View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Warwick Research Archives Portal Repository

Original citation:
Rolls, Edmund T.. (2017) A scientific theory of ars memoriae : spatial view cells in a continuous attractor network with linked items. Hippocampus.

Permanent WRAP URL:
http://wrap.warwick.ac.uk/87568

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:
This is the peer reviewed version of the following article: Rolls, E. T. (2017), A scientific theory of Ars Memoriae: Spatial view cells in a continuous attractor network with linked items. Hippocampus. doi:10.1002/hipo.22713, which has been published in final form at http://dx.doi.org/10.1002/hipo.22713 . This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

A note on versions:
The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

A scientific theory of *ars memoriae*: spatial view cells in a continuous attractor network with linked items

Edmund T Rolls

Oxford Centre for Computational Neuroscience, Oxford, UK (www.oxcns.org) and
University of Warwick, Department of Computer Science, Coventry, UK

Key words: memory method of loci; spatial view cells; place cells; attractor network; object-place memory; episodic memory

Running head: ars memoriae and spatial view cells

Please address correspondence to Professor E. T. Rolls

Email: Edmund.Rolls@oxcns.org.
Web: www.oxcns.org

Abstract

The art of memory (*ars memoriae*) used since classical times includes using a well-known scene to associate each view or part of the scene with a different item in a speech. This memory technique is also known as the 'method of loci'. The new theory is proposed that this type of memory is implemented in the CA3 region of the hippocampus where there are spatial view cells in primates that allow a particular view to be associated with a particular object in an event or episodic memory. Given that the CA3 cells with their extensive recurrent collateral system connecting different CA3 cells, and associative synaptic modifiability, form an autoassociation or attractor network, the spatial view cells with their approximately Gaussian view fields become linked in a continuous attractor network. As the view space is traversed continuously (for example by self-motion or imagined self-motion across the scene), the views are therefore successively recalled in the correct order, with no view missing, and with low interference between the items to be recalled. Given that each spatial view has been associated with a different discrete item, the items are recalled in the correct order, with none missing. This is the first neuroscience theory of *ars memoriae*. The theory provides a foundation for understanding how a key feature of *ars memoriae*, the ability to use a spatial scene to encode a sequence of items to be remembered, is implemented.

1. Introduction

Simonides of Ceos lived to tell the story of how when a banquet hall collapsed in an earthquake, he could identify all the victims by recalling from each place at the table who had been sitting there (Cicero, 55 BC). This way of remembering items was developed into what has become known as *ars memoriae* by Roman senators, who presented the steps of a complex legal argument in a speech that might last a whole day by associating each step in the argument with a location in a spatial scene through which their memory could progress from one end to the other during the speech to recall each item in the correct order (Yates, 1992). The procedure is also known as the 'method of loci'. Phrases such as 'in the first place', 'in the second place', probably refer to this method. Empirical work has demonstrated that the method of loci is efficacious (De Beni and Cornoldi, 1985; Moe and De Beni, 2005).

Why is *ars memoriae* so successful in helping to remember complex series of points, or arguments, or people, or objects? The aim of this paper is to provide a scientific theory of why *ars memoriae* is very effective. I develop the theory by describing first the empirical and theoretical foundations for the theory (sections 2 and 3), and then I present the theory (section 4).

2. Empirical foundations – spatial view cells in the primate hippocampus

We can start with the well-known place cells discovered in the rat hippocampus by O'Keefe and colleagues (O'Keefe, 1990; O'Keefe and Dostrovsky, 1971), and which were recognized in the award of the Nobel prize in 2015 to John O'Keefe, and to Edvard and May-Britt Moser for the discovery of entorhinal cortex grid cells, which fire to repeated places in the environment as the rat traverses the places (Hafting et al., 2005; Moser et al., 2014). Unfortunately these rat place and grid cells will not easily help to explain the memory of items in a spatial scene when a human is in one place moving the eyes across the whole spatial scene, for no place or grid cells would be altering their firing if the subject was not moving, and the subject was staying at one place.

More promising are the hippocampal spatial view cells which respond when a macaque is stationary at one place, but respond when the monkey looks at one point in a spatial scene (Feigenbaum and Rolls, 1991; Georges-François et al., 1999; Robertson et al., 1998; Rolls et al., 1989; Rolls and O'Mara, 1995; Rolls et al., 1997; Rolls et al., 1998; Rolls and Xiang, 2006) (see example of a typical spatial view neuron in Fig. 1). Each hippocampal neuron responds to a different spatial view, that is when a different part of the spatial scene is looked at (Rolls et al., 1998). Each spatial view cell has an approximately Gaussian shape of its spatial view field, with the peak of activity at one point in a scene, and gradually decreasing the further away from its peak is fixated (Georges-François et al., 1999; Robertson et al., 1998; Rolls et al., 1997; Rolls et al., 1998). Provided that the monkey is looking at a given part of the scene, the exact place where the monkey is has little effect (Georges-François et al., 1999; Robertson et al., 1998; Rolls et al., 1997; Rolls et al., 1998). That is, the neurons encode spatial view, not place. Many of these spatial view neurons respond when a scene is being remembered, for example when the scene is obscured with curtains and the lights are turned off (Robertson et al., 1998), or when a position in a spatial scene is recalled from an object, or vice versa (Rolls and Xiang, 2006).

The presence of spatial view cells in the macaque hippocampus that respond to landmarks being viewed has recently been confirmed during a virtual reality spatial navigation task (Wirth et al., 2017), though it was not possible in that study to fully separate the effects of spatial view and place in a factorial design in which spatial view is one factor, and the second is the place where the actor is located.

Now these spatial view cells do appear to be involved in memory, for in a task in which the location in the scene where an object was seen must be remembered, some macaque hippocampal neurons respond to the place in the scene where the object is shown, some to the object, and some to a combination of the two (Rolls et al., 2005). Moreover, some of these neurons respond in a one-trial object-place memory task when a place in a scene is recalled from an object, or when an object is recalled from a place (Rolls and Xiang, 2006). Primate hippocampal neurons also associate spatial view with the reward available at a viewed location (Rolls and Xiang, 2005).

