Neurobiology of Learning and Memory 145 (2017) 205–221

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

journal homepage: www.elsevier.com/locate/ynlme

Computations in the deep vs superficial layers of the cerebral cortex

Edmund T. Rolls[a,b,*], W. Patrick C. Mills[a]

[a] Oxford Centre for Computational Neuroscience, Oxford, UK
[b] University of Warwick, Department of Computer Science, Coventry, UK

ARTICLE INFO

Keywords:
Cerebral neocortex
Deep layers
Superficial layers
Attractor networks
Recurrent collaterals
Memory

ABSTRACT

A fundamental question is how the cerebral neocortex operates functionally, computationally. The cerebral neocortex with its superficial and deep layers and highly developed recurrent collateral systems that provide a basis for memory-related processing might perform somewhat different computations in the superficial and deep layers. Here we take into account the quantitative connectivity within and between laminae. Using integrate-and-fire neuronal network simulations that incorporate this connectivity, we first show that attractor networks implemented in the deep layers that are activated by the superficial layers could be partly independent in that the deep layers might have a different time course, which might because of adaptation be more transient and useful for outputs from the neocortex. In contrast the superficial layers could implement more prolonged firing, useful for slow learning and for short-term memory. Second, we show that a different type of computation could in principle be performed in the superficial and deep layers, by showing that the superficial layers could operate as a discrete attractor network useful for categorisation and feeding information forward up a cortical hierarchy, whereas the deep layers could operate as a continuous attractor network useful for providing a spatially and temporally smooth output to output systems in the brain. A key advance is that we draw attention to the functions of the recurrent collateral connections between cortical pyramidal cells, often omitted in canonical models of the neocortex, and address principles of operation of the neocortex by which the superficial and deep layers might be specialized for different types of attractor-related memory functions implemented by the recurrent collaterals.

1. Introduction

1.1. Conceptual introduction

A key architectural feature of the cerebral neocortex is the presence of short-range recurrent collateral excitatory connections between pyramidal cells (Harris & Shepherd, 2015; Lefort, Tomm, Floyd Sarria, & Petersen, 2009; Rolls, 2016). With their associatively modifiable synapses, there is considerable evidence that this recurrent collateral connectivity implements autoassociation or attractor networks that are fundamental in learning and memory (Rolls, 2016). The maintenance of activity in the local recurrent networks provides the basis for short-term memory, and planning (Goldman-Rakic, 1996; Rolls, 2016). The autoassociation aspect of the local connectivity provides for long-term memory, in which the whole of a memory can be recalled from one part (Rolls, 2016). Competition between the representations in cortical attractor networks provides the basis for decision-making, and for maintaining the neuronal activity in the winner to guide the implementation of the decision (Deco, Rolls, Albantakis, & Romo, 2013; Rolls, 2016; Rolls & Deco, 2010).

However, the superficial and deep pyramidal cell layers of the cerebral neocortex have somewhat different sets of recurrent collateral connections (Harris & Shepherd, 2015; Lefort et al., 2009; Rolls, 2016). A key issue that therefore arises in understanding the operation of the neocortex, in for example learning and memory, is whether the deep layers (especially layer 5) perform somewhat different computations to the superficial layers (2 and 3). The aim of this paper is to consider some of the different types of computation that may be performed in the superficial and deep layers in the light of their anatomy and physiology; and then to explore the implications of these issues using simulations of the operation of the attractor networks in the superficial and deep layers of the neocortex. The aim here is to formulate and then to explore some concepts about how the different connectivity of the superficial and deep layers may contribute to the major issue of how the cerebral neocortex computes, and how its computations contribute to learning and memory.

The issues that are explored here are whether different attractor dynamics implement somewhat different computation in the time

* Corresponding author at: Oxford Centre for Computational Neuroscience, Oxford, UK.
E-mail address: Edmund.Rolls@oxcns.org (E.T. Rolls).
URL: http://www.oxcns.org (E.T. Rolls).

http://dx.doi.org/10.1016/j.nlm.2017.10.011
Received 8 August 2017; Received in revised form 7 October 2017; Accepted 10 October 2017
Available online 16 October 2017

domain (e.g. longer in the superficial layers to promote slow learning useful for learning transform-invariant representations of objects (Rolls, 2016, 2012b; Wiskott & Sejnowski, 2002) vs more temporally precise and limited for the deep layers to support precise motor function); and in the spatial domain with for example more continuous spatial representations in the deep layers which may be useful for smooth motor outputs from the neocortex. Further, if the deep layers operated more as a continuous attractor network, this could facilitate smooth transitions between outputs that would be useful especially in low-dimensional motor spaces; could facilitate the feedback of a more general rather than very discrete signal for top-down attention and recall to previous cortical areas; and might be quite stable even with stochastically spiking integrate-and-fire neurons.

1.2. Cortical connectivity, and recurrent collateral connections within the superficial and deep layers

1.2.1. Partially separate attractor networks in the superficial and deep layers of the cerebral cortex?

Both the superficial and the deep layers of the cerebral cortex have a highly developed excitatory recurrent collateral system, with thousands of synapses on the dendrites of each neuron for the recurrent collaterals from nearby neurons (Harris & Shepherd, 2015; Lefort et al., 2009; Rolls, 2016) (see Fig. 1). Evidence that these recurrent collaterals can support attractor states that help to implement short-term memory, long-term memory, and decision-making has been described (Rolls, 2016).

A fundamental question is why there are somewhat separate superficial and deep layers of the cerebral cortex. The advantages include the different outputs that may be appropriate for sending to the next stage of the cortical hierarchy to build higher level representations (with their origin in the superficial layer 2 and 3 pyramidal cells), whereas the deep layer pyramidal cells in layer 5 may provide an output more suitable for driving the often motor-related target systems such as the striatum and in the case of V1 the superior colliculus (Rolls, 2016). For example, the representations in the superficial pyramidal cells (L2/L3) may be more sparse than the deep pyramidal cells (L5/L6) (Harris & Shepherd, 2015), and Rolls (2016) hypothesizes that this helps the neocortex to increase the memory capacity of what can be stored in discrete autoassociation networks in the superficial layers. The superficial cortical layers are hypothesized to perform the main computationally useful functions of the cerebral neocortex, which involve feedforward operations to form non-linear combinations of the inputs from previous cortical areas utilizing the convergence from stage to construct useful combinations essential for cortical computation, and to implement using the attractor properties of the recurrent collaterals useful functions such as information storage and retrieval (long-term memory, short-term memory, decision-making, etc.). According to the hypotheses being developed, these feedforward competitive learning computations performed by the superficial layers are the computationally useful aspects of the design of the neocortex for building new representations (Rolls, 2016).

The L5A pyramidal cells have a well-developed recurrent collateral system, and the hypothesis has been proposed that the deep layers may operate as a partly separate attractor from that in the superficial layers, for the L2/L3 pyramidal cell dendrites do not descend into the deep layers of the cerebral cortex in which the deep layer neuron recurrent collaterals make their main local connections with nearby layer 5 pyramidal cells (Rolls, 2016) (Fig. 1).

1.2.2. The connectivity of the superficial and deep layers of the neocortex

Quantitative evidence is now becoming available on the distribution of cortical connections within and between cortical laminae (Hooks et al., 2011; Holmgren, Harkany, Svennenfors, & Zilberter, 2003; Lefort et al., 2009). Some of the evidence is shown in Fig. 2, obtained by multiple simultaneous single neuron recording in the mouse whisker barrel cortex (Lefort et al., 2009). Fig. 2a shows that there is a relatively high probability of a connection between pyramidal neurons in layer 3 (0.19), and between neurons in layer 5A (0.19), and that this is higher than the connection probability from L3 neurons to L5A neurons (0.06). Fig. 2b estimates the overall strength of the synaptic effects from neurons in one layer (the columns) to neurons in another layer (the rows), by weighting the probability of synaptic connections between pyramidal cells in different laminae by the strength of the synaptic connections between neurons in the different laminae, based on data from Lefort et al. (2009). The quantitative aspects of this connectivity are

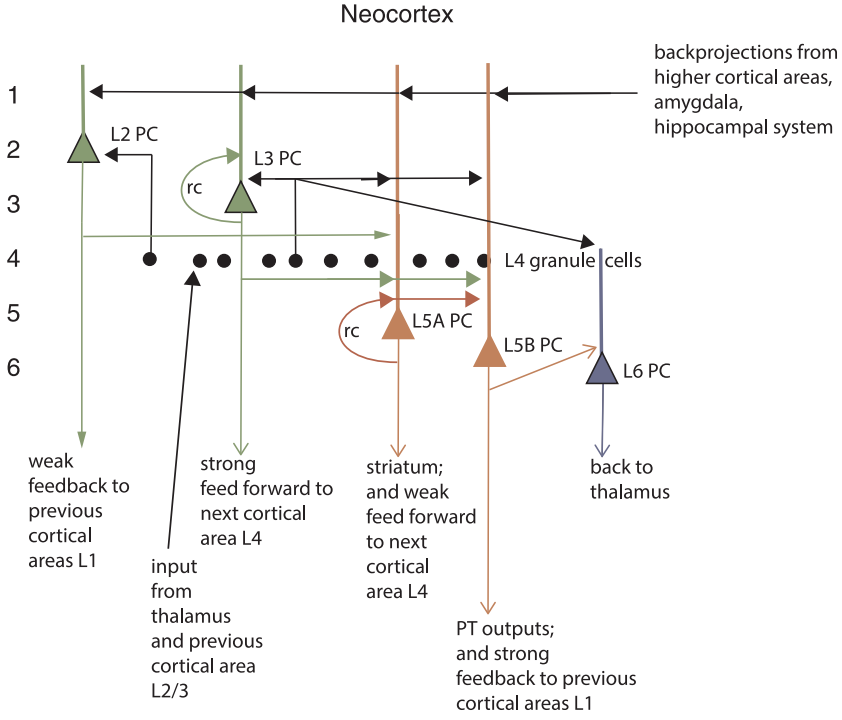

Fig. 1. Functional canonical microcircuit of the neocortex (see text). Recurrent collateral connections (rc) are shown as a loop back to a particular population of cells, of which just one neuron is shown. The dendrites are shown as thick lines above the cell bodies. It should be noted that the dendrites of the superficial (L2 and L3) pyramidal cells do not descend into the deep layers (L5, L6) of the neocortex. In primates the cortico-cortical feedforward projection neurons are concentrated in L3; and the main cortico-cortical feedback projection neurons are in Lower L5 (L5B), and there is weak cortico-cortical feedback from some L2 neurons. Some L6 cortico-thalamic neurons send a projection to L4. Further information is provided in the text and by Rolls (2016).

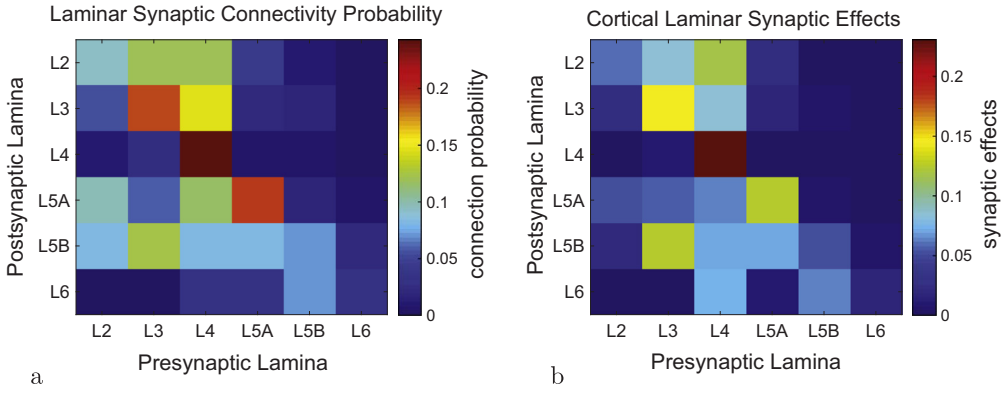

Fig. 2. (a) The probability of synaptic connections between pyramidal cells in different laminae, based on data from Lefort et al. (2009). The probability of synaptic inputs from a neuron in any presynaptic lamina (the columns) to a neuron in a postsynaptic lamina (shown as rows) can be read where the column intersects with a row. (b) The overall strength of the synaptic effects from neurons in one layer (the columns) to neurons in another layer (the rows), based on data from Lefort et al. (2009, Table 2). The overall strength was estimated from the probability of synaptic connections between pyramidal cells in different laminae weighted by (using a dot product) the strength of the synaptic connections measured by the mean EPSP amplitude for the synaptic connections between the relevant layers.

consistent with the hypothesis that the superficial pyramidal cells with their recurrent collateral connections could form an attractor network, and that the deep pyramidal cells could form a separate attractor network, which receives its inputs from the L3 network in such a way that the L5A network can be driven into activity by the superficial network, and that then the deep network can operate somewhat independently as the coupling between the networks is weaker than the coupling within the networks (Renart, Parga, & Rolls, 1999a).

