dentate gyrus and CA3 of the hippocampus, *Science* **315**: 961–966.

Leutgeb, S. and Leutgeb, J. K. (2007). Pattern separation, pattern completion, and new neuronal codes within a continuous CA3 map, *Learning and Memory* **14**: 745–757.

Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M. B. and Moser, E. I. (2004). Distinct ensemble codes in hippocampal areas CA3 and CA1, *Science* **305**: 1295–1298.

Leutgeb, S., Leutgeb, J. K., Treves, A., Meyer, R., Barnes, C. A., McNaughton, B. L., Moser, M.-B. and Moser, E. I. (2005). Progressive transformation of hippocampal neuronal representations in "morphed" environments, *Neuron* **48**: 345–358.

LeVay, S., Wiesel, T. N. and Hubel, D. H. (1980). The development of ocular dominance columns in normal and visually deprived monkeys, *Journal of Comparative Neurology* **191**: 1–51.

Levitt, J. B., Lund, J. S. and Yoshioka, T. (1996). Anatomical substrates for early stages in cortical processing of visual information in the macaque monkey, *Behavioural Brain Research* **76**: 5–19.

Levy, W. B. (1985). Associative changes in the synapse: LTP in the hippocampus, *in* W. B. Levy, J. A. Anderson and S. Lehmkuhle (eds), *Synaptic Modification, Neuron Selectivity, and Nervous System Organization*, Erlbaum, Hillsdale, NJ, chapter 1, pp. 5–33.

Levy, W. B. (1989). A computational approach to hippocampal function, *in* R. D. Hawkins and G. H. Bower (eds), *Computational Models of Learning in Simple Neural Systems*, Academic Press, San Diego, pp. 243–305.

Levy, W. B. (1996). A sequence predicting ca3 is a flexible associator that learns and uses context to solve hippocampal-like tasks, *Hippocampus* **6**: 579–590.

Levy, W. B. and Baxter, R. A. (1996). Energy efficient neural codes, *Neural Computation* **8**: 531–543.

Levy, W. B. and Desmond, N. L. (1985). The rules of elemental synaptic plasticity, *in* W. B. Levy, J. A. Anderson and S. Lehmkuhle (eds), *Synaptic Modification, Neuron Selectivity, and Nervous System Organization*, Erlbaum, Hillsdale, NJ, chapter 6, pp. 105–121.

Levy, W. B., Colbert, C. M. and Desmond, N. L. (1990). Elemental adaptive processes of neurons and synapses: a statistical/computational perspective, *in* M. Gluck and D. Rumelhart (eds), *Neuroscience and Connectionist Theory*, Erlbaum, Hillsdale, NJ, chapter 5, pp. 187–235.

Levy, W. B., Wu, X. and Baxter, R. A. (1995). Unification of hippocampal function via computational/encoding considerations, *International Journal of Neural Systems* **6, Suppl.**: 71–80.

Lewis, D. A. (2014). Inhibitory neurons in human cortical circuits: substrate for cognitive dysfunction in schizophrenia, *Current Opinion in Neurobiology* **26**: 22–6.

Lewis, M. D. (2005). Bridging emotion theory and neurobiology through dynamic systems modeling, *Behavioral and Brain Sciences* **28**: 169–245.

Li, C. S. and Cho, Y. K. (2006). Efferent projection from the bed nucleus of the stria terminalis suppresses activity of taste-responsive neurons in the hamster parabrachial nuclei, *American Journal of Physiology Regul Integr Comp Physiol* **291**: R914–R926.

Li, C. S., Cho, Y. K. and Smith, D. V. (2002). Taste responses of neurons in the hamster solitary nucleus are modulated by the central nucleus of the amygdala, *Journal of Neurophysiology* **88**: 2979–2992.

Li, H., Matsumoto, K. and Watanabe, H. (1999). Different effects of unilateral and bilateral hippocampal lesions in rats on the performance of radial maze and odor-paired associate tasks, *Brain Research Bulletin* **48**: 113–119.

Libet, B. (2002). The timing of mental events: Libet's experimental findings and their implications, *Consciousmess and Cognition* **11**: 291–299; discussion 304–333.

Liddle, P. F. (1987). The symptoms of chronic schizophrenia: a re-examination of the positive-negative dichotomy, *British Journal of Psychiatry* **151**: 145–151.

Lieberman, J. A., Perkins, D., Belger, A., Chakos, M., Jarskog, F., Boteva, K. and Gilmore, J. (2001). The early stages of schizophrenia: speculations on pathogenesis, pathophysiology, and therapeutic approaches, *Biological Psychiatry* **50**: 884–897.

Lisberger, S. G. and Thach, W. T. (2013). The cerebellum, *in* E. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum and A. J. Hudspeth (eds), *Principles of Neural Science*, 5th edn, McGraw-Hill, New York, chapter 24, pp. 960–981.

Lisman, J. E. and Idiart, M. A. (1995). Storage of 7 ± 2 short-term memories in oscillatory subcycles, *Science* **267**: 1512–1515.

Lisman, J. E. and Idiart, M. A. (1999). Relating hippocampal circuitry to function: recall of memory sequences by reciprocal dentate-CA3 interactions, *Neuron* **22**: 233–242.

Lisman, J. E., Fellous, J. M. and Wang, X. J. (1998). A role for NMDA-receptor channels in working memory, *Nature Neuroscience* **1**: 273–275.

Lisman, J. E., Talamini, L. M. and Raffone, A. (2005). Recall of memory sequences by interaction of the dentate and CA3: a revised model of the phase precession, *Neural Networks* **18**: 1191–1201.

Little, W. A. (1974). The existence of persistent states in the brain, *Mathematical Bioscience* **19**: 101–120.

Liu, Y. and Wang, X.-J. (2001). Spike-frequency adaptation of a generalized leaky integrate-and-fire model neuron, *Journal of Computational Neuroscience* **10**: 25–45.

Liu, Y. H. and Wang, X.-J. (2008). A common cortical circuit mechanism for perceptual categorical discrimination and veridical judgment, *PLoS Computational Biology* p. e1000253.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI, *Nature* **453**: 869–878.

Logothetis, N. K. and Sheinberg, D. L. (1996). Visual object recognition, *Annual Review of Neuroscience* **19**: 577–621.

Logothetis, N. K., Pauls, J., Bulthoff, H. H. and Poggio, T. (1994). View-dependent object recognition by monkeys, *Current Biology* **4**: 401–414.

Logothetis, N. K., Pauls, J. and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys, *Current Biology* **5**: 552–563.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal, *Nature* **412**: 150–157.

Loh, M., Rolls, E. T. and Deco, G. (2007a). A dynamical systems hypothesis of schizophrenia, *PLoS Computational Biology* **3**: e228. doi:10.1371/journal.pcbi.0030228.

Loh, M., Rolls, E. T. and Deco, G. (2007b). Statistical fluctuations in attractor networks related to schizophrenia, *Pharmacopsychiatry* **40**: S78–84.

Lowe, D. (1985). *Perceptual Organization and Visual Recognition*, Kluwer, Boston.

Lozano, A. M., Giacobbe, P., Hamani, C., Rizvi, S. J., Kennedy, S. H., Kolivakis, T. T., Debonnel, G., Sadikot, A. F., Lam, R. W., Howard, A. K., Ilcewicz-Klimek, M., Honey, C. R. and Mayberg, H. S. (2012). A multicenter pilot study of subcallosal cingulate area deep brain stimulation for treatment-resistant depression, *Journal of*

Neurosurgery **116**: 315–322.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*, Oxford University Press, New York.

Luck, S. J., Chelazzi, L., Hillyard, S. A. and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex, *Journal of Neurophysiology* **77**: 24–42.

Lujan, J. L., Chaturvedi, A., Choi, K. S., Holtzheimer, P. E., Gross, R. E., Mayberg, H. S. and McIntyre, C. C. (2013). Tractography-activation models applied to subcallosal cingulate deep brain stimulation, *Brain Stimulation* **6**: 737–739.

Luk, C. H. and Wallis, J. D. (2009). Dynamic encoding of responses and outcomes by neurons in medial prefrontal cortex, *Journal of Neuroscience* **29**: 7526–7539.

Lund, J. S. (1984). Spiny stellate neurons, *in* A. Peters and E. Jones (eds), *Cerebral Cortex, Vol. 1, Cellular Components of the Cerebral Cortex*, Plenum, New York, chapter 7, pp. 255–308.

Lundy, R. F., J. and Norgren, R. (2004). Activity in the hypothalamus, amygdala, and cortex generates bilateral and convergent modulation of pontine gustatory neurons, *Journal of Neurophysiology* **91**: 1143–1157.

Luo, Q., Ge, T., Grabenhorst, F., Feng, J. and Rolls, E. T. (2013). Attention-dependent modulation of cortical taste circuits revealed by Granger causality with signal-dependent noise, *PLoS Computational Biology* **9**: e1003265.

Luskin, M. B. and Price, J. L. (1983). The topographic organization of associational fibers of the olfactory system in the rat, including centrifugal fibers to the olfactory bulb, *Journal of Comparative Neurology* **216**: 264–291.

Lycan, W. G. (1997). Consciousness as internal monitoring, *in* N. Block, O. Flanagan and G. Guzeldere (eds), *The Nature of Consciousness: Philosophical Debates*, MIT Press, Cambridge, MA, pp. 755–771.

Lynch, G. and Gall, C. M. (2006). AMPAkines and the threefold path to cognitive enhancement, *Trends in Neuroscience* **29**: 554–562.

Lynch, M. A. (2004). Long-term potentiation and memory, *Physiological Reviews* **84**: 87–136.

Ma, Y. (2015). Neuropsychological mechanism underlying antidepressant effect: a systematic meta-analysis, *Molecular Psychiatry* **20**: 311–319.

Maaswinkel, H., Jarrard, L. E. and Whishaw, I. Q. (1999). Hippocampectomized rats are impaired in homing by path integration, *Hippocampus* **9**: 553–561.

MacDonald, C. J., Lepage, K. Q., Eden, U. T. and Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events, *Neuron* **71**: 737–749.

MacGregor, R. J. (1987). *Neural and Brain Modelling*, Academic Press, San Diego, CA.

MacKay, D. M. and McCulloch, W. S. (1952). The limiting information capacity of a neuronal link, *Bulletin of Mathematical Biophysics* **14**: 127–135.

Madsen, J. and Kesner, R. P. (1995). The temporal-distance effect in subjects with dementia of the Alzheimer type, *Alzheimer's Disease and Associated Disorders* **9**: 94–100.

Maguire, E. A. (2014). Memory consolidation in humans: new evidence and opportunities, *Experimental Physiology* **99**: 471–486.

Maier, A., Logothetis, N. K. and Leopold, D. A. (2005). Global competition dictates local suppression in pattern rivalry, *Journal of Vision* **5**: 668–677.

Malhotra, A. K., Pinals, D. A., Weingartner, H., Sirocco, K., Missar, C. D., Pickar, D. and Breier, A. (1996). NMDA receptor function and human cognition: the effects of ketamine in healthy volunteers, *Neuropsychopharmacology* **14**: 301–307.

Malkova, L. and Mishkin, M. (2003). One-trial memory for object-place associations after separate lesions of hippocampus and posterior parahippocampal region in the monkey, *Journal of Neuroscience* **23**: 1956–1965.

Malkova, L., Bachevalier, J., Mishkin, M. and Saunders, R. C. (2001). Neurotoxic lesions of perirhinal cortex impair visual recognition memory in rhesus monkeys, *Neuroreport* **12**: 1913–1917.

Malsburg, C. v. d. (1973). Self-organization of orientation-sensitive columns in the striate cortex, *Kybernetik* **14**: 85–100.

Malsburg, C. v. d. (1990a). A neural architecture for the representation of scenes, *in* J. L. McGaugh, N. M. Weinburger and G. Lynch (eds), *Brain Organization and Memory: Cells, Systems and Circuits*, Oxford University Press, Oxford, chapter 18, pp. 356–372.

Malsburg, C. v. d. (1990b). A neural architecture for the representation of scenes, *in* J. L. McGaugh, N. M. Weinberger and G. Lynch (eds), *Brain Organization and Memory: Cells, Systems and Circuits*, Oxford University Press, New York, chapter 19, pp. 356–372.

Mangia, S., Giove, F., Tkac, I., Logothetis, N. K., Henry, P. G., Olman, C. A., Maraviglia, B., Di Salle, F. and Ugurbil, K. (2009). Metabolic and hemodynamic events after changes in neuronal activity: current hypotheses, theoretical predictions and in vivo NMR experimental findings, *Journal of Cerebral Blood Flow and Metabolism* **29**: 441–463.

Manwani, A. and Koch, C. (2001). Detecting and estimating signals over noisy and unreliable synapses: information-theoretic analysis, *Neural Computation* **13**: 1–33.

Markov, N. T. and Kennedy, H. (2013). The importance of being hierarchical, *Current Opinion in Neurobiology* **23**: 187–194.

Markov, N. T., Ercsey-Ravasz, M., Lamy, C., Ribeiro Gomes, A. R., Magrou, L., Misery, P., Giroud, P., Barone, P.,

Dehay, C., Toroczkai, Z., Knoblauch, K., Van Essen, D. C. and Kennedy, H. (2013a). The role of long-range connections on the specificity of the macaque interareal cortical network, *Proceedings of the National Academy of Sciences USA* **110**: 5187–5192.

Markov, N. T., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., Toroczkai, Z. and Kennedy, H. (2013b). Cortical high-density counterstream architectures, *Science* **342**: 1238406.

Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., Misery, P., Falchier, A., Quilodran, R., Gariel, M. A., Sallet, J., Gamanut, R., Huissoud, C., Clavagnier, S., Giroud, P., Sappey-Marinier, D., Barone, P., Dehay, C., Toroczkai, Z., Knoblauch, K., Van Essen, D. C. and Kennedy, H. (2014a). A weighted and directed interareal connectivity matrix for macaque cerebral cortex, *Cerebral Cortex* **24**: 17–36.

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., Barone, P., Dehay, C., Knoblauch, K. and Kennedy, H. (2014b). Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex, *Journal of Comparative Neurology* **522**: 225–259.

Markram, H. and Segal, M. (1990). Acetylcholine potentiates responses to N-methyl-D-aspartate in the rat hippocampus, *Neurosci Letters* **113**: 62–65.

Markram, H. and Segal, M. (1992). The inositol 1,4,5-triphosphate pathway mediates cholinergic potentiation of rat hippocampal neuronal responses to NMDA, *Journal of Physiology* **447**: 513–533.

Markram, H. and Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons, *Nature* **382**: 807–810.

Markram, H., Lübke, J., Frotscher, M. and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs, *Science* **275**: 213–215.

Markram, H., Pikus, D., Gupta, A. and Tsodyks, M. (1998). Information processing with frequency-dependent synaptic connections, *Neuropharmacology* **37**: 489–500.

Markram, H., Gerstner, W. and Sjöström, P. J. (2012). Spike-timing-dependent plasticity: a comprehensive overview, *Frontiers in Synaptic Neuroscience* **4**: 2.

Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., Kahou, G. A., Berger, T. K., Bilgili, A., Buncic, N., Chalimourda, A., Chindemi, G., Courcol, J. D., Delalondre, F., Delattre, V., Druckmann, S., Dumusc, R., Dynes, J., Eilemann, S., Gal, E., Gevaert, M. E., Ghobril, J. P., Gidon, A., Graham, J. W., Gupta, A., Haenel, V., Hay, E., Heinis, T., Hernando, J. B., Hines, M., Kanari, L., Keller, D., Kenyon, J., Khazen, G., Kim, Y., King, J. G., Kisvarday, Z., Kumbhar, P., Lasserre, S., Le Be, J. V., Magalhaes, B. R., Merchan-Perez, A., Meystre, J., Morrice, B. R., Muller, J., Munoz-Cespedes, A., Muralidhar, S., Muthurasa, K., Nachbaur, D., Newton, T. H., Nolte, M., Ovcharenko, A., Palacios, J., Pastor, L., Perin, R., Ranjan, R., Riachi, I., Rodriguez, J. R., Riquelme, J. L., Rossert, C., Sfyrakis, K., Shi, Y., Shillcock, J. C., Silberg, G., Silva, R., Tauheed, F., Telefont, M., Toledo-Rodriguez, M., Trankler, T., Van Geit, W., Diaz, J. V., Walker, R., Wang, Y., Zaninetta, S. M., DeFelipe, J., Hill, S. L., Segev, I. and Schurmann, F. (2015). Reconstruction and simulation of neocortical microcircuitry, *Cell* **163**: 456–92.

Markus, E. J., Qin, Y. L., Leonard, B., Skaggs, W., McNaughton, B. L. and Barnes, C. A. (1995). Interactions between location and task affect the spatial and directional firing of hippocampal neurons, *Journal of Neuroscience* **15**: 7079–7094.

Marr, D. (1969). A theory of cerebellar cortex, *Journal of Physiology* **202**: 437–470.

Marr, D. (1970). A theory for cerebral cortex, *Proceedings of The Royal Society of London, Series B* **176**: 161–234.

Marr, D. (1971). Simple memory: a theory for archicortex, *Philosophical Transactions of The Royal Society of London, Series B* **262**: 23–81.

Marr, D. (1982). *Vision*, Freeman, San Francisco.

Marr, D. and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional structure, *Proceedings of the Royal Society of London B* **200**: 269–294.

Marshuetz, C. (2005). Order information in working memory: an integrative review of evidence from brain and behavior, *Psychological Bulletin* **131**: 323–339.

Marti, D., Deco, G., Del Giudice, P. and Mattia, M. (2006). Reward-biased probabilistic decision-making: mean-field predictions and spiking simulations, *Neurocomputing* **39**: 1175–1178.

Marti, D., Deco, G., Mattia, M., Gigante, G. and Del Giudice, P. (2008). A fluctuation-driven mechanism for slow decision processes in reverberant networks, *PLoS ONE* **3**: e2534. doi:10.1371/journal.pone.0002534.

Martin, K. A. C. (1984). Neuronal circuits in cat striate cortex, *in* E. Jones and A. Peters (eds), *Cerebral Cortex, Vol. 2, Functional Properties of Cortical Cells*, Plenum, New York, chapter 9, pp. 241–284.

Martin, S. J., Grimwood, P. D. and Morris, R. G. (2000). Synaptic plasticity and memory: an evaluation of the hypothesis, *Annual Review of Neuroscience* **23**: 649–711.

Martinez, C. O., Do, V. H., Martinez, J. L. J. and Derrick, B. E. (2002). Associative long-term potentiation (LTP) among extrinsic afferents of the hippocampal CA3 region in vivo, *Brain Research* **940**: 86–94.

Martinez-Garcia, M., Rolls, E. T., Deco, G. and Romo, R. (2011). Neural and computational mechanisms of postponed decisions, *Proceedings of the National Academy of Sciences* **108**: 11626–11631.

Martinez-Trujillo, J. and Treue, S. (2002). Attentional modulation strength in cortical area MT depends on stimulus contrast, *Neuron* **35**: 365–370.

Mascaro, M. and Amit, D. J. (1999). Effective neural response function for collective population states, *Network* **10**: 351–373.

Mason, A. and Larkman, A. (1990). Correlations between morphology and electrophysiology of pyramidal neurones in slices of rat visual cortex. I. Electrophysiology, *Journal of Neuroscience* **10**: 1415–1428.

Masuda, N. and Aihara, K. (2003). Ergodicity of spike trains: when does trial averaging make sense?, *Neural Computation* **15**: 1341–1372.

Matsumoto, M. and Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals, *Nature* **459**: 837–841.

Matsumoto, M., Matsumoto, K., Abe, H. and Tanaka, K. (2007). Medial prefrontal selectivity signalling prediction errors of action values, *Nature Neuroscience* **10**: 647–656.

Matsumura, N., Nishijo, H., Tamura, R., Eifuku, S., Endo, S. and Ono, T. (1999). Spatial- and task-dependent neuronal responses during real and virtual translocation in the monkey hippocampal formation, *Journal of Neuroscience* **19**: 2318–2393.

Mattia, M. and Del Giudice, P. (2002). Population dynamics of interacting spiking neurons, *Physical Review E* **66**: 051917.

Mattia, M. and Del Giudice, P. (2004). Finite-size dynamics of inhibitory and excitatory interacting spiking neurons, *Physical Review E* **70**: 052903.

Maunsell, J. H. R. (1995). The brain's visual world: representation of visual targets in cerebral cortex, *Science* **270**: 764–769.

Mayberg, H. S. (2003). Positron emission tomography imaging in depression: a neural systems perspective, *Neuro-imaging Clinics of North America* **13**: 805–815.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*, Cambridge University Press, Cambridge.

Maynard Smith, J. (1984). Game theory and the evolution of behaviour, *Behavioral and Brain Sciences* **7**: 95–125.

McAdams, C. and Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4, *Journal of Neuroscience* **19**: 431–441.

McCabe, C. and Rolls, E. T. (2007). Umami: a delicious flavor formed by convergence of taste and olfactory pathways in the human brain, *European Journal of Neuroscience* **25**: 1855–1864.

McCabe, C., Rolls, E. T., Bilderbeck, A. and McGlone, F. (2008). Cognitive influences on the affective representation of touch and the sight of touch in the human brain, *Social, Cognitive and Affective Neuroscience* **3**: 97–108.

McClelland, J. L. and Rumelhart, D. E. (1986). A distributed model of human learning and memory, *in* J. L. McClelland and D. E. Rumelhart (eds), *Parallel Distributed Processing*, Vol. 2, MIT Press, Cambridge, MA, chapter 17, pp. 170–215.

McClelland, J. L. and Rumelhart, D. E. (1988). *Explorations in Parallel Distributed Processing*, MIT Press, Cambridge, MA.

McClelland, J. L., McNaughton, B. L. and 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.

McClure, S. M., Laibson, D. I., Loewenstein, G. and Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards, *Science* **306**: 503–507.

McCormick, D. A. and Westbrook, G. L. (2013). Sleep and dreaming, *in* E. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum and A. J. Hudspeth (eds), *Principles of Neural Science*, 5th edn, McGraw-Hill, New York, chapter 51, pp. 1140–1158.

McGaugh, J. L. (2000). Memory - a century of consolidation, *Science* **287**: 248–251.

McGinty, D. and Szymusiak, R. (1988). Neuronal unit activity patterns in behaving animals: brainstem and limbic system, *Annual Review of Psychology* **39**: 135–168.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices, *Nature* **264**: 746–748.

McLeod, P., Plunkett, K. and Rolls, E. T. (1998). *Introduction to Connectionist Modelling of Cognitive Processes*, Oxford University Press, Oxford.