For clarification, the evidence on spatial view cells in primates is that this is a representation provided in the primate hippocampus that has not been found in the rat hippocampus; but also it has been made clear that the primate hippocampal spatial view representation may not be entirely independent of the place where the primate is located. For example, in a population of hippocampal spatial view cells in which spatial view vs place encoding was carefully analysed by measuring the information provided when the firing was measured in several places each with several very similar spatial views (a *sine qua non* for any such investigation), the information in a population of cells about spatial view was 0.327 bits, and about place was 0.026 bits, using rigorous Shannon information theoretic measures of the stimulus-specific information (Georges-François et al., 1999). Thus the encoding of spatial view by primate hippocampal neurons is not entirely independent of the encoding of place. In addition, at least some primate hippocampal spatial view cells do have their spatial view firing modulated by the place where the macaque is located (Rolls and O'Mara, 1995). Further, I note that if there is a cell responding to the spatial view of, for example a landmark such as a trolley, table or cup in the testing environment, then that spatial view cell will fire when the primate is at the place of that landmark, provided of course that when at that place the landmark still looks similar to its appearance from other places (Robertson et al., 1998; Rolls et al., 1997). An example is provided in Fig. 1 of Robertson et al (1998), in which a neuron responded to a landmark, a table (T2) when the monkey was distant from the table, and when the monkey was at the table, provided in both cases that the monkey was looking at the table (Robertson et al., 1998). Thus a primate spatial view cell can respond to a spatial view or landmark when the primate is at the place of the landmark (Robertson et al., 1998; Rolls et al., 1997).

3. Theoretical foundations – a theory of the hippocampal CA3 system as an object-spatial view memory system

Fig. 2 shows how object representations, from for example the temporal lobes, are brought together with spatial representations, from for example the parietal lobe, in especially the CA3 hippocampal recurrent collateral network. This network potentially allows objects to be associated with places (Kesner and Rolls, 2015; Rolls, 1989a; Rolls, 1990; Rolls, 2016; Rolls and Kesner, 2006; Rolls and Treves, 1994; Treves and Rolls, 1994).

3.1 The functional architecture of the CA3 recurrent collateral network

A quantitative theory has been developed of how the hippocampus helps to implement episodic memory, for example the memory of a particular scene with people or objects in it (Kesner and Rolls, 2015; Rolls, 1987; Rolls, 1989a; Rolls, 1989b; Rolls, 1990; Rolls, 1996; Rolls, 2008; Rolls, 2016; Rolls and Kesner, 2006; Rolls and Stringer, 2005; Rolls et al., 2002; Treves and Rolls, 1992; Treves and Rolls, 1994). Within this theory, the CA3 pyramidal cell to CA3 pyramidal cell recurrent collateral system which is highly developed in primates (Kondo et al., 2009) provides the basis for an autoassociation or attractor network. In such a network, a particular object activating a small subset of neurons becomes associated with a particular place by associative synaptic modification. Later, presentation of the object can recall the place using the strengthened synapses; or a place can be recalled from the object.

More formally, many of the synapses in the hippocampus show associative modification as shown by long-term potentiation, and this synaptic modification appears to be involved in learning (see Andersen et al., 2007; Jackson, 2013; Lynch, 2004; Morris, 2003; Morris et al., 2003; Nakazawa et al., 2004; Nakazawa et al., 2003; Wang and Morris, 2010). The architecture of an autoassociation network is shown in Fig. 3, and the learning rule is as shown in Eqn. (1) below (Rolls, 2016; Rolls and Treves, 1998). The operation and properties of autoassociation or attractor networks have been described in detail elsewhere (Hertz et al., 1991; Hopfield, 1982; Rolls, 2008; Rolls, 2016; Rolls and Treves, 1998; Samsonovich and McNaughton, 1997; Treves and Rolls, 1991). Neuronal network software to illustrate the properties of attractor networks is available (Rolls, 2016) (see http://www.oxcns.org).

The hypothesis is that because the CA3 operates effectively as a single autoassociation network, it can allow arbitrary associations between inputs originating from very different parts of the cerebral cortex to be formed. These might involve associations between information originating in the temporal visual cortex about the presence of an object, and information originating in the parietal cortex about where it is. I note that although there is some spatial gradient in the CA3 recurrent connections, so that the connectivity is not fully uniform (Ishizuka et al., 1990; Witter, 2007), nevertheless the network will still have the properties of a single interconnected autoassociation network allowing associations between arbitrary neurons to be formed, given the presence of many long-range connections which overlap from different CA3 cells. It is very interesting indeed that in primates (macaques), the associational projections from CA3 to CA3 travel extensively along the longitudinal axis, and overall the radial, transverse, and longitudinal gradients of CA3 fiber distribution, clear in the rat, are much more subtle in the nonhuman primate brain (Kondo et al., 2009). The implication is that in primates, the CA3 network operates even more as a single network than in rodents.

A fundamental property of the autoassociation model of the CA3 recurrent collateral network is that the recall can be symmetric, that is, the whole of the memory can be retrieved from any part. For example, in an object-place autoassociation memory, an object could be recalled from a place retrieval cue, and vice versa (Rolls, 2016).

3.2 Continuous spatial patterns and CA3 representations

The fact that spatial patterns, which imply continuous representations of space such as those provided by spatial view cells in primates and place cells in rodents, are represented in the hippocampus has led to the application of continuous attractor models to help understand hippocampal function (Rolls

and Stringer, 2005; Samsonovich and McNaughton, 1997; Stringer et al., 2002a; Stringer et al., 2002b; Zhang, 1996). This has been necessary, because space is inherently continuous, because the firing of place and spatial view cells is approximately Gaussian as a function of the distance away from the preferred spatial location, because these cells have spatially overlapping fields, and because the theory is that these cells in CA3 are connected by Hebb-modifiable synapses (Kesner and Rolls, 2015; Rolls, 2008; Rolls, 2016).

A "Continuous Attractor" neural network (CANN) can maintain the firing of its neurons to represent any location along a continuous physical dimension such as spatial position, head direction, etc (Amari, 1977; Battaglia and Treves, 1998; Rolls, 2016; Rolls and Stringer, 2005; Stringer and Rolls, 2002; Stringer et al., 2004; Stringer et al., 2005; Stringer et al., 2002a; Stringer et al., 2002b). It uses excitatory recurrent collateral connections between the neurons (as are present in CA3) to reflect the distance between the neurons in the state space of the animal (e.g. place or spatial view). These networks can maintain the bubble or bump of neural activity constant for long periods wherever it is started to represent the current state (spatial view, place, etc) of the animal, and are likely to be involved in many aspects of spatial processing and memory, including spatial vision (Rolls, 2016).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 (Rolls, 2016), and have the same architecture, as illustrated in Fig. 3. 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 neuron is varied, and with different neurons having tuning that overlaps throughout the space. Such tuning is illustrated in Fig. 4. For comparison, autoassociation networks normally 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. 4). 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.