Now, this is the case even for rodent whisker barrel somatosensory cortex (Lefort et al., 2009), which is highly dominated by the layer 4 inputs, and in higher cortical areas of primates involved for example in memory the recurrent collateral connections in the superficial and deep layers may be much more highly developed (Elston et al., 2006). In this context, it is interesting to note from Fig. 2 that in whisker barrel cortex, layer 4 intralaminar connectivity dominates the intralaminar connectivity, but this will not be the case in most higher cortical areas, for layer 4 become progressively more minimal as one progresses further up a cortical hierarchy (Pandya, Seltzer, Petrides, & Cipolloni, 2015). It is likely that when frontal and related areas of the cortex involved in memory are investigated, the dominance of the recurrent collateral systems within a cortical layer will become even more apparent, and it will then be of interest to assess quantitatively the relative value of the connectivity from superficial to deep layers compared to that within cortical layers.

Putting together what is shown in Fig. 2 (based on data from Lefort et al. (2009)) with key findings on cortical connectivity (Harris & Shepherd, 2015; Hooks et al., 2011; Holmgren et al., 2003; Lefort et al., 2009; Markov et al., 2014; Pandya et al., 2015; Rockland & Pandya, 1979) leads to what is shown in Fig. 1, which is amplified in the following text. Some authors refer to L2 in Fig. 1 as L2/3A; and to L3 as L3B (Harris & Shepherd, 2015). The quantitative aspects of the connectivity in higher cortical areas are likely to be much less influenced by L4 than what is shown in Fig. 2, which is for rodent whisker barrel (primary somatosensory) cortex (Lefort et al., 2009).

The **L4 granule cells** receive strong thalamic inputs, which also reach to a smaller extent neurons in other cortical layers. L4 projects on to neurons in other layers as shown in Fig. 2. L4 becomes progressively smaller up each cortical hierarchy (Pandya et al., 2015), for as one progresses up the hierarchy, the inputs that dominate the processing come more and more from neurons in L3 of the preceding cortical area. L4 has strong onward connections to L3 and L2, and somewhat weaker to neurons in deeper cortical layers (Fig. 2). As shown in Fig. 2 the L4 neurons in whisker barrel cortex have many connections with each other. This may reflect the operation of a L4 attractor network used to help minimize noise and perform constraint satisfaction (Rolls, 2016) to produce a population code that best represents the inputs from the sensory receptors associated with a single whisker. In whisker barrel cortex, L6 does not project back to L4, though it does in some other cortical areas (Harris & Shepherd, 2015).

L2 receives its inputs from L4 > L3 > L2 (where > should be read as 'more strongly than'). L2 projects to L5A and L2 > L3 and L5B. L2 neurons provide weak cortico-cortical feedback to L1 of the preceding cortical area in a hierarchy.

L3 has a highly developed recurrent collateral system to other L3 neurons, and this provides the basis for a cortical attractor network in the superficial layers of the neocortex (Rolls, 2016). L3 has strong forward projections to the middle layers of the next layer in the cortical hierarchy, and these are precisely directed to primarily the next cortical area. L3 has strong connections to L5A and L5B pyramidal cells.

L5A has a highly developed recurrent collateral system to other L5A neurons, and this provides the basis for a cortical attractor network in the deep layers of the neocortex (Rolls, 2016). Because L5A does not have strong projections to the superficial layers, it could not contribute to and be fully part of any attractor process implemented by the L3 neurons. L5A receives its main inputs from L2, L3, and L4. It has weak forward projections to the middle layers of the next layer in the hierarchy, and these are precisely directed to the next cortical area. L5A neurons have outputs to the striatum (caudate, putamen, and ventral striatum), which is an evolutionarily old region. An interesting hypothesis is that in the evolution of the cortex, the deep layers evolved early in order to provide essential outputs for the cortex; and that the superficial layers developed progressively later to provide for feedforward cortico-cortical computation up the hierarchy as new areas were added in evolution at the end of the existing hierarchy.

L5B is unlikely to implement an attractor network, as it does not have a high probability of recurrent collateral connections to other L5B neurons (Fig. 2). The main inputs to L5B pyramidal neurons are from L3 > L4 = L5A. (This L5B does receive from an attractor, L3.) L5B has pyramidal tract (PT) outputs, which may be evolutionarily newer than the striatal outputs of L5A. L5B may act as a relay for L3, L4, and L5A (which are both attractors, so it may not need to be an attractor in its own right). L5B is the source of the strong cortico-cortical backprojections, which end mainly in layer 1 of the preceding cortical area, but all backprojections tend to be diffuse in that they project also to earlier cortical areas in the hierarchy (Markov et al., 2014; Pandya et al., 2015).

L6 is cortico-thalamic. Its inputs come from L4 and L5B. So L6 is not strongly influenced by an attractor. It may primarily implement a negative feedback to the thalamus, implementing gain control.

The average connectivity between pyramidal cells in mouse whisker barrel cortex is 0.10%, but this is a small column of approximately 300 μm in diameter (Lefort et al., 2009), and in more typical cortex the connectivity may not be restricted by the presence of a barrel devoted to a single whisker. Indeed, in rat visual and somatosensory neocortex layers L2–L3 the probability of a connection between pyramidal cells in

these superficial layers decreases from 0.09 to 0.01 over a radial distance of 25–200 μm (Holmgren et al., 2003). The connections with inhibitory interneurons are somewhat more widely distributed (Holmgren et al., 2003), consistent with the hypothesis that the interneurons provide for feedback inhibition to maintain the stability of the excitatory networks and to control the sparseness of the representation, that is, the proportion of the neurons that are active to any one stimulus or event (Rolls, 2016).

Further evidence on somewhat separate computations in superficial and deep layers is that different tuning has been found for neurons in superficial vs deep layers of the macaque striate cortex, with the deep layer neurons suggested to be more appropriate for generating eye movements such as vergence eye movements (Bauer & Dow, 1991).

Further evidence on somewhat separate computations in superficial and deep layers is that the deep layers are described as showing greater adaptation than superficial layers, with adaptation differing between different classes of cortical neuron Markram et al. (2015). Consistent with this, more adaptation was found in deep cortical layers than in superficial layers of the rat auditory cortex to tones (Szymanski, Garcia-Lazaro, & Schnupp, 2009).

1.3. Hypotheses: partly separate attractor networks in superficial and deep layers

The evidence just described and other evidence led to the hypothesis that the deep layers may operate as a partly separate attractor network from that in the superficial layers (Rolls, 2016). What computational function might a separate attractor network perform in the deep layers?

1.3.1. Different time courses in the superficial and deep layers

One hypothesis is that an attractor network in the deep layers might have a different time course, with the potential for a shorter time course of firing, consistent with the more bursty firing sometimes reported in the deep layers (Harris & Shepherd, 2015), and which may be more suitable for precise temporally accurate motor control (Rolls, 2016). The hypothesis is that the deep layers might be temporally faster, as the premium is now on what may sometimes be a requirement for fast but brief output to suddenly drive motor systems when necessary. In contrast, the superficial layers may benefit from slower integration of information, to help build useful new categories, and to allow representations to be influenced by not only the recurrent collaterals, but also by what may be being backprojected from the next cortical stage in the hierarchy (Rolls, 2016, 2012b). This hypothesis is investigated in Experiment 1 by introducing adaptation into the neurons in the deep layer attractor network. The hypothesis that is tested is that the L5A attractor network may have more neuronal adaptation, or synaptic adaptation in its recurrent collaterals, than is the case for L3 neurons, and that this may enable L5A neurons to have shorter periods of attractor-related firing to implement temporally exact motor control; whereas the L3 attractor network with less adaptation in its recurrent collaterals may produce a typically longer time course of its firing, helping to provide a more prolonged input to help the next cortical area learn and to implement processes such as slow learning which are believed to be important in learning transform-invariant representations, and also short-term memory by prolonged firing (Rolls, 2016, 2012b).

1.3.2. Discrete vs spatially continuous attractors in the superficial and deep layers

A second hypothesis is that layer 5A operates as an attractor network, but of a different type to the attractor network in the L3 superficial pyramidal cells. For example, the L3 pyramidal neurons may often need to support discrete attractor states, such as whether it is this object, or that object, especially in the ventral cortical areas involved in 'what' representations for visual, auditory, taste, and olfactory stimuli (Rolls, 2016). (Discrete representations involve different subsets of neurons firing to represent different objects. The subsets may overlap, but separate objects are represented, without smooth continuity over the representational space (Rolls, 2016; Rolls & Treves, 2011).) The following proposal follows from the evidence that the pyramidal cells of L5A project to motor structures such as the striatum (Fig. 1).

The space of movements is essentially continuous: an arm can move to any position in 3D space, and the eyes can move to any position in their 3D space, without discrete stable positions. The suggestion on this basis is that the L5A pyramidal cells support a separate attractor network from that in L3, and in any case does not contribute to the L3 attractor network. One possibility is that L5A operates as a continuous attractor network, which as described by Rolls (2016) can be stable at any position in a continuous space. A potential advantage of having a continuous attractor here is that all the neurons that are active for any position in the space can become associated with each other, thus smoothing the transitions through the space by making sure that all neurons relevant to a particular position in the space based on previous experience are recruited into activity.

In this context, it should be noted that networks with associatively modifiable recurrent collateral synapses do not necessarily need to have a positive feedback gain sufficiently high to maintain the activity in a high firing rate attractor state even when there is no input. The network with the structure of an attractor network can nevertheless perform constraint satisfaction, by providing a moderate input from other neurons in the network to add to and smooth whatever inputs are providing the main driving input to the network, which in this case would be a L2/L3 input driving the L5 pyramidal cell firing. This would help to provide a smooth trajectory through the continuous attractor space of movements, without jerks and incompatible motor commands being produced, as described by Rolls (2016).

An interesting implication of this hypothesis is that the L3 attractor networks which may often (especially in the 'what' and semantic memory cortical areas) support discrete attractors, may operate as a largely separate attractor system from the L5A continuous attractor network system, as the computational spaces that they encode are different. Consistent with this at least partial separation of superficial and deep layer attractor network system, the dendrites of deep pyramidal cells are largely confined to the deep layers of the neocortex apart from the branches in layer 1 where backprojections are received; and the dendrites of superficial layer pyramidal cells are largely confined to the superficial layers of the cortex, as illustrated in Fig. 1. The main relation it is proposed between the superficial layer discrete and the deep layer continuous attractor systems of the neocortex is that the superficial cortical layers should provide the main driving input to the deep cortical layers. Treves (2003) and Montagnini and Treves (2003) also pointed to the computational advantage to having cortical layers with sparse distributed encoding of a discrete or 'what' representation of stimulus identity, separate from a 'where' topographically mapped more spatially continuous representation. The relation between what and where information in a cortical patch has also been considered by Roudi and Treves (2008), though in that study they did not consider the possibility of separate attractor networks in the superficial and deep layers of the neocortex.