McNaughton, B. L. (1991). Associative pattern completion in hippocampal circuits: new evidence and new questions, *Brain Research Reviews* **16**: 193–220.

McNaughton, B. L. and Morris, R. G. M. (1987). Hippocampal synaptic enhancement and information storage within a distributed memory system, *Trends in Neuroscience* **10**: 408–415.

McNaughton, B. L. and Nadel, L. (1990). Hebb-Marr networks and the neurobiological representation of action in space, *in* M. A. Gluck and D. E. Rumelhart (eds), *Neuroscience and Connectionist Theory*, Erlbaum, Hillsdale, NJ, pp. 1–64.

McNaughton, B. L., Barnes, C. A. and O'Keefe, J. (1983). The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats, *Experimental Brain Research* **52**: 41–49.

McNaughton, B. L., Barnes, C. A., Meltzer, J. and Sutherland, R. J. (1989). Hippocampal granule cells are necessary for normal spatial learning but not for spatially selective pyramidal cell discharge, *Experimental Brain Research* **76**: 485–496.

McNaughton, B. L., Chen, L. L. and Markus, E. J. (1991). "Dead reckoning", landmark learning, and the sense of direction: a neurophysiological and computational hypothesis, *Journal of Cognitive Neuroscience* **3**: 190–202.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. and Moser, M.-B. (2006). Path integration and the neural basis of the hippocampal map, *Nature Reviews Neuroscience* **7**: 663–676.

Mease, R. A., Krieger, P. and Groh, A. (2014). Cortical control of adaptation and sensory relay mode in the thalamus, *Proceedings of the National Academy of Sciences* **111**: 6798–6803.

Medin, D. L. and Schaffer, M. M. (1978). Context theory of classification learning, *Psychological Review* **85**: 207–238.

Mel, B. W. (1997). SEEMORE: Combining color, shape, and texture histogramming in a neurally-inspired approach to visual object recognition, *Neural Computation* **9**: 777–804.

Mel, B. W. and Fiser, J. (2000). Minimizing binding errors using learned conjunctive features, *Neural Computation* **12**: 731–762.

Mel, B. W., Ruderman, D. L. and Archie, K. A. (1998). Translation-invariant orientation tuning in visual "complex" cells could derive from intradendritic computations, *Journal of Neuroscience* **18**(11): 4325–4334.

Menzies, L., Chamberlain, S. R., Laird, A. R., Thelen, S. M., Sahakian, B. J. and Bullmore, E. T. (2008). Integrating evidence from neuroimaging and neuropsychological studies of obsessive-compulsive disorder: The orbitofronto-striatal model revisited, *Neuroscience and Biobehavioral Reviews* **32**: 525–549.

Mesulam, M.-M. (1990). Human brain cholinergic pathways, *Progress in Brain Research* **84**: 231–241.

Mesulam, M. M. (1998). From sensation to cognition, *Brain* **121**: 1013–1084.

Meunier, M., Bachevalier, J. and Mishkin, M. (1997). Effects of orbital frontal and anterior cingulate lesions on object and spatial memory in rhesus monkeys, *Neuropsychologia* **35**: 999–1015.

Middleton, F. A. and Strick, P. L. (1996a). New concepts about the organization of the basal ganglia., *in* J. A. Obeso (ed.), *Advances in Neurology: The Basal Ganglia and the Surgical Treatment for Parkinson's Disease*, Raven, New York.

Middleton, F. A. and Strick, P. L. (1996b). The temporal lobe is a target of output from the basal ganglia, *Proceedings of the National Academy of Sciences of the USA* **93**: 8683–8687.

Middleton, F. A. and Strick, P. L. (2000). Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies, *Brain and Cognition* **42**: 183–200.

Mikami, A., Nakamura, K. and Kubota, K. (1994). Neuronal responses to photographs in the superior temporal sulcus of the rhesus monkey, *Behavioural Brain Research* **60**: 1–13.

Miller, E. K. and Buschman, T. J. (2013). Cortical circuits for the control of attention, *Current Opinion in Neurobiology* **23**: 216–222.

Miller, E. K. and Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory, *Science* **263**: 520–522.

Miller, E. K., Gochin, P. M. and Gross, C. G. (1993a). Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus, *Brain Research* **616**: 25–29.

Miller, E. K., Li, L. and Desimone, R. (1993b). Activity of neurons in anterior inferior temporal cortex during a short-term memory task, *Journal of Neuroscience* **13**: 1460–1478.

Miller, E. K., Nieder, A., Freedman, D. J. and Wallis, J. D. (2003). Neural correlates of categories and concepts, *Current Opinion in Neurobiology* **13**: 198–203.

Miller, G. A. (1955). Note on the bias of information estimates, *Information Theory in Psychology; Problems and Methods II-B* pp. 95–100.

Miller, J. F., Neufang, M., Solway, A., Brandt, A., Trippel, M., Mader, I., Hefft, S., Merkow, M., Polyn, S. M., Jacobs, J., Kahana, M. J. and Schulze-Bonhage, A. (2013). Neural activity in human hippocampal formation reveals the spatial context of retrieved memories, *Science* **342**: 1111–1114.

Miller, K. D. (1994). Models of activity-dependent neural development, *Progress in Brain Research* **102**: 303–308.

Miller, K. D. (2016). Canonical computations of cerebral cortex, *Current Opinion in Neurobiology* **37**: 75–84.

Miller, P. and Wang, X.-J. (2006). Power-law neuronal fluctuations in a recurrent network model of parametric working memory, *Journal of Neurophysiology* **95**: 1099–1114.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories: New Foundation for Realism*, MIT Press, Cambridge, MA.

Milner, A. (2008). Conscious and unconscious visual processing in the human brain, *in* L. Weiskrantz and M. Davies (eds), *Frontiers of Consciousness*, Oxford University Press, Oxford, chapter 5, pp. 169–214.

Milner, A. D. and Goodale, M. A. (1995). *The Visual Brain in Action*, Oxford University Press, Oxford.

Milner, P. (1974). A model for visual shape recognition, *Psychological Review* **81**: 521–535.

Milton, A. L., Lee, J. L. L., Butler, V. J., Gardner, R. and Everitt, B. J. (2008). Intra-amygdala and systemic antagonism of NMDA receptors prevents the reconsolidation of drug-associated memory and impairs subsequently both novel and previously acquired drug-seeking behaviors, *Journal of Neuroscience* **28**: 8230–8237.

Minai, A. A. and Levy, W. B. (1993). Sequence learning in a single trial, *International Neural Network Society World Congress of Neural Networks* **2**: 505–508.

Minsky, M. L. and Papert, S. A. (1969). *Perceptrons*, expanded 1988 edn, MIT Press, Cambridge, MA.

Miyamoto, S., Duncan, G. E., Marx, C. E. and Lieberman, J. A. (2005). Treatments for schizophrenia: a critical review of pharmacology and mechanisms of action of antipsychotic drugs, *Molecular Psychiatry* **10**: 79–104.

Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex, *Nature* **335**: 817–820.

Miyashita, Y. (1993). Inferior temporal cortex: where visual perception meets memory, *Annual Review of Neuroscience* **16**: 245–263.

Miyashita, Y. and Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex, *Nature* **331**: 68–70.

Miyashita, Y., Rolls, E. T., Cahusac, P. M. B., Niki, H. and Feigenbaum, J. D. (1989). Activity of hippocampal neurons in the monkey related to a conditional spatial response task, *Journal of Neurophysiology* **61**: 669–678.

Miyashita, Y., Okuno, H., Tokuyama, W., Ihara, T. and Nakajima, K. (1996). Feedback signal from medial temporal lobe mediates visual associative mnemonic codes of inferotemporal neurons, *Cognitive Brain Research* **5**: 81–86.

Miyashita, Y., Kameyama, M., Hasegawa, I. and Fukushima, T. (1998). Consolidation of visual associative long-term memory in the temporal cortex of primates, *Neurobiology of Learning and Memory* **1**: 197–211.

Mizumori, S. J. and Tryon, V. L. (2015). Integrative hippocampal and decision-making neurocircuitry during goal-relevant predictions and encoding, *Progress in Brain Research* **219**: 217–242.

Mizumori, S. J., Ragozzino, K. E., Cooper, B. G. and Leutgeb, S. (1999). Hippocampal representational organization and spatial context, *Hippocampus* **9**: 444–451.

Moller, H. J. (2005). Antipsychotic and antidepressive effects of second generation antipsychotics: two different pharmacological mechanisms?, *European Archives of Psychiatry and Clinical Neuroscience* **255**: 190–201.

Molnar, Z., Kaas, J. H., de Carlos, J. A., Hevner, R. F., Lein, E. and Nemec, P. (2014). Evolution and development of the mammalian cerebral cortex, *Brain, Behavior and Evolution* **83**: 126–39.

Mombaerts, P. (2006). Axonal wiring in the mouse olfactory system, *Annual Review of Cell and Developmental Biology* **22**: 713–737.

Monaghan, D. T. and Cotman, C. W. (1985). Distribution of N-methyl-D-aspartate-sensitive L-[3H]glutamate-binding sites in the rat brain, *Journal of Neuroscience* **5**: 2909–2919.

Mongillo, G., Barak, O. and Tsodyks, M. (2008). Synaptic theory of working memory, *Science* **319**: 1543–1546.

Montagnini, A. and Treves, A. (2003). The evolution of mammalian cortex, from lamination to arealization, *Brain Research Bulletin* **60**: 387–393.

Montague, P. R., Gally, J. A. and Edelman, G. M. (1991). Spatial signalling in the development and function of neural connections, *Cerebral Cortex* **1**: 199–220.

Montague, P. R., King-Casas, B. and Cohen, J. D. (2006). Imaging valuation models in human choice, *Annual Review of Neuroscience* **29**: 417–448.

Moore, D. B., Lee, P., Paiva, M., Walker, D. W. and Heaton, M. B. (1998). Effects of neonatal ethanol exposure on cholinergic neurons of the rat medial septum, *Alcohol* **15**: 219–226.

Mora, F., Rolls, E. T. and Burton, M. J. (1976). Modulation during learning of the responses of neurones in the lateral hypothalamus to the sight of food, *Experimental Neurology* **53**: 508–519.

Mora, F., Avrith, D. B., Phillips, A. G. and Rolls, E. T. (1979). Effects of satiety on self-stimulation of the orbitofrontal cortex in the monkey, *Neuroscience Letters* **13**: 141–145.

Mora, F., Avrith, D. B. and Rolls, E. T. (1980). An electrophysiological and behavioural study of self-stimulation in the orbitofrontal cortex of the rhesus monkey, *Brain Research Bulletin* **5**: 111–115.

Moran, J. and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex, *Science* **229**: 782–784.

Morecraft, R. J. and Tanji, J. (2009). Cingulofrontal interactions and the cingulate motor areas, *in* B. Vogt (ed.), *Cingulate Neurobiology and Disease*, Oxford University Press, Oxford, chapter 5, pp. 113–144.

Morecraft, R. J., Geula, C. and Mesulam, M.-M. (1992). Cytoarchitecture and neural afferents of orbitofrontal cortex in the brain of the monkey, *Journal of Comparative Neurology* **323**: 341–358.

Moreno-Bote, R., Rinzel, J. and Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability, *Journal of Neurophysiology* **98**: 1125–1139.

Morin, E. L., Hadj-Bouziane, F., Stokes, M., Ungerleider, L. G. and Bell, A. H. (2015). Hierarchical encoding of social cues in primate inferior temporal cortex, *Cerebral Cortex* **25**: 3036–45.

Mormann, F., Dubois, J., Kornblith, S., Milosavljevic, M., Cerf, M., Ison, M., Tsuchiya, N., Kraskov, A., Quiroga, R. Q., Adolphs, R., Fried, I. and Koch, C. (2011). A category-specific response to animals in the right human amygdala, *Nature Neuroscience* **14**: 1247–1249.

Morris, A. M., Churchwell, J. C., Kesner, R. P. and Gilbert, P. E. (2012). Selective lesions of the dentate gyrus produce disruptions in place learning for adjacent spatial locations, *Neurobiology of Learning and Memory* **97**: 326–31.

Morris, J. S., Fritch, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J. and Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy face expressions, *Nature* **383**: 812–815.

Morris, R. G., Moser, E. I., Riedel, G., Martin, S. J., Sandin, J., Day, M. and O'Carroll, C. (2003). Elements of a neurobiological theory of the hippocampus: the role of activity-dependent synaptic plasticity in memory, *Philosophical Transactions of the Royal Society of London B* **358**: 773–786.

Morris, R. G. M. (1989). Does synaptic plasticity play a role in information storage in the vertebrate brain?, *in* R. G. M. Morris (ed.), *Parallel Distributed Processing: Implications for Psychology and Neurobiology*, Oxford

University Press, Oxford, chapter 11, pp. 248–285.

Morris, R. G. M. (2003). Long-term potentiation and memory, *Philosophical Transactions of the Royal Society of London B* **358**: 643–647.

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., McAndrews, M. P., Levine, B., Black, S., Winocur, G. and Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory, *Journal of Anatomy* **207**: 35–66.

Moser, E. I., Moser, M. B. and Roudi, Y. (2014a). Network mechanisms of grid cells, *Philosophical Transactions of the Royal Society of London B Biological Science* **369**: 20120511.

Moser, E. I., Roudi, Y., Witter, M. P., Kentros, C., Bonhoeffer, T. and Moser, M. B. (2014b). Grid cells and cortical representation, *Nature Reviews Neuroscience* **15**: 466–481.

Moser, M. B., Rowland, D. C. and Moser, E. I. (2015). Place cells, grid cells, and memory, *Cold Spring Harbor Perspectives in Biology* **7**: a021808.

Motter, B. (1994). Neural correlates of attentive selection for colours or luminance in extrastriate area V4, *Journal of Neuroscience* **14**: 2178–2189.

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli, *Journal of Neurophysiology* **70**: 909–919.

Mountcastle, V. B. (1957). Modality and topographic properties of single neurons of cat's somatosensory cortex, *Journal of Neurophysiology* **20**: 408–434.

Mountcastle, V. B. (1984). Central nervous mechanisms in mechanoreceptive sensibility, *in* I. Darian-Smith (ed.), *Handbook of Physiology, Section 1: The Nervous System, Vol III, Sensory Processes, Part 2*, American Physiological Society, Bethesda, MD, pp. 789–878.

Movshon, J. A., Adelson, E. H., Gizzi, M. S. and Newsome, W. T. (1985). The analysis of moving visual patterns, *in* C. Chagas, R. Gattass and C. G. Gross (eds), *Pattern Recognition Mechanisms*, Springer-Verlag, New York, pp. 117–151.

Moyer, J. R. J., Deyo, R. A. and Disterhoft, J. F. (1990). Hippocampectomy disrupts trace eye-blink conditioning in rabbits, *Behavioral Neuroscience* **104**: 243–252.

Mozer, M. C. (1991). *The Perception of Multiple Objects: A Connectionist Approach*, MIT Press, Cambridge, MA.

Mueser, K. T. and McGurk, S. R. (2004). Schizophrenia, *Lancet* **363**: 2063–2072.

Muir, J. L., Everitt, B. J. and Robbins, T. W. (1994). AMPA-induced excitotoxic lesions of the basal forebrain: a significant role for the cortical cholinergic system in attentional function, *Journal of Neuroscience* **14**: 2313–2326.

Muller, R. U., Kubie, J. L., Bostock, E. M., Taube, J. S. and Quirk, G. J. (1991). Spatial firing correlates of neurons in the hippocampal formation of freely moving rats, *in* J. Paillard (ed.), *Brain and Space*, Oxford University Press, Oxford, pp. 296–333.

Muller, R. U., Ranck, J. B. and Taube, J. S. (1996). Head direction cells: properties and functional significance, *Current Opinion in Neurobiology* **6**: 196–206.

Mundy, J. and Zisserman, A. (1992). Introduction – towards a new framework for vision, *in* J. Mundy and A. Zisserman (eds), *Geometric Invariance in Computer Vision*, MIT Press, Cambridge, MA, pp. 1–39.

Murray, E. A. and Izquierdo, A. (2007). Orbitofrontal cortex and amygdala contributions to affect and action in primates, *Annals of the New York Academy of Sciences* **1121**: 273–296.

Murray, E. A., Gaffan, D. and Mishkin, M. (1993). Neural substrates of visual stimulus–stimulus association in rhesus monkeys, *Journal of Neuroscience* **13**: 4549–4561.

Murray, E. A., Baxter, M. G. and Gaffan, D. (1998). Monkeys with rhinal cortex damage or neurotoxic hippocampal lesions are impaired on spatial scene learning and object reversals, *Behavioral Neuroscience* **112**: 1291–1303.

Murray, E. A., Moylan, E. J., Saleem, K. S., Basile, B. M. and Turchi, J. (2015). Specialized areas for value updating and goal selection in the primate orbitofrontal cortex, *Elife*.

Mutch, J. and Lowe, D. G. (2008). Object class recognition and localization using sparse features with limited receptive fields, *International Journal of Computer Vision* **80**: 45–57.

Naber, P. A., Lopes da Silva, F. H. and Witter, M. P. (2001). Reciprocal connections between the entorhinal cortex and hippocampal fields CA1 and the subiculum are in register with the projections from CA1 to the subiculum, *Hippocampus* **11**: 99–104.

Nadal, J. P., Toulouse, G., Changeux, J. P. and Dehaene, S. (1986). Networks of formal neurons and memory palimpsests, *Europhysics Letters* **1**: 535–542.

Nagahama, Y., Okada, T., Katsumi, Y., Hayashi, T., Yamauchi, H., Oyanagi, C., Konishi, J., Fukuyama, H. and Shibasaki, H. (2001). Dissociable mechanisms of attentional control within the human prefrontal cortex, *Cerebral Cortex* **11**: 85–92.

Nakazawa, K., Quirk, M. C., Chitwood, R. A., Watanabe, M., Yeckel, M. F., Sun, L. D., Kato, A., Carr, C. A., Johnston, D., Wilson, M. A. and Tonegawa, S. (2002). Requirement for hippocampal CA3 NMDA receptors in associative memory recall, *Science* **297**: 211–218.

Nakazawa, K., Sun, L. D., Quirk, M. C., Rondi-Reig, L., Wilson, M. A. and Tonegawa, S. (2003). Hippocampal CA3 NMDA receptors are crucial for memory acquisition of one-time experience, *Neuron* **38**: 305–315.

Nakazawa, K., McHugh, T. J., Wilson, M. A. and Tonegawa, S. (2004). NMDA receptors, place cells and hippocampal

spatial memory, *Nature Reviews Neuroscience* **5**: 361–372.

Nanry, K. P., Mundy, W. R. and Tilson, H. A. (1989). Colchicine-induced alterations of reference memory in rats: role of spatial versus non-spatial task components, *Behavioral Brain Research* **35**: 45–53.

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G. and Tootell, R. B. (2011). Scene-selective cortical regions in human and nonhuman primates, *Journal of Neuroscience* **31**: 13771–85.

Nelken, I., Prut, Y., Vaadia, E. and Abeles, M. (1994). Population responses to multifrequency sounds in the cat auditory cortex: one- and two-parameter families of sounds, *Hearing Research* **72**: 206–222.

Nesse, R. M. and Lloyd, A. T. (1992). The evolution of psychodynamic mechanisms, *in* J. H. Barkow, L. Cosmides and J. Tooby (eds), *The Adapted Mind*, Oxford University Press, New York, pp. 601–624.

Newcomer, J. W., Farber, N. B., Jevtovic-Todorovic, V., Selke, G., Melson, A. K., Hershey, T., Craft, S. and Olney, J. W. (1999). Ketamine-induced NMDA receptor hypofunction as a model of memory impairment and psychosis, *Neuropsychopharmacology* **20**: 106–118.

Newman, E. L., Gupta, K., Climer, J. R., Monaghan, C. K. and Hasselmo, M. E. (2012). Cholinergic modulation of cognitive processing: insights drawn from computational models, *Frontiers in Behavioural Neuroscience* **6**: 24.

Newsome, W. T., Britten, K. H. and Movshon, J. A. (1989). Neuronal correlates of a perceptual decision, *Nature* **341**: 52–54.

Nicoll, R. A. (1988). The coupling of neurotransmitter receptors to ion channels in the brain, *Science* **241**: 545–551.

Niv, Y., Duff, M. O. and Dayan, P. (2005). Dopamine, uncertainty, and TD learning, *Behavioral and Brain Functions* **1**: 6.

Norgren, R. (1974). Gustatory afferents to ventral forebrain, *Brain Research* **81**: 285–295.

Norgren, R. (1976). Taste pathways to hypothalamus and amygdala, *Journal of Comparative Neurology* **166**: 17–30.

Norgren, R. (1990). Gustatory system, *in* G. Paxinos (ed.), *The Human Nervous System*, Academic Press, San Diego, pp. 845–861.

Norgren, R. and Leonard, C. M. (1971). Taste pathways in rat brainstem, *Science* **173**: 1136–1139.

Norgren, R. and Leonard, C. M. (1973). Ascending central gustatory pathways, *Journal of Comparative Neurology* **150**: 217–238.

Nowak, L. and Bullier, J. (1997). The timing of information transfer in the visual system, *in* K. Rockland, J. Kaas and A. Peters (eds), *Cerebral Cortex: Extrastriate Cortex in Primate*, Plenum, New York, p. 870.

Nugent, A. C., Milham, M. P., Bain, E. E., Mah, L., Cannon, D. M., Marrett, S., Zarate, C. A., Pine, D. S., Price, J. L. and Drevets, W. C. (2006). Cortical abnormalities in bipolar disorder investigated with mri and voxel-based morphometry, *Neuroimage* **30**: 485–497.

Nusslock, R., Young, C. B. and Damme, K. S. (2014). Elevated reward-related neural activation as a unique biological marker of bipolar disorder: assessment and treatment implications, *Behaviour Research and Therapy* **62**: 74–87.

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J. and Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex, *Nature Neuroscience* **4**: 95–102.

Oja, E. (1982). A simplified neuron model as a principal component analyzer, *Journal of Mathematical Biology* **15**: 267–273.

O'Kane, D. and Treves, A. (1992). Why the simplest notion of neocortex as an autoassociative memory would not work, *Network* **3**: 379–384.

O'Keefe, J. (1979). A review of the hippocampal place cells, *Progress in Neurobiology* **13**: 419–439.

O'Keefe, J. (1984). Spatial memory within and without the hippocampal system, *in* W. Seifert (ed.), *Neurobiology of the Hippocampus*, Academic Press, London, pp. 375–403.

O'Keefe, J. (1990). A computational theory of the hippocampal cognitive map, *Progress in Brain Research* **83**: 301–312.

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely moving rat, *Brain Research* **34**: 171–175.