Evidence that there is a continuous attractor network in CA3 includes the following. In rats, hippocampal place cells show different place fields in different environments with remapping between different environments (Alme et al., 2014; Lee et al., 2004; Leutgeb and Leutgeb, 2007; Muller and Kubie, 1987; Wills et al., 2005), consistent with the theory that the places for one environment are mapped into one continuous spatial chart or configuration, and for another environment into another continuous spatial chart (Battaglia and Treves, 1998). Further, in rats that are placed in an environment that is ambiguous between two different environments, hippocampal CA3 place cells sometimes respond as if the animal was in one environment, and sometimes as if the animal was in the second environment (Jezek et al., 2011). Rather than sliding through a continuum of intermediate activity states, the CA3 network undergoes a short period of competitive flickering between preformed representations of the past and present environment before settling on the latter. The place cells can even flicker stochastically between representing one and then the other environment in different theta cycles (Jezek et al., 2011). Thus there is considerable evidence

supporting the attractor theory of CA3 operation (Kesner and Rolls, 2015; McNaughton and Morris, 1987; Rolls, 1987; Rolls, 1989a).

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. place, spatial view 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 & de Araujo (2002b) showed how the appropriate synaptic 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, during learning spatial view cells are forced to fire by visual cues in the environment that produce Gaussian firing as a function of the spatial view from an optimal spatial view for each neuron. The learning rule is simply that the weights w_{ij} from spatial view neuron j with firing rate r_j to spatial view neuron i with firing rate r_i are updated according to an associative (Hebb) rule that is consistent with findings from long-term potentiation

$$\delta w_{ij} = k r_i \, r_j \qquad (1)$$

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 of each spatial view neuron i might be the following Gaussian function of the distance of the spatial view from the optimal firing view of the neuron

$$r_i = e^{-s^2/2\sigma^2}, \qquad (2)$$

where s is the difference between the spatial view x (in degrees) of the agent and the spatial view x_i for neuron i, and σ is the standard deviation. Stringer, Trappenberg, Rolls & de Araujo (2002b) showed that after training at all positions in the space, the synaptic connections develop strengths that are an almost Gaussian function of the distance between the cells in the space. This shows how cells such as spatial view cells could be associated together in CA3 to form a continuous attractor network.

3.3 Combined continuous and discrete memory representations in the same (e.g. CA3) network, and episodic memory

Space is continuous, and object representations are discrete. If these representations are to be combined in for example an object-place memory, then we need to understand the operation of networks that combine these representations. It has now been shown that attractor networks can store both continuous patterns and discrete patterns (as illustrated in Fig. 4), and can thus be used to store for example the location in (continuous, physical) space (e.g. the place "out there" in a room represented by spatial view cells) where an object (a discrete item) is present (Rolls et al., 2002).

3.4 Idiothetic update by path integration

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. Path integration can be implemented in recurrent attractor networks as described elsewhere for hippocampal CA3 (Kesner and Rolls, 2015; Rolls, 2016; Rolls and Stringer, 2005; Samsonovich and McNaughton, 1997; Stringer and Rolls, 2002; Stringer et al., 2004; Stringer et al., 2002a; Stringer et al., 2002b) and for the entorhinal cortex (Giocomo et al., 2011; Zilli, 2012). In our approach, the movement signal, in this case eye position and head direction, are used as inputs to the spatial view continuous attractor network, with the appropriate synaptic strengths set up by self-organizing learning (Stringer et al., 2005). The effect is that there can be a steady trajectory through the space of spatial views, and potentially remembered spatial views, produced by head and eye movements (Stringer et al., 2005).

4. The theory of *ars memoriae*

4.1 The theory

Building on the above evidence, I now describe a theory, which uses the above components, of how it is that attaching items to different parts of a remembered but familiar scene provides a good strategy to remember the order of the items.

An essential feature of the proposal comes from the way in which a smooth continuous trajectory through a space can be produced by a small input that pushes one through the space, in the way described above for spatial view cells (Stringer et al., 2005). Because the spatial view fields have overlapping approximately Gaussian receptive fields, when the pushing input moves one in a certain direction, say left to right, across the scene, the firing of a set of spatial view cells for one view in the scene automatically recruits the next set of spatial view cells for the next view in the scene. That is, the continuous attractor network effectively embodies the structure of the space, that is the order of the different views in the scene, because cells that represent the same view in the space (even though not topologically together in the CANN) have strong synaptic connections between each other, and fairly strong connections only with other neurons in the network that represent nearby views in the scene. That is the effect of the learning mechanism described above. Therefore as we move from one part of the scene to another sliding across the continuous space, the CANN automatically 'looks up' the next adjacent spatial view. Thus the order of the spatial views is implicit in the structure of the CANN. Which spatial view is next to another is what is encoded in the CANN, which effectively represents the distances between the spatial views in the strengths of the synapses between the neurons.

The proposal is that *ars memoriae* takes advantage of this spatial structure and order, with a different item or object associated with each spatial view in the scene, in exactly the way that our research indicates happens for object-place associations in hippocampal CA3. Thus as the person sweeps continuously from one spatial view in a scene to another, the correct items are recalled in the correct order. One strategy is thus to start at the left of the scene, and then imagine moving one's view continuously from the left to the right. After one space has been traversed, in *ars memoriae* sometimes a different space, for

example another room, is continuously traversed next, allowing the objects or arguments linked by object-place learning to each view in the next room to be recalled in the correct order (Yates, 1992).

The capacity of the proposed combined continuous and discrete attractor network in hippocampal CA3 is sufficient to implement *ars memoriae* with many different discrete points to be remembered when traversing a continuous space. For example, in the simulations of continuous and discrete items associated together in a single attractor network, 1000 neurons were devoted to the continuous attractor representation in the single network (with the standard deviation σ of the Gaussian connectivity supporting the packet of activity 55 neurons), 500 neurons were devoted to the discrete attractor representation in the same attractor network, and it was shown that with 10 different place-object associations stored in the network, the recall of each of the 10 objects from each of the 10 places was perfect (Rolls et al., 2002). More details are provided elsewhere (Rolls et al., 2002) of just how exact the recall of an object is from the spatial position retrieval cue. Further, it has been shown that in the rat CA3, it should be possible to store approximately 100 different charts, with each chart of a different environment having many different positions represented in the continuous map for any one environment (Battaglia and Treves, 1998). Thus the capacity of the system proposed in the theory would be sufficient for ars memoriae.

This sweeping in a continuous trajectory through a CANN state space can be produced by eye and head movements, which can move the bubble of activity continuously across the continuous attractor space from one position in the space to another (Stringer et al., 2005). We have shown that this can occur for some hippocampal cells in the primate, when the eyes and/or head move in conditions when the views themselves cannot be seen, so that effectively a remembered position in the spatial view space is being recalled (Robertson et al., 1998). The only difference is that for *ars memoriae*, it is possible that the eyes and head do not actually have to be moved (though they may move!), but that one just thinks of moving across the scene to the next position in the spatial view space. Similarly, it is just possible that the same theory would apply to a place cell representation, if one were to assume that each object or step of the argument was associated with a different place, and it was possible to traverse the imagined places in a continuous spatial sequence.