In this context, the second hypothesis investigated here in Experiment 2 by integrate-and-fire simulations is that the superficial and deep layers support somewhat different attractors, with the superficial layers being specialised for feedforward processing involving categorisation into discrete representations using competitive learning to discover new categories, based on the convergence of inputs from different parts of the preceding cortical area. As a result, the superficial layers would implement discrete attractor networks with the recurrent collaterals. In contrast, the deep layers, in particular layer 5A with its outputs to motor systems including via the striatum/ basal ganglia may need a different type of attractor, which though initiated by the superficial layers, might be smoother spatially to make movements smooth, by implementing what is more like a continuous attractor

network. In Experiment 2, the way in which the deep layers of the neocortex are set up to produce these somewhat different types of representation is investigated. Although tested with continuous attractor dynamics in the deep layer, the aim of Experiment 2 is to show more generally whether the superficial and deep layers could support somewhat different representations implemented by differences in the connectivity between the neurons.

2. Experiment 1: The effects of adaptation in the deep layers of the neocortex on the time course of the firing

The faster time course of the deeper layers might be implemented by having relatively fast synaptic or neuronal adaptation (Rolls (2016) and Markram et al. (2015) who show that adaptation varies greatly between different classes of cortical neuron). In this experiment, both the superficial and the deep layers were discrete attractor networks, but neuronal adaptation was present in the deep layer attractor network to investigate the hypothesis. A faster time course in deep layers might also be produced by a relatively higher proportion of AMPA or kainate receptors with their short time constant (5–10 ms) in the deep layers, with relatively more NMDA receptors with their relatively long time constants (100 ms) in superficial layers. Consistent with this hypothesis, NMDA receptors are present in relatively higher densities in superficial layers (L2/L3), and kainate receptors in relatively higher densities in deep layers (L5 and L6) of the human fusiform gyrus, a ventral stream visual area involved in face and object processing (Caspers et al., 2015). In addition, the pyramidal cells in the deep layers probably have different time constants, for they tend to be large, related in part to the fact that many of their axons have to travel long distances to reach their motor targets and therefore need fast conduction velocities.

2.1. Methods

Two attractor networks were implemented as illustrated in Fig. 3. The two networks are connected by synapses from the superficial to the deep layer network with strength w_{SD}.

The aim is to investigate the operation of the system in a biophysically realistic attractor framework, so that the properties of receptors, synaptic currents and the statistical effects related to the probabilistic spiking of the neurons can be part of the model. We use a minimal architecture, a single attractor or autoassociation network (Amit, 1989; Hertz, Krogh, & Palmer, 1991; Hopfield, 1982; Rolls, 2008; Rolls & Deco, 2002; Rolls & Treves, 1998) for each module. A recurrent (attractor) integrate-and-fire network model which includes synaptic channels for AMPA, NMDA and $GABA_A$ receptors (Brunel & Wang, 2001; Rolls & Deco, 2010) was used.

The single superficial and the deep discrete attractor networks each contained 800 excitatory, and 200 inhibitory neurons, which is consistent with the observed proportions of pyramidal cells and interneurons in the cerebral cortex (Abeles, 1991; Braitenberg & Schütz, 1991). The connection strengths are adjusted using mean-field analysis (Brunel & Wang, 2001; Deco & Rolls, 2006; Rolls & Deco, 2010; Rolls, 2016), so that the excitatory and inhibitory neurons exhibit a spontaneous activity of 3 Hz and 9 Hz, respectively (Koch & Fuster, 1989; Wilson, O'Scalaidhe, & Goldman-Rakic, 1994). The recurrent excitation mediated by the AMPA and NMDA receptors is dominated by the NMDA current to avoid instabilities during delay periods (Wang, 2002).

The architecture of the cortical network module which is the architecture of both the superficial and the deep attractor networks as illustrated in Fig. 4 has 10 selective pools each with 80 excitatory neurons. The connection weights between the neurons within each pool or population are called the intra-pool connection strengths $w+$, which were set to values in the range 1.9–2.2 for the simulations described. All other weights including w_{inh} were set to 1.

All the excitatory neurons in each attractor pool S1, S2,…,SN receive an external bias input $\lambda_1,\lambda_2,...,\lambda_N$. This external input consists of

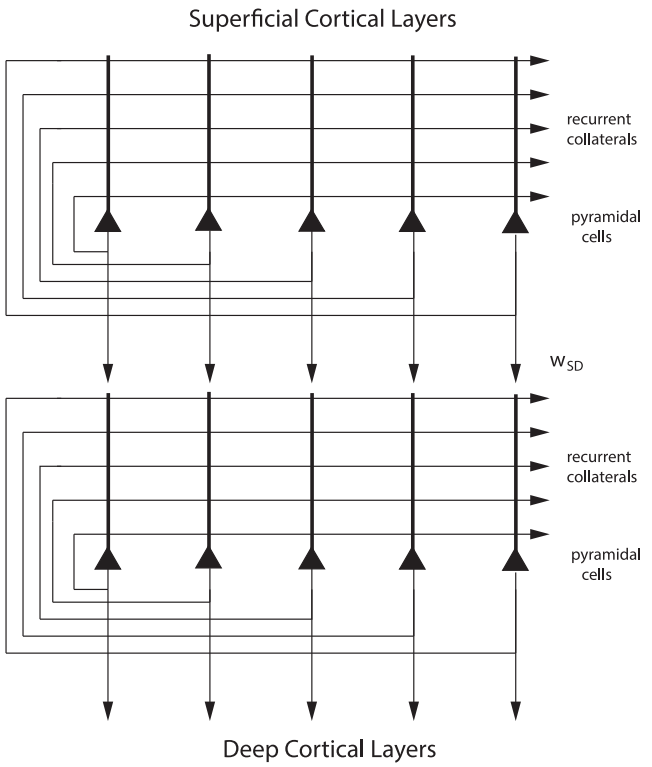

Fig. 3. Schematic drawing of the cortical architecture simulated in Experiment 1 with a discrete attractor network in the superficial layers connecting to a discrete attractor network in the deep layers. One local patch or module or column of neocortex was modelled. Pyramidal cells are represented by filled triangles, their dendrites are indicated by thick lines above the cell body, and the axons are represented by thin lines. Synapses are present wherever an axon crosses a dendrite. The superficial layers (2 and 3) were modelled as a fully connected discrete attractor network, which connected via synapses of strength w_SD to the network in the deep layers also modelled as a fully connected discrete attractor network. The difference between the two networks was that there was adaptation of the neuronal firing in the deep layer network. Both networks were integrate-and-fire attractor networks with n possible attractor states. Five populations of neurons, each implementing one of the attractor states are signified by the five pyramidal cells drawn for each network. For the simulations described here there were $n = 10$ orthogonal attractor states in the each network, each implemented by a population of neurons with strong excitatory connections between the neurons with value w_+ for the NMDA and AMPA synapses. There were $N_s = 1000$ neurons in each network, of which 0.8 were excitatory (pyramidal cells), and 0.2 were inhibitory (not illustrated).

Poisson external input spikes via AMPA receptors which are envisioned to originate from 800 external neurons. One component of this bias which is present by default arrives at an average spontaneous firing rate of 3 Hz from each external neuron onto each of the 800 synapses for external inputs, consistent with the spontaneous activity observed in the cerebral cortex (Rolls, 2016; Rolls & Treves, 1998; Wilson et al., 1994).

Both excitatory and inhibitory neurons are represented by a leaky integrate-and-fire model (Tuckwell, 1988). The basic state variable of a single model neuron is the membrane potential. It decays in time when the neurons receive no synaptic input down to a resting potential. When synaptic input causes the membrane potential to reach a threshold, a spike is emitted and the neuron is set to the reset potential at which it is kept for the refractory period. The emitted action potential is propagated to the other neurons in the network. The excitatory neurons transmit their action potentials via the glutamatergic receptors AMPA and NMDA which are both modeled by their effect in producing exponentially decaying currents in the postsynaptic neuron. The rise time of the AMPA current is neglected, because it is typically very short (<1 ms). The NMDA channel is modeled with an alpha function including both a rise and a decay term. In addition, the synaptic function of the NMDA current includes a voltage dependence controlled by the

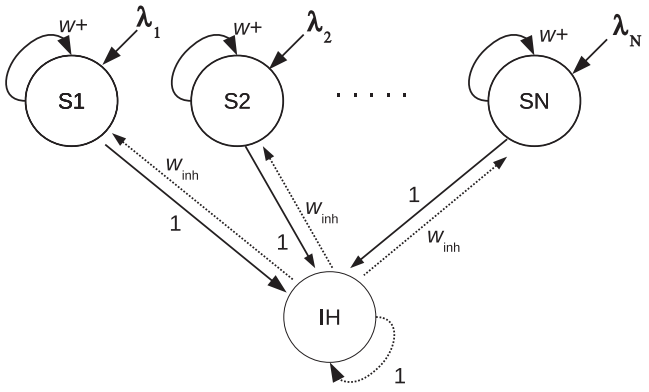

Fig. 4. The architecture of a superficial or deep module containing one fully connected discrete attractor network. The excitatory neurons are divided into N = 10 selective pools or neuronal populations S1–SN of which three are shown, S1, S2 and SN. The synaptic connections have strengths that are consistent with associative learning. In particular, there are strong intra-pool connection strengths w_+. The excitatory neurons receive inputs from the inhibitory neurons with synaptic connection strength w_{inh} = 1. The other connection strengths are 1. The integrate-and-fire spiking module contained 500 neurons, with 40 in each of the 10 non-overlapping excitatory pools, and 100 in the inhibitory pool IH. Each neuron in the network also receives external Poisson inputs λ_{ext} from 800 external neurons at a typical rate of 3 Hz/synapse to simulate the effect of inputs coming from other brain areas.

extracellular magnesium concentration (Jahr & Stevens, 1990). The inhibitory postsynaptic potential is mediated by a $GABA_A$ receptor model and is described by a decay term. A detailed mathematical description is provided in the Appendix.

A property of cortical neurons is that they tend to adapt with repeated input (Abbott, Varela, Sen, & Nelson, 1997; Fuhrmann, Markram, & Tsodyks, 2002). The mechanism is understood as follows. The afterpolarization (AHP) that follows the generation of a spike in a neuron is primarily mediated by two calcium-activated potassium currents, I_{AHP} and the sI_{AHP} (Sah & Faber, 2002), which are activated by calcium influx during action potentials. The I_{AHP} current is mediated by small conductance calcium-activated potassium (SK) channels, and its time course primarily follows cytosolic calcium, rising rapidly after action potentials and decaying with a time constant of 50 to several hundred milliseconds (Sah & Faber, 2002). In contrast, the kinetics of the sI_{AHP} are slower, exhibiting a distinct rising phase and decaying with a time constant of 1–2 s (Sah, 1996). A variety of neuromodulators, including acetylcholine (ACh) acting via a muscarinic receptor, noradrenaline, and glutamate acting via G-protein-coupled receptors, suppress the sI_{AHP} and thus reduce spike-frequency adaptation (Nicoll, 1988).

When recordings are made from single neurons operating in physiological conditions in the awake behaving monkey, peristimulus time histograms of inferior temporal cortex neurons to visual stimuli show only limited adaptation. There is typically an onset of the neuronal response at 80–100 ms after the stimulus, followed within 50 ms by the highest firing rate. There is after that some reduction in the firing rate, but the firing rate is still typically more than half-maximal 500 ms later (see example in Tovee, Rolls, Treves, & Bellis (1993)). Thus under normal physiological conditions, some, but limited, firing rate adaptation can occur.

The effects of this adaptation can be studied by including a time-varying intrinsic (potassium-like) conductance in the cell membrane (Brown, Gähwiler, Griffith, & Halliwell, 1990; Rolls, 2016; Treves, 1993). This can be done by specifying that this conductance, which if 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, & Andersen (1984), Mason & Larkman (1990)). The details of the implementation used are described in Section A.2.