O'Keefe, J. and Nadel, L. (1978). *The Hippocampus as a Cognitive Map*, Clarendon Press, Oxford.

O'Keefe, J. and Speakman, A. (1987). Single unit activity in the rat hippocampus during a spatial memory task, *Experimental Brain Research* **68**(1): 1–27.

Okubo, Y., Suhara, T., Sudo, Y. and Toru, M. (1997a). Possible role of dopamine D1 receptors in schizophrenia, *Molecular Psychiatry* **2**: 291–292.

Okubo, Y., Suhara, T., Suzuki, K., Kobayashi, K., Inoue, O., Terasaki, O., Someya, Y., Sassa, T., Sudo, Y., Matsushima, E., Iyo, M., Tateno, Y. and Toru, M. (1997b). Decreased prefrontal dopamine D1 receptors in schizophrenia revealed by PET, *Nature* **385**: 634–636.

Olshausen, B. A. and Field, D. J. (1997). Sparse coding with an incomplete basis set: a strategy employed by V1, *Vision Research* **37**: 3311–3325.

Olshausen, B. A. and Field, D. J. (2004). Sparse coding of sensory inputs, *Current Opinion in Neurobiology* **14**: 481–487.

Olshausen, B. A., Anderson, C. H. and Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information, *Journal of Neuroscience* **13**: 4700–4719.

Olshausen, B. A., Anderson, C. H. and Van Essen, D. C. (1995). A multiscale dynamic routing circuit for forming size- and position-invariant object representations, *Journal of Computational Neuroscience* **2**: 45–62.

O'Mara, S. M., Rolls, E. T., Berthoz, A. and Kesner, R. P. (1994). Neurons responding to whole-body motion in the primate hippocampus, *Journal of Neuroscience* **14**: 6511–6523.

O'Neill, M. J. and Dix, S. (2007). AMPA receptor potentiators as cognitive enhancers, *IDrugs* **10**: 185–192.

Ongur, D. and Price, J. L. (2000). The organisation of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans, *Cerebral Cortex* **10**: 206–219.

Ongur, D., Ferry, A. T. and Price, J. L. (2003). Architectonic subdivision of the human orbital and medial prefrontal cortex, *Journal of Comparative Neurology* **460**: 425–449.

Ono, T., Nakamura, K., Nishijo, H. and Eifuku, S. (1993). Monkey hippocampal neurons related to spatial and nonspatial functions, *Journal of Neurophysiology* **70**: 1516–1529.

Optican, L. M. and Richmond, B. J. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex: III. Information theoretic analysis, *Journal of Neurophysiology* **57**: 162–178.

Optican, L. M., Gawne, T. J., Richmond, B. J. and Joseph, P. J. (1991). Unbiased measures of transmitted information and channel capacity from multivariate neuronal data, *Biological Cybernetics* **65**: 305–310.

Oram, M. W. and Perrett, D. I. (1994). Modeling visual recognition from neurophysiological constraints, *Neural Networks* **7**: 945–972.

O'Regan, J. K., Rensink, R. A. and Clark, J. J. (1999). Change-blindness as a result of 'mudsplashes', *Nature* **398**: 1736–1753.

O'Reilly, J. and Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience*, MIT Press, Cambridge, MA.

O'Reilly, R. C. (2006). Biologically based computational models of high-level cognition, *Science* **314**: 91–94.

O'Reilly, R. C. and Rudy, J. W. (2001). Conjunctive representations in learning and memory: principles of cortical and hippocampal function, *Psychological Review* **108**: 311–345.

Otto, T. and Eichenbaum, H. (1992). Neuronal activity in the hippocampus during delayed non-match to sample performance in rats: evidence for hippocampal processing in recognition memory, *Hippocampus* **2**: 323–334.

Owen, M. J., Sawa, A. and Mortensen, P. B. (2016). Schizophrenia, *Lancet*.

Padoa-Schioppa, C. (2011). Neurobiology of economic choice: a good-based model, *Annual Review of Neuroscience* **34**: 333–359.

Padoa-Schioppa, C. and Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value, *Nature* **441**: 223–226.

Palomero-Gallagher, N. and Zilles, K. (2004). Isocortex, *in* G. Paxinos (ed.), *The Rat Nervous System*, Elsevier Academic Press, San Diego, pp. 729–757.

Pandya, D. N., Seltzer, B., Petrides, M. and Cipolloni, P. B. (2015). *Cerebral Cortex: Architecture, Connections, and the Dual Origin Concept*, Oxford University Press, Oxford.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*, Oxford University Press, New York.

Pantelis, C., Barber, F. Z., Barnes, T. R., Nelson, H. E., Owen, A. M. and Robbins, T. W. (1999). Comparison of set-shifting ability in patients with chronic schizophrenia and frontal lobe damage, *Schizophrenia Research* **37**: 251–270.

Panzeri, S. and Treves, A. (1996). Analytical estimates of limited sampling biases in different information measures, *Network* **7**: 87–107.

Panzeri, S., Biella, G., Rolls, E. T., Skaggs, W. E. and Treves, A. (1996). Speed, noise, information and the graded nature of neuronal responses, *Network* **7**: 365–370.

Panzeri, S., Schultz, S. R., Treves, A. and Rolls, E. T. (1999a). Correlations and the encoding of information in the nervous system, *Proceedings of the Royal Society B* **266**: 1001–1012.

Panzeri, S., Treves, A., Schultz, S. and Rolls, E. T. (1999b). On decoding the responses of a population of neurons from short time windows, *Neural Computation* **11**: 1553–1577.

Panzeri, S., Rolls, E. T., Battaglia, F. and Lavis, R. (2001). Speed of feedforward and recurrent processing in multilayer networks of integrate-and-fire neurons, *Network: Computation in Neural Systems* **12**: 423–440.

Papp, G. and Treves, A. (2008). Network analysis of the significance of hippocampal subfields, *in* S. J. Y. Mizumori (ed.), *Hippocampal Place Fields: Relevance to Learning and Memory.*, Oxford University Press, New York, chapter 20, pp. 328–342.

Parga, N. and Rolls, E. T. (1998). Transform invariant recognition by association in a recurrent network, *Neural Computation* **10**: 1507–1525.

Parisi, G. (1986). A memory which forgets, *Journal of Physics A* **19**: L617–L619.

Park, H. and Poo, M.-M. (2013). Neurotrophin regulation of neural circuit development and function, *Nature Reviews Neuroscience* **14**: 7–23.

Parker, A. J. (2007). Binocular depth perception and the cerebral cortex, *Nature Reviews Neuroscience* **8**: 379–391.

Parker, A. J., Cumming, B. G. and Dodd, J. V. (2000). Binocular neurons and the perception of depth, *in* M. Gazzaniga (ed.), *The New Cognitive Neurosciences, Second Edition*, MIT Press, Cambridge, MA, chapter 18, pp. 263–277.

Parkinson, J. K., Murray, E. A. and Mishkin, M. (1988). A selective mnemonic role for the hippocampus in monkeys: memory for the location of objects, *Journal of Neuroscience* **8**: 4059–4167.

Passingham, R. E. P. and Wise, S. P. (2012). *The Neurobiology of the Prefrontal Cortex*, Oxford University Press,

Oxford.

Pauls, D. L., Abramovitch, A., Rauch, S. L. and Geller, D. A. (2014). Obsessive-compulsive disorder: an integrative genetic and neurobiological perspective, *Nature Reviews: Neuroscience* **15**: 410–24.

Pearce, J. M. and Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli., *Psychological Review* **87**: 532.

Peitgen, H.-O., Jürgens, H. and Saupe, D. (2004). *Chaos and Fractals: New Frontiers of Science*, Springer, New York.

Peled, A. (2004). From plasticity to complexity: a new diagnostic method for psychiatry, *Medical Hypotheses* **63**: 110–114.

Penades, R., Catalan, R., Andres, S., Salamero, M. and Gasto, C. (2005). Executive function and nonverbal memory in obsessive-compulsive disorder, *Psychiatry Research* **133**: 81–90.

Penades, R., Catalan, R., Rubia, K., Andres, S., Salamero, M. and Gasto, C. (2007). Impaired response inhibition in obsessive compulsive disorder, *European Psychiatry* **22**: 404–410.

Peng, H. C., Sha, L. F., Gan, Q. and Wei, Y. (1998). Energy function for learning invariance in multilayer perceptron, *Electronics Letters* **34**(3): 292–294.

Percheron, G., Yelnik, J. and François, C. (1984a). The primate striato-pallido-nigral system: an integrative system for cortical information, *in* J. S. McKenzie, R. E. Kemm and L. N. Wilcox (eds), *The Basal Ganglia: Structure and Function*, Plenum, New York, pp. 87–105.

Percheron, G., Yelnik, J. and François, C. (1984b). A Golgi analysis of the primate globus pallidus. III. Spatial organization of the striato-pallidal complex, *Journal of Comparative Neurology* **227**: 214–227.

Percheron, G., Yelnik, J., François, C., Fenelon, G. and Talbi, B. (1994). Informational neurology of the basal ganglia related system, *Revue Neurologique (Paris)* **150**: 614–626.

Perez Castillo, I. and Skantzos, N. S. (2004). The Little-Hopfield model on a sparse random graph, *Journal of Physics A: Math Gen* **37**: 9087–9099.

Perin, R., Berger, T. K. and Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups, *Proc Natl Acad Sci U S A* **108**: 5419–5424.

Perin, R., Telefont, M. and Markram, H. (2013). Computing the size and number of neuronal clusters in local circuits, *Frontiers in Neuroanatomy* **7**: 1.

Perrett, D. I., Rolls, E. T. and Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex, *Experimental Brain Research* **47**: 329–342.

Perrett, D. I., Smith, P. A. J., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., Broennimann, R., Milner, A. D. and Jeeves, M. A. (1985a). Visual analysis of body movements by neurons in the temporal cortex of the macaque monkey: a preliminary report, *Behavioural Brain Research* **16**: 153–170.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, D. and Jeeves, M. A. (1985b). Visual cells in temporal cortex sensitive to face view and gaze direction, *Proceedings of the Royal Society of London, Series B* **223**: 293–317.

Perry, C. J. and Fallah, M. (2014). Feature integration and object representations along the dorsal stream visual hierarchy, *Frontiers in Computational Neuroscience*.

Perry, G., Rolls, E. T. and Stringer, S. M. (2006). Spatial vs temporal continuity in view invariant visual object recognition learning, *Vision Research* **46**: 3994–4006.

Perry, G., Rolls, E. T. and Stringer, S. M. (2010). Continuous transformation learning of translation invariant representations, *Experimental Brain Research* **204**: 255–270.

Personnaz, L., Guyon, I. and Dreyfus, G. (1985). Information storage and retrieval in spin-glass-like neural networks, *Journal de Physique Lettres (Paris)* **46**: 359–365.

Pessoa, L. (2009). How do emotion and motivation direct executive control?, *Trends in Cognitive Sciences* **13**: 160–166.

Pessoa, L. and Adolphs, R. (2010). Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance, *Nature Reviews Neuroscience* **11**: 773–783.

Pessoa, L. and Padmala, S. (2005). Quantitative prediction of perceptual decisions during near-threshold fear detection, *Proceedings of the National Academy of Sciences USA* **102**: 5612–5617.

Peters, A. (1984a). Bipolar cells, *in* A. Peters and E. G. Jones (eds), *Cerebral Cortex, Vol. 1, Cellular Components of the Cerebral Cortex*, Plenum, New York, chapter 11, pp. 381–407.

Peters, A. (1984b). Chandelier cells, *in* A. Peters and E. G. Jones (eds), *Cerebral Cortex, Vol. 1, Cellular Components of the Cerebral Cortex*, Plenum, New York, chapter 10, pp. 361–380.

Peters, A. and Jones, E. G. (eds) (1984). *Cerebral Cortex, Vol. 1, Cellular Components of the Cerebral Cortex*, Plenum, New York.

Peters, A. and Regidor, J. (1981). A reassessment of the forms of nonpyramidal neurons in area 17 of the cat visual cortex, *Journal of Comparative Neurology* **203**: 685–716.

Peters, A. and Saint Marie, R. L. (1984). Smooth and sparsely spinous nonpyramidal cells forming local axonal plexuses, *in* A. Peters and E. G. Jones (eds), *Cerebral Cortex, Vol. 1, Cellular Components of the Cerebral Cortex*, New York, Plenum, chapter 13, pp. 419–445.

Peterson, C. and Anderson, J. R. (1987). A mean field theory learning algorithm for neural networks, *Complex*

Systems **1**: 995–1015.

Petrides, M. (1985). Deficits on conditional associative-learning tasks after frontal- and temporal-lobe lesions in man, *Neuropsychologia* **23**: 601–614.

Petrides, M. (1996). Specialized systems for the processing of mnemonic information within the primate frontal cortex, *Philosophical Transactions of the Royal Society of London B* **351**: 1455–1462.

Phelps, E. A. and LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior, *Neuron* **48**: 175–187.

Phillips, W. A., Kay, J. and Smyth, D. (1995). The discovery of structure by multi-stream networks of local processors with contextual guidance, *Network* **6**: 225–246.

Pickens, C. L., Saddoris, M. P., Setlow, B., Gallagher, M., Holland, P. C. and Schoenbaum, G. (2003). Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task, *Journal of Neuroscience* **23**: 11078–11084.

Pinker, S. and Bloom, P. (1992). Natural language and natural selection, *in* J. H. Barkow, L. Cosmides and J. Tooby (eds), *The Adapted Mind*, Oxford University Press, New York, chapter 12, pp. 451–493.

Pinto, N., Cox, D. D. and DiCarlo, J. J. (2008). Why is real-world visual object recognition hard?, *PLoS Computational Biology* **4**: e27.

Pinto, N., Doukhan, D., DiCarlo, J. J. and Cox, D. D. (2009). A high-throughput screening approach to discovering good forms of biologically inspired visual representation, *PLoS Computational Biology* **5**: e1000579.

Pirmoradian, S. and Treves, A. (2013). Encoding words into a Potts attractor network, *in* J. Mayor and P. Gomez (eds), *Computational Models of Cognitive Processes: Proceedings of the 13th Neural Computation and Psychology Workshop (NCPW13)*, World Scientific Press, Singapore, pp. 29–42.

Pitkanen, A., Kelly, J. L. and Amaral, D. G. (2002). Projections from the lateral, basal, and accessory basal nuclei of the amygdala to the entorhinal cortex in the macaque monkey, *Hippocampus* **12**: 186–205.

Pittenger, C. (2015). Glutamatergic agents for ocd and related disorders, *Curr Treat Options Psychiatry* **2**: 271–283.

Pittenger, C., Krystal, J. H. and Coric, V. (2006). Glutamate-modulating drugs as novel pharmacotherapeutic agents in the treatment of obsessive-compulsive disorder, *NeuroRx* **3**(1): 69–81.

Pittenger, C., Bloch, M. H. and Williams, K. (2011). Glutamate abnormalities in obsessive compulsive disorder: Neurobiology, pathophysiology, and treatment, *Pharmacology and Therapeutics* **132**: 314–332.

Poggio, T. and Edelman, S. (1990). A network that learns to recognize three-dimensional objects, *Nature* **343**: 263–266.

Poggio, T. and Girosi, F. (1990a). Networks for approximation and learning, *Proceedings of the IEEE* **78**: 1481–1497.

Poggio, T. and Girosi, F. (1990b). Regularization algorithms for learning that are equivalent to multilayer networks, *Science* **247**: 978–982.

Pollen, D. and Ronner, S. (1981). Phase relationship between adjacent simple cells in the visual cortex, *Science* **212**: 1409–1411.

Posner, M. I. and Keele, S. W. (1968). On the genesis of abstract ideas, *Journal of Experimental Psychology* **77**: 353–363.

Poucet, B. (1989). Object exploration, habituation, and response to a spatial change in rats following septal or medial frontal cortical damage, *Behavioral Neuroscience* **103**: 1009–1016.

Powell, T. P. S. (1981). Certain aspects of the intrinsic organisation of the cerebral cortex, *in* O. Pompeiano and C. Ajmone Marsan (eds), *Brain Mechanisms and Perceptual Awareness*, Raven Press, New York, pp. 1–19.

Power, J. M. and Sah, P. (2008). Competition between calcium-activated K+ channels determines cholinergic action on firing properties of basolateral amygdala projection neurons, *Journal of Neuroscience* **28**: 3209–3220.

Preston, A. R. and Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory, *Current Biology* **23**: R764–73.

Preuss, T. M. (1995). Do rats have prefrontal cortex? The Rose-Woolsey-Akert program reconsidered, *Journal of Cognitive Neuroscience* **7**: 1–24.

Price, J. (2006). Connections of orbital cortex, *in* D. H. Zald and S. L. Rauch (eds), *The Orbitofrontal Cortex*, Oxford University Press, Oxford, chapter 3, pp. 39–55.

Price, J. L. and Drevets, W. C. (2012). Neural circuits underlying the pathophysiology of mood disorders, *Trends in Cognitive Science* **16**: 61–71.

Price, J. L., Carmichael, S. T., Carnes, K. M., Clugnet, M.-C. and Kuroda, M. (1991). Olfactory input to the prefrontal cortex, *in* J. L. Davis and H. Eichenbaum (eds), *Olfaction: A Model System for Computational Neuroscience*, MIT Press, Cambridge, MA, pp. 101–120.

Pritchard, T. C., Hamilton, R. B., Morse, J. R. and Norgren, R. (1986). Projections of thalamic gustatory and lingual areas in the monkey, *Journal of Comparative Neurology* **244**: 213–228.

Pritchard, T. C., Hamilton, R. B. and Norgren, R. (1989). Neural coding of gustatory information in the thalamus of macaca mulatta, *Journal of Neurophysiology* **61**: 1–14.

Pryce, C. R., Azzinnari, D., Spinelli, S., Seifritz, E., Tegethoff, M. and Meinlschmidt, G. (2011). Helplessness: a systematic translational review of theory and evidence for its relevance to understanding and treating depression, *Pharmacol Ther* **132**: 242–267.

Quinlan, P. T. and Humphreys, G. W. (1987). Visual search for targets defined by combinations of color, shape, and

size: An examination of the task constraints on feature and conjunction searches, *Perception and Psychophysics* **41**: 455–472.

Quiroga, R. Q. (2012). Concept cells: the building blocks of declarative memory functions, *Nature Reviews Neuroscience* **13**: 587–597.

Quiroga, R. Q. (2013). Gnostic cells in the 21st century, *Acta Neurobiol Exp (Warsaw)* **73**: 463–471.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C. and Fried, I. (2005). Invariant visual representation by single neurons in the human brain, *Nature* **453**: 1102–1107.

Quiroga, R. Q., Kreiman, G., Koch, C. and Fried, I. (2008). Sparse but not 'grandmother-cell' coding in the medial temporal lobe, *Trends in Cognitive Sciences* **12**: 87–91.

Rabinovich, M., Huerta, R. and Laurent, G. (2008). Transient dynamics for neural processing, *Science* **321**: 48–50.

Rabinovich, M. I., Varona, P., Tristan, I. and Afraimovich, V. S. (2014). Chunking dynamics: heteroclinics in mind, *Frontiers in Computational Neuroscience* **8**: 22.

Rachlin, H. (1989). *Judgement, Decision, and Choice: A Cognitive/Behavioural Synthesis*, Freeman, New York.

Rada, P., Mark, G. P. and Hoebel, B. G. (1998). Dopamine in the nucleus accumbens released by hypothalamic stimulation-escape behavior, *Brain Research* **782**: 228–234.

Ranck, Jr., J. B. (1985). Head direction cells in the deep cell layer of dorsolateral presubiculum in freely moving rats, *in* G. Buzsáki and C. H. Vanderwolf (eds), *Electrical Activity of the Archicortex*, Akadémiai Kiadó, Budapest.

Rao, S. C., Rainer, G. and Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex, *Science* **276**: 821–824.

Ratcliff, R. and Rouder, J. F. (1998). Modeling response times for two-choice decisions, *Psychological Science* **9**: 347–356.

Ratcliff, R., Zandt, T. V. and McKoon, G. (1999). Connectionist and diffusion models of reaction time, *Psychological Reviews* **106**: 261–300.

Rauch, A., La Camera, G., Luescher, H.-R., Senn, W. and Fusi, S. (2003). Neocortical pyramidal cells respond as integrate-and-fire neurons to in vivo-like input currents, *Journal of Neurophysiology* **90**: 1598–1612.

Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language, *Frontiers in Evolutionary Neuroscience* **4**: 7.

Rauschecker, J. P. and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing, *Nature Neuroscience* **12**: 718–724.

Rawlins, J. N. P. (1985). Associations across time: the hippocampus as a temporary memory store, *Behavioral Brain Science* **8**: 479–496.

Redgrave, P., Prescott, T. J. and Gurney, K. (1999). Is the short-latency dopamine response too short to signal reward error?, *Trends in Neuroscience* **22**: 146–151.

Reisenzein, R. (1983). The Schachter theory of emotion: two decades later, *Psychological Bulletin* **94**: 239–264.

Renart, A., Parga, N. and Rolls, E. T. (1999a). Associative memory properties of multiple cortical modules, *Network* **10**: 237–255.

Renart, A., Parga, N. and Rolls, E. T. (1999b). Backprojections in the cerebral cortex: implications for memory storage, *Neural Computation* **11**: 1349–1388.

Renart, A., Parga, N. and Rolls, E. T. (2000). A recurrent model of the interaction between the prefrontal cortex and inferior temporal cortex in delay memory tasks, *in* S. Solla, T. Leen and K.-R. Mueller (eds), *Advances in Neural Information Processing Systems*, Vol. 12, MIT Press, Cambridge, MA, pp. 171–177.

Renart, A., Moreno, R., Rocha, J., Parga, N. and Rolls, E. T. (2001). A model of the IT–PF network in object working memory which includes balanced persistent activity and tuned inhibition, *Neurocomputing* **38–40**: 1525–1531.

Rensink, R. A. (2000). Seeing, sensing, and scrutinizing, *Vision Research* **40**: 1469–1487.

Rensink, R. A. (2014). Limits to the usability of iconic memory, *Frontiers in Psychology* **5**: 971.

Rescorla, R. A. and Wagner, A. R. (1972). A theory of Pavlovian conditioning: the effectiveness of reinforcement and non-reinforcement, *Classical Conditioning II: Current Research and Theory*, Appleton-Century-Crofts, New York, pp. 64–69.

Reynolds, J. and Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem, *Neuron* **24**: 19–29.

Reynolds, J. and Desimone, R. (2003). Interacting roles of attention and visual saliency in V4, *Neuron* **37**: 853–863.

Reynolds, J. H., Chelazzi, L. and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4, *Journal of Neuroscience* **19**: 1736–1753.