4.2 Forgetting the previous day's items, and producing a new order for different items

The CA3 network is a single network, in that the recurrent collaterals reach throughout the CA3 region, and make synaptic contacts with other CA3 neurons in all parts of the CA3 region (Kondo et al., 2009). This architecture enables any object to be associated with any spatial view / place (Kesner and Rolls, 2015; Rolls, 1989a; Rolls, 2013). However, the capacity of an autoassociation net is limited, mainly by the number of recurrent collaterals onto any one CA3 neuron (Rolls, 2016; Treves and Rolls, 1991). We have shown that with a sparse representation, the number of memories that can be stored is in the order of the number of recurrent collateral connections onto any one neuron (Rolls and Treves, 1994; Treves and Rolls, 1991; Treves and Rolls, 1994), which is in the order of 12,000 in the rat. If the limit on the capacity of an attractor network is exceeded, then the ability to recall memories from the network will be very degraded (Amit, 1989; Hopfield, 1982). For this reason, and given that new episodic memories are being formed, some form of forgetting is required in the hippocampus, and there are several possible mechanisms (Rolls, 2016).

One mechanism is decay of synaptic strength. The simple forgetting mechanism is just an

exponential decay of the synaptic value back to its baseline, which may be exponential in time or in the number of learning changes incurred (Nadal et al., 1986). Another form of forgetting is implemented by setting limits to the range allowed for each synaptic strength or weight (Parisi, 1986). A third mechanism for forgetting is overwriting of previously stored memories, which will happen as a result of heterosynaptic long-term depression. If a postsynaptic neuron is activated during the formation of a new memory, then any inactive synaptic inputs, from other memories, will become weaker, and effect termed heterosynaptic long-term depression, and this will tend to weaken previously stored memories, and thus gradual forgetting of old memories occurs (Rolls, 2008; Rolls, 2016).

These neural mechanisms contribute to allowing new items for a different speech or occasion to be added onto a well-memorized scene, which was used in the practice of *ars memoriae* (Cicero, 55 BC; Yates, 1992). This is the new scientific theory of *ars memoriae* proposed in this paper. Indeed, this is the first scientific theory of *ars memoriae* based on knowledge about the firing of hippocampal neurons in primates, and on an understanding of continuous attractor networks and how a position in such a continuous attractor network can be associated with an object (Rolls, 2016).

5. Discussion

The power of the mechanism described here that it is proposed underlies *ars memoriae* (the art of memory) is that space is inherently continuous, and when mapped into a continuous attractor network in a brain region such as the hippocampus where there are spatial view neurons with approximately Gaussian receptive fields, the proximity of different views of a spatial scene is represented by the strengths of the synaptic weights between the neurons. Thus when the bubble or packet of neuronal activity in the CANN moves, whether by idiothetic (self-motion) input, or by moving in thought from one view of a scene in say a clockwise direction, then the next adjacent view in the scene is automatically retrieved form the network. In this way, a set of views can be recalled in the correct order. If each view is associated with a different discrete item (e.g. an object or thought), then the items are retrieved in the correct order, and, moreover, none is forgotten. It is this order inherent in the spatial representation in the brain that helps to provide the system with its power. The items themselves have no order, for each item or object is represented just by a randomly chosen set of neurons, that is, by a discrete representation in which each item is uncorrelated with the other items, as illustrated in Fig. 4.

The mechanism as described was implemented with spatial view cells with Gaussian receptive fields. The exact shape of the receptive field does not matter, provided that it has a peaks at the centre of the view field, and a decreasing firing rate as one moves away from the view at which a neuron has its peak firing rate. The reason is that the continuous attractor network just learns proximity in its synaptic strengths by an overlap of firing, and the exact shape of the overlap does not matter. The view fields of spatial view neurons do have the required properties (Georges-François et al., 1999; Robertson et al., 1998; Rolls et al., 1997; Rolls et al., 1998).

This is a new theory, and it is new in a number of ways. First, no previous theory has used as a foundation the spatial view cells present in primates. These are an essential component of the theory, for a human can stand in one place, and remember a series of locations in a scene “out there in space”, each one of which is associated with a different object or event. That is an important aspect of ars memoriae, and

how it is used, for example to deliver a speech when standing in one place. This function could not naturally be performed by rat place cells (Andersen et al., 2007; Hartley et al., 2014; O'Keefe, 1990; O'Keefe and Dostrovsky, 1971), for the place where the human is located while learning the scene-object associations, or later recalling them, is not changing. Second, no previous theory of *ars memoriae* that I know has used a continuous attractor network to help account for the retrieval of the items (objects or events) in the correct order. A continuous attractor network is an important component of the present theory, because it provides for a continuously linked set of spatial locations that is formed in the continuous attractor network. The network, because of the inherent continuity of space, and of the way that spatial proximity operates in a continuous attractor network, results in a continuous trajectory through the state space during recall, 'automatically' (i.e. mechanistically) leading to the items being recalled in the correct order. Thus the continuous attractor network is an important part of the new theory. Third, an essential component of the theory is that the system, as implicated in the primate hippocampus, is a memory system, with spatial to object, or time to object, associations being important. These associations are implemented by the firing of primate hippocampal spatial view cells, as shown by object-spatial view neurons and reward-spatial view neurons. Fourth, I know of no other neuroscience theory of *ars memoriae* based on a fundamental neurophysiological analysis of the relevant properties of neurons involved in memory, and of the computational neuroscience of how these neurons could implement memory (Kesner and Rolls, 2015; Rolls, 1989a; Rolls, 1990; Rolls, 1996; Rolls, 2016; Rolls and Kesner, 2006; Rolls and Treves, 1994; Treves and Rolls, 1994).

In *ars memoriae*, an extended argument (or series of points to be made) may be implemented by associating the first set of items with the different views of say a first room in a building, then moving to a second room and associating further items with each view in that room, etc. The theory accounts for that well, for the different rooms are linked of course by their proximity, and the last view in one room can become associated with the first view in a second room, all followed in a systematic order (e.g. left to right, or clockwise) in *ars memoriae*.