2.2. Results

In the first experiment to illustrate the effects of adaptation in the deep discrete attractor network, after 500 ms of spontaneous activity, pool 5 was activated for 100 ms, and its firing stopped within 200 ms, as $w+$ was set to 1.9 to help the superficial attractor to stop firing relatively soon after stimulus offset (Fig. 5). Through the connection to the deep discrete attractor network with strength w_{SD} = 0.15, pool 5 in

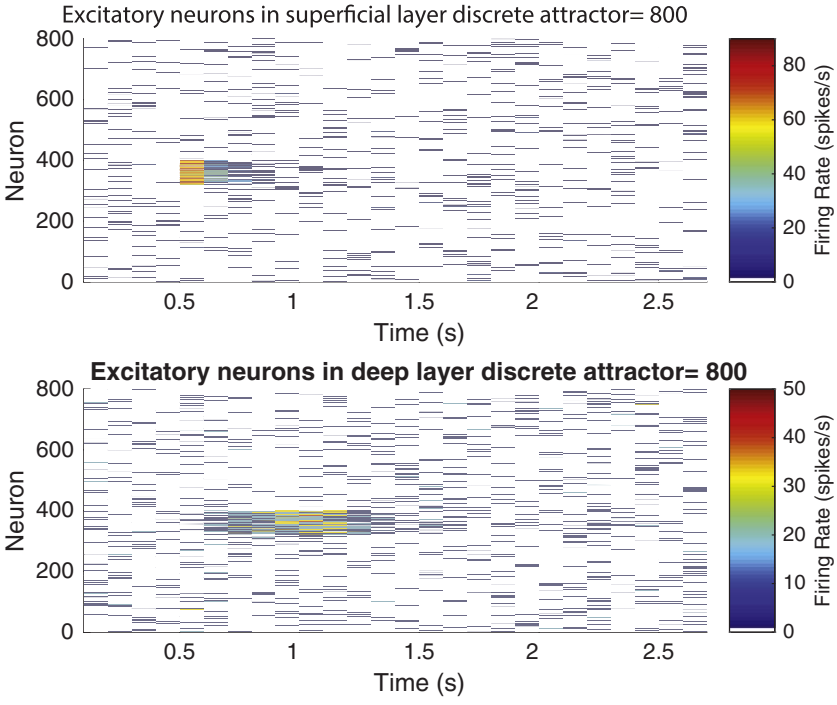

Fig. 5. Effects of adaptation on the time course of the deep layer discrete attractor in Experiment 1. Simulation of the operation of the superficial discrete attractor network (above) and the deep discrete attractor network (below), both of which contained 800 excitatory neurons. Time from 0 to 2.7 s is shown along the abscissa. Neurons 0–800 above are the excitatory neurons in the 10 pools in the superficial module discrete attractor network. Neurons 0–800 below are the excitatory neurons in the 10 pools in the deep module discrete attractor network. The colour indicates the firing rate of each neuron, with the calibration bar showing the rate in spikes/s measured over 100 ms epochs. After 500 ms of spontaneous activity, at time 500–600 ms pool 5 of the superficial discrete attractor network received an external input, and after a short period, it entered a high firing rate attractor state which died off after the stimulus to pool 5 was removed. Via the connections from the superficial to the deep module, neurons in the corresponding pool 5 of the deep discrete attractor network started firing, showing that the deep network was operating as an attractor. However, with the neuronal adaptation, the firing in the deep network continued for only approximately 800 ms. Without the neuronal adaptation the firing in the deep network continued indefinitely. For parameters, see text. Additional parameters: $w+$ for deep attractor = 2.2; AHPCaD = 1.0; inh0n = 1.00; inhS = 1.05; inhD = 1.07. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

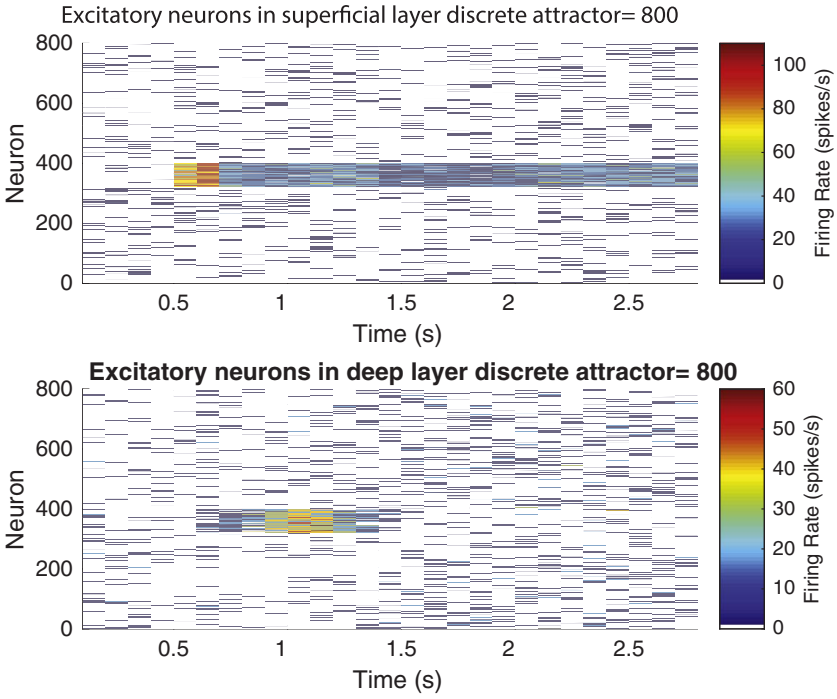

Fig. 6. Effects of adaptation on the time course of the deep layer discrete attractor in Experiment 1. Simulation of the operation of the superficial discrete attractor network (above) and the deep discrete attractor network (below), both of which contained 800 excitatory neurons. Time from 0 to 2.7 s is shown along the abscissa. Neurons 0–800 above are the excitatory neurons in the 10 pools in the superficial module discrete attractor network. Neurons 0–800 below are the excitatory neurons in the 10 pools in the deep module discrete attractor network. The colour indicates the firing rate of each neuron, with the calibration bar showing the rate in spikes/s measured over 100 ms epochs. After 500 ms of spontaneous activity, at time 500–700 ms pool 5 of the superficial discrete attractor network received an external input, and after a short period, it entered a high firing rate attractor state which continued indefinitely after the stimulus to pool 5 was removed because its $w+$ was 2.05. Via the connections from the superficial to the deep module, neurons in the corresponding pool 5 of the deep discrete attractor network started firing. However, with the neuronal adaptation, the firing in the deep network continued for only approximately 700 ms. Without the neuronal adaptation the firing in the deep network continued indefinitely. For parameters, see text. Additional parameters: $w+$ for deep attractor = 2.2; AHPCaD = 0.4; inh0n = 1.00; inhS = 1.02; inhD = 1.08. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the deep attractor started firing, and continued firing for some time with its $w+$ =2.2. This demonstrates that the deep network was operating as a discrete attractor network that could maintain its firing for longer than the input from the superficial layers was firing. However, with the neuronal adaptation switched on (with τCa = 150 ms and AHP-Ca = 1.0), the deep attractor network did stop firing after approximately 800 ms of activity. If the calcium-implemented neuronal adaptation was switched off, the deep attractor continued to fire indefinitely.

This simulation thus shows that adaptation can set the time course of the deep network when it has been demonstrated to be operating as an attractor (Fig. 5).

In the second experiment to illustrate the effects of adaptation in the deep discrete attractor network, after 500 ms of spontaneous activity, pool 5 was activated for 200 ms, and its firing continued indefinitely, as $w+$ was set to 2.05 (Fig. 6). Through the connection to the deep discrete attractor network with strength w_{SD} = 0.1, pool 5 in the deep attractor started firing, and continued firing for some time with its $w+$ =2.2. However, with the neuronal adaptation switched on (with τCa = 1000 ms and AHP-Ca = 0.4), the deep attractor network did stop firing after approximately 700 ms of activity. If the calcium-implemented neuronal adaptation was switched off, the deep attractor continued to fire indefinitely.

This simulation thus shows that adaptation can set the time course of the deep network even when it continues to receive input from the superficial attractor network (Fig. 6).

Another key issue investigated was the strength of the coupling between the attractor networks in the superficial and deep layers that would allow them to show the partially separate operation described here. For the adaptation effects illustrated in Fig. 6, the coupling term was set to w_{SD} = 0.1, which corresponds to the ratio of the excitatory synaptic weights received by a neuron in the deep layer from the superficial layer to the weights from all recurrent excitatory connections from the other neurons in the deep layers. This value for w_{SD} of 0.1 was just sufficient to enable an attractor state in the deep layer network to be started from the spontaneous state by firing in one of the attractors in the superficial layer network. At the time point in the simulation when the superficial attractor had started, but the deep attractor had not yet started, the ratio of the currents was 0.21. That therefore helps to set a minimal value for the currents that must be received from the superficial layer attractor by the neurons in a deep layer attractor network. If the value of w_{SD} was increased to 0.5, then at the same time in the simulation, the ratio of the currents was 0.74, and yet the deep attractor could still operate partly independently of the superficial attractor in that the deep attractor showed adaptation, whereas the superficial attractor did not. Another key condition for w_{SD} and the currents that it can induce is that it must be sufficiently large to switch an attractor state in the deep layers of the cortex to another state when the superficial layer attractor changes from one population of neurons to another. In a deep network without adaptation, it was found that w_{SD} had to be approximately 1.0 to reliably switch the deep attractor, and the current ratio (input currents from the superficial network via w_{SD}/deep net recurrent collateral currents) was 0.21. This higher value for w_{SD} is expected because once a network is in an attractor, it is stable and difficult to perturb (Rolls, 2016). This analysis thus helps to provide a framework with which to assess, given the connectivity values shown in Fig. 2, whether the superficial and deep networks are connected so that they can operate as coupled but somewhat independent attractor networks (see Discussion).

3. Experiment 2: A discrete attractor network for the superficial layers and a continuous attractor network for the deep layers

In Experiment 2 the aim was to investigate whether the deep and superficial layer recurrent collaterals could implement different types of attractor network and computations. This was investigated with an integrate-and-fire network of a small patch or module or column of neocortex in which the superficial layers were modelled as a fully connected discrete attractor network, which connected with stronger forward than backward connections to the deep layers modelled as an integrate-and-fire continuous attractor network with strong local connectivity between nearby neurons, as illustrated in Fig. 7.