Reynolds, J. H., Pastemak, T. and Desimone, R. (2000). Attention increases sensitivity of V4 neurons, *Neuron* **26**: 703–714.

Rhodes, P. (1992). The open time of the NMDA channel facilitates the self-organisation of invariant object responses in cortex, *Society for Neuroscience Abstracts* **18**: 740.

Ridley, M. (1993). *The Red Queen: Sex and the Evolution of Human Nature*, Penguin, London.

Rieke, F., Warland, D. and Bialek, W. (1993). Coding efficiency and information rates in sensory neurons, *Europhysics Letters* **22**: 151–156.

Rieke, F., Warland, D., de Ruyter van Steveninck, R. R. and Bialek, W. (1997). *Spikes: Exploring the Neural Code*, MIT Press, Cambridge, MA.

Riesenhuber, M. and Poggio, T. (1998). Just one view: Invariances in inferotemporal cell tuning, *in* M. I. Jordan, M. J. Kearns and S. A. Solla (eds), *Advances in Neural Information Processing Systems*, Vol. 10, MIT Press, Cambridge, MA, pp. 215–221.

Riesenhuber, M. and Poggio, T. (1999a). Are cortical models really bound by the "binding problem"?, *Neuron* **24**: 87–93.

Riesenhuber, M. and Poggio, T. (1999b). Hierarchical models of object recognition in cortex, *Nature Neuroscience* **2**: 1019–1025.

Riesenhuber, M. and Poggio, T. (2000). Models of object recognition, *Nature Neuroscience Supplement* **3**: 1199–1204.

Riley, M. A. and Turvey, M. T. (2001). The self-organizing dynamics of intentions and actions, *American Journal of Psychology* **114**: 160–169.

Risold, P. Y. and Swanson, L. W. (1997). Connections of the rat lateral septal complex, *Brain Research Reviews* **24**: 115–195.

Robertson, R. G., Rolls, E. T. and Georges-François, P. (1998). Spatial view cells in the primate hippocampus: Effects of removal of view details, *Journal of Neurophysiology* **79**: 1145–1156.

Robertson, R. G., Rolls, E. T., Georges-François, P. and Panzeri, S. (1999). Head direction cells in the primate pre-subiculum, *Hippocampus* **9**: 206–219.

Robins, L. N., Helzer, J. E., Weissman, M. M., Orvaschel, H., Gruenberg, E., Burke, J. D. J. and Regier, D. A. (1984). Lifetime prevalence of specific psychiatric disorders in three sites, *Archives of General Psychiatry* **41**: 949–958.

Robinson, L. and Rolls, E. T. (2015). Invariant visual object recognition: biologically plausibile approaches, *Biological Cybernetics* **109**: 505–535.

Roelfsema, P. R., Lamme, V. A. and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey, *Nature* **395**: 376–381.

Roesch, M. R., Esber, G. R., Li, J., Daw, N. D. and Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce–Hall and Rescorla–Wagner coexist within the brain, *European Journal of Neuroscience* **35**: 1190–1200.

Rogers, J. L. and Kesner, R. P. (2003). Cholinergic modulation of the hippocampus during encoding and retrieval, *Neurobiology of Learning and Memory* **80**: 332–342.

Rogers, J. L. and Kesner, R. P. (2004). Cholinergic modulation of the hippocampus during encoding and retrieval of tone/shock-induced fear conditioning, *Learning and Memory* **11**: 102–107.

Rogers, J. L., Hunsaker, M. R. and Kesner, R. P. (2006). Effects of ventral and dorsal CA1 subregional lesions on trace fear conditioning, *Neurobiology of Learning and Memory* **86**: 72–81.

Rogers, R. D., Andrews, T. C., Grasby, P. M., Brooks, D. J. and Robbins, T. W. (2000). Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans, *Journal of Cognitive Neuroscience* **12**: 142–162.

Roitman, J. D. and Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task, *Journal of Neuroscience* **22**: 9475–9489.

Roland, P. E. and Friberg, L. (1985). Localization of cortical areas activated by thinking, *Journal of Neurophysiology* **53**: 1219–1243.

Rolls, E. T. (1975). *The Brain and Reward*, Pergamon Press, Oxford.

Rolls, E. T. (1981a). Central nervous mechanisms related to feeding and appetite, *British Medical Bulletin* **37**: 131–134.

Rolls, E. T. (1981b). Processing beyond the inferior temporal visual cortex related to feeding, learning, and striatal function, *in* Y. Katsuki, R. Norgren and M. Sato (eds), *Brain Mechanisms of Sensation*, Wiley, New York, chapter 16, pp. 241–269.

Rolls, E. T. (1982). Neuronal mechanisms underlying the formation and disconnection of associations between visual stimuli and reinforcement in primates, *in* C. D. Woody (ed.), *Conditioning: Representation of Involved Neural Functions*, Plenum, New York, pp. 363–373.

Rolls, E. T. (1986a). Neural systems involved in emotion in primates, *in* R. Plutchik and H. Kellerman (eds), *Emotion: Theory, Research, and Experience*, Vol. 3: Biological Foundations of Emotion, Academic Press, New York, chapter 5, pp. 125–143.

Rolls, E. T. (1986b). Neuronal activity related to the control of feeding, *in* R. Ritter, S. Ritter and C. Barnes (eds), *Feeding Behavior: Neural and Humoral Controls*, Academic Press, New York, chapter 6, pp. 163–190.

Rolls, E. T. (1986c). A theory of emotion, and its application to understanding the neural basis of emotion, *in* Y. Oomura (ed.), *Emotions. Neural and Chemical Control*, Japan Scientific Societies Press; and Karger, Tokyo; and Basel, pp. 325–344.

Rolls, E. T. (1987a). Information representation, processing and storage in the brain: analysis at the single neuron level, *in* J.-P. Changeux and M. Konishi (eds), *The Neural and Molecular Bases of Learning*, Wiley, Chichester, pp. 503–540.

Rolls, E. T. (1987b). Information representation, processing and storage in the brain: analysis at the single neuron level, *in* J. P. Changeux and M. Konishi (eds), *The Neural and Molecular Bases of Learning*, Wiley, Chichester, pp. 503–540.

Rolls, E. T. (1989a). Functions of neuronal networks in the hippocampus and cerebral cortex in memory, *in* R. Cotterill

(ed.), *Models of Brain Function*, Cambridge University Press, Cambridge, pp. 15–33.

Rolls, E. T. (1989b). Functions of neuronal networks in the hippocampus and neocortex in memory, *in* J. H. Byrne and W. O. Berry (eds), *Neural Models of Plasticity: Experimental and Theoretical Approaches*, Academic Press, San Diego, CA, chapter 13, pp. 240–265.

Rolls, E. T. (1989c). Information processing and basal ganglia function, *in* C. Kennard and M. Swash (eds), *Hierarchies in Neurology*, Springer-Verlag, London, chapter 15, pp. 123–142.

Rolls, E. T. (1989d). Information processing in the taste system of primates, *Journal of Experimental Biology* **146**: 141–164.

Rolls, E. T. (1989e). Parallel distributed processing in the brain: implications of the functional architecture of neuronal networks in the hippocampus, *in* R. G. M. Morris (ed.), *Parallel Distributed Processing: Implications for Psychology and Neurobiology*, Oxford University Press, Oxford, chapter 12, pp. 286–308.

Rolls, E. T. (1989f). The representation and storage of information in neuronal networks in the primate cerebral cortex and hippocampus, *in* R. Durbin, C. Miall and G. Mitchison (eds), *The Computing Neuron*, Addison-Wesley, Wokingham, England, chapter 8, pp. 125–159.

Rolls, E. T. (1990a). Functions of the primate hippocampus in spatial processing and memory, *in* D. S. Olton and R. P. Kesner (eds), *Neurobiology of Comparative Cognition*, L. Erlbaum, Hillsdale, NJ, chapter 12, pp. 339–362.

Rolls, E. T. (1990b). Theoretical and neurophysiological analysis of the functions of the primate hippocampus in memory, *Cold Spring Harbor Symposia in Quantitative Biology* **55**: 995–1006.

Rolls, E. T. (1990c). A theory of emotion, and its application to understanding the neural basis of emotion, *Cognition and Emotion* **4**: 161–190.

Rolls, E. T. (1992a). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas, *Philosophical Transactions of the Royal Society* **335**: 11–21.

Rolls, E. T. (1992b). Neurophysiology and functions of the primate amygdala, *in* J. P. Aggleton (ed.), *The Amygdala*, Wiley-Liss, New York, chapter 5, pp. 143–165.

Rolls, E. T. (1993). The neural control of feeding in primates, *in* D. Booth (ed.), *Neurophysiology of Ingestion*, Pergamon, Oxford, chapter 9, pp. 137–169.

Rolls, E. T. (1994a). Brain mechanisms for invariant visual recognition and learning, *Behavioural Processes* **33**: 113–138.

Rolls, E. T. (1994b). Neurophysiological and neuronal network analysis of how the primate hippocampus functions in memory, *in* J. Delacour (ed.), *The Memory System of the Brain*, World Scientific, London, pp. 713–744.

Rolls, E. T. (1995a). Learning mechanisms in the temporal lobe visual cortex, *Behavioural Brain Research* **66**: 177–185.

Rolls, E. T. (1995b). A model of the operation of the hippocampus and entorhinal cortex in memory, *International Journal of Neural Systems* **6, Supplement**: 51–70.

Rolls, E. T. (1995c). A theory of emotion and consciousness, and its application to understanding the neural basis of emotion, *in* M. S. Gazzaniga (ed.), *The Cognitive Neurosciences*, MIT Press, Cambridge, MA, chapter 72, pp. 1091–1106.

Rolls, E. T. (1996a). The orbitofrontal cortex, *Philosophical Transactions of the Royal Society B* **351**: 1433–1444.

Rolls, E. T. (1996b). Roles of long term potentiation and long term depression in neuronal network operations in the brain, *in* M. S. Fazeli and G. L. Collingridge (eds), *Cortical Plasticity: LTP and LTD*, Bios, Oxford, chapter 11, pp. 223–250.

Rolls, E. T. (1996c). A theory of hippocampal function in memory, *Hippocampus* **6**: 601–620.

Rolls, E. T. (1997a). Brain mechanisms of vision, memory, and consciousness, *in* M. Ito, Y. Miyashita and E. Rolls (eds), *Cognition, Computation, and Consciousness*, Oxford University Press, Oxford, chapter 6, pp. 81–120.

Rolls, E. T. (1997b). Consciousness in neural networks?, *Neural Networks* **10**: 1227–1240.

Rolls, E. T. (1997c). A neurophysiological and computational approach to the functions of the temporal lobe cortical visual areas in invariant object recognition, *in* M. Jenkin and L. Harris (eds), *Computational and Psychophysical Mechanisms of Visual Coding*, Cambridge University Press, Cambridge, chapter 9, pp. 184–220.

Rolls, E. T. (1997d). Taste and olfactory processing in the brain and its relation to the control of eating, *Critical Reviews in Neurobiology* **11**: 263–287.

Rolls, E. T. (1999a). *The Brain and Emotion*, Oxford University Press, Oxford.

Rolls, E. T. (1999b). The functions of the orbitofrontal cortex, *Neurocase* **5**: 301–312.

Rolls, E. T. (1999c). Spatial view cells and the representation of place in the primate hippocampus, *Hippocampus* **9**: 467–480.

Rolls, E. T. (2000a). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition, *Neuron* **27**: 205–218.

Rolls, E. T. (2000b). Memory systems in the brain, *Annual Review of Psychology* **51**: 599–630.

Rolls, E. T. (2000c). Neurophysiology and functions of the primate amygdala, and the neural basis of emotion, *in* J. P. Aggleton (ed.), *The Amygdala: Second Edition. A Functional Analysis*, Oxford University Press, Oxford, chapter 13, pp. 447–478.

Rolls, E. T. (2000d). The orbitofrontal cortex and reward, *Cerebral Cortex* **10**: 284–294.

Rolls, E. T. (2001). The rules of formation of the olfactory representations found in the orbitofrontal cortex olfactory

areas in primates, *Chemical Senses* **26**: 595–604.

Rolls, E. T. (2003). Consciousness absent and present: a neurophysiological exploration, *Progress in Brain Research* **144**: 95–106.

Rolls, E. T. (2004a). The functions of the orbitofrontal cortex, *Brain and Cognition* **55**: 11–29.

Rolls, E. T. (2004b). A higher order syntactic thought (HOST) theory of consciousness, *in* R. J. Gennaro (ed.), *Higher Order Theories of Consciousness*, John Benjamins, Amsterdam, chapter 7, pp. 137–172.

Rolls, E. T. (2005). *Emotion Explained*, Oxford University Press, Oxford.

Rolls, E. T. (2006a). Consciousness absent and present: a neurophysiological exploration of masking, *in* H. Ogmen and B. G. Breitmeyer (eds), *The First Half Second*, MIT Press, Cambridge, MA, chapter 6, pp. 89–108.

Rolls, E. T. (2006b). The neurophysiology and functions of the orbitofrontal cortex, *in* D. H. Zald and S. L. Rauch (eds), *The Orbitofrontal Cortex*, Oxford University Press, Oxford, chapter 5, pp. 95–124.

Rolls, E. T. (2007a). The affective neuroscience of consciousness: higher order syntactic thoughts, dual routes to emotion and action, and consciousness, *in* P. D. Zelazo, M. Moscovitch and E. Thompson (eds), *Cambridge Handbook of Consciousness*, Cambridge University Press, New York, chapter 29, pp. 831–859.

Rolls, E. T. (2007b). A computational neuroscience approach to consciousness, *Neural Networks* **20**: 962–982.

Rolls, E. T. (2007c). Invariant representations of objects in natural scenes in the temporal cortex visual areas, *in* S. Funahashi (ed.), *Representation and Brain*, Springer, Tokyo, chapter 3, pp. 47–102.

Rolls, E. T. (2007d). Memory systems: multiple systems in the brain and their interactions, *in* H. L. Roediger, Y. Dudai and S. M. Fitzpatrick (eds), *Science of Memory: Concepts*, Oxford University Press, New York, chapter 59, pp. 345–351.

Rolls, E. T. (2007e). The representation of information about faces in the temporal and frontal lobes of primates including humans, *Neuropsychologia* **45**: 124–143.

Rolls, E. T. (2007f). Sensory processing in the brain related to the control of food intake, *Proceedings of the Nutrition Society* **66**: 96–112.

Rolls, E. T. (2008a). Emotion, higher order syntactic thoughts, and consciousness, *in* L. Weiskrantz and M. Davies (eds), *Frontiers of Consciousness*, Oxford University Press, Oxford, chapter 4, pp. 131–167.

Rolls, E. T. (2008b). Face representations in different brain areas, and critical band masking, *Journal of Neuropsychology* **2**: 325–360.

Rolls, E. T. (2008c). Functions of the orbitofrontal and pregenual cingulate cortex in taste, olfaction, appetite and emotion, *Acta Physiologica Hungarica* **95**: 131–164.

Rolls, E. T. (2008d). *Memory, Attention, and Decision-Making. A Unifying Computational Neuroscience Approach*, Oxford University Press, Oxford.

Rolls, E. T. (2008e). The primate hippocampus and episodic memory, *in* E. Dere, A. Easton, L. Nadel and J. P. Huston (eds), *Handbook of Episodic Memory*, Elsevier, Amsterdam, chapter 4.2, pp. 417–438.

Rolls, E. T. (2008f). Top-down control of visual perception: attention in natural vision, *Perception* **37**: 333–354.

Rolls, E. T. (2009a). The anterior and midcingulate cortices and reward, *in* B. Vogt (ed.), *Cingulate Neurobiology and Disease*, Oxford University Press, Oxford, chapter 8, pp. 191–206.

Rolls, E. T. (2009b). Functional neuroimaging of umami taste: what makes umami pleasant, *American Journal of Clinical Nutrition* **90**: 803S–814S.

Rolls, E. T. (2009c). The neurophysiology and computational mechanisms of object representation, *in* S. Dickinson, M. Tarr, A. Leonardis and B. Schiele (eds), *Object Categorization: Computer and Human Vision Perspectives*, Cambridge University Press, Cambridge, chapter 14, pp. 257–287.

Rolls, E. T. (2010a). The affective and cognitive processing of touch, oral texture, and temperature in the brain, *Neuroscience and Biobehavioral Reviews* **34**: 237–245.

Rolls, E. T. (2010b). A computational theory of episodic memory formation in the hippocampus, *Behavioural Brain Research* **215**: 180–196.

Rolls, E. T. (2010c). Noise in the brain, decision-making, determinism, free will, and consciousness, *in* E. Perry, D. Collerton, F. LeBeau and H. Ashton (eds), *New Horizons in the Neuroscience of Consciousness*, John Benjamins, Amsterdam, pp. 113–120.

Rolls, E. T. (2011a). Chemosensory learning in the cortex, *Frontiers in Systems Neuroscience* **5**: 78 (1–13).

Rolls, E. T. (2011b). Consciousness, decision-making, and neural computation, *in* V. Cutsuridis, A. Hussain and J. G. Taylor (eds), *Perception-Action Cycle: Models, architecture, and hardware*, Springer, Berlin, chapter 9, pp. 287–333.

Rolls, E. T. (2011c). David Marr's Vision: floreat computational neuroscience, *Brain* **134**: 913–916.

Rolls, E. T. (2011d). Face neurons, *in* A. J. Calder, G. Rhodes, M. H. Johnson and J. V. Haxby (eds), *The Oxford Handbook of Face Perception*, Oxford University Press, Oxford, chapter 4, pp. 51–75.

Rolls, E. T. (2011e). Taste, olfactory, and food texture reward processing in the brain and obesity, *International Journal of Obesity* **35**: 550–561.

Rolls, E. T. (2012a). Advantages of dilution in the connectivity of attractor networks in the brain, *Biologically Inspired Cognitive Architectures* **1**: 44–54.

Rolls, E. T. (2012b). Glutamate, obsessive-compulsive disorder, schizophrenia, and the stability of cortical attractor neuronal networks, *Pharmacology, Biochemistry and Behavior* **100**: 736–751.

Rolls, E. T. (2012c). Invariant visual object and face recognition: neural and computational bases, and a model, VisNet, *Frontiers in Computational Neuroscience* **6**(35): 1–70.

Rolls, E. T. (2012d). *Neuroculture: On the Implications of Brain Science*, Oxford University Press, Oxford.

Rolls, E. T. (2012e). Taste, olfactory, and food texture reward processing in the brain and the control of appetite, *Proceedings of the Nutrition Society* **71**: 488–501.

Rolls, E. T. (2012f). Willed action, free will, and the stochastic neurodynamics of decision-making, *Frontiers in Integrative Neuroscience* **6**: 68.

Rolls, E. T. (2013a). A biased activation theory of the cognitive and attentional modulation of emotion, *Frontiers in Human Neuroscience* **7**: 74.

Rolls, E. T. (2013b). The mechanisms for pattern completion and pattern separation in the hippocampus, *Frontiers in Systems Neuroscience* **7**: 74.

Rolls, E. T. (2013c). On the relation between the mind and the brain: a neuroscience perspective, *Philosophia Scientiae* **17**: 31–70.

Rolls, E. T. (2013d). A quantitative theory of the functions of the hippocampal CA3 network in memory, *Frontiers in Cellular Neuroscience* **7**: 98.

Rolls, E. T. (2013e). What are emotional states, and why do we have them?, *Emotion Review* **5**: 241–247.

Rolls, E. T. (2014a). *Emotion and Decision-Making Explained*, Oxford University Press, Oxford.

Rolls, E. T. (2014b). Emotion and decision-making explained: Précis, *Cortex* **59**: 185–193.

Rolls, E. T. (2014c). Neuroculture: art, aesthetics, and the brain, *Rendiconti Lincei Scienze Fisiche e Naturali* **25**: 291–307.

Rolls, E. T. (2015a). Central neural integration of taste, smell and other sensory modalities, *in* R. L. Doty (ed.), *Handbook of Olfaction and Gustation*, third edn, Wiley, New York, chapter 44, pp. 1027–1048.

Rolls, E. T. (2015b). Diluted connectivity in pattern association networks facilitates the recall of information from the hippocampus to the neocortex, *Progress in Brain Research* **219**: 21–43.

Rolls, E. T. (2015c). Functions of the anterior insula in taste, autonomic, and related functions, *Brain and Cognition* p. doi: 10.1016/j.bandc.2015.07.002.

Rolls, E. T. (2015d). Limbic systems for emotion and for memory, but no single limbic system, *Cortex* **62**: 119–157.

Rolls, E. T. (2015e). The neuronal representation of information in the human brain. Review, *Brain* **138**: 3459–3462.

Rolls, E. T. (2015f). Taste, olfactory, and food reward value processing in the brain, *Progress in Neurobiology* **127–128**: 64–90.

Rolls, E. T. (2016a). *Cerebral Cortex: Principles of Operation*, Oxford University Press, Oxford.

Rolls, E. T. (2016b). Cortical coding, *Language, Cognition and Neuroscience* p. doi: 10.1080/23273798.2016.1203443.

Rolls, E. T. (2016c). Motivation Explained: Ultimate and proximate accounts of hunger and appetite, *Advances in Motivation Science* **3**: 187–249.

Rolls, E. T. (2016d). Neurobiological foundations of aesthetics and art, *New Ideas in Psychology* p. in press.

Rolls, E. T. (2016e). A non-reward attractor theory of depression, *Neuroscience and Biobehavioral Reviews* **68**: 47–58.

Rolls, E. T. (2016f). Pattern separation, completion, and categorisation in the hippocampus and neocortex, *Neurobiology of Learning and Memory* **129**: 4–28.

Rolls, E. T. (2016g). Reward systems in the brain and nutrition, *Annual Review of Nutrition* **36**: 14.1–14.36.

Rolls, E. T. (2016h). A scientific theory of *ars memoriae*: spatial view cells in a continuous attractor network with linked items.

Rolls, E. T. and Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey, *Experimental Brain Research* **65**: 38–48.

Rolls, E. T. and Baylis, L. L. (1994). Gustatory, olfactory and visual convergence within the primate orbitofrontal cortex, *Journal of Neuroscience* **14**: 5437–5452.

Rolls, E. T. and Deco, G. (2002). *Computational Neuroscience of Vision*, Oxford University Press, Oxford.

Rolls, E. T. and Deco, G. (2006). Attention in natural scenes: neurophysiological and computational bases, *Neural Networks* **19**: 1383–1394.

Rolls, E. T. and Deco, G. (2010). *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*, Oxford University Press, Oxford.

Rolls, E. T. and Deco, G. (2011a). A computational neuroscience approach to schizophrenia and its onset, *Neuroscience and Biobehavioral Reviews* **35**: 1644–1653.