Also, for *ars memoriae*, a prediction is that it is likely to be useful to utilize scenes that are distant from an observer, for then the views in the scene are linked continuously as one scans steadily across the scene from a single place (from which one might be giving a lecture or speech). A scene with an object in the middle around which the observer walks produces sudden and unsystematic changes in the views and the relations between the views, and is likely to be less efficient for *ars memoriae*. Indeed, when analysing the properties of spatial view cells, we did not have a landmark in the middle of the room, because it has no relatively fixed spatial relation to other landmarks when walking around any central landmark. A second prediction is therefore that the neural encoding of such landmarks that do not bear a fixed relation to other landmarks may be different from the continuous attractor mechanism described in this paper. Such central 'landmarks' (such as Nelson's column in Trafalgar Square around which one can walk) may be treated more like objects, and associated only transiently with the relatively fixed and distant scene landmarks when viewed from a particular place. The point here is that the relationship of a central landmark in a space does not have a fixed relationship to the distant fixed elements of a scene, as one moves round a central landmark. Indeed, it is part of the theory of hippocampal function that it is best suited to incorporate allocentric spatial relations, for then the elements of a scene or a series of places do have a fixed relationship to each other, and can therefore be learned in a continuous attractor network (Kesner and Rolls, 2015; Rolls,

1989a; Rolls, 1990; Rolls, 1996; Rolls, 2016; Rolls and Kesner, 2006; Rolls and Treves, 1994; Treves and Rolls, 1994).

The theory described here is, as far as I know, the first neuroscience theory of *ars memoriae*, an art that has been of interest to scholars since classical times more than two thousand years ago (Yates, 1992). The theory provides a foundation for understanding how a key feature of *ars memoriae*, the ability to use a spatial scene to encode a sequence of items to be remembered, is implemented. The theory also provides an interesting example of cross-fertilization between the fields of classics and the arts, and neuroscience.

Figure Legends

Fig. 1. Examples of the firing of a hippocampal spatial view cell when the monkey was walking around the laboratory. (a) The firing of the cell is indicated by the spots in the outer set of 4 rectangles, each of which represents one of the walls of the room. There is one spot in the *outer rectangle* for each action potential. The base of the wall is towards the centre of each rectangle. The positions on the walls fixated during the recording sessions are indicated by points in the *inner set of 4 rectangles*, each of which also represents a wall of the room. The central square is a plan view of the room, with a triangle printed every 250 ms to indicate the position of the monkey, thus showing that many different places were visited during the recording sessions. A spot is printed every 250 ms in the inner rectangles to show that a wide range of spatial locations was viewed. (b) A similar representation of the same 3 recording sessions as in (a), but modified to indicate more fully the range of places when the cell fired at more than 12 spikes/s, to indicate that this is not a place cell, but a spatial view cell. The triangle indicates the current position of the monkey. The same data are shown in the inner and outer rectangles, and each dot represents one action potential. c1 – c4 are cups in which food may be found. T1 and T2 are trolleys at fixed positions within the space within which the monkey can walk freely. The four walls of the room are situated 1-3 mm from the 3x3 m space in which the monkey can walk, and are part of a rich laboratory environment with windows, door, apparatus etc. (After Georges-Francois, Rolls and Robertson 1999.) (hipsvc5.eps)

Fig. 2. Forward connections (solid lines) from areas of cerebral association neocortex via the parahippocampal gyrus and perirhinal cortex, and entorhinal cortex, to the hippocampus; and backprojections (dashed lines) via the hippocampal CA1 pyramidal cells, subiculum, and parahippocampal gyrus to the neocortex. There is great convergence in the forward connections down to the single network implemented in the CA3 pyramidal cells; and great divergence again in the backprojections. Left: block diagram. Right: more detailed representation of some of the principal excitatory neurons in the pathways. Abbreviations - D: Deep pyramidal cells. DG: Dentate Granule cells. F: Forward inputs to areas of the association cortex from preceding cortical areas in the hierarchy. mf: mossy fibres. PHG: parahippocampal gyrus and perirhinal cortex. pp: perforant path. rc: recurrent collateral of the CA3 hippocampal pyramidal cells. S: Superficial pyramidal cells. 2: pyramidal cells in layer 2 of the entorhinal cortex. 3: pyramidal cells in layer 3 of the entorhinal cortex. The thick lines above the cell bodies represent the dendrites. (hipconns2.eps).

Fig. 3. The architecture of a continuous attractor neural network (CANN). The architecture is the same as that of a discrete attractor neural network. During learning, external inputs e_i with Gaussian spatial fields force the output neurons to fire with rates r_i, the recurrent collaterals produce the same rates r_j as the presynaptic inputs to the neurons, and, the synapses w_{ij} become associatively modified. Many different inputs each corresponding to a different spatial representation are applied during learning, and the synapses between every pair of neurons come to represent the distance between the positions represented by each pair of neurons. The neurons shown are excitatory, and inhibitory neurons maintain the average firing so that one the neurons that correspond to one part of the space are firing, they keep each other firing by the excitatory synaptic connections between them, to provide a packet of bubble of neuronal activity, as

illustrated in Fig. 4. Because of the Gaussian spatial response profile of each neuron, the space defined is continuous. a continuous attractor network need not have local connectivity, but the global connectivity illustrated in Fig. 3. It is the associations between nearby positions and the repeated trajectories in the same order through the space that result in a continuous attractor as a result of the associative learning between co-active neurons, with nearby positions in the space not represented necessarily by nearby neurons in the continuous attractor (Rolls, 2016). Thus as the space is traversed from one end to the other, the bubble of activity moves continuously through the space. (cannarchi.eps)

Fig. 4. The types of firing patterns stored in continuous attractor networks are illustrated for the patterns present on neurons 1-1000 for Memory 1 (when the firing is that produced when the spatial state represented is that for location 300), and for Memory 2 (when the firing is that produced when the spatial state represented is that for location 500). The continuous nature of the spatial representation results from the fact that each neuron has a Gaussian firing rate that peaks at its optimal location. This particular mixed network also contains discrete representations that consist of discrete subsets of active binary firing rate neurons in the range 1001-1500. The firing of these latter neurons can be thought of as representing the discrete events that occur at the location. Continuous attractor networks by definition contain only continuous representations, but this particular network can store mixed continuous and discrete representations, and is illustrated to show the difference of the firing patterns normally stored in separate continuous attractor and discrete attractor networks. For this particular mixed network, during learning, Memory 1 is stored in the synaptic weights, then Memory 2, etc, and each memory contains part that is continuously distributed to represent physical space, and part that represents a discrete event or object. (hip_so_cann6_fig2r.eps)

Acknowledgements: Different parts of the research described here were supported by Programme Grants from the UK Medical Research Council; by Human Frontier Science program grants; by an EEC BRAIN grant; by the MRC Oxford Interdisciplinary Research Centre in Cognitive Neuroscience; and by the Oxford McDonnell-Pew Centre in Cognitive Neuroscience. The author declares no conflict of interest.