3.1. Method

The superficial layers were modelled as a discrete attractor network as illustrated in Fig. 4. The network is an integrate-and-fire attractor network with n possible attractor states. For the simulations described

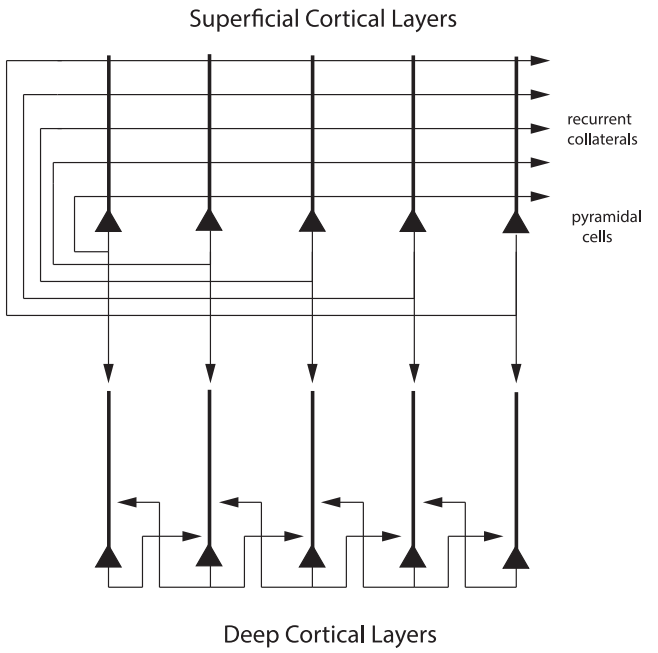

Fig. 7. Schematic drawing of the cortical architecture simulated in Experiment 2. One local patch or module or column of neocortex was modelled. Pyramidal cells are represented by filled triangles, their dendrites are indicated by thick lines above the cell body, and the axons are represented by thin lines. Synapses are present in the superficial layers wherever an axon crosses a dendrite; and in the deep layers by where the arrows at the end of each axon point to a dendrite. The superficial layers (2 and 3) were modelled as a fully connected discrete attractor network. This is an integrate-and-fire attractor network with n possible attractor states. Five populations of neurons, each implementing one of the attractor states are signified by the five pyramidal cells drawn. For the simulations described here there were $n = 10$ orthogonal attractor states in the superficial layer of the module, each implemented by a population of neurons with strong excitatory connections between the neurons with value $w_+ = 2.1$ for the NMDA and AMPA synapses. There were always $N_s = 500$ neurons in the superficial layer of the module, of which 0.8 were excitatory (pyramidal cells), and 0.2 were inhibitory (not illustrated). The deep layers (layer 5) were modelled as an integrate-and-fire continuous attractor network. The main difference was that the (excitatory) recurrent collaterals connected more strongly to their neighbours than to more distant neurons, as shown schematically. The number of neurons in the deep, continuous attractor, network was typically $N = 500$ neurons, but in some investigations was set to $N = 125$, 250, 1000 or 1500 neurons. Each of the $n = 10$ local regions of the deep layers received inputs from one of the n discrete attractor populations of neurons in the superficial layers, in a way that could be set up by self-organizing topographic map formation.

here there were $n = 10$ orthogonal attractor states in the superficial layer discrete attractor network, each implemented by a population of neurons with strong excitatory connections between the neurons with value $w_+ = 2.1$ for the NMDA and AMPA synapses. There were always $N_s = 500$ neurons in the module, of which 0.8 were excitatory, and 0.2 were inhibitory. The parameters in the module were set so that the spontaneous firing state with no input applied was stable, and so that only one of its n possible attractor states was active at any time when inputs were applied (Deco et al., 2013; Rolls, 2016; Rolls & Deco, 2010) (see Appendix).

The deep layer module was modelled as an integrate-and-fire continuous attractor network. The main difference was that the (excitatory) recurrent collaterals connected more strongly to their neighbours than to more distant neurons (see Fig. 7), with a Gaussian function of width σ setting the width of the profile. σ was typically 15 neurons, but other values in the range 5–140 were investigated where specified. The synaptic strengths of these recurrent collaterals were scaled to be proportional to the height of the Gaussian distribution. These Gaussian connections were set to wrap from one end of the array of neurons back to the beginning, so that this was a ring continuous attractor network. The sum of all the excitatory input synaptic strengths on each neuron was set to the same value as the sum for the equivalent integrate-and-fire discrete attractor network described above, in order to set the network to operate in the same regime of balanced excitation and inhibition (Brunel & Wang, 2001; Deco et al., 2013; Rolls, 2016; Rolls & Deco, 2010). The number of neurons in the deep, continuous attractor, network was typically $N = 500$ neurons, but in some investigations was set to $N = 125$, 250, 1000 or 1500 neurons. In the deep continuous attractor network, 0.8 of the neurons were excitatory and 0.2 were inhibitory.

The deep module, the continuous attractor network, operated with similar integrate-and-fire neurons. The inhibition in the continuous attractor network was set to 1.03, and a target parameter that adjusted the equivalent of $w+$ for the deep network was set to a value that enabled the continuous attractor network to maintain its activity once started (and in the absence of more than spontaneous firing in the superficial network) but with a peak firing rate still in the biological range of less than approximately 100 spikes/s (Rolls, 2016; Rolls & Treves, 2011).

It is noted that this type of continuous attractor network is described as having metric connectivity, in that the recurrent connections between the pyramidal cells are short range (Roudi & Treves, 2008). This can be set up by self-organising learning, as illustrated by Rolls (2016). A different scenario for a continuous attractor network is when the connectivity is global, and the continuous attractor network is set up by learning which of the neurons with a Gaussian tuning profile are close together in the space being represented while the agent or animal traverses the space repeatedly (Rolls, 2016).

An important difference in these discrete and continuous attractor networks is in the learning. In a discrete attractor network it is assumed that a random set of neurons is caused to fire by the inputs being received. Competitive learning may be involved in this process (Rolls, 2016). The recurrent collateral connections between the neurons then show associative (Hebbian) synaptic modification. The set of currently active neurons thereby form a discrete attractor, as described by Rolls (2016). A large number of independent representations can be set up in this way. In contrast, in a continuous attractor network, neurons that are nearby in a space tend to excite each other depending on their distance from each other in the space. One way in which such firing might be set up is if the input being received is in some continuous space, such as a space of head direction, or spatial view, or movements being made in 3D space. In this case, Hebbian synaptic modification allows the continuous space to be learned by the continuous attractor network, and near points in the space need not be represented by nearby neurons. Another way in which such continuous representations can be set up is by short-range recurrent collaterals, which, because they are short range and their density falls off with distance across the cortex, tend to make nearby neurons in the cortex have activity to similar stimuli. This is how topographic maps can be formed in the cortex, with an example of such a map being that in the primary visual cortex. These concepts are described in more detail by Rolls (2016). For the present work, which of these two methods is used to set up the continuous attractor network does not matter in principle, but in practice we set up the continuous attractor network by setting up the recurrent collaterals to have a strength of connection to neighbouring pyramidal cells that falls off as a function of distance between the pyramidal cells, which is a property of neocortex (Holmgren et al., 2003).

In the model, there are forward connections from excitatory neurons in the superficial (discrete attractor) module to excitatory neurons in the deep (continuous attractor) module, with strength w_{SD}. The role of these forward connections is to provide the input to the deep module. This mapping was by default a 1–1 mapping, so that with 400 excitatory neurons in the superficial module, there were 1:1 connections to the 400 excitatory neurons in the deep module. If there were 800 neurons in the deep module, there was still a regular mapping to the deep module, but each superficial neuron connected to two adjacent deep layer neurons. If there were fewer neurons in the deep continuous

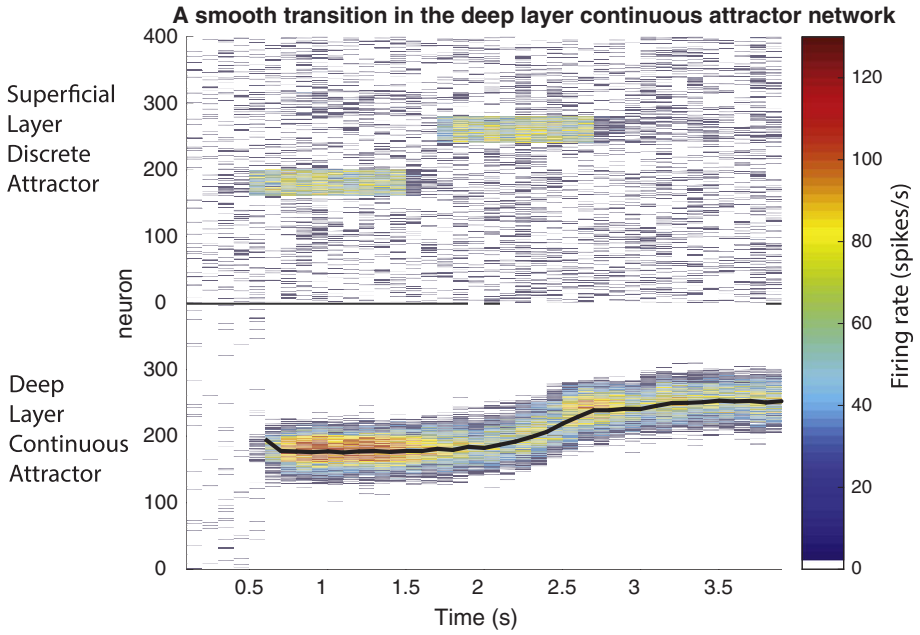

Fig. 8. A smooth transition in the deep continuous attractor network. Simulation of the operation of the superficial discrete attractor network (above) and the deep continuous attractor network (below). Time from 0 to 4 s is shown along the abscissa. Neurons 0–400 above are the 10 pools in the superficial module discrete attractor network. Neurons 0–400 below are the excitatory neurons in the deep module continuous attractor network. The colour indicates the firing rate of each neuron, with the calibration bar showing the rate in spikes/s measured over 100 ms epochs. At time 500–1500 ms pool 5 of the superficial discrete attractor network received an external input, and after a short period, it entered a high firing rate attractor state which died off after the stimulus to pool 5 was removed. Via the connections from the superficial to the deep module, neurons in the corresponding part of the continuous attractor network started firing. The width (standard deviation) of the Gaussian profile for the recurrent connectivity synaptic strength was 60 neurons. After a delay of 200 ms, pool 7 was activated by an external input, from time = 1700–2700 ms. The discrete attractor pool 7 entered a high firing rate attractor state, which died off some time after the stimulus to pool 7 was removed. The figure shows that the deep continuous attractor packet or bubble of activity moved somewhat continuously from its position when it was being activated by the input from superficial pool 5 to the position when it was being activated by the input from superficial pool 7. The centre of gravity of the deep (continuous attractor) pool is shown by the black line. The figure shows that no such continuous transition occurred in the superficial, discrete attractor, network. The spontaneous firing rate of the neurons in the deep layer not in the packet of activity was inhibited in this and the following figures by inhibition received through the inhibitory neurons from the excitatory neurons firing fast in the continuous attractor. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

attractor network, there was still a regular mapping to the deep attractor, but superficial neurons were skipped. For example, if there were half as many neurons in the deep attractor, then every other superficial neuron had a connection to a deep neuron. These connections were used in the stability investigations only to start the deep attractor, after which the nature of the drift was explored without any further input from the superficial network. As was the case for the superficial network, the neurons in the deep network always had an external Poisson input to each of 800 synapses on each neuron set to a default external spontaneous rate of 3 spikes/s on every one of these 800 synapses.

The stability analyses were performed as follows. The deep module continuous attractor was set up as a ring attractor, that is, the end of the excitatory firing rate array was connected back to the start with the local recurrent collateral connections. There were 10 sectors in the ring attractor, in that each of the ten superficial discrete attractors was connected to one tenth of the excitatory neurons in the deep attractor. The amount of drift was measured by the number of the 10 sectors that the centre of gravity of the bubble of activity drifted in 12 s (or by the number of neurons over which the drift occurs). As will be shown, the drift was typically less than the width of one of the 10 sectors in the ring attractor. Simulations were performed with different numbers of neurons in the deep module, to investigate how the size of the continuous attractor and therefore the finite size noise influenced the drift and stability. In addition to the total drift over 12 s, a second measure was the short-term movement, again measured in sectors or in neurons, that occurred between the centre of gravity of the bubble measured every 100 ms.

During operation, one attractor pool in the superficial module received an increased input for 1000 ms starting after an initial 500 ms period of spontaneous firing. That started firing in the deep module.