Rolls, E. T. and Deco, G. (2011b). Prediction of decisions from noise in the brain before the evidence is provided, *Frontiers in Neuroscience* **5**: 33.

Rolls, E. T. and Deco, G. (2015a). Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain, *Brain Research* **1621**: 316–334.

Rolls, E. T. and Deco, G. (2015b). A stochastic neurodynamics approach to the changes in cognition and memory in aging, *Neurobiology of Learning and Memory* **118**: 150–161.

Rolls, E. T. and Deco, G. (2016). Non-reward neural mechanisms in the orbitofrontal cortex, *Cortex* p. doi:10.1016/j.cortex.2016.06.023.

Rolls, E. T. and Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: from affect to decision-making,

Progress in Neurobiology **86**: 216–244.

Rolls, E. T. and Johnstone, S. (1992). Neurophysiological analysis of striatal function, *in* G. Vallar, S. Cappa and C. Wallesch (eds), *Neuropsychological Disorders Associated with Subcortical Lesions*, Oxford University Press, Oxford, chapter 3, pp. 61–97.

Rolls, E. T. and Kesner, R. P. (2006). A theory of hippocampal function, and tests of the theory, *Progress in Neurobiology* **79**: 1–48.

Rolls, E. T. and Kesner, R. P. (2016). Pattern separation and pattern completion in the hippocampal system, *Neurobiology of Learning and Memory* **129**: 1–3.

Rolls, E. T. and McCabe, C. (2007). Enhanced affective brain representations of chocolate in cravers vs non-cravers, *European Journal of Neuroscience* **26**: 1067–1076.

Rolls, E. T. and Milward, T. (2000). A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures, *Neural Computation* **12**: 2547–2572.

Rolls, E. T. and O'Mara, S. (1993). Neurophysiological and theoretical analysis of how the hippocampus functions in memory, *in* T. Ono, L. Squire, M. Raichle, D. Perrett and M. Fukuda (eds), *Brain Mechanisms of Perception and Memory: From Neuron to Behavior*, Oxford University Press, New York, chapter 17, pp. 276–300.

Rolls, E. T. and O'Mara, S. M. (1995). View-responsive neurons in the primate hippocampal complex, *Hippocampus* **5**: 409–424.

Rolls, E. T. and Rolls, B. J. (1973). Altered food preferences after lesions in the basolateral region of the amygdala in the rat, *Journal of Comparative and Physiological Psychology* **83**: 248–259.

Rolls, E. T. and Rolls, J. H. (1997). Olfactory sensory-specific satiety in humans, *Physiology and Behavior* **61**: 461–473.

Rolls, E. T. and Scott, T. R. (2003). Central taste anatomy and neurophysiology, *in* R. Doty (ed.), *Handbook of Olfaction and Gustation*, 2nd edn, Dekker, New York, chapter 33, pp. 679–705.

Rolls, E. T. and Stringer, S. M. (2000). On the design of neural networks in the brain by genetic evolution, *Progress in Neurobiology* **61**: 557–579.

Rolls, E. T. and Stringer, S. M. (2001a). Invariant object recognition in the visual system with error correction and temporal difference learning, *Network: Computation in Neural Systems* **12**: 111–129.

Rolls, E. T. and Stringer, S. M. (2001b). A model of the interaction between mood and memory, *Network: Computation in Neural Systems* **12**: 89–109.

Rolls, E. T. and Stringer, S. M. (2005). Spatial view cells in the hippocampus, and their idiothetic update based on place and head direction, *Neural Networks* **18**: 1229–1241.

Rolls, E. T. and Stringer, S. M. (2006). Invariant visual object recognition: a model, with lighting invariance, *Journal of Physiology – Paris* **100**: 43–62.

Rolls, E. T. and Stringer, S. M. (2007). Invariant global motion recognition in the dorsal visual system: a unifying theory, *Neural Computation* **19**: 139–169.

Rolls, E. T. and Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking, *Proceedings of the Royal Society, B* **257**: 9–15.

Rolls, E. T. and Tovee, M. J. (1995a). The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the visual field, *Experimental Brain Research* **103**: 409–420.

Rolls, E. T. and Tovee, M. J. (1995b). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex, *Journal of Neurophysiology* **73**: 713–726.

Rolls, E. T. and Treves, A. (1990). The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain, *Network* **1**: 407–421.

Rolls, E. T. and Treves, A. (1998). *Neural Networks and Brain Function*, Oxford University Press, Oxford.

Rolls, E. T. and Treves, A. (2011). The neuronal encoding of information in the brain, *Progress in Neurobiology* **95**: 448–490.

Rolls, E. T. and Webb, T. J. (2012). Cortical attractor network dynamics with diluted connectivity, *Brain Research* **1434**: 212–225.

Rolls, E. T. and Webb, T. J. (2014). Finding and recognising objects in natural scenes: complementary computations in the dorsal and ventral visual systems, *Frontiers in Computational Neuroscience* **8**: 85.

Rolls, E. T. and Williams, G. V. (1987a). Neuronal activity in the ventral striatum of the primate, *in* M. B. Carpenter and A. Jayamaran (eds), *The Basal Ganglia II – Structure and Function – Current Concepts*, Plenum, New York, pp. 349–356.

Rolls, E. T. and Williams, G. V. (1987b). Sensory and movement-related neuronal activity in different regions of the primate striatum, *in* J. S. Schneider and T. I. Lidsky (eds), *Basal Ganglia and Behavior: Sensory Aspects and Motor Functioning*, Hans Huber, Bern, pp. 37–59.

Rolls, E. T. and Xiang, J.-Z. (2005). Reward–spatial view representations and learning in the primate hippocampus, *Journal of Neuroscience* **25**: 6167–6174.

Rolls, E. T. and Xiang, J.-Z. (2006). Spatial view cells in the primate hippocampus, and memory recall, *Reviews in the Neurosciences* **17**: 175–200.

Rolls, E. T., Burton, M. J. and Mora, F. (1976). Hypothalamic neuronal responses associated with the sight of food,

Brain Research **111**: 53–66.

Rolls, E. T., Judge, S. J. and Sanghera, M. (1977). Activity of neurones in the inferotemporal cortex of the alert monkey, *Brain Research* **130**: 229–238.

Rolls, E. T., Sanghera, M. K. and Roper-Hall, A. (1979). The latency of activation of neurons in the lateral hypothalamus and substantia innominata during feeding in the monkey, *Brain Research* **164**: 121–135.

Rolls, E. T., Burton, M. J. and Mora, F. (1980). Neurophysiological analysis of brain-stimulation reward in the monkey, *Brain Research* **194**: 339–357.

Rolls, E. T., Perrett, D. I., Caan, A. W. and Wilson, F. A. W. (1982). Neuronal responses related to visual recognition, *Brain* **105**: 611–646.

Rolls, E. T., Rolls, B. J. and Rowe, E. A. (1983a). Sensory-specific and motivation-specific satiety for the sight and taste of food and water in man, *Physiology and Behavior* **30**: 185–192.

Rolls, E. T., Thorpe, S. J. and Maddison, S. P. (1983b). Responses of striatal neurons in the behaving monkey. 1. Head of the caudate nucleus, *Behavioural Brain Research* **7**: 179–210.

Rolls, E. T., Thorpe, S. J., Boytim, M., Szabo, I. and Perrett, D. I. (1984). Responses of striatal neurons in the behaving monkey. 3. Effects of iontophoretically applied dopamine on normal responsiveness, *Neuroscience* **12**: 1201–1212.

Rolls, E. T., Baylis, G. C. and Leonard, C. M. (1985). Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus, *Vision Research* **25**: 1021–1035.

Rolls, E. T., Baylis, G. C. and Hasselmo, M. E. (1987). The responses of neurons in the cortex in the superior temporal sulcus of the monkey to band-pass spatial frequency filtered faces, *Vision Research* **27**: 311–326.

Rolls, E. T., Scott, T. R., Sienkiewicz, Z. J. and Yaxley, S. (1988). The responsiveness of neurones in the frontal opercular gustatory cortex of the macaque monkey is independent of hunger, *Journal of Physiology* **397**: 1–12.

Rolls, E. T., Baylis, G. C., Hasselmo, M. and Nalwa, V. (1989a). The representation of information in the temporal lobe visual cortical areas of macaque monkeys, *in* J. Kulikowski, C. Dickinson and I. Murray (eds), *Seeing Contour and Colour*, Pergamon, Oxford.

Rolls, E. T., Miyashita, Y., Cahusac, P. M. B., Kesner, R. P., Niki, H., Feigenbaum, J. and Bach, L. (1989b). Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown, *Journal of Neuroscience* **9**: 1835–1845.

Rolls, E. T., Sienkiewicz, Z. J. and Yaxley, S. (1989c). Hunger modulates the responses to gustatory stimuli of single neurons in the caudolateral orbitofrontal cortex of the macaque monkey, *European Journal of Neuroscience* **1**: 53–60.

Rolls, E. T., Yaxley, S. and Sienkiewicz, Z. J. (1990). Gustatory responses of single neurons in the orbitofrontal cortex of the macaque monkey, *Journal of Neurophysiology* **64**: 1055–1066.

Rolls, E. T., Cahusac, P. M. B., Feigenbaum, J. D. and Miyashita, Y. (1993). Responses of single neurons in the hippocampus of the macaque related to recognition memory, *Experimental Brain Research* **93**: 299–306.

Rolls, E. T., Hornak, J., Wade, D. and McGrath, J. (1994a). Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage, *Journal of Neurology, Neurosurgery and Psychiatry* **57**: 1518–1524.

Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L. and Azzopardi, P. (1994b). The responses of neurons in the temporal cortex of primates, and face identification and detection, *Experimental Brain Research* **101**: 474–484.

Rolls, E. T., Critchley, H. D. and Treves, A. (1996a). The representation of olfactory information in the primate orbitofrontal cortex, *Journal of Neurophysiology* **75**: 1982–1996.

Rolls, E. T., Critchley, H. D., Mason, R. and Wakeman, E. A. (1996b). Orbitofrontal cortex neurons: role in olfactory and visual association learning, *Journal of Neurophysiology* **75**: 1970–1981.

Rolls, E. T., Critchley, H. D., Wakeman, E. A. and Mason, R. (1996c). Responses of neurons in the primate taste cortex to the glutamate ion and to inosine 5′-monophosphate, *Physiology and Behavior* **59**: 991–1000.

Rolls, E. T., Robertson, R. G. and Georges-François, P. (1997a). Spatial view cells in the primate hippocampus, *European Journal of Neuroscience* **9**: 1789–1794.

Rolls, E. T., Treves, A. and Tovee, M. J. (1997b). The representational capacity of the distributed encoding of information provided by populations of neurons in the primate temporal visual cortex, *Experimental Brain Research* **114**: 149–162.

Rolls, E. T., Treves, A., Foster, D. and Perez-Vicente, C. (1997c). Simulation studies of the CA3 hippocampal subfield modelled as an attractor neural network, *Neural Networks* **10**: 1559–1569.

Rolls, E. T., Treves, A., Tovee, M. and Panzeri, S. (1997d). Information in the neuronal representation of individual stimuli in the primate temporal visual cortex, *Journal of Computational Neuroscience* **4**: 309–333.

Rolls, E. T., Critchley, H. D., Browning, A. and Hernadi, I. (1998a). The neurophysiology of taste and olfaction in primates, and umami flavor, *Annals of the New York Academy of Sciences* **855**: 426–437.

Rolls, E. T., Treves, A., Robertson, R. G., Georges-François, P. and Panzeri, S. (1998b). Information about spatial view in an ensemble of primate hippocampal cells, *Journal of Neurophysiology* **79**: 1797–1813.

Rolls, E. T., Critchley, H. D., Browning, A. S., Hernadi, A. and Lenard, L. (1999a). Responses to the sensory properties of fat of neurons in the primate orbitofrontal cortex, *Journal of Neuroscience* **19**: 1532–1540.

Rolls, E. T., Tovee, M. J. and Panzeri, S. (1999b). The neurophysiology of backward visual masking: information

analysis, *Journal of Cognitive Neuroscience* **11**: 335–346.

Rolls, E. T., Stringer, S. M. and Trappenberg, T. P. (2002). A unified model of spatial and episodic memory, *Proceedings of The Royal Society B* **269**: 1087–1093.

Rolls, E. T., Aggelopoulos, N. C. and Zheng, F. (2003a). The receptive fields of inferior temporal cortex neurons in natural scenes, *Journal of Neuroscience* **23**: 339–348.

Rolls, E. T., Franco, L., Aggelopoulos, N. C. and Reece, S. (2003b). An information theoretic approach to the contributions of the firing rates and the correlations between the firing of neurons, *Journal of Neurophysiology* **89**: 2810–2822.

Rolls, E. T., Kringelbach, M. L. and De Araujo, I. E. T. (2003c). Different representations of pleasant and unpleasant odours in the human brain, *European Journal of Neuroscience* **18**: 695–703.

Rolls, E. T., O'Doherty, J., Kringelbach, M. L., Francis, S., Bowtell, R. and McGlone, F. (2003d). Representations of pleasant and painful touch in the human orbitofrontal and cingulate cortices, *Cerebral Cortex* **13**: 308–317.

Rolls, E. T., Verhagen, J. V. and Kadohisa, M. (2003e). Representations of the texture of food in the primate orbitofrontal cortex: neurons responding to viscosity, grittiness, and capsaicin, *Journal of Neurophysiology* **90**: 3711–3724.

Rolls, E. T., Aggelopoulos, N. C., Franco, L. and Treves, A. (2004). Information encoding in the inferior temporal visual cortex: contributions of the firing rates and the correlations between the firing of neurons, *Biological Cybernetics* **90**: 19–32.

Rolls, E. T., Browning, A. S., Inoue, K. and Hernadi, S. (2005a). Novel visual stimuli activate a population of neurons in the primate orbitofrontal cortex, *Neurobiology of Learning and Memory* **84**: 111–123.

Rolls, E. T., Franco, L. and Stringer, S. M. (2005b). The perirhinal cortex and long-term familiarity memory, *Quarterly Journal of Experimental Psychology B* **58**: 234–245.

Rolls, E. T., Xiang, J.-Z. and Franco, L. (2005c). Object, space and object-space representations in the primate hippocampus, *Journal of Neurophysiology* **94**: 833–844.

Rolls, E. T., Critchley, H. D., Browning, A. S. and Inoue, K. (2006a). Face-selective and auditory neurons in the primate orbitofrontal cortex, *Experimental Brain Research* **170**: 74–87.

Rolls, E. T., Franco, L., Aggelopoulos, N. C. and Jerez, J. M. (2006b). Information in the first spike, the order of spikes, and the number of spikes provided by neurons in the inferior temporal visual cortex, *Vision Research* **46**: 4193–4205.

Rolls, E. T., Stringer, S. M. and Elliot, T. (2006c). Entorhinal cortex grid cells can map to hippocampal place cells by competitive learning, *Network: Computation in Neural Systems* **17**: 447–465.

Rolls, E. T., Grabenhorst, F. and Parris, B. (2008a). Warm pleasant feelings in the brain, *Neuroimage* **41**: 1504–1513.

Rolls, E. T., Grabenhorst, F., Margot, C., da Silva, M. and Velazco, M. I. (2008b). Selective attention to affective value alters how the brain processes olfactory stimuli, *Journal of Cognitive Neuroscience* **20**: 1815–1826.

Rolls, E. T., Loh, M. and Deco, G. (2008c). An attractor hypothesis of obsessive-compulsive disorder, *European Journal of Neuroscience* **28**: 782–793.

Rolls, E. T., Loh, M., Deco, G. and Winterer, G. (2008d). Computational models of schizophrenia and dopamine modulation in the prefrontal cortex, *Nature Reviews Neuroscience* **9**: 696–709.

Rolls, E. T., McCabe, C. and Redoute, J. (2008e). Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task, *Cerebral Cortex* **18**: 652–663.

Rolls, E. T., Tromans, J. M. and Stringer, S. M. (2008f). Spatial scene representations formed by self-organizing learning in a hippocampal extension of the ventral visual system, *European Journal of Neuroscience* **28**: 2116–2127.

Rolls, E. T., Grabenhorst, F. and Franco, L. (2009). Prediction of subjective affective state from brain activations, *Journal of Neurophysiology* **101**: 1294–1308.

Rolls, E. T., Critchley, H., Verhagen, J. V. and Kadohisa, M. (2010a). The representation of information about taste and odor in the primate orbitofrontal cortex, *Chemosensory Perception* **3**: 16–33.

Rolls, E. T., Grabenhorst, F. and Deco, G. (2010b). Choice, difficulty, and confidence in the brain, *Neuroimage* **53**: 694–706.

Rolls, E. T., Grabenhorst, F. and Deco, G. (2010c). Decision-making, errors, and confidence in the brain, *Journal of Neurophysiology* **104**: 2359–2374.

Rolls, E. T., Grabenhorst, F. and Parris, B. A. (2010d). Neural systems underlying decisions about affective odors, *Journal of Cognitive Neuroscience* **10**: 1068–1082.

Rolls, E. T., Webb, T. J. and Deco, G. (2012). Communication before coherence, *European Journal of Neuroscience* **36**: 2689–2709.

Rolls, E. T., Dempere-Marco, L. and Deco, G. (2013). Holding multiple items in short term memory: a neural mechanism, *PLoS One* **8**: e61078.

Rolls, E. T., Cheng, W., Qiu, J., Liu, W., Tang, Y., Huang, C.-C., Wang, X., Zhang, J., Lin, W., Zheng, L., Pu, J., Tsai, S.-J., Yang, A. C., Lin, C.-P., Wang, F., Xie, P. and Feng, J. (2016a). Medial reward and lateral non-reward orbitofrontal cortex circuits change in opposite directions in depression, *Federation of European Neuroscience Societies, Abstracts 2016* p. 3147.

Rolls, E. T., Lu, W., Wan, L., Liddle, P., Palaniyappan, L., Li, M., Guo, S., Zhou, Q., Wang, Q., Ma, X., Deng, W.,

Zhao, L., Guo, W., Deng, W., Li, T. and Feng, J. (2016b). Individual differences in schizophrenia.

Romanski, L. M., Averbeck, B. B. and Diltz, M. (2005). Neural representation of vocalizations in the primate ventrolateral prefrontal cortex, *Journal of Neurophysiology* **93**: 734–747.

Romo, R. and Salinas, E. (2003). Flutter discrimination: Neural codes, perception, memory and decision making, *Nature Reviews Neuroscience* **4**: 203–218.

Romo, R., Hernandez, A. and Zainos, A. (2004). Neuronal correlates of a perceptual decision in ventral premotor cortex, *Neuron* **41**: 165–173.

Rondi-Reig, L., Libbey, M., Eichenbaum, H. and Tonegawa, S. (2001). Ca1-specific n-methyl-d-aspartate receptor knockout mice are deficient in solving a nonspatial transverse patterning task, *Proceedings of the National Academy of Sciences of the United States of America* **98**: 3543–8.

Rosch, E. (1975). Cognitive representations of semantic categories, *Journal of Experimental Psychology: General* **104**: 192–233.

Rosenberg, D. R., MacMaster, F. P., Keshavan, M. S., Fitzgerald, K. D., Stewart, C. M. and Moore, G. J. (2000). Decrease in caudate glutamatergic concentrations in pediatric obsessive-compulsive disorder patients taking paroxetine, *Journal of the American Academy of Child and Adolescent Psychiatry* **39**: 1096–1103.

Rosenberg, D. R., MacMillan, S. N. and Moore, G. J. (2001). Brain anatomy and chemistry may predict treatment response in paediatric obsessive–compulsive disorder, *International Journal of Neuropsychopharmacology* **4**: 179–190.

Rosenberg, D. R., Mirza, Y., Russell, A., Tang, J., Smith, J. M., Banerjee, S. P., Bhandari, R., Rose, M., Ivey, J., Boyd, C. and Moore, G. J. (2004). Reduced anterior cingulate glutamatergic concentrations in childhood OCD and major depression versus healthy controls, *Journal of the American Academy of Child and Adolescent Psychiatry* **43**: 1146–1153.

Rosenblatt, F. (1961). *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, Spartan, Washington, DC.

Rosenkilde, C. E., Bauer, R. H. and Fuster, J. M. (1981). Single unit activity in ventral prefrontal cortex in behaving monkeys, *Brain Research* **209**: 375–394.

Rosenthal, D. (1990). A theory of consciousness, *ZIF Report 40/1990. Zentrum für Interdisziplinaire Forschung, Bielefeld*. Reprinted in Block, N., Flanagan, O. and Guzeldere, G. (eds.) (1997) *The Nature of Consciousness: Philosophical Debates*. MIT Press, Cambridge MA, pp. 729–853.

Rosenthal, D. M. (1986). Two concepts of consciousness, *Philosophical Studies* **49**: 329–359.

Rosenthal, D. M. (1993). Thinking that one thinks, *in* M. Davies and G. W. Humphreys (eds), *Consciousness*, Blackwell, Oxford, chapter 10, pp. 197–223.

Rosenthal, D. M. (2004). Varieties of higher order theory, *in* R. J. Gennaro (ed.), *Higher Order Theories of Consciousness*, John Benjamins, Amsterdam, pp. 17–44.

Rosenthal, D. M. (2005). *Consciousness and Mind*, Oxford University Press, Oxford.

Rosenthal, D. M. (2012). Higher-order awareness, misrepresentation, and function, *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 1424–1438.

Rossi, A. F., Pessoa, L., Desimone, R. and Ungerleider, L. G. (2009). The prefrontal cortex and the executive control of attention, *Experimental Brain Research* **192**: 489–497.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T. and Eickhoff, S. B. (2012). Modelling neural correlates of working memory: a coordinate-based meta-analysis, *Neuroimage* **60**: 830–846.

Roudi, Y. and Treves, A. (2006). Localized activity profiles and storage capacity of rate-based autoassociative networks, *Physical Review E* **73**: 061904.

Roudi, Y. and Treves, A. (2008). Representing where along with what information in a model of a cortical patch, *PLoS Computational Biology* **4**(3): e1000012.

Roy, R. D., Stefan, M. I. and Rosenmund, C. (2014). Biophysical properties of presynaptic short-term plasticity in hippocampal neurons: insights from electrophysiology, imaging and mechanistic models, *Frontiers in Cellular Neuroscience* **8**: 141.

Rudebeck, P. H. and Murray, E. A. (2011). Dissociable effects of subtotal lesions within the macaque orbital prefrontal cortex on reward-guided behavior, *Journal of Neuroscience* **31**: 10569–10578.

Rudebeck, P. H., Behrens, T. E., Kennerley, S. W., Baxter, M. G., Buckley, M. J., Walton, M. E. and Rushworth, M. F. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli, *Journal of Neuroscience* **28**: 13775–13785.