Fig. 1. (hipsvc5.eps)

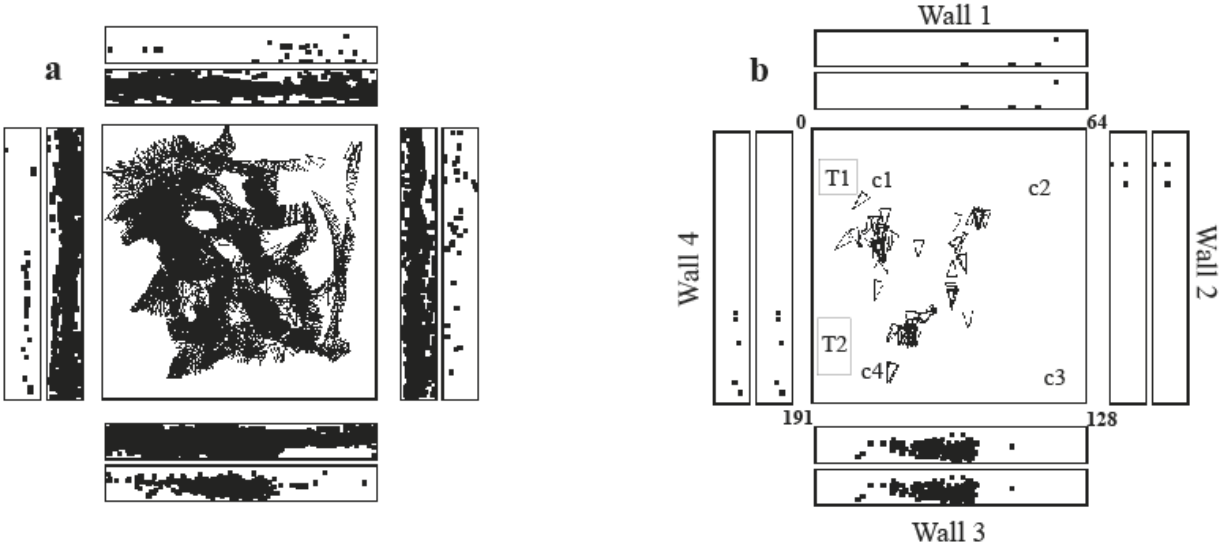


Fig. 2. (hipconns2.eps)

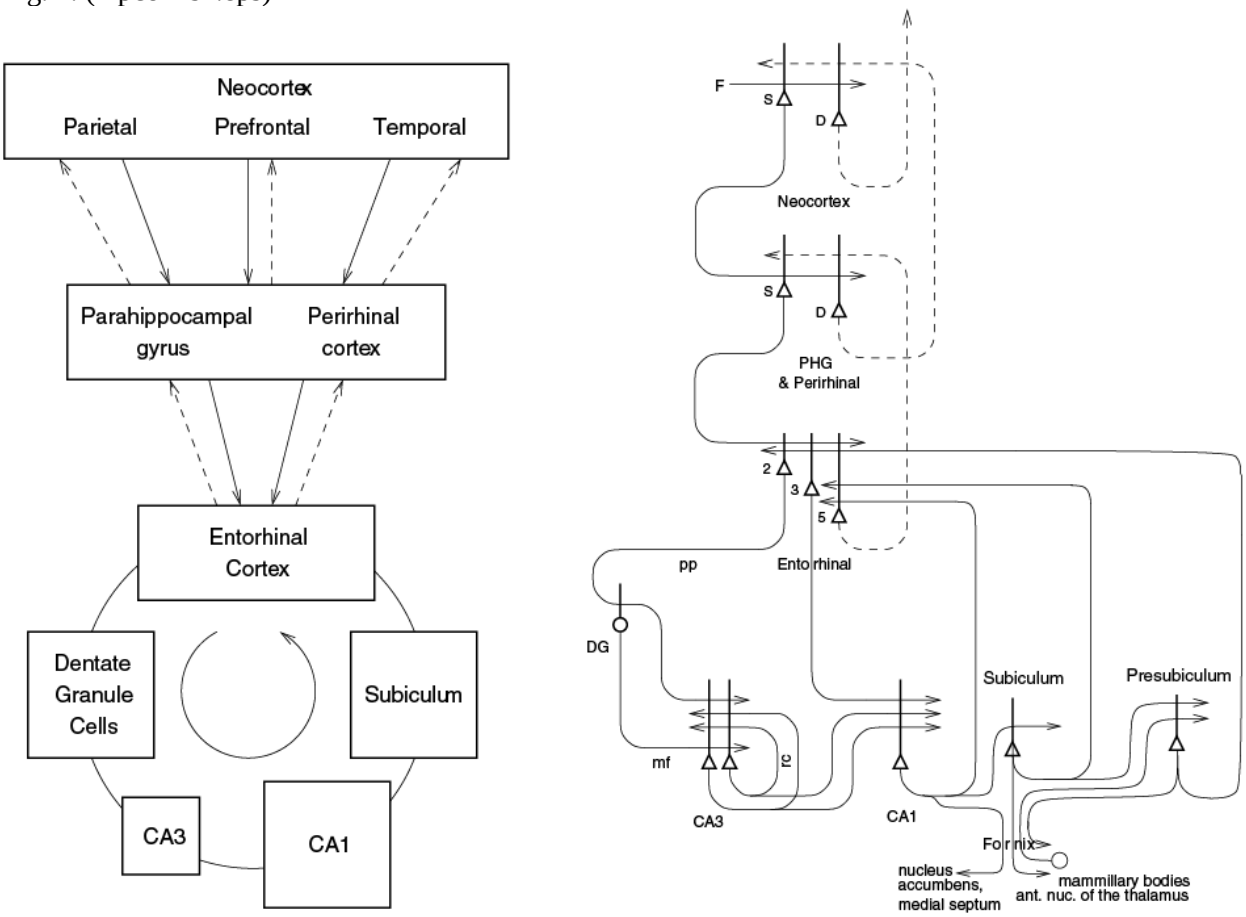


Fig. 3. (cannarchi.eps)