3.2. A hypothesis: adaptation must be minimal in a continuous attractor network

It is hypothesized here that any adaptation in a continuous attractor network will impair its stability, because when neurons in the packet of activity adapt, the lower firing neurons at the edge of the activity packet will be less adapted, and the packet of activity will drift towards the less adapted neurons. This is hypothesized here to be an important constraint on the operation of continuous attractor networks in the brain. An implication and prediction is that states in which adaptation is more likely to occur, including drowsiness when cholinergic neuronal firing may be reduced, or the use of a cholinergic blocker such as scopolamine, may impair tasks that involve idiothetic (i.e. self-motion) update of position such as head direction, eye position and how far one has travelled. Cholinergic projections from the basal forebrain are important in minimizing cortical adaptation, and the basal forebrain neurons are activated by rewarding, punishing, and novel stimuli (Burton, Rolls, & Mora, 1976; Mora, Rolls, & Burton, 1976; Rolls, 2016, 2014; Rolls & Deco, 2015b; Rolls, Burton, & Mora, 1976; Rolls, Sanghera, & Roper-Hall, 1979; Wilson & Rolls, 1990b, 1990a, 1990c), all of which help to maintain a non-adapted cortex. (Scopolamine is frequently prescribed for travel sickness, and it is suggested might operate by reducing the stability of continuous attractor networks, so that there would be less discrepancy between them and visual motion cues, in that the continuous attractor networks may be more easily driven by visual inputs.)

To investigate the effect of adaptation on the operation of a continuous attractor network, a spike-frequency adaptation mechanism was implemented in some of the simulations using Ca^{2+}-activated K^{+} hyper-polarizing currents (Liu & Wang, 2001), as described in detail in the Appendix in Section A.2. Its parameters were chosen to produce spike frequency adaptation similar in timecourse to that found in the inferior temporal visual cortex of the behaving macaque (Tovee et al., 1993). In particular, $[Ca^{2+}]$ is initially set to be 0 μM, τ_{Ca} = 300 ms, α = 0.002, V_K = −80 mV and g_{AHP} = 200 nS unless otherwise specified. (We note that there are a number of other biological mechanisms that might implement the slow transitions from one attractor state to another, some investigated by Deco & Rolls (2005), and that we use the spike frequency adaptation mechanism to illustrate the principles of operation of the networks.)

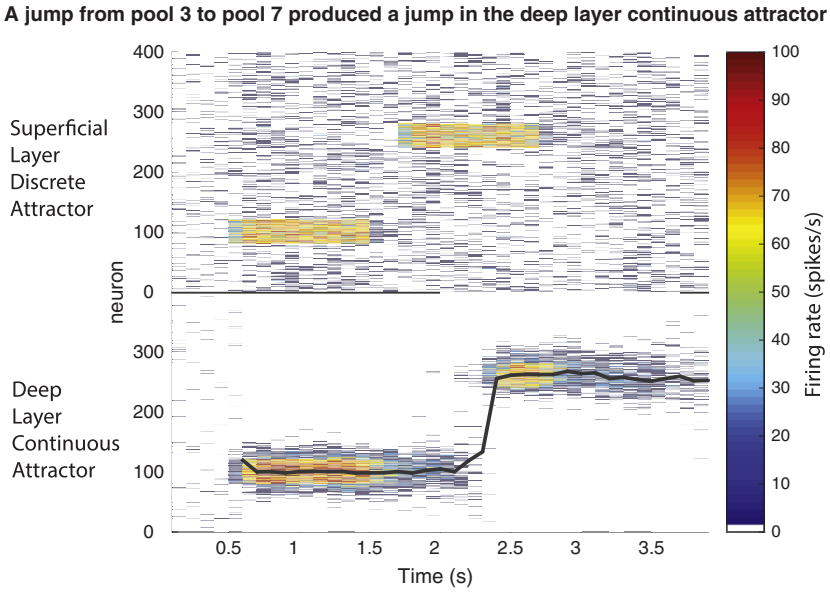

Fig. 9. A jump transition in the deep continuous attractor network. Simulation of the operation of the superficial discrete attractor network (above) and the deep continuous attractor network (below). Time from 0 to 4 s is shown along the abscissa. Neurons 0–400 above are the 10 pools in the superficial module discrete attractor network. Neurons 0–400 below are the excitatory neurons in the deep module continuous attractor network. The colour indicates the firing rate of each neuron, with the calibration bar showing the rate in spikes/s measured over 100 ms epochs. At time 500–1500 ms pool 3 of the superficial discrete attractor network received an external input, and after a short period, it entered a high firing rate attractor state which died off after the stimulus to pool 3 was removed. Via the connections from the superficial to the deep module, neurons in the corresponding part of the continuous attractor network started firing. The width (standard deviation) of the Gaussian profile for the recurrent connectivity synaptic strength was 60 neurons. After a delay of 200 ms, discrete superficial pool 7 was activated by an external input, from time = 1700–2700 ms. The discrete attractor pool 7 entered a high firing rate attractor state, which died off some time after the stimulus to pool 7 was removed. The figure shows that the deep continuous attractor packet or bubble of activity jumped without a continuous transition from its position when it was being activated by the input from superficial pool 3 to the position when it was being activated by the input from superficial pool 7. The centre of gravity of the deep (continuous attractor) pool is shown by the black line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

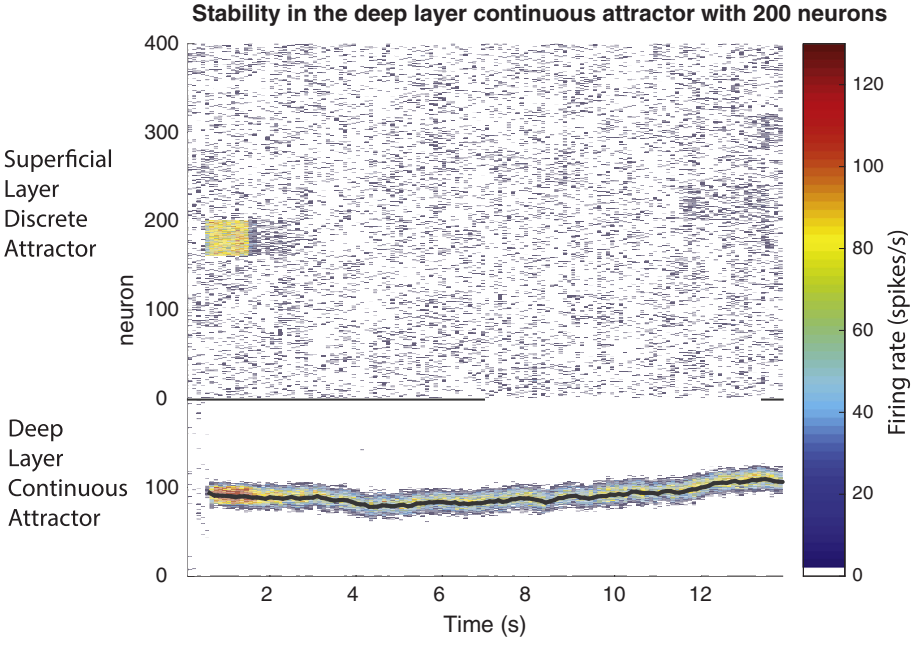

Fig. 10. Stability of the continuous attractor network with stochastic integrate-and-fire neurons. Simulation of the operation of the superficial discrete attractor network (above) and the deep continuous attractor network (below) for a case where there is a relatively small number of excitatory neurons, 200, in the deep attractor network. Time from 0 to 14 s is shown along the abscissa. Neurons 0–400 above are the 10 pools in the superficial module discrete attractor network. Neurons 0–200 below are the excitatory neurons in the deep module continuous attractor network. The colour indicates the firing rate of each neuron, with the calibration bar showing the rate in spikes/s measured over 100 ms epochs. At time 500–1500 ms pool 5 of the superficial discrete attractor network received an external input, and after a short period, it entered a high firing rate attractor state which died off after the stimulus to pool 5 was removed. Via the connections from the superficial to the deep module, neurons in the continuous attractor network started firing. The width (standard deviation) of the Gaussian profile for the recurrent connectivity synaptic strength was 15 neurons. The centre of gravity of the deep (continuous attractor) pool is shown by the black line. The figure shows that there was some short-term noise in the continuous attractor network, and some long-term drift over the 12 s period in which it received no external input. (Target = 0.35). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Results

3.3.1. Smooth transitions in the deep network continuous attractor module

If the deep network made a transition from one position corresponding to one of the superficial pools, to another position corresponding to another superficial pool, the transition under some circumstances of the deep pool could be continuous. This is illustrated in Fig. 8, in which a smooth transition occurred in the continuous attractor network when the superficial attractor network jumped from pool 3 (neurons 80–120 given that there were 400 excitatory neurons divided into 10 pools) to pool 5 (neurons 160–200). This was found to occur primarily when the width of the bubble spanned the first and second parts of the space over which a transition had to be made.

This may be helpful biologically, by smoothing transitions between nearby states.

When the transition being driven by the superficial network was larger than the width of the bubble, the bubble simply stopped in the first position, and started in the new position, with what appeared to be an abrupt jump. This is illustrated in Fig. 9, when the superficial network jumped from pool 3 to pool 7. This resulted in a jump in the deep continuous attractor network. This may be biologically plausible, for it may not be appropriate for the deep network to transition through all the intermediate positions when two distant positions for the superficial network are successively active.

A third scenario occurs when the deep continuous attractor network has entered a region of attraction, i.e. firing has started. The deep continuous attractor is then more stable than when its neurons are just firing at a spontaneous rate. Under these conditions, only a strong input from the superficial attractor will move the deep attractor to a new position. This may provide some form of noise tolerance for the deep, continuous attractor, network. The higher the firing rate of the continuous attractor, the more resistant it is to being perturbed by an

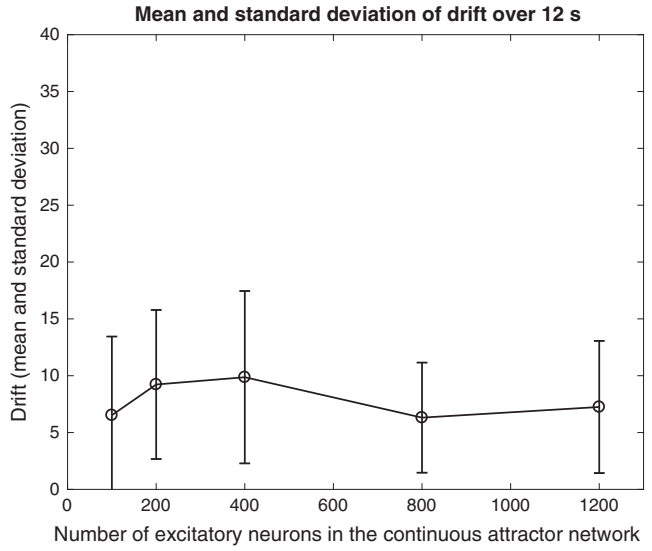

Fig. 11. Stability of the continuous attractor network with stochastic integrate-and-fire neurons. The mean and standard deviation of the movements of the centre of gravity of the packet of activity in the continuous attractor network from the beginning to the end of the 12 s period in which it received only spontaneous firing from the superficial network. The units are in neurons by which the absolute value of the centre of gravity moved. The X-axis shows the number of excitatory neurons in the continuous attractor network. The plot shows that there is little effect of the size of the continuous attractor network on the drift produced by noise due to the almost Poisson firing times of the neurons. This was because the packet of activity involved the same number of active neurons when the network was scaled up.

external input.

3.3.2. Stability of the integrate-and-fire stochastic spiking continuous attractor network

The stability of the integrate-and-fire continuous attractor network was measured in a 12 s period starting after time = 2 s in a simulation during which the bubble of activity was maintained. The start time was thus after the end of the period in which the superficial attractor had been active, as illustrated in Fig. 10. In this case, there were only 200 excitatory neurons in the deep continuous attractor network, and both short-term noise due to the statistical fluctuation in the numbers of spikes in the different neurons was evident, as well as long-term drift.

The effects of the size of the continuous attractor network on its long term drift were investigated as follows. The measure of stability was the long-term drift of the centre of gravity of the continuous attractor network, again after the driving input from the superficial attractor had been removed. The measure was the absolute value of the number of neurons over which the centre of gravity of the bubble had drifted in a 12 s period. Considerable stability was evident, as illustrated in Fig. 11. Indeed, the typical drift over 12 s was rather small (a drift of between 5 and 10 neurons, with standard deviation of approximately 5). There was no significant effect of N on the long-term drift. Although there may be short-term statistical fluctuations with the bubble moving left or right in a short time period in the order of 100 ms, these short term drifts largely cancel out over long periods.