Rumelhart, D. E. and McClelland, J. L. (1986). *Parallel Distributed Processing*, Vol. 1: Foundations, MIT Press, Cambridge, MA.

Rumelhart, D. E. and Zipser, D. (1985). Feature discovery by competitive learning, *Cognitive Science* **9**: 75–112.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986a). Learning internal representations by error propagation, *in* D. E. Rumelhart, J. L. McClelland and the PDP Research Group (eds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, MIT Press, Cambridge, MA, chapter 8, pp. 318–362.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986b). Learning representations by back-propagating errors, *Nature* **323**: 533–536.

Rupniak, N. M. J. and Gaffan, D. (1987). Monkey hippocampus and learning about spatially directed movements,

Journal of Neuroscience **7**: 2331–2337.

Rushworth, M. F., Noonan, M. P., Boorman, E. D., Walton, M. E. and Behrens, T. E. (2011). Frontal cortex and reward-guided learning and decision-making, *Neuron* **70**: 1054–1069.

Rushworth, M. F., Kolling, N., Sallet, J. and Mars, R. B. (2012). Valuation and decision-making in frontal cortex: one or many serial or parallel systems?, *Current Opinion in Neurobiology* **22**: 946–955.

Rushworth, M. F. S., Walton, M. E., Kennerley, S. W. and Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex, *Trends in Cognitive Sciences* **8**: 410–417.

Rushworth, M. F. S., Behrens, T. E., Rudebeck, P. H. and Walton, M. E. (2007a). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour, *Trends in Cognitive Sciences* **11**: 168–176.

Rushworth, M. F. S., Buckley, M. J., Behrens, T. E., Walton, M. E. and Bannerman, D. M. (2007b). Functional organization of the medial frontal cortex, *Current Opinion in Neurobiology* **17**: 220–227.

Russchen, F. T., Amaral, D. G. and Price, J. L. (1985). The afferent connections of the substantia innominata in the monkey, Macaca fascicularis, *Journal of Comparative Neurology* **242**: 1–27.

Rust, N. C., Mante, V., Simoncelli, E. P. and Movshon, J. A. (2006). How mt cells analyze the motion of visual patterns, *Nature Neuroscience* **9**: 1421–31.

Sah, P. (1996). Ca^{2+}-activated K^+ currents in neurones: types, physiological roles and modulation, *Trends in Neuroscience* **19**: 150–154.

Sah, P. and Faber, E. S. (2002). Channels underlying neuronal calcium-activated potassium currents, *Progress in Neurobiology* **66**: 345–353.

Saint-Cyr, J. A., Ungerleider, L. G. and Desimone, R. (1990). Organization of visual cortical inputs to the striatum and subsequent outputs to the pallido-nigral complex in the monkey, *Journal of Comparative Neurology* **298**: 129–156.

Sakai, K. and Miyashita, Y. (1991). Neural organisation for the long-term memory of paired associates, *Nature* **354**(6349): 152–155.

Salakhutdinov, R. and Larochelle, H. (2010). Efficient learning of deep boltzmann machines, *International Conference on Artificial Intelligence and Statistics*, pp. 693–700.

Salinas, E. and Abbott, L. F. (1996). A model of multiplicative neural responses in parietal cortex, *Proceedings of the National Academy of Sciences USA* **93**: 11956–11961.

Salinas, E. and Abbott, L. F. (1997). Invariant visual responses from attentional gain fields, *Journal of Neurophysiology* **77**: 3267–3272.

Salinas, E. and Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information, *Nat Rev Neurosci* **2**: 539–550.

Samsonovich, A. and McNaughton, B. (1997). Path integration and cognitive mapping in a continuous attractor neural network model, *Journal of Neuroscience* **17**: 5900–5920.

Sanghera, M. K., Rolls, E. T. and Roper-Hall, A. (1979). Visual responses of neurons in the dorsolateral amygdala of the alert monkey, *Experimental Neurology* **63**: 610–626.

Saper, C. B., Swanson, L. W. and Cowan, W. M. (1979). An autoradiographic study of the efferent connections of the lateral hypothalamic area in the rat, *Journal of Comparative Neurology* **183**: 689–706.

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M. B. and Moser, E. I. (2006). Conjunctive representation of position, direction, and velocity in entorhinal cortex, *Science* **312**: 758–762.

Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in macaque, *Experimental Brain Research* **77**: 23–30.

Save, E., Guazzelli, A. and Poucet, B. (2001). Dissociation of the effects of bilateral lesions of the dorsal hippocampus and parietal cortex on path integration in the rat, *Behavaioral Neuroscience* **115**: 1212–1223.

Sawaguchi, T. and Goldman-Rakic, P. S. (1991). D1 dopamine receptors in prefrontal cortex: Involvement in working memory, *Science* **251**: 947–950.

Sawaguchi, T. and Goldman-Rakic, P. S. (1994). The role of D1-dopamine receptor in working memory: local injections of dopamine antagonists into the prefrontal cortex of rhesus monkeys performing an oculomotor delayed-response task, *Journal of Neurophysiology* **71**: 515–528.

Schachter, S. and Singer, J. (1962). Cognitive, social and physiological determinants of emotional state, *Psychological Review* **69**: 378–399.

Scheuerecker, J., Ufer, S., Zipse, M., Frodl, T., Koutsouleris, N., Zetzsche, T., Wiesmann, M., Albrecht, J., Bruckmann, H., Schmitt, G., Moller, H. J. and Meisenzahl, E. M. (2008). Cerebral changes and cognitive dysfunctions in medication-free schizophrenia – An fMRI study, *Journal of Psychiatric Research* **42**: 469–476.

Schiller, D., Monfils, M. H., Raio, C. M., Johnson, D. C., LeDoux, J. E. and Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms, *Nature* **463**: 49–53.

Schliebs, R. and Arendt, T. (2006). The significance of the cholinergic system in the brain during aging and in Alzheimer's disease, *Journal of Neural Transmission* **113**: 1625–1644.

Schmolesky, M., Wang, Y., Hanes, D., Thompson, K., Leutgeb, S., Schall, J. and Leventhal, A. (1998). Signal timing across the macaque visual system, *Journal of Neurophysiology* **79**: 3272–3277.

Schnupp, J., Nelken, I. and King, A. (2011). *Auditory Neuroscience: Making Sense of Sound*, MIT Press.

Schoenbaum, G., Roesch, M. R., Stalnaker, T. A. and Takahashi, Y. K. (2009). A new perspective on the role of the

orbitofrontal cortex in adaptive behaviour, *Nature Reviews Neuroscience* **10**: 885–892.

Schultz, S. and Rolls, E. T. (1999). Analysis of information transmission in the schaffer collaterals, *Hippocampus* **9**: 582–598.

Schultz, W. (2004). Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology, *Current Opinion in Neurobiology* **14**: 139–147.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward, *Annual Review of Psychology* **57**: 87–115.

Schultz, W. (2013). Updating dopamine reward signals, *Current Opinion in Neurobiology* **23**: 229–238.

Schultz, W., Romo, R., Ljunberg, T., Mirenowicz, J., Hollerman, J. R. and Dickinson, A. (1995). Reward-related signals carried by dopamine neurons, *in* J. C. Houk, J. L. Davis and D. G. Beiser (eds), *Models of Information Processing in the Basal Ganglia*, MIT Press, Cambridge, MA, chapter 12, pp. 233–248.

Schultz, W., Dayan, P. and Montague, P. R. (1997). A neural substrate of prediction and reward, *Science* **275**: 1593–1599.

Schwindel, C. D. and McNaughton, B. L. (2011). Hippocampal-cortical interactions and the dynamics of memory trace reactivation, *Progress in Brain Research* **193**: 163–177.

Scott, T. R. and Giza, B. K. (1987). A measure of taste intensity discrimination in the rat through conditioned taste aversions, *Physiology and Behaviour* **41**: 315–320.

Scott, T. R. and Plata-Salaman, C. R. (1999). Taste in the monkey cortex, *Physiology and Behavior* **67**: 489–511.

Scott, T. R. and Small, D. M. (2009). The role of the parabrachial nucleus in taste processing and feeding, *Annals of the New York Academy of Sciences* **1170**: 372–377.

Scott, T. R., Yaxley, S., Sienkiewicz, Z. J. and Rolls, E. T. (1986a). Gustatory responses in the frontal opercular cortex of the alert cynomolgus monkey, *Journal of Neurophysiology* **56**: 876–890.

Scott, T. R., Yaxley, S., Sienkiewicz, Z. J. and Rolls, E. T. (1986b). Taste responses in the nucleus tractus solitarius of the behaving monkey, *Journal of Neurophysiology* **55**: 182–200.

Scott, T. R., Yan, J. and Rolls, E. T. (1995). Brain mechanisms of satiety and taste in macaques, *Neurobiology* **3**: 281–292.

Seamans, J. K. and Yang, C. R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex, *Progress in Neurobiology* **74**: 1–58.

Seamans, J. K., Gorelova, N., Durstewitz, D. and Yang, C. R. (2001). Bidirectional dopamine modulation of GABAergic inhibition in prefrontal cortical pyramidal neurons, *Journal of Neuroscience* **21**: 3628–3638.

Seeman, P. and Kapur, S. (2000). Schizophrenia: more dopamine, more D2 receptors, *Proceeedings of the National Academy of Sciences USA* **97**: 7673–7675.

Seeman, P., Weinshenker, D., Quirion, R., Srivastava, L. K., Bhardwaj, S. K., Grandy, D. K., Premont, R. T., Sotnikova, T. D., Boksa, P., El-Ghundi, M., O'Dowd, B. F., George, S.-R., Perreault, M. L., Mannisto, P. T., Robinson, S., Palmiter, R. D. and Tallerico, T. (2005). Dopamine supersensitivity correlates with D2 High states, implying many paths to psychosis, *Proceedings of the National Academy of Sciences USA* **102**: 3513–3518.

Seeman, P., Schwarz, J., Chen, J. F., Szechtman, H., Perreault, M., McKnight, G. S., Roder, J. C., Quirion, R., Boksa, P., Srivastava, L. K., Yanai, K., Weinshenker, D. and Sumiyoshi, T. (2006). Psychosis pathways converge via D2 high dopamine receptors, *Synapse* **60**: 319–346.

Seigel, M. and Auerbach, J. M. (1996). Neuromodulators of synaptic strength, *in* M. S. Fazeli and G. L. Collingridge (eds), *Cortical Plasticity*, Bios, Oxford, chapter 7, pp. 137–148.

Seleman, L. D. and Goldman-Rakic, P. S. (1985). Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey, *Journal of Neuroscience* **5**: 776–794.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning, *in* D. Blake and A.Uttley (eds), *The Mechanization of Thought Processes*, H. M. Stationery Office, London, pp. 511–529.

Seltzer, B. and Pandya, D. N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey, *Brain Research* **149**: 1–24.

Senn, W., Markram, H. and Tsodyks, M. (2001). An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timing, *Neural Computation* **13**: 35–67.

Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U. and Poggio, T. (2007a). A quantitative theory of immediate visual recognition, *Progress in Brain Research* **165**: 33–56.

Serre, T., Oliva, A. and Poggio, T. (2007b). A feedforward architecture accounts for rapid categorization, *Proceedings of the National Academy of Sciences* **104**: 6424–6429.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M. and Poggio, T. (2007c). Robust object recognition with cortex-like mechanisms, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**: 411–426.

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J. and Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex, *Nature Reviews Neuroscience* **12**: 154–167.

Shadlen, M. and Newsome, W. (1995). Is there a signal in the noise?, *Current Opinion in Neurobiology* **5**: 248–250.

Shadlen, M. and Newsome, W. (1998). The variable discharge of cortical neurons: implications for connectivity, computation and coding, *Journal of Neuroscience* **18**: 3870–3896.

Shadlen, M. N. and Kiani, R. (2013). Decision making as a window on cognition, *Neuron* **80**: 791–806.

Shadlen, M. N. and Movshon, J. A. (1999). Synchrony unbound: A critical evaluation of the temporal binding hypothesis, *Neuron* **24**: 67–77.

Shallice, T. and Burgess, P. (1996). The domain of supervisory processes and temporal organization of behaviour, *Philosophical Transactions of the Royal Society B,* **351**: 1405–1411.

Shannon, C. E. (1948). A mathematical theory of communication, *AT&T Bell Laboratories Technical Journal* **27**: 379–423.

Sharp, P. E. (1996). Multiple spatial–behavioral correlates for cells in the rat postsubiculum: multiple regression analysis and comparison to other hippocampal areas, *Cerebral Cortex* **6**: 238–259.

Sharp, P. E. (ed.) (2002). *The Neural Basis of Navigation: Evidence From Single Cell Recording*, Kluwer, Boston.

Shashua, A. (1995). Algebraic functions for recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **17**: 779–789.

Sheinberg, D. L. and Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: The role of temporal cortical neurons in natural vision, *Journal of Neuroscience* **21**: 1340–1350.

Shepherd, G. M. (2004). *The Synaptic Organisation of the Brain*, 5th edn, Oxford University Press, Oxford.

Shepherd, G. M. and Grillner, S. (eds) (2010). *Handbook of Brain Microcircuits*, Oxford University Press, Oxford.

Shergill, S. S., Brammer, M. J., Williams, S. C., Murray, R. M. and McGuire, P. K. (2000). Mapping auditory hallucinations in schizophrenia using functional magnetic resonance imaging, *Archives of General Psychiatry* **57**: 1033–1038.

Sherman, S. M. and Guillery, R. W. (2013). *Functional Connections of Cortical Areas: A New View from the Thalamus*, MIT Press, Cambridge, MA.

Shevelev, I. A., Novikova, R. V., Lazareva, N. A., Tikhomirov, A. S. and Sharaev, G. A. (1995). Sensitivity to cross-like figures in cat striate neurons, *Neuroscience* **69**: 51–57.

Shizgal, P. and Arvanitogiannis, A. (2003). Gambling on dopamine, *Science* **299**: 1856–1858.

Si, B. and Treves, A. (2009). The role of competitive learning in the generation of dg fields from ec inputs, *Cognitive Neurodynamics* **3**: 177–87.

Sidhu, M. K., Stretton, J., Winston, G. P., Bonelli, S., Centeno, M., Vollmar, C., Symms, M., Thompson, P. J., Koepp, M. J. and Duncan, J. S. (2013). A functional magnetic resonance imaging study mapping the episodic memory encoding network in temporal lobe epilepsy, *Brain* **136**: 1868–1888.

Sigala, N. and Logothetis, N. K. (2002). Visual categorisation shapes feature selectivity in the primate temporal cortex, *Nature* **415**: 318–320.

Sikström, S. (2007). Computational perspectives on neuromodulation of aging, *Acta Neurochirurgica, Supplement* **97**: 513–518.

Silberberg, G., Wu, C. and Markram, H. (2004). Synaptic dynamics control the timing of neuronal excitation in the activated neocortical microcircuit, *Journal of Physiology* **556**: 19–27.

Silberberg, G., Grillner, S., LeBeau, F. E. N., Maex, R. and Markram, H. (2005). Synaptic pathways in neural microcircuits, *Trends in neurosciences* **28**: 541–551.

Sillito, A. M. (1984). Functional considerations of the operation of GABAergic inhibitory processes in the visual cortex, *in* E. G. Jones and A. Peters (eds), *Cerebral Cortex, Vol. 2, Functional Properties of Cortical Cells*, Plenum, New York, chapter 4, pp. 91–117.

Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J. and Davis, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities, *Nature* **378**: 492–496.

Simmen, M. W., Rolls, E. T. and Treves, A. (1996a). On the dynamics of a network of spiking neurons, *in* F. Eekman and J. Bower (eds), *Computations and Neuronal Systems: Proceedings of CNS95*, Kluwer, Boston.

Simmen, M. W., Treves, A. and Rolls, E. T. (1996b). Pattern retrieval in threshold-linear associative nets, *Network* **7**: 109–122.

Singer, W. (1987). Activity-dependent self-organization of synaptic connections as a substrate for learning, *in* J.-P. Changeux and M. Konishi (eds), *The Neural and Molecular Bases of Learning*, Wiley, Chichester, pp. 301–335.

Singer, W. (1995). Development and plasticity of cortical processing architectures, *Science* **270**: 758–764.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations?, *Neuron* **24**: 49–65.

Singer, W. (2000). Response synchronisation: A universal coding strategy for the definition of relations, *in* M. Gazzaniga (ed.), *The New Cognitive Neurosciences*, 2nd edn, MIT Press, Cambridge, MA, chapter 23, pp. 325–338.

Singer, W. and Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis, *Annual Review of Neuroscience* **18**: 555–586.

Singer, W., Gray, C., Engel, A., Konig, P., Artola, A. and Brocher, S. (1990). Formation of cortical cell assemblies, *Cold Spring Harbor Symposium on Quantitative Biology* **55**: 939–952.

Sjöström, P. J., Turrigiano, G. G. and Nelson, S. B. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity, *Neuron* **32**: 1149–1164.

Skaggs, W. E. and McNaughton, B. L. (1992). Quantification of what it is that hippocampal cell firing encodes, *Society for Neuroscience Abstracts* **18**: 1216.

Skaggs, W. E., McNaughton, B. L., Gothard, K. and Markus, E. (1993). An information theoretic approach to deciphering the hippocampal code, *in* S. Hanson, J. D. Cowan and C. L. Giles (eds), *Advances in Neural Information Processing Systems*, Vol. 5, Morgan Kaufmann, San Mateo, CA, pp. 1030–1037.

Skaggs, W. E., Knierim, J. J., Kudrimoti, H. S. and McNaughton, B. L. (1995). A model of the neural basis of the rat's sense of direction, *in* G. Tesauro, D. S. Touretzky and T. K. Leen (eds), *Advances in Neural Information*

Processing Systems, Vol. 7, MIT Press, Cambridge, MA, pp. 173–180.

Sloper, J. J. and Powell, T. P. S. (1979a). An experimental electron microscopic study of afferent connections to the primate motor and somatic sensory cortices, *Philosophical Transactions of the Royal Society of London, Series B* **285**: 199–226.

Sloper, J. J. and Powell, T. P. S. (1979b). A study of the axon initial segment and proximal axon of neurons in the primate motor and somatic sensory cortices, *Philosophical Transactions of the Royal Society of London, Series B* **285**: 173–197.

Small, D. M. (2010). Taste representation in the human insula, *Brain Structre and Function* **214**: 551–561.

Small, D. M. and Scott, T. R. (2009). Symposium overview: What happens to the pontine processing? Repercussions of interspecies differences in pontine taste representation for tasting and feeding, *Annals of the New York Academy of Science* **1170**: 343–346.

Small, D. M., Zald, D. H., Jones-Gotman, M., Zatorre, R. J., Petrides, M. and Evans, A. C. (1999). Human cortical gustatory areas: a review of functional neuroimaing data, *NeuroReport* **8**: 3913–3917.

Smerieri, A., Rolls, E. T. and Feng, J. (2010). Decision time, slow inhibition, and theta rhythm, *Journal of Neuroscience* **30**: 14173–14181.

Smith, M. L. and Milner, B. (1981). The role of the right hippocampus in the recall of spatial location, *Neuropsychologia* **19**: 781–793.

Smith-Swintosky, V. L., Plata-Salaman, C. R. and Scott, T. R. (1991). Gustatory neural coding in the monkey cortex: stimulus quality, *Journal of Neurophysiology* **66**: 1156–1165.

Snyder, A. Z. and Raichle, M. E. (2012). A brief history of the resting state: The washington university perspective, *Neuroimage* **62**: 902–910.

Soltani, A. and Koch, C. (2010). Visual saliency computations: mechanisms, constraints, and the effect of feedback, *Journal of Neuroscience* **30**: 12831–12843.

Soltani, A. and Wang, X.-J. (2006). A biophysically based neural model of matching law behavior: melioration by stochastic synapses, *Journal of Neuroscience* **26**: 3731–3744.

Somogyi, P. (2010). Hippocampus: intrinsic organisation, *in* G. M. Shepherd and S. Grillner (eds), *Handbook of Brain Microcircuits*, Oxford University Press, Oxford, chapter 15, pp. 148–164.

Somogyi, P. and Cowey, A. C. (1984). Double bouquet cells, *in* A. Peters and E. G. Jones (eds), *Cerebral Cortex, Vol. 1, Cellular Components of the Cerebral Cortex*, Plenum, New York, chapter 9, pp. 337–360.

Somogyi, P., Kisvarday, Z. F., Martin, K. A. C. and Whitteridge, D. (1983). Synaptic connections of morphologically identified and physiologically characterized large basket cells in the striate cortex of the cat, *Neuroscience* **10**: 261–294.

Somogyi, P., Tamas, G., Lujan, R. and Buhl, E. H. (1998). Salient features of synaptic organisation in the cerebral cortex, *Brain Research Reviews* **26**: 113–135.

Sompolinsky, H. (1987). The theory of neural networks: the Hebb rule and beyond, *in* L. van Hemmen and I. Morgenstern (eds), *Heidelberg Colloquium on Glassy Dynamics*, Vol. 275, Springer, New York, pp. 485–527.

Sompolinsky, H. and Kanter, I. (1986). Temporal association in asymmetric neural networks, *Physical Review Letters* **57**: 2861–2864.

Song, S., Yao, H. and Treves, A. (2014). A modular latching chain, *Cognitive neurodynamics* **8**: 37–46.

Spiridon, M., Fischl, B. and Kanwisher, N. (2006). Location and spatial profile of category-specific regions in human extrastriate cortex, *Human Brain Mapping* **27**: 77–89.

Spitzer, H., Desimone, R. and Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance, *Science* **240**: 338–340.

Spoerer, C. J., Eguchi, A. and Stringer, S. M. (2016). A computational exploration of complementary learning mechanisms in the primate ventral visual pathway, *Vision Research* **119**: 16–28.

Spruston, N., Jonas, P. and Sakmann, B. (1995). Dendritic glutamate receptor channel in rat hippocampal CA3 and CA1 pyramidal neurons, *Journal of Physiology* **482**: 325–352.

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., Stark, C. E. L. and Clark, R. E. (2004). The medial temporal lobe, *Annual Review of Neuroscience* **27**: 279–306.

Stankiewicz, B. and Hummel, J. (1994). Metricat: A representation for basic and subordinate-level classification, *in* G. W. Cottrell (ed.), *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, Erlbaum, San Diego, pp. 254–259.

Stefanacci, L., Suzuki, W. A. and Amaral, D. G. (1996). Organization of connections between the amygdaloid complex and the perirhinal and parahippocampal cortices in macaque monkeys, *Journal of Comparative Neurology* **375**: 552–582.