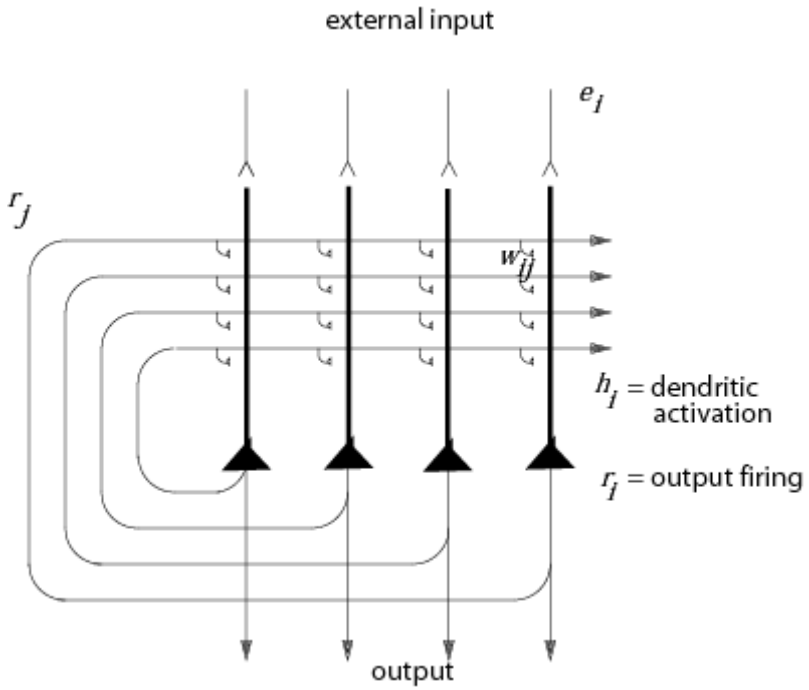
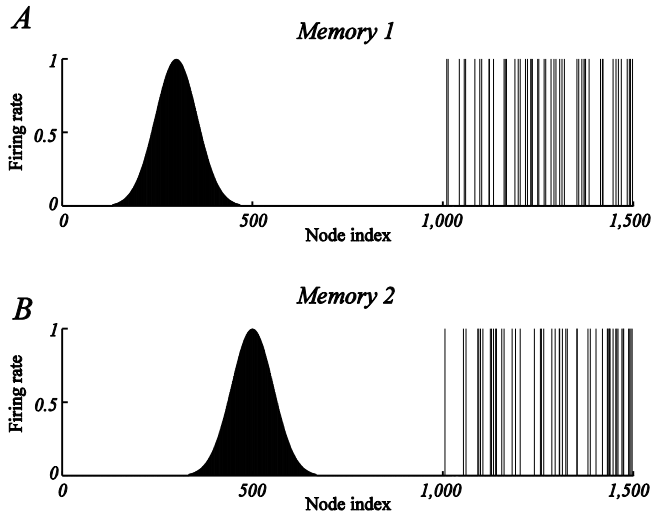


Fig. 4. (hip_so_cann6_fig2r.eps)



References

Alme CB, Miao C, Jezek K, Treves A, Moser EI, Moser MB. 2014. Place cells in the hippocampus: eleven maps for eleven rooms. Proc Natl Acad Sci U S A 111(52):18428-35.

Amari S. 1977. Dynamics of pattern formation in lateral-inhibition type neural fields. Biological Cybernetics 27:77-87.

Amit DJ. 1989. Modeling Brain Function. Cambridge: Cambridge University Press.

Andersen P, Morris RGM, Amaral DG, Bliss TVP, O'Keefe J. 2007. The Hippocampus Book. London: Oxford University Press.

Battaglia FP, Treves A. 1998. Attractor neural networks storing multiple space representations: a model for hippocampal place fields. Physical Review E 58:7738-7753.

Cicero MT. 55 BC. De Oratore II. Rome: Cicero.

De Beni R, Cornoldi C. 1985. Effects of the mnemotechnique of loci in the memorization of concrete words. Acta Psychologica 60(1):11-24.

Feigenbaum JD, Rolls ET. 1991. Allocentric and egocentric spatial information processing in the hippocampal formation of the behaving primate. Psychobiology 19:21-40.

Georges-François P, Rolls ET, Robertson RG. 1999. Spatial view cells in the primate hippocampus: allocentric view not head direction or eye position or place. Cerebral Cortex 9:197-212.

Giocomo LM, Moser MB, Moser EI. 2011. Computational models of grid cells. Neuron 71(4):589-603.

Hafting T, Fyhn M, Molden S, Moser MB, Moser EI. 2005. Microstructure of a spatial map in the entorhinal cortex. Nature 436:801-806.

Hartley T, Lever C, Burgess N, O'Keefe J. 2014. Space in the brain: how the hippocampal formation supports spatial cognition. Philos Trans R Soc Lond B Biol Sci 369(1635):20120510.

Hertz J, Krogh A, Palmer RG. 1991. An Introduction to the Theory of Neural Computation. Wokingham: Addison-Wesley.

Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Science USA 79:2554-2558.

Ishizuka N, Weber J, Amaral DG. 1990. Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. Journal of Comparative Neurology 295:580-623.

Jackson MB. 2013. Recall of spatial patterns stored in a hippocampal slice by long-term potentiation. J Neurophysiol 110(11):2511-9.

Jezek K, Henriksen EJ, Treves A, Moser EI, Moser MB. 2011. Theta-paced flickering between place-cell maps in the hippocampus. Nature 478:246-249.

Kesner RP, Rolls ET. 2015. A computational theory of hippocampal function, and tests of the theory: new developments. Neuroscience and Biobehavioral Reviews 48:92-147.

Kondo H, Lavenex P, Amaral DG. 2009. Intrinsic connections of the macaque monkey hippocampal formation: II. CA3 connections. J Comp Neurol 515(3):349-77.

Lee I, Yoganarasimha D, Rao G, Knierim JJ. 2004. Comparison of population coherence of place cells in hippocampal subfields CA1 and CA3. Nature 430:456-9.

Leutgeb S, Leutgeb JK. 2007. Pattern separation, pattern completion, and new neuronal codes within a continuous CA3 map. Learn Mem 14(11):745-57.

Lynch MA. 2004. Long-term potentiation and memory. Physiol Rev 84:87-136.

McNaughton BL, Morris RGM. 1987. Hippocampal synaptic enhancement and information storage within a distributed memory system. Trends in Neurosciences 10(10):408-415.

Moe A, De Beni R. 2005. Stressing the efficacy of the loci method: Oral presentation and the subject-generation of the loci pathway with expository passages. Applied Cognitive Psychology 19(1):95-106.

Morris RG. 2003. Long-term potentiation and memory. Philos Trans R Soc Lond B Biol Sci 358:643-647.

Morris RG, Moser EI, Riedel G, Martin SJ, Sandin J, Day M, O'Carroll C. 2003. Elements of a neurobiological theory of the hippocampus: the role of activity-dependent synaptic plasticity in memory. Philos Trans R Soc Lond B Biol Sci 358:773-786.

Moser EI, Moser MB, Roudi Y. 2014. Network mechanisms of grid cells. Philos Trans R Soc Lond B Biol Sci 369(1635):20120511.

Muller RU, Kubie JL. 1987. The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. J Neurosci 7(7):1951-68.