It is notable in Fig. 11 that there was relatively little effect of the number of excitatory neurons N in the continuous attractor network on the number of neurons over which the packet of activity fluctuated or drifted. The implication is that if the continuous attractor ring represents the whole of a space, then having more neurons in the ring continuous attractor results in more stability of the representation expressed as the fraction of the ring (and of the whole space) over which the centre of gravity moves.

The factors that lead to the stability of even relatively small continuous attractor networks with spiking neurons are considered in the Discussion.

3.3.3. Effects of adaptation on the stability of the deep layer continuous attractor

Both synaptic and neuronal adaptation can occur, and can be modelled (Deco & Rolls, 2005). The hypothesis proposed is that adaptation would affect the stability of the continuous attractor network, for as neurons fire less because of adaptation, other neurons in the fringes of the continuous attractor bubble with lower firing rates but still being activated by the local recurrent collaterals would be less adapted, so would start to fire more than neurons at the centre of the activity packet. This was expected to cause the deep continuous attractor network to drift continuously across the space being represented.

This hypothesis was tested by implementing neuronal firing rate adaptation as described in Section A.2. This had the effect of causing the packet of activity to become unstable, and to drift continuously across the space being represented, as illustrated in Fig. 12. In this

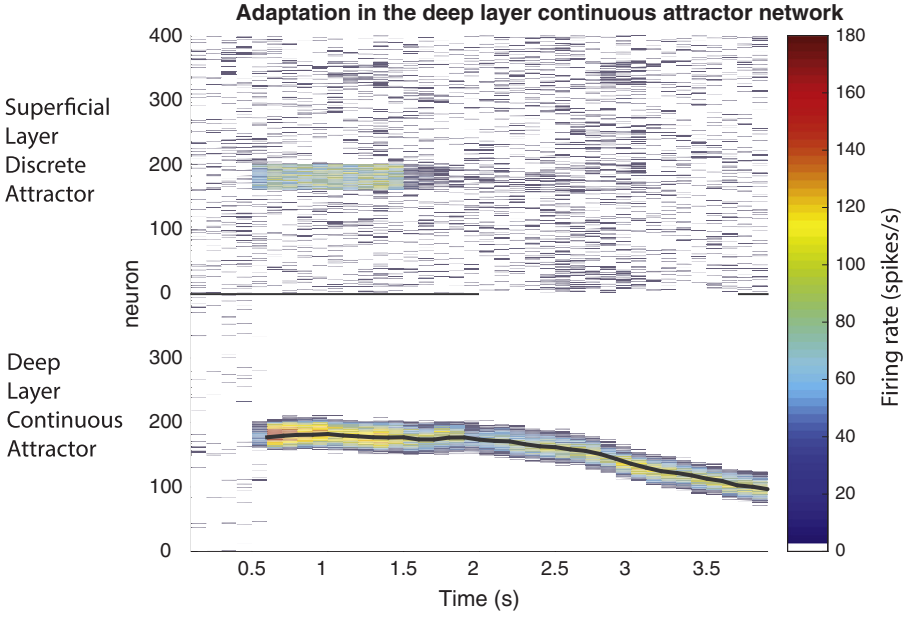

Fig. 12. Effects of adaptation on a continuous attractor network. Simulation of the operation of the superficial discrete attractor network (above) and the deep continuous attractor network (below). Time from 0 to 4 s is shown along the abscissa. Neurons 0–400 above are the 10 pools in the superficial module discrete attractor network. Neurons 0–400 below are the excitatory neurons in the deep module continuous attractor network. The colour indicates the firing rate of each neuron, with the calibration bar showing the rate in spikes/s measured over 100 ms epochs. At time 500–1500 ms pool 5 of the superficial discrete attractor network received an external input, and after a short period, it entered a high firing rate attractor state which died off after the stimulus to pool 5 was removed. Via the connections from the superficial to the deep module, neurons in the continuous attractor network started firing. The width (standard deviation) of the Gaussian profile for the recurrent connectivity synaptic strength was 15 neurons. The figure shows that when it was no longer clamped by the firing in the superficial network, the deep continuous attractor packet or bubble of activity drifted continuously from its position when it was being activated by the input from superficial pool 5 across the space mapped by the continuous attractor network. The centre of gravity of the deep (continuous attractor) pool is shown by the black line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

simulation τ_{Ca} = 300 ms, and the target parameter for the deep w+ was 3.3. On different trials with different random seeds the drift was in different directions, but once started, continued in the same direction, right round the ring attractor if left for sufficiently long. Any adaptation can thus impair the stability of a continuous attractor network.

4. Discussion

The results presented here for Experiment 1 show how the time course of the neuronal activity might be usefully different in the deep layers of the neocortex compared to the superficial layers. First it was shown that adaptation can set the time course of the deep network when it has been demonstrated to be operating as an attractor, that is, maintaining its firing for longer than the superficial layers which provide the input to the deep layers (Fig. 5). Second, it was shown that adaptation can limit to short firing the time course of the deep network even when it continues to receive input from the superficial attractor network (Fig. 6). Both these effects affects were implemented by having more adaptation in the deep layers of the cortex, consistent with experimental observations (Markram et al., 2015). Having a time course limited in this way for the a deep layer discrete attractor network could be useful if precisely timed outputs are needed to for example drive motor and related outputs. This would leave the superficial layers able to maintain their attractor state for longer, which could be useful in the training of the next layers up in the hierarchy, using for example slow learning (Rolls, 2012b, 2016).

A key issue in understanding whether the superficial and deep layers of the neocortex could operate as separate attractors with the superficial influencing the deep attractor networks but not fully determining their states is the nature of the connectivity from the superficial layers to the deep layers. One key condition is that the number of connections from neurons in the superficial layers to any single neuron in the deep layers must be in the order of the number of connections onto any one neuron in the superficial layers from the other neurons in the superficial layers via the recurrent collaterals. The quantitative argument is the same as for the number of backprojections in the cerebral cortex onto any one neuron compared to the number of connections between hippocampal CA3 neurons, and arises because if one can represent a certain number of memory states with the recurrent collaterals, then one needs to be able to transmit the same number of states (Rolls, 2016; Treves & Rolls, 1994). This condition appears to be approximately satisfied by the number of connections from L3 to each L5 neuron relative to the number of recurrent collaterals onto any one L3 neuron from other L3 neurons (Fig. 2a).

A second key condition is that the strength of the synaptic connections from the superficial layers to the deep layers must be sufficiently strong that these connections can start up an attractor state in the deep layers, or shift it from one attractor state to another. It is shown in Section 2.2 that for these effects, w_{SD} needs to be in the range of approximately 0.1–1.0, with the corresponding current ratios of forward to recurrent currents greater than 0.2. At the same time, some independence of operation of the two networks was found with w_{SD} = 0.5 and the current ratio 0.74, in that the deep but not the superficial network could then adapt out to produce a burst of firing in the deep layers but with a longer period of firing being maintained in the superficial layers. The calculated currents between neurons in different layers illustrated in Fig. 2b are fully consistent with the results described in the present investigations. (We note that the issue of whether the connectivity is diluted in each of the superficial and deep attractor networks is not the issue here. Dilution helps by reducing the distortion of the basins of attraction that would be produced if by chance there were many pairs of neurons with two or more recurrent collateral synapses between any pair of neurons (Rolls, 2012a). The issue here is of the ratio of the connectivity between and within coupled attractor networks (Rolls, 2016).)

Further, it must be remembered that what is shown in Fig. 2 is for the mouse primary somatosensory cortex, and the quantitative aspects of the recurrent collateral connectivity are likely to be considerably better for recurrent dynamics in higher cortical areas where the numbers of recurrent collaterals are so much larger, L4 is so much reduced, and when the currents between the neurons involved are measured when the networks are operating normally. We note that low values for the coupling between coupled attractors enables the weak interactions that are beneficial (Rolls, 2016; Renart, Parga, & Rolls, 1999b). A further line of evidence about whether the superficial and deep layers could operate as coupled attractors, not as a single attractor (Rolls, 2016), is that the superficial layers project to the deep layers, but not vice versa (Fig. 2a). This shows that the superficial and deep layers could not operate as a single attractor network, in that the deep layers do not participate at all strongly in the states reached by the neurons in the superficial layers. Instead, the evidence described here is consistent with the hypothesis that the superficial and deep layers of the neocortex could operate as partly separate attractor networks, with the superficial layers acting as an attractor network coupled to influence but not dominate the operation of an attractor network in the deep layers of the neocortex.

The results presented here for Experiment 2 show how the deep layers might perform a different type of computation to the superficial layers. In the example used to illustrate this potential principle of operation of the cerebral cortex, a continuous attractor network implemented by local recurrent collateral synaptic connections in the deep layers could enable smooth changes from one representation to another when the representations are relatively close in the space. It is also shown that in the integrate-and-fire continuous attractor network with neurons that fire stochastically the packet of neuronal firing is relatively stable. This may be because the short-term statistical fluctuations caused by the random spiking times of the neurons for a given mean rate tend to cancel out over periods of 100 ms or more. It is also shown that any neuronal adaptation would cause drift across the space being represented, with the implication that some continuous attractor networks in the brain should have little neuronal, or synaptic, adaptation, in order to provide for path integration (Rolls, 2016). On the other hand, in some parts of the cortex, some adaptation and therefore drift might have evolved as a mechanism for forward movement, rhythm, music, song and speech (Rolls & Deco, 2015a).

In relation to the stability of the continuous attractor network implemented with integrate-and-fire neurons, it was found that the long-term drift did not alter significantly as N increased (Fig. 11). An implication is that if a ring continuous attractor had to represent a certain physical space, say 360 degrees of space, then the net effect of having more neurons (i.e. increasing N) would mean less drift per unit of physical space. In effect, the measure of drift expressed as the proportion of the physical space being represented decreases as the size of the continuous attractor network, N, increases.

One possibility was that the drift in the continuous attractor network might have been greater with small networks, for then the statistical fluctuations in the firing rates would be larger. In discrete attractor networks, this variability can influence the decision-making (Deco & Rolls, 2006; Rolls, Grabenhorst, & Deco, 2010; Rolls, 2016), and has been observed in continuous attractor networks (Cerasti & Treves, 2013). However, in the present simulations there was not more drift in the smaller continuous attractor networks than the large networks (Fig. 11). The reason for our finding is that we kept the width of the bubble of activity constant when we increased the number of neurons in the network. Thus the number of neurons active in our simulations, and generating the statistical fluctuations in firing rates, was constant when we increased the size of the network, and that is the reason we propose why the stability measured by the drift did not alter significantly when we increased the size on the network. In contrast, Cerasti and Treves (2013) kept the spatial geometry of their inputs constant when they increased the size of the network, so that more neurons were allocated to a packet of activity as the number of neurons

in the network was increased. We hypothesize that it was because in their simulations more neurons were active as the size of the network increased that they observed a reduction of noise, evident as drift, as they increased the size of the network.

These new findings emphasise the following concepts. If more neurons are in a continuous attractor network, one scenario is that more spatial positions can be represented in the network, provided that the packet of activity does not involve more neurons when the network is scaled up, as in Fig. 11. This might correspond to more spatial scenes, or more environments, maps, or charts (Battaglia & Treves, 1998) being represented. This would greatly help the capacity of the system considered as an object-location associative memory system, in that many more such memories could be stored in such an attractor network with mixed discrete (for objects) and continuous (for scenes and places) representations (Rolls, 2016, 2017; Rolls, Stringer, & Trappenberg, 2002). This would be important in for example a hippocampal memory system (Rolls, 2016, 2017; Rolls et al., 2002). In so far as the hippocampal system involves associations between discrete objects and continuous spaces, we note that it would not operate well as a continuous attractor for spatial navigation, because the space would be distorted by the discrete object representations. A second scenario is that if there are more neurons in a continuous attractor network, then the number of places or scenes or maps might remain constant, with the consequence that there would then be more neurons in each packet of activity, which would reduce the drift. That is the situation investigated by Cerasti and Treves (2013). We note that even with the quite small networks investigated here, stability over 12 s was not a major issue, and so we believe that the ability to represent more locations, rather than to increase stability, is likely to be more biologically important.