Stein, D. J., Kogan, C. S., Atmaca, M., Fineberg, N. A., Fontenelle, L. F., Grant, J. E., Matsunaga, H., Reddy, Y. C., Simpson, H. B., Thomsen, P. H., van den Heuvel, O. A., Veale, D., Woods, D. W. and Reed, G. M. (2016). The classification of obsessive-compulsive and related disorders in the icd-11, *Journal of Affective Disorders* **190**: 663–74.

Stella, F., Cerasti, E. and Treves, A. (2013). Unveiling the metric structure of internal representations of space, *Frontiers in Neural Circuits* **7**: 81.

Stent, G. S. (1973). A psychological mechanism for Hebb's postulate of learning, *Proceedings of the National Academy of Sciences USA* **70**: 997–1001.

Stephan, K. E. and Roebroeck, A. (2012). A short history of causal modeling of fmri data, *Neuroimage* **62**: 856–863.

Stephan, K. E., Baldeweg, T. and Friston, K. J. (2006). Synaptic plasticity and dysconnection in schizophrenia, *Biological Psychiatry* **59**: 929–939.

Stewart, S. E., Fagerness, J. A., Platko, J., Smoller, J. W., Scharf, J. M., Illmann, C., Jenike, E., Chabane, N., Leboyer, M., Delorme, R., Jenike, M. A. and Pauls, D. L. (2007). Association of the SLC1A1 glutamate transporter gene and obsessive-compulsive disorder, *American Journal of Medical Genetics B: Neuropsychiatric Genetics* **144**: 1027–1033.

Storm-Mathiesen, J., Zimmer, J. and Ottersen, O. P. (1990). Understanding the brain through the hippocampus, *Progress in Brain Research* **83**: 1–.

Strick, P. L., Dum, R. P. and Picard, N. (1995). Macro-organization of the circuits connecting the basal ganglia with the cortical motor areas, *in* J. C. Houk, J. L. Davis and D. G. Beiser (eds), *Models of Information Processing in the Basal Ganglia*, MIT Press, Cambridge, MA, chapter 6, pp. 117–130.

Stringer, S. M. and Rolls, E. T. (2000). Position invariant recognition in the visual system with cluttered environments, *Neural Networks* **13**: 305–315.

Stringer, S. M. and Rolls, E. T. (2002). Invariant object recognition in the visual system with novel views of 3D objects, *Neural Computation* **14**: 2585–2596.

Stringer, S. M. and Rolls, E. T. (2006). Self-organizing path integration using a linked continuous attractor and competitive network: path integration of head direction, *Network: Computation in Neural Systems* **17**: 419–445.

Stringer, S. M. and Rolls, E. T. (2007). Hierarchical dynamical models of motor function, *Neurocomputing* **70**: 975–990.

Stringer, S. M. and Rolls, E. T. (2008). Learning transform invariant object recognition in the visual system with multiple stimuli present during training, *Neural Networks* **21**: 888–903.

Stringer, S. M., Rolls, E. T., Trappenberg, T. P. and De Araujo, I. E. T. (2002a). Self-organizing continuous attractor networks and path integration: Two-dimensional models of place cells, *Network: Computation in Neural Systems* **13**: 429–446.

Stringer, S. M., Trappenberg, T. P., Rolls, E. T. and De Araujo, I. E. T. (2002b). Self-organizing continuous attractor networks and path integration: One-dimensional models of head direction cells, *Network: Computation in Neural Systems* **13**: 217–242.

Stringer, S. M., Rolls, E. T. and Trappenberg, T. P. (2004). Self-organising continuous attractor networks with multiple activity packets, and the representation of space, *Neural Networks* **17**: 5–27.

Stringer, S. M., Rolls, E. T. and Trappenberg, T. P. (2005). Self-organizing continuous attractor network models of hippocampal spatial view cells, *Neurobiology of Learning and Memory* **83**: 79–92.

Stringer, S. M., Perry, G., Rolls, E. T. and Proske, J. H. (2006). Learning invariant object recognition in the visual system with continuous transformations, *Biological Cybernetics* **94**: 128–142.

Stringer, S. M., Rolls, E. T. and Taylor, P. (2007a). Learning movement sequences with a delayed reward signal in a hierarchical model of motor function, *Neural Networks* **29**: 172–181.

Stringer, S. M., Rolls, E. T. and Tromans, J. M. (2007b). Invariant object recognition with trace learning and multiple stimuli present during training, *Network: Computation in Neural Systems* **18**: 161–187.

Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. R. and Bialek, W. (1998). Entropy and information in neural spike trains, *Physical Review Letters* **80**: 197–200.

Sugase, Y., Yamane, S., Ueno, S. and Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex, *Nature* **400**: 869–873.

Sugrue, L. P., Corrado, G. S. and Newsome, W. T. (2005). Choosing the greater of two goods: neural currencies for valuation and decision making, *Nature Reviews Neuroscience* **6**: 363–375.

Suri, R. E. and Schultz, W. (1998). Learning of sequential movements by neural network model with dopamine-like reinforcement signal, *Experimental Brain Research* **121**: 350–354.

Sutherland, N. S. (1968). Outline of a theory of visual pattern recognition in animal and man, *Proceedings of the Royal Society, B* **171**: 297–317.

Sutherland, R. J. and Rudy, J. W. (1991). Exceptions to the rule of space, *Hippocampus* **1**: 250–252.

Sutherland, R. J., Whishaw, I. Q. and Kolb, B. (1983). A behavioural analysis of spatial localization following electrolytic, kainate- or colchicine-induced damage to the hippocampal formation in the rat, *Behavioral Brain Research* **7**: 133–153.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences, *Machine Learning* **3**: 9–44.

Sutton, R. S. and Barto, A. G. (1981). Towards a modern theory of adaptive networks: expectation and prediction, *Psychological Review* **88**: 135–170.

Sutton, R. S. and Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement, *in* M. Gabriel and J. Moore (eds), *Learning and Computational Neuroscience*, MIT Press, Cambridge, MA, pp. 497–537.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning*, MIT Press, Cambridge, MA.

Suzuki, W. A. and Amaral, D. G. (1994a). Perirhinal and parahippocampal cortices of the macaque monkey – cortical afferents, *Journal of Comparative Neurology* **350**: 497–533.

Suzuki, W. A. and Amaral, D. G. (1994b). Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices, *Journal of Neuroscience* **14**: 1856–1877.

Swann, A. C. (2009). Impulsivity in mania, *Current Psychiatry Reports* **11**: 481–487.

Swanson, L. W. and Cowan, W. M. (1977). An autoradiographic study of the organization of the efferent connections of the hippocampal formation in the rat, *Journal of Comparative Neurology* **172**: 49–84.

Swash, M. (1989). John Hughlings Jackson: a historical introduction, *in* C. Kennard and M. Swash (eds), *Hierarchies in Neurology*, Springer, London, chapter 1, pp. 3–10.

Szabo, M., Almeida, R., Deco, G. and Stetter, M. (2004). Cooperation and biased competition model can explain attentional filtering in the prefrontal cortex, *European Journal of Neuroscience* **19**: 1969–1977.

Szabo, M., Deco, G., Fusi, S., Del Giudice, P., Mattia, M. and Stetter, M. (2006). Learning to attend: Modeling the shaping of selectivity in infero-temporal cortex in a categorization task, *Biological Cybernetics* **94**: 351–365.

Szentagothai, J. (1978). The neuron network model of the cerebral cortex: a functional interpretation, *Proceedings of the Royal Society of London, Series B* **201**: 219–248.

Szymanski, F. D., Garcia-Lazaro, J. A. and Schnupp, J. W. (2009). Current source density profiles of stimulus-specific adaptation in rat auditory cortex, *Journal of Neurophysiology* **102**: 1483–90.

Tabuchi, E., Mulder, A. B. and Wiener, S. I. (2003). Reward value invariant place responses and reward site associated activity in hippocampal neurons of behaving rats, *Hippocampus* **13**: 117–132.

Tagamets, M. and Horwitz, B. (1998). Integrating electrophysical and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain study, *Cerebral Cortex* **8**: 310–320.

Takeuchi, T., Duszkiewicz, A. J. and Morris, R. G. (2014). The synaptic plasticity and memory hypothesis: encoding, storage and persistence, *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **369**: 20130288.

Tanaka, K. (1993). Neuronal mechanisms of object recognition, *Science* **262**: 685–688.

Tanaka, K. (1996). Inferotemporal cortex and object vision, *Annual Review of Neuroscience* **19**: 109–139.

Tanaka, K., Saito, C., Fukada, Y. and Moriya, M. (1990). Integration of form, texture, and color information in the inferotemporal cortex of the macaque, *in* E. Iwai and M. Mishkin (eds), *Vision, Memory and the Temporal Lobe*, Elsevier, New York, chapter 10, pp. 101–109.

Tanaka, K., Saito, H., Fukada, Y. and Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey, *Journal of Neurophysiology* **66**: 170–189.

Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y. and Yamawaki, S. (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops, *Nature Neuroscience* **7**: 887–893.

Tanila, H. (1999). Hippocampal place cells can develop distinct representations of two visually identical environments, *Hippocampus* **9**: 235–246.

Taube, J. S., Muller, R. U. and Ranck, Jr., J. B. (1990a). Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis, *Journal of Neuroscience* **10**: 420–435.

Taube, J. S., Muller, R. U. and Ranck, Jr., J. B. (1990b). Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations, *Journal of Neuroscience* **10**: 436–447.

Taube, J. S., Goodridge, J. P., Golob, E. G., Dudchenko, P. A. and Stackman, R. W. (1996). Processing the head direction signal: a review and commentary, *Brain Research Bulletin* **40**: 477–486.

Taylor, J. G. (1999). Neural 'bubble' dynamics in two dimensions: foundations, *Biological Cybernetics* **80**: 393–409.

Teyler, T. J. and DiScenna, P. (1986). The hippocampal memory indexing theory, *Behavioral Neuroscience* **100**: 147–154.

Thibault, O., Porter, N. M., Chen, K. C., Blalock, E. M., Kaminker, P. G., Clodfelter, G. V., Brewer, L. D. and Landfield, P. W. (1998). Calcium dysregulation in neuronal aging and AlzheimerŠs disease: history and new directions, *Cell Calcium* **24**: 417–433.

Thomson, A. M. and Deuchars, J. (1994). Temporal and spatial properties of local circuits in neocortex, *Trends in Neurosciences* **17**: 119–126.

Thomson, A. M. and Lamy, C. (2007). Functional maps of neocortical local circuitry, *Frontiers in Neuroscience* **1**: 19–42.

Thorpe, S. J. and Imbert, M. (1989). Biological constraints on connectionist models, *in* R. Pfeifer, Z. Schreter and F. Fogelman-Soulie (eds), *Connectionism in Perspective*, Elsevier, Amsterdam, pp. 63–92.

Thorpe, S. J., Maddison, S. and Rolls, E. T. (1979). Single unit activity in the orbitofrontal cortex of the behaving monkey, *Neuroscience Letters* **S3**: S77.

Thorpe, S. J., Rolls, E. T. and Maddison, S. (1983). Neuronal activity in the orbitofrontal cortex of the behaving monkey, *Experimental Brain Research* **49**: 93–115.

Thorpe, S. J., O'Regan, J. K. and Pouget, A. (1989). Humans fail on XOR pattern classification problems, *in* L. Personnaz and G. Dreyfus (eds), *Neural Networks: From Models to Applications*, I.D.S.E.T., Paris, pp. 12–25.

Thorpe, S. J., Fize, D. and Marlot, C. (1996). Speed of processing in the human visual system, *Nature* **381**: 520–522.

Thorpe, S. J., Delorme, A. and Van Rullen, R. (2001). Spike-based strategies for rapid processing, *Neural Networks* **14**: 715–725.

Tompa, T. and Sary, G. (2010). A review on the inferior temporal cortex of the macaque, *Brain Res Rev* **62**: 165–82.

Tonegawa, S., Nakazawa, K. and Wilson, M. A. (2003). Genetic neuroscience of mammalian learning and memory, *Philosophical Transactions of the Royal Society of London B Biological Sciences* **358**: 787–795.

Tononi, G. and Cirelli, C. (2014). Sleep and the price of plasticity: from synaptic and cellular homeostasis to memory consolidation and integration, *Neuron* **81**: 12–34.

Tou, J. T. and Gonzalez, A. G. (1974). *Pattern Recognition Principles*, Addison-Wesley, Reading, MA.

Tovee, M. J. and Rolls, E. T. (1992). Oscillatory activity is not evident in the primate temporal visual cortex with static stimuli, *Neuroreport* **3**: 369–372.

Tovee, M. J. and Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex, *Visual Cognition* **2**: 35–58.

Tovee, M. J., Rolls, E. T., Treves, A. and Bellis, R. P. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex, *Journal of Neurophysiology* **70**: 640–654.

Tovee, M. J., Rolls, E. T. and Azzopardi, P. (1994). Translation invariance and the responses of neurons in the temporal visual cortical areas of primates, *Journal of Neurophysiology* **72**: 1049–1060.

Tovee, M. J., Rolls, E. T. and Ramachandran, V. S. (1996). Rapid visual learning in neurones of the primate temporal visual cortex, *NeuroReport* **7**: 2757–2760.

Trantham-Davidson, H., Neely, L. C., Lavin, A. and Seamans, J. K. (2004). Mechanisms underlying differential D1 versus D2 dopamine receptor regulation of inhibition in prefrontal cortex, *Journal of Neuroscience* **24**: 10652–10659.

Trappenberg, T. P., Rolls, E. T. and Stringer, S. M. (2002). Effective size of receptive fields of inferior temporal visual cortex neurons in natural scenes, *in* T. G. Dietterich, S. Becker and Z. Gharamani (eds), *Advances in Neural Information Processing Systems*, Vol. 14, MIT Press, Cambridge, MA, pp. 293–300.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects, *Journal of Experimental Psychology: Human Perception and Performance* **8**: 194–214.

Treisman, A. (1988). Features and objects: The fourteenth Barlett memorial lecture, *The Quarterly Journal of Experimental Psychology* **40A**: 201–237.

Treisman, A. and Gelade, G. (1980). A feature-integration theory of attention, *Cognitive Psychology* **12**: 97–136.

Treisman, A. and Sato, S. (1990). Conjunction search revisited, *Journal of Experimental Psychology: Human Perception and Performance,* **16**: 459–478.

Tremblay, L. and Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex, *Nature* **398**: 704–708.

Tremblay, L. and Schultz, W. (2000). Modifications of reward expectation-related neuronal activity during learning in primate orbitofrontal cortex, *Journal of Neurophysiology* **83**: 1877–1885.

Treves, A. (1990). Graded-response neurons and information encodings in autoassociative memories, *Physical Review A* **42**: 2418–2430.

Treves, A. (1991a). Are spin-glass effects relevant to understanding realistic auto-associative networks?, *Journal of Physics A* **24**: 2645–2654.

Treves, A. (1991b). Dilution and sparse encoding in threshold-linear nets, *Journal of Physics A: Mathematical and General* **24**: 327–335.

Treves, A. (1993). Mean-field analysis of neuronal spike dynamics, *Network* **4**: 259–284.

Treves, A. (1995). Quantitative estimate of the information relayed by the Schaffer collaterals, *Journal of Computational Neuroscience* **2**: 259–272.

Treves, A. (1997). On the perceptual structure of face space, *Biosystems* **40**: 189–196.

Treves, A. (2003). Computational constraints that may have favoured the lamination of sensory cortex, *Journal of computational neuroscience* **14**: 271–282.

Treves, A. (2005). Frontal latching networks: a possible neural basis for infinite recursion, *Cognitive Neuropsychology* **22**: 276–291.

Treves, A. (2016). The dentate gyrus, defining a new memory of David Marr, *in* L. M. Vaina and R. E. Passingham (eds), *Computational Theories and their Implementation in the Brain: The Legacy of David Marr*, Oxford University Press, Oxford.

Treves, A. and Panzeri, S. (1995). The upward bias in measures of information derived from limited data samples, *Neural Computation* **7**: 399–407.

Treves, A. and Rolls, E. T. (1991). What determines the capacity of autoassociative memories in the brain?, *Network* **2**: 371–397.

Treves, A. and Rolls, E. T. (1992a). Computational analysis of the operation of a real neuronal network in the brain: the role of the hippocampus in memory, *in* I. Aleksander and J. Taylor (eds), *Artificial Neural Networks*, Vol. 2, North-Holland, Amsterdam, pp. 891–898.

Treves, A. and Rolls, E. T. (1992b). Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network, *Hippocampus* **2**: 189–199.

Treves, A. and Rolls, E. T. (1994). A computational analysis of the role of the hippocampus in memory, *Hippocampus*

4: 374–391.

Treves, A., Rolls, E. T. and Simmen, M. (1997). Time for retrieval in recurrent associative memories, *Physica D* **107**: 392–400.

Treves, A., Panzeri, S., Rolls, E. T., Booth, M. and Wakeman, E. A. (1999). Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli, *Neural Computation* **11**: 601–631.

Treves, A., Tashiro, A., Witter, M. P. and Moser, E. I. (2008). What is the mammalian dentate gyrus good for?, *Neuroscience* **154**: 1155–72.

Trivers, R. L. (1976). Foreword, *The Selfish Gene by R. Dawkins*, Oxford University Press, Oxford.

Trivers, R. L. (1985). *Social Evolution*, Benjamin, Cummings, CA.

Tsao, D. (2014). The macaque face patch system: A window into object representation, *Cold Spring Harbor Symposia on Quantitative Biology* **79**: 109–14.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. and Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells, *Science* **311**: 617–618.

Tsodyks, M. V. and Feigel'man, M. V. (1988). The enhanced storage capacity in neural networks with low activity level, *Europhysics Letters* **6**: 101–105.

Tuckwell, H. (1988). *Introduction to Theoretical Neurobiology*, Cambridge University Press, Cambridge.

Turner, B. H. (1981). The cortical sequence and terminal distribution of sensory related afferents to the amygdaloid complex of the rat and monkey, *in* Y. Ben-Ari (ed.), *The Amygdaloid Complex*, Elsevier, Amsterdam, pp. 51–62.

Turova, T. (2012). The emergence of connectivity in neuronal networks: From bootstrap percolation to auto-associative memory, *Brain Research* **1434**: 277–284.

Ullman, S. (1996). *High-Level Vision. Object Recognition and Visual Cognition*, Bradford/MIT Press, Cambridge, MA.

Ullman, S., Assif, L., Fetaya, E. and Harari, D. (2016). Atoms of recognition in human and computer vision, *Proceedings of the National Academy of Sciences of the United States of America* **113**: 2744–9.

Ungerleider, L. G. (1995). Functional brain imaging studies of cortical mechanisms for memory, *Science* **270**: 769–775.

Ungerleider, L. G. and Haxby, J. V. (1994). 'What' and 'Where' in the human brain, *Current Opinion in Neurobiology* **4**: 157–165.

Ungerleider, L. G. and Mishkin, M. (1982). Two cortical visual systems, *in* D. Ingle, M. A. Goodale and R. J. W. Mansfield (eds), *Analysis of Visual Behaviour*, MIT Press, Cambridge, MA.

Usher, M. and McClelland, J. (2001). On the time course of perceptual choice: the leaky competing accumulator model, *Psychological Reviews* **108**: 550–592.

Usher, M. and Niebur, E. (1996). Modelling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention, *Journal of Cognitive Neuroscience* **8**: 311–327.

van den Heuvel, O. A., Veltman, D. J., Groenewegen, H. J., Cath, D. C., van Balkom, A. J., van Hartskamp, J., Barkhof, F. and van Dyck, R. (2005). Frontal-striatal dysfunction during planning in obsessive-compulsive disorder, *Archives of General Psychiatry* **62**: 301–309.

Van Essen, D., Anderson, C. H. and Felleman, D. J. (1992). Information processing in the primate visual system: an integrated systems perspective, *Science* **255**: 419–423.

van Haeften, T., Baks-te Bulte, L., Goede, P. H., Wouterlood, F. G. and Witter, M. P. (2003). Morphological and numerical analysis of synaptic interactions between neurons in deep and superficial layers of the entorhinal cortex of the rat, *Hippocampus* **13**: 943–952.

Van Hoesen, G. W. (1981). The differential distribution, diversity and sprouting of cortical projections to the amygdala in the rhesus monkey, *in* Y. Ben-Ari (ed.), *The Amygdaloid Complex*, Elsevier, Amsterdam, pp. 77–90.

Van Hoesen, G. W. (1982). The parahippocampal gyrus. New observations regarding its cortical connections in the monkey, *Trends in Neurosciences* **5**: 345–350.

Van Hoesen, G. W., Yeterian, E. H. and Lavizzo-Mourey, R. (1981). Widespread corticostriate projections from temporal cortex of the rhesus monkey, *Journal of Comparative Neurology* **199**: 205–219.

Van Ooyen, A. (2011). Using theoretical models to analyse neural development, *Nature Reviews Neuroscience* **12**: 311–326.

van Strien, N. M., Cappaert, N. L. and Witter, M. P. (2009). The anatomy of memory: an interactive overview of the parahippocampal-hippocampal network, *Nature Reviews Neuroscience* **10**: 272–282.

Vann, S. D. and Aggleton, J. P. (2004). The mammillary bodies: two memory systems in one?, *Nature Reviews Neuroscience* **5**: 35–44.

VanRullen, R., Guyonneau, R. and Thorpe, S. J. (2005). Spike times make sense, *Trends in Neuroscience* **28**: 1–4.

Varela, F., Lachaux, J., Rodriguez, E. and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration, *Nature Reviews Neuroscience* **2**: 229–239.

Veale, D. M., Sahakian, B. J., Owen, A. M. and Marks, I. M. (1996). Specific cognitive deficits in tests sensitive to frontal lobe dysfunction in obsessive-compulsive disorder, *Psychological Medicine* **26**: 1261–1269.

Verhagen, J. V., Rolls, E. T. and Kadohisa, M. (2003). Neurons in the primate orbitofrontal cortex respond to fat texture independently of viscosity, *Journal of Neurophysiology* **90**: 1514–1525.

Verhagen, J. V., Kadohisa, M. and Rolls, E. T. (2004). The primate insular taste cortex: neuronal representations of the viscosity, fat texture, grittiness, and the taste of foods in the mouth, *Journal of Neurophysiology* **92**: 1685–1699.

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H. and Cappa, S. F. (2011). Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies, *Neuroscience and Biobehavioral Reviews* **35**: 407–426.

Vinje, W. E. and Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision, *Science* **287**: 1273–1276.

Vogels, R. and Biederman, I. (2002). Effects of illumination intensity and direction on object coding in macaque inferior temporal cortex, *Cerebral Cortex* **12**: 756–766.