Nadal JP, Toulouse G, Changeux JP, Dehaene S. 1986. Networks of formal neurons and memory palimpsests. EPL (Europhysics Letters) 1(10):535.

Nakazawa K, McHugh TJ, Wilson MA, Tonegawa S. 2004. NMDA receptors, place cells and hippocampal spatial memory. Nature Reviews Neuroscience 5:361-372.

Nakazawa K, Sun LD, Quirk MC, Rondi-Reig L, Wilson MA, Tonegawa S. 2003. Hippocampal CA3 NMDA receptors are crucial for memory acquisition of one-time experience. Neuron 38:305-315.

O'Keefe J. 1990. A computational theory of the hippocampal cognitive map. Progress in Brain Research 83:301-312.

O'Keefe J, Dostrovsky J. 1971. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely moving rat. Brain Research 34:171-175.

Parisi G. 1986. A memory which forgets. Journal of Physics A: Mathematical and General 19(10):L617.

Robertson RG, Rolls ET, Georges-François P. 1998. Spatial view cells in the primate hippocampus: Effects of removal of view details. Journal of Neurophysiology 79:1145-1156.

Rolls ET. 1987. Information representation, processing and storage in the brain: analysis at the single neuron level. In: Changeux J-P, Konishi M, editors. The Neural and Molecular Bases of Learning. Chichester: Wiley. p 503-540.

Rolls ET. 1989a. Functions of neuronal networks in the hippocampus and neocortex in memory. In: Byrne JH, Berry WO, editors. Neural Models of Plasticity: Experimental and Theoretical Approaches. San Diego: Academic Press. p 240-265.

Rolls ET. 1989b. Parallel distributed processing in the brain: Implications of the functional architecture of neuronal networks in the hippocampus. In: Morris RGM, editor. Parallel Distributed Processing: Implications for Psychology and Neurobiology. Oxford: Oxford University Press. p 286-308.

Rolls ET. 1990. Theoretical and neurophysiological analysis of the functions of the primate hippocampus in memory. Cold Spring Harbor Symposia in Quantitative Biology 55:995-1006.

Rolls ET. 1996. A theory of hippocampal function in memory. Hippocampus 6:601-620.

Rolls ET. 2008. Memory, Attention, and Decision-Making: A Unifying Computational Neuroscience Approach. Oxford: Oxford University Press.

Rolls ET. 2013. A quantitative theory of the functions of the hippocampal CA3 network in memory. Front Cell Neurosci 7:98.

Rolls ET. 2016. Cerebral Cortex: Principles of Operation. Oxford: Oxford University Press.

Rolls ET, Kesner RP. 2006. A computational theory of hippocampal function, and empirical tests of the theory. Progress in Neurobiology 79:1-48.

Rolls ET, Miyashita Y, Cahusac PMB, Kesner RP, Niki H, Feigenbaum J, Bach L. 1989. Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. Journal of Neuroscience 9:1835-1845.

Rolls ET, O'Mara SM. 1995. View-responsive neurons in the primate hippocampal complex. Hippocampus 5:409-424.

Rolls ET, Robertson RG, Georges-François P. 1997. Spatial view cells in the primate hippocampus. European Journal of Neuroscience 9:1789-1794.

Rolls ET, Stringer SM. 2005. Spatial view cells in the hippocampus, and their idiothetic update based on place and head direction. Neural Networks 18:1229-1241.

Rolls ET, Stringer SM, Trappenberg TP. 2002. A unified model of spatial and episodic memory. Proceedings of the Royal Society of London B 269:1087-1093.

Rolls ET, Treves A. 1994. Neural networks in the brain involved in memory and recall. Progress in Brain Research 102:335-341.

Rolls ET, Treves A. 1998. Neural Networks and Brain Function. Oxford: Oxford University Press.

Rolls ET, Treves A, Robertson RG, Georges-François P, Panzeri S. 1998. Information about spatial view in an ensemble of primate hippocampal cells. Journal of Neurophysiology 79:1797-1813.

Rolls ET, Xiang J-Z. 2005. Reward-spatial view representations and learning in the hippocampus. Journal of Neuroscience 25:6167– 6174.

Rolls ET, Xiang J-Z. 2006. Spatial view cells in the primate hippocampus, and memory recall. Reviews in the Neurosciences 17(1-2):175-200.

Rolls ET, Xiang J-Z, Franco L. 2005. Object, space and object-space representations in the primate hippocampus. Journal of Neurophysiology 94(1):833-844.

Samsonovich A, McNaughton BL. 1997. Path integration and cognitive mapping in a continuous attractor neural network model. Journal of Neuroscience 17:5900-5920.

Stringer SM, Rolls ET. 2002. Invariant object recognition in the visual system with novel views of 3D objects. Neural Computation 14:2585-2596.

Stringer SM, Rolls ET, Trappenberg TP. 2004. Self-organising continuous attractor networks with multiple activity packets, and the representation of space. Neural Networks 17:5-27.

Stringer SM, Rolls ET, Trappenberg TP. 2005. Self-organizing continuous attractor network models of hippocampal spatial view cells. Neurobiology of Learning and Memory 83:79-92.

Stringer SM, Rolls ET, Trappenberg TP, Araujo IET. 2002a. Self-organizing continuous attractor networks and path integration. Two-dimensional models of place cells. Network: Computation in Neural Systems 13:429-446.

Stringer SM, Trappenberg TP, Rolls ET, Araujo IET. 2002b. Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells. Network: Computation in Neural Systems 13:217-242.

Treves A, Rolls ET. 1991. What determines the capacity of autoassociative memories in the brain? Network 2:371-397.

Treves A, Rolls ET. 1992. Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. Hippocampus 2:189-199.

Treves A, Rolls ET. 1994. A computational analysis of the role of the hippocampus in memory. Hippocampus 4:374-391.

Wang SH, Morris RG. 2010. Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. Annu Rev Psychol 61:49-79, C1-4.

Wills TJ, Lever C, Cacucci F, Burgess N, O'Keefe J. 2005. Attractor dynamics in the hippocampal representation of the local environment. Science 308:873-876.

Wirth W, Baraduc P, Plante A, Pinede S, Duhamel J-R. 2017. Gaze-informed, task-situated representation of space in primate hippocampus during virtual navigation. PLoS Biology.

Witter MP. 2007. Intrinsic and extrinsic wiring of CA3: indications for connectional heterogeneity.

Learning and Memory 14(11):705-13.

Yates FA. 1992. The Art of Memory. Chicago, Ill.: University of Chicago Press.

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 EA. 2012. Models of grid cell spatial firing published 2005-2011. Front Neural Circuits 6:16.