Vogt, B. A. (ed.) (2009). *Cingulate Neurobiology and Disease*, Oxford University Press, Oxford.

Von Bonin, G. and Bailey, P. (1947). *The Neocortex of Macaca mulatta*, University of Illinois Press, Urbana.

Von der Malsburg, C. and Bienenstock, E. (1986). Statistical coding and short-term synaptic plasticity: a scheme for knowledge representation in the brain, *in* E. Bienenstock, F. Fogelman-Soulie and G. Weisbach (eds), *Disordered Systems and Biological Organization, NATO ASI Series*, Vol. F20, Springer, Berlin, pp. 247–272.

Waelti, P., Dickinson, A. and Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory, *Nature* **412**: 43–48.

Walker, M. P. and Stickgold, R. (2006). Sleep, memory, and plasticity, *Annual Review of Psychology* **57**: 139–166.

Wallace, D. G. and Whishaw, I. Q. (2003). NMDA lesions of Ammon's horn and the dentate gyrus disrupt the direct and temporally paced homing displayed by rats exploring a novel environment: evidence for a role of the hippocampus in dead reckoning, *European Journal of Neuroscience* **18**: 513–523.

Wallenstein, G. V. and Hasselmo, M. E. (1997). GABAergic modulation of hippocampal population activity: sequence learning, place field development, and the phase precession effect, *Journal of Neurophysiology* **78**: 393–408.

Wallis, G. and Baddeley, R. (1997). Optimal unsupervised learning in invariant object recognition, *Neural Computation* **9**: 883–894.

Wallis, G. and Bülthoff, H. (1999). Learning to recognize objects, *Trends in Cognitive Sciences* **3**: 22–31.

Wallis, G. and Rolls, E. T. (1997). Invariant face and object recognition in the visual system, *Progress in Neurobiology* **51**: 167–194.

Wallis, G., Rolls, E. T. and Földiák, P. (1993). Learning invariant responses to the natural transformations of objects, *International Joint Conference on Neural Networks* **2**: 1087–1090.

Wallis, J. D. and Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task, *European Journal of Neuroscience* **18**: 2069–2081.

Wallis, J. D., Anderson, K. C. and Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules, *Nature* **411**: 953–956.

Wallisch, P., Lusignan, M., Benayoun, M., Baker, T. L., Dickey, A. S. and Hatsoupoulos, N. G. (2009). *MATLAB for Neuroscientists*, Academic Press, Burlington, MA.

Wang, M., Ramos, B. P., Paspalas, C. D., Shu, Y., Simen, A., Duque, A., Vijayraghavan, S., Brennan, A., Dudley, A., Nou, E., Mazer, J. A., McCormick, D. A. and Arnsten, A. F. (2007). Alpha2a-adrenoceptors strengthen working memory networks by inhibiting camp-hcn channel signaling in prefrontal cortex, *Cell* **129**: 397–410.

Wang, S.-H. and Morris, R. G. M. (2010). Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation, *Annual Review of Psychology* **61**: 49–79.

Wang, X. J. (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory, *Journal of Neuroscience* **19**: 9587–9603.

Wang, X. J. (2001). Synaptic reverberation underlying mnemonic persistent activity, *Trends in Neurosciences* **24**: 455–463.

Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits, *Neuron* **36**: 955–968.

Wang, X.-J. (2008). Decision making in recurrent neuronal circuits, *Neuron* **60**: 215–234.

Wang, X. J. and Kennedy, H. (2016). Brain structure and dynamics across scales: in search of rules, *Current Opinion in Neurobiology* **37**: 92–98.

Wang, Y., Markram, H., Goodman, P. H., Berger, T. K., Ma, J. and Goldman-Rakic, P. S. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex, *Nature Neuroscience* **9**: 534–542.

Warrington, E. K. and Weiskrantz, L. (1973). An analysis of short-term and long-term memory defects in man, *in* J. A. Deutsch (ed.), *The Physiological Basis of Memory*, Academic Press, New York, chapter 10, pp. 365–395.

Wasserman, E., Kirkpatrick-Steger, A. and Biederman, I. (1998). Effects of geon deletion, scrambling, and movement on picture identification in pigeons, *Journal of Experimental Psychology – Animal Behavior Processes* **24**: 34–46.

Watanabe, S., Lea, S. E. G. and Dittrich, W. H. (1993). What can we learn from experiments on pigeon discrimination?, *in* H. P. Zeigler and H.-J. Bischof (eds), *Vision, Brain, and Behavior in Birds*, MIT Press, Cambridge, MA, pp. 351–376.

Watkins, L. H., Sahakian, B. J., Robertson, M. M., Veale, D. M., Rogers, R. D., Pickard, K. M., Aitken, M. R. and Robbins, T. W. (2005). Executive function in Tourette's syndrome and obsessive-compulsive disorder, *Psychological Medicine* **35**: 571–582.

Webb, T. J. and Rolls, E. T. (2014). Deformation-specific and deformation-invariant visual object recognition: pose

vs identity recognition of people and deforming objects, *Frontiers in Computational Neuroscience* **8**: 37.

Webb, T. J., Rolls, E. T., Deco, G. and Feng, J. (2011). Noise in attractor networks in the brain produced by graded firing rate representations, *PLoS One* **6**: e23620.

Weiner, K. S. and Grill-Spector, K. (2015). The evolution of face processing networks, *Trends in Cognitive Sciences* **19**: 240–1.

Weiskrantz, L. (1997). *Consciousness Lost and Found*, Oxford University Press, Oxford.

Weiskrantz, L. (1998). *Blindsight*, 2nd edn, Oxford University Press, Oxford.

Weiskrantz, L. (2009). Is blindsight just degraded normal vision?, *Experimental Brain Research* **192**: 413–416.

Weiss, A. P. and Heckers, S. (1999). Neuroimaging of hallucinations: a review of the literature, *Psychiatry Research* **92**: 61–74.

Weiss, C., Bouwmeester, H., Power, J. M. and Disterhoft, J. F. (1999). Hippocampal lesions prevent trace eyeblink conditioning in the freely moving rat, *Behavioural Brain Research* **99**: 123–132.

Weissman, M. M., Bland, R. C., Canino, G. J., Greenwald, S., Hwu, H. G., Lee, C. K., Newman, S. C., Oakley-Browne, M. A., Rubio-Stipec, M., Wickramaratne, P. J. et al. (1994). The cross national epidemiology of obsessive compulsive disorder. The Cross National Collaborative Group, *Journal of Clinical Psychiatry* **55 Suppl**: 5–10.

Welford, A. T. (ed.) (1980). *Reaction Times*, Academic Press, London.

Wessa, M., Kanske, P. and Linke, J. (2014). Bipolar disorder: a neural network perspective on a disorder of emotion and motivation, *Restorative Neurology and Neuroscience* **32**: 51–62.

Whalen, P. J. and Phelps, E. A. (2009). *The Human Amygdala*, Guilford, New York.

Wheeler, E. Z. and Fellows, L. K. (2008). The human ventromedial frontal lobe is critical for learning from negative feedback, *Brain* **131**: 1323–1331.

Whelan, R., Conrod, P. J., Poline, J.-B., Lourdusamy, A., Banaschewski, T., Barker, G. J., Bellgrove, M. A., Büchel, C., Byrne, M., Cummins, T. D. R. et al. (2012). Adolescent impulsivity phenotypes characterized by distinct brain networks, *Nature Neuroscience* **15**: 920–925.

Whishaw, I. Q., Hines, D. J. and Wallace, D. G. (2001). Dead reckoning (path integration) requires the hippocampal formation: evidence from spontaneous exploration and spatial learning tasks in light (allothetic) and dark (idiothetic) tests, *Behavioral Brain Research* **127**: 49–69.

Whitehouse, P., Price, D., Struble, R., Clarke, A., Coyle, J. and Delong, M. (1982). Alzheimer's disease in senile dementia: loss of neurones in the basal forebrain, *Science* **215**: 1237–1239.

Whiten, A. and Byrne, R. W. (1997). *Machiavellian Intelligence II: Extensions and Evaluations*, Cambridge University Press, Cambridge.

Whittlesea, B. W. A. (1983). *Representation and Generalization of Concepts: the Abstractive and Episodic Perspectives Evaluated*, Unpublished doctoral dissertation, MacMaster University.

Widrow, B. and Hoff, M. E. (1960). Adaptive switching circuits, *1960 IRE WESCON Convention Record, Part 4 (Reprinted in Anderson and Rosenfeld, 1988)*, IRE, New York, pp. 96–104.

Widrow, B. and Stearns, S. D. (1985). *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ.

Wiener, N. (1956). The theory of prediction, *in* E. Beckenbach (ed.), *Modern Mathematics for Engineers*, McGraw-Hill, New York, chapter 8, pp. 165–190.

Wiener, S. I. and Taube, J. (eds) (2005). *Head Direction Cells and the Neural Mechanisms Underlying Directional Orientation*, MIT Press, Cambridge, MA.

Williams, G. V., Rolls, E. T., Leonard, C. M. and Stern, C. (1993). Neuronal responses in the ventral striatum of the behaving macaque, *Behavioural Brain Research* **55**: 243–252.

Wills, T. J., Lever, C., Cacucci, F., Burgess, N. and O'Keefe, J. (2005). Attractor dynamics in the hippocampal representation of the local environment, *Science* **308**: 873–876.

Willshaw, D. J. (1981). Holography, associative memory, and inductive generalization, *in* G. E. Hinton and J. A. Anderson (eds), *Parallel Models of Associative Memory*, Erlbaum, Hillsdale, NJ, chapter 3, pp. 83–104.

Willshaw, D. J. and Buckingham, J. T. (1990). An assessment of Marr's theory of the hippocampus as a temporary memory store, *Philosophical Transactions of The Royal Society of London, Series B* **329**: 205–215.

Willshaw, D. J. and Longuet-Higgins, H. C. (1969). The holophone – recent developments, *in* D. Mitchie (ed.), *Machine Intelligence*, Vol. 4, Edinburgh University Press, Edinburgh.

Willshaw, D. J. and von der Malsburg, C. (1976). How patterned neural connections can be set up by self-organization, *Proceedings of The Royal Society of London, Series B* **194**: 431–445.

Willshaw, D. J., Dayan, P. and Morris, R. G. (2015). Memory, modelling and marr: a commentary on marr (1971) 'simple memory: a theory of archicortex', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*.

Wilson, C. J. (1995). The contribution of cortical neurons to the firing pattern of striatal spiny neurons, *in* J. C. Houk, J. L. Davis and D. G. Beiser (eds), *Models of Information Processing in the Basal Ganglia*, MIT Press, Cambridge, MA, chapter 3, pp. 29–50.

Wilson, D. A. and Sullivan, R. M. (2011). Cortical processing of odor objects, *Neuron* **72**: 506–519.

Wilson, D. A., Xu, W., Sadrian, B., Courtiol, E., Cohen, Y. and Barnes, D. C. (2014). Cortical odor processing in health and disease, *Progress in Brain Research* **208**: 275–305.

Wilson, F. A., O'Scalaidhe, S. P. and Goldman-Rakic, P. S. (1994a). Functional synergism between putative gamma-aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex, *Proceedings of the Natlional Academy of Sciences U S A* **91**: 4009–4013.

Wilson, F. A. W. and Rolls, E. T. (1990a). Learning and memory are reflected in the responses of reinforcement-related neurons in the primate basal forebrain, *Journal of Neuroscience* **10**: 1254–1267.

Wilson, F. A. W. and Rolls, E. T. (1990b). Neuronal responses related to reinforcement in the primate basal forebrain, *Brain Research* **509**: 213–231.

Wilson, F. A. W. and Rolls, E. T. (1990c). Neuronal responses related to the novelty and familiarity of visual stimuli in the substantia innominata, diagonal band of broca and periventricular region of the primate, *Experimental Brain Research* **80**: 104–120.

Wilson, F. A. W. and Rolls, E. T. (1993). The effects of stimulus novelty and familiarity on neuronal activity in the amygdala of monkeys performing recognition memory tasks, *Experimental Brain Research* **93**: 367–382.

Wilson, F. A. W. and Rolls, E. T. (2005). The primate amygdala and reinforcement: a dissociation between rule-based and associatively-mediated memory revealed in amygdala neuronal activity, *Neuroscience* **133**: 1061–1072.

Wilson, F. A. W., Riches, I. P. and Brown, M. W. (1990). Hippocampus and medial temporal cortex: neuronal responses related to behavioural responses during the performance of memory tasks by primates, *Behavioral Brain Research* **40**: 7–28.

Wilson, F. A. W., O'Sclaidhe, S. P. and Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex, *Science* **260**: 1955–1958.

Wilson, F. A. W., O'Scalaidhe, S. P. and Goldman-Rakic, P. (1994b). Functional synergism between putative gamma-aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex, *Proceedings of the National Academy of Science* **91**: 4009–4013.

Wilson, H. R. (1999). *Spikes, Decisions and Actions: Dynamical Foundations of Neuroscience*, Oxford University Press, Oxford.

Wilson, H. R. and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localised populations of model neurons, *Biophysics Journal* **12**: 1–24.

Wilson, M. A. (2002). Hippocampal memory formation, plasticity, and the role of sleep, *Neurobiology of Learning and Memory* **78**: 565–569.

Wilson, M. A. and McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep, *Science* **265**: 603–604.

Winkielman, P. and Berridge, K. C. (2003). What is an unconscious emotion?, *Cognition and Emotion* **17**: 181–211.

Winkielman, P. and Berridge, K. C. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value, *Personality and Social Psychology Bulletin* **31**: 111–135.

Winocur, G. and Moscovitch, M. (2011). Memory transformation and systems consolidation, *Journal of the International Neuropsychological Society* **17**: 766–80.

Winston, P. H. (1975). Learning structural descriptions from examples, *in* P. H. Winston (ed.), *The Psychology of Computer Vision*, McGraw-Hill, New York, pp. 157–210.

Winterer, G. and Weinberger, D. R. (2004). Genes, dopamine and cortical signal-to-noise ratio in schizophrenia, *Trends in Neurosciences* **27**: 683–690.

Winterer, G., Ziller, M., Dorn, H., Frick, K., Mulert, C., Wuebben, Y., Herrmann, W. M. and Coppola, R. (2000). Schizophrenia: reduced signal-to-noise ratio and impaired phase-locking during information processing, *Clinical Neurophysiology* **111**: 837–849.

Winterer, G., Coppola, R., Goldberg, T. E., Egan, M. F., Jones, D. W., Sanchez, C. E. and Weinberger, D. R. (2004). Prefrontal broadband noise, working memory, and genetic risk for schizophrenia, *American Journal of Psychiatry* **161**: 490–500.

Winterer, G., Musso, F., Beckmann, C., Mattay, V., Egan, M. F., Jones, D. W., Callicott, J. H., Coppola, R. and Weinberger, D. R. (2006). Instability of prefrontal signal processing in schizophrenia, *American Journal of Psychiatry* **163**: 1960–1968.

Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N. and Suzuki, W. A. (2003). Single neurons in the monkeys hippocampus and learning of new associations, *Science* **300**: 1578–1581.

Wise, S. P. (2008). Forward frontal fields: phylogeny and fundamental function, *Trends in Neuroscience* **31**: 599–608.

Wiskott, L. (2003). Slow feature analysis: A theoretical analysis of optimal free responses, *Neural Computation* **15**: 2147–2177.

Wiskott, L. and Sejnowski, T. J. (2002). Slow feature analysis: unsupervised learning of invariances, *Neural Computation* **14**: 715–770.

Witter, M. P. (1993). Organization of the entorhinal–hippocampal system: a review of current anatomical data, *Hippocampus* **3**: 33–44.

Witter, M. P. (2007). Intrinsic and extrinsic wiring of ca3: indications for connectional heterogeneity, *Learning and Memory* **14**: 705–713.

Witter, M. P., Groenewegen, H. J., Lopes da Silva, F. H. and Lohman, A. H. M. (1989a). Functional organization of the extrinsic and intrinsic circuitry of the parahippocampal region, *Progress in Neurobiology* **33**: 161–254.

Witter, M. P., Van Hoesen, G. W. and Amaral, D. G. (1989b). Topographical organisation of the entorhinal projection to the dentate gyrus of the monkey, *Journal of Neuroscience* **9**: 216–228.

Witter, M. P., Naber, P. A., van Haeften, T., Machielsen, W. C., Rombouts, S. A., Barkhof, F., Scheltens, P. and Lopes da Silva, F. H. (2000a). Cortico-hippocampal communication by way of parallel parahippocampal-subicular pathways, *Hippocampus* **10**: 398–410.

Witter, M. P., Wouterlood, F. G., Naber, P. A. and Van Haeften, T. (2000b). Anatomical organization of the parahippocampal-hippocampal network, *Annals of the New York Academcy of Sciences* **911**: 1–24.

Wolfe, J. M., Cave, K. R. and Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search, *Journal of Experimental Psychology: Human Perception and Performance* **15**: 419–433.

Wolkin, A., Sanfilipo, M., Wolf, A. P., Angrist, B., Brodie, J. D. and Rotrosen, J. (1992). Negative symptoms and hypofrontality in chronic schizophrenia, *Archives of General Psychiatry* **49**: 959–965.

Womelsdorf, T., Fries, P., Mitra, P. and Desimone, R. (2006). Gamma-band synchronization in visual cortex predicts speed of change detection, *Nature* **439**: 733–736.

Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K. and Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization, *Science* **316**: 1609–1612.

Wong, K. and Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions, *Journal of Neuroscience* **26**: 1314–1328.

Wong, K. F., Huk, A. C., Shadlen, M. N. and Wang, X.-J. (2007). Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making, *Frontiers in Computational Neuroscience* **1**: 6.

Wood, E. R., Dudchenko, P. A. and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity, *Nature* **397**: 613–616.

Wood, E. R., Agster, K. M. and Eichenbaum, H. (2004). One-trial odor–reward association: a form of event memory not dependent on hippocampal function, *Behavioral Neuroscience* **118**: 526–539.

Wu, X., Baxter, R. A. and Levy, W. B. (1996). Context codes and the effect of noisy learning on a simplified hippocampal CA3 model, *Biological Cybernetics* **74**: 159–165.

Wyss, R., Konig, P. and Verschure, P. F. (2006). A model of the ventral visual system based on temporal stability and local memory, *PLoS Biology* **4**: e120.

Xavier, G. F., Oliveira-Filho, F. J. and Santos, A. M. (1999). Dentate gyrus-selective colchicine lesion and disruption of performance in spatial tasks: difficulties in "place strategy" because of a lack of flexibility in the use of environmental cues?, *Hippocampus* **9**: 668–681.

Xiang, J. Z. and Brown, M. W. (1998). Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe, *Neuropharmacology* **37**: 657–676.

Xiang, J. Z. and Brown, M. W. (2004). Neuronal responses related to long-term recognition memory processes in prefrontal cortex, *Neuron* **42**: 817–829.

Yamane, S., Kaji, S. and Kawano, K. (1988). What facial features activate face neurons in the inferotemporal cortex of the monkey?, *Experimental Brain Research* **73**: 209–214.

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D. and DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex, *Proceedings of the National Academy of Sciences of the U S A* **111**: 8619–8624.

Yates, F. A. (1992). *The Art of Memory*, University of Chicago Press, Chicago, Ill.

Yaxley, S., Rolls, E. T., Sienkiewicz, Z. J. and Scott, T. R. (1985). Satiety does not affect gustatory activity in the nucleus of the solitary tract of the alert monkey, *Brain Research* **347**: 85–93.

Yaxley, S., Rolls, E. T. and Sienkiewicz, Z. J. (1988). The responsiveness of neurones in the insular gustatory cortex of the macaque monkey is independent of hunger, *Physiology and Behavior* **42**: 223–229.

Yaxley, S., Rolls, E. T. and Sienkiewicz, Z. J. (1990). Gustatory responses of single neurons in the insula of the macaque monkey, *Journal of Neurophysiology* **63**: 689–700.

Yelnik, J. (2002). Functional anatomy of the basal ganglia, *Movement Disorders* **17 Suppl 3**: S15–S21.

Yoganarasimha, D., Yu, X. and Knierim, J. J. (2006). Head direction cell representations maintain internal coherence during conflicting proximal and distal cue rotations: comparison with hippocampal place cells, *Journal of Neuroscience* **26**: 622–631.

Zanos, P., Moaddel, R., Morris, P. J., Georgiou, P., Fischell, J., Elmer, G. I., Alkondon, M., Yuan, P., Pribut, H. J., Singh, N. S., Dossou, K. S., Fang, Y., Huang, X. P., Mayo, C. L., Wainer, I. W., Albuquerque, E. X., Thompson, S. M., Thomas, C. J., Zarate, C. A., J. and Gould, T. D. (2016). Nmdar inhibition-independent antidepressant actions of ketamine metabolites, *Nature* **533**: 481–6.

Zeidman, P. and Maguire, E. A. (2016). Anterior hippocampus: the anatomy of perception, imagination and episodic memory, *Nature Reviews Neuroscience* p. 10.1038/nrn.2015.24.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory, *Journal of Neuroscience* **16**: 2112–2126.

Zilli, E. A. (2012). Models of grid cell spatial firing published 2005–2011, *Frontiers in Neural Circuits* **6**: 16.

Zink, C. F., Pagnoni, G., Martin, M. E., Dhamala, M. and Berns, G. S. (2003). Human striatal responses to salient nonrewarding stimuli, *Journal of Neuroscience* **23**: 8092–8097.

Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C. and Berns, G. S. (2004). Human striatal responses

to monetary reward depend on saliency, *Neuron* **42**: 509–517.

Zipser, K., Lamme, V. and Schiller, P. (1996). Contextual modulation in primary visual cortex, *Journal of Neuroscience* **16**: 7376–7389.

Zohary, E., Shadlen, M. N. and Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance, *Nature* **370**: 140–143.

Zola-Morgan, S., Squire, L. R., Amaral, D. G. and Suzuki, W. A. (1989). Lesions of perirhinal and parahippocampal cortex that spare the amygdala and hippocampal formation produce severe memory impairment, *Journal of Neuroscience* **9**: 4355–4370.

Zola-Morgan, S., Squire, L. R. and Ramus, S. J. (1994). Severity of memory impairment in monkeys as a function of locus and extent of damage within the medial temporal lobe memory system, *Hippocampus* **4**: 483–494.

Zucker, R. S. and Regehr, W. G. (2002). Short-term synaptic plasticity, *Annual Review of Physiology* **64**: 355–405.

Zucker, S. W., Dobbins, A. and Iverson, L. (1989). Two stages of curve detection suggest two styles of visual computation, *Neural Computation* **1**: 68–81.

Index