In terms of physics, factors that influence the stability of a continuous attractor network include the following. The first effect is the statistical fluctuations in the firing rate referred to as “fast noise” (Cerasti & Treves, 2013). This shows finite size effects, with more noise (or fluctuations) of the firing rate that may produce a drift in activity of the population of neurons present in small networks. This can increase the drift in a continuous attractor network, or lead to more noisy decision-making, memory retrieval, etc. in a discrete attractor network (Deco et al., 2013; Rolls, 2016; Rolls & Deco, 2010). The second effect is the presence of “slow” or “quenched” noise which refers to the fact that with small N, the synaptic weights in the continuous attractor network have small discontinuities that result in “holes” in the attractor landscape, and these tend to restrict the free movement of the packet of activity when N is relatively small. This effect would tend to reduce drift in the continuous attractor network with small N. The slow and fast noise can thus operate in opposite directions to influence the stability of a continuous attractor network. The fast noise may be the more important factor, to judge from the results of Cerasti and Treves (2013).

Continuous attractors (in the deep layers) may be especially useful for cortical areas that represent low-dimensional spaces (such as motor areas), for then the continuity of output may help to produce smooth movements. The essence of a continuous attractor is that it can be learned when there is a fixed trajectory through the firing rate state space that is repeated frequently. This might arise for example if three movements regularly followed each other. If running in a 2-dimensional environment occurred, then grid cells might arise in this low-dimensional space with the hexagonal arrangement reflecting the efficient packing possible in this low-dimensional space (Kropff & Treves, 2008; Stella & Treves, 2015). On the other hand, in regions such as the high order visual cortical areas including the inferior temporal visual cortex where neurons represent a large number of relatively separate objects, and faces (Rolls, 2012b, 2016), a fixed order trajectory through this high dimensional state space is unlikely to occur naturally. (That is, 1000 objects are unlikely to be presented in the same sequence regularly!) Thus in high dimensional spaces the utility of some continuity of representation such as that produced by a continuous attractor network in the deep cortical layers is less clear. However, even in these areas, there may be some similarity structure between stimuli, with some objects more similar to each other than other objects. In this situation, some continuity in the representation may be useful, for then a new object intermediate between the two similar objects can produce an appropriate output correctly mapped into the space without any further learning.

A continuous attractor network might be as simple as a topographic map, but have the advantage just described of enabling new objects to be placed correctly in a space with some continuities. Such a topographic map (described further by Rolls (2016)) need not be capable of sustained activity which is generally a property of attractor networks. But such a continuous map, perhaps more realised in the deep cortical layers, might still have the desirable property of providing a continuous representation, so that any incoming stimulus not actually seen before can be placed at the correct place in the space being represented.

Indeed, the superficial layers might have in part the properties of a discrete attractor network using the recurrent collaterals; but also might have some mixed continuous properties because of the fact that the recurrent collaterals in the neocortex are relatively short-range, spreading approximately 2 mm. In this scenario, the deep layers might implement a further move towards a continuous attractor, a ‘second stage’.

However, for the feedforward processing from stage to stage in a neocortical hierarchy, discrete representations may be useful, so as not to throw away information about exactly which stimulus is present. The connectivity of the superficial layers seems to provide for this, with the input to any one stage being received predominantly into layer 4 and the superficial layers 2 and 3, which in turn project forward to the superficial layers of the next cortical area in the hierarchy (Rolls, 2016).

In this scenario, how might more continuous representations in the deep layers be computationally useful, apart from facilitating smooth movements and trajectories through state spaces? One interesting possibility is that in language production, in which syntactic role may be encoded in a fixed order such as subject-verb-object, a continuous attractor network for this trajectory through the state space (Rolls & Deco, 2015a) might be implemented in the deep cortical layers. It may be more likely that this particular sequence is encoded in the deep layers, for this may be more biologically plausible than that the order is implemented by the feedforward projections between connected cortical areas within a speech production region such as Broca’s area (Rolls & Deco, 2015a). Another possibility is that phoneme production in cortical motor areas (Bouchard, Mesgarani, Johnson, & Chang, 2013) might be implemented in deep layer attractor networks (A.Treves, personal communication). Another possibility is that for the cortico-cortical backprojections, which originate predominantly in primates from the deep layer pyramidal cells in especially layer 5 (Pandya et al., 2015; Rolls, 2016), a more continuous representation, possibly less sparse, may be useful, for then the top-down modulation of attention and of cognition would then bias on all the neurons in a preceding cortical area in the hierarchy that might be being activated by a particular discrete incoming bottom-up stimulus. This is a computational argument for the backprojections, from mainly the deep layers of the cortex, having a more continuous, and probably less sparse, representation than the superficial layers of the neocortex (Rolls, 2016).

In addition to these top-down attentional and cognitive effects, the cortico-cortical backprojections may be part of a cascade gain control system that helps to improve stability. The inner fast loop would be implemented by the GABA inhibitory neurons within a cortical area. The backprojections may help to control gain by feeding back to the preceding cortical area, in an outer loop. In such a cascade control system, the inner loop should be several times faster than the outside loop for stability and effectiveness (Marlin, 2000; Morari & Zafiriou, 1989). This longer time constant for the outer loop might be implemented by the backprojections from the deep neocortical layers, which tend to end on the apical dendrites of pyramidal cells in the

previous cortical area (Pandya et al., 2015; Rolls, 2016). Because these synaptic terminals are so far from the cell body, they tend to act with a longer time constant than terminals would closer to the cell body.

In this research, we have explored hypotheses about possible differences in the computations performed by the superficial and deep layers of the neocortex. We have demonstrated some possible principles of operation of the deep layers of the neocortex if they implement, in part, continuous attractor networks. These principles include being able to produce smooth outputs, and producing continuous outputs less sparse than those of the superficial layers which may be useful for top-down backprojected information for attentional and cognitive modulation. (The sparser representations that may be more evident in the superficial layers may be more useful for storing large numbers of different, discrete, representations in a discrete attractor network, and for looking up particular objects using forward cortico-cortical processing in a hierarchy.) Nevertheless there are other computational advantages of having somewhat computationally different superficial and deep layers of the neocortex, and we regard the present work as an exploration of possibilities and possible implementations, rather than proving that the deep layers have different functions computationally to the superficial layers. Moreover, we note that the hypotheses that are explored here may not be compatible with each other in an implementation of the neocortex, as each hypothesis requires different parameters, especially with respect to adaptation.

Indeed, there are a number of other possible reasons for separate superficial and deep layers of the neocortex (Rolls, 2016). One is that the superficial layers 2 and 3 tend to project forward up the hierarchy which tends to be close topologically and so does not require large pyramidal cells with large axons, whereas layer 5 projects to motor and related structures including the striatum, brainstem, and spinal cord, for which large axons for fast conduction over longer distances that require large pyramidal cells may be useful. A second is that the molecular recognition by classes of neurons is simplified if the superficial neurons are set up genetically to make synaptic connections with the next cortical area (using genetically specified recognition of targets); and if the deep neurons are set up genetically to make synaptic connections with striatal, brainstem and motor areas (for layer 5), and with the thalamus for layer 6 (Rolls, 2016). In the context of recurrent collateral connections, another difference is that the superficial pyramidal cells with their dendrites limited primarily to the superficial cortical layers are well set up to form a superficial attractor network; whereas the deep pyramidal cells with their prolific dendrites in deep cortical layers, appear well set up to implement somewhat separate attractors implemented by their recurrent collaterals which may be especially dense in the deep cortical layers. In this context, the apical dendrites of deep pyramidal cells may rise especially to layer 1, so that the deep pyramidal cells can receive some backprojected information from the next cortical area up in the hierarchy, and not necessarily from the superficial pyramidal cells' recurrent collaterals (see Fig. 1) (Rolls, 2016).

In summary, the research described in this paper shows how the operation of the deep layers of the neocortex could be useful because of different temporal properties, and because different types of computation might be implemented. For example, it is shown that the deep layers might operate with some of the properties of a continuous attractor network, which could facilitate smooth transitions between outputs that would be useful especially in low-dimensional motor spaces; could facilitate the feedback of a more general rather than very discrete signal for top-down attention and recall to previous cortical areas; would be quite stable even with stochastically spiking integrate-and-fire neurons; and would not tolerate much adaptation.

Acknowledgements

Professor Gordon M.G. Shepherd (Northwestern University) is warmly thanked for stimulating discussions. The authors have no competing interests to declare.

Appendix A

A.1. Implementation of neural and synaptic dynamics

We use the mathematical formulation of the integrate-and-fire neurons and synaptic currents described by Brunel and Wang (2001). Here we provide a brief summary of this framework.

The dynamics of the sub-threshold membrane potential V of a neuron are given by the equation:

$$C_m \frac{dV(t)}{dt} = -g_m(V(t)-V_L)-I_{syn}(t), \tag{1}$$

Both excitatory and inhibitory neurons have a resting potential $V_L = -70$ mV, a firing threshold $V_{thr} = -50$ mV and a reset potential $V_{reset} = -55$ mV. The membrane parameters are different for both types of neurons: Excitatory (Inhibitory) neurons are modeled with a membrane capacitance $C_m = 0.5$ nF (0.2 nF), a leak conductance $g_m = 25$ nS (20 nS), a membrane time constant $\tau_m = 20$ ms (10 ms), and a refractory period $t_{ref} = 2$ ms (1 ms). Values are extracted from McCormick, Connors, Lighthall, and Prince (1985).

When the threshold membrane potential V_{thr} is reached, the neuron is set to the reset potential V_{reset} at which it is kept for a refractory period τ_{ref} and the action potential is propagated to the other neurons.

The network is fully connected with $N_E = 800$ excitatory neurons and $N_I = 200$ inhibitory neurons, which is consistent with the observed proportions of the pyramidal neurons and interneurons in the cerebral cortex (Abeles, 1991; Braitenberg & Schütz, 1991). The synaptic current impinging on each neuron is given by the sum of recurrent excitatory currents ($I_{AMPA,rec}$ and $I_{NMDA,rec}$), the external excitatory current($I_{AMPA,ext}$) the inhibitory current (I_{GABA}):

$$I_{syn}(t) = I_{AMPA,ext}(t) + I_{AMPA,rec}(t) + I_{NMDA,rec}(t) + I_{GABA}(t). \tag{2}$$

The recurrent excitation is mediated by the AMPA and NMDA receptors, inhibition by GABA receptors. In addition, the neurons are exposed to external Poisson input spike trains mediated by AMPA receptors at a rate of 2.4 kHz. These can be viewed as originating from $N_{ext} = 800$ external neurons at an average rate of 3 Hz per neuron, consistent with the spontaneous activity observed in the cerebral cortex (Rolls & Treves, 1998; Wilson et al., 1994). The currents are defined by:

$$I_{AMPA,ext}(t) = g_{AMPA,ext}(V(t)-V_E)\sum_{j=1}^{N_{ext}} s_j^{AMPA,ext}(t) \tag{3}$$

$$I_{AMPA,rec}(t) = g_{AMPA,rec}(V(t)-V_E)\sum_{j=1}^{N_E} w_{ji}^{AMPA} s_j^{AMPA,rec}(t) \quad (4)$$

$$I_{NMDA,rec}(t) = \frac{g_{NMDA}(V(t)-V_E)}{1+[Mg^{++}]exp(-0.062V(t))/3.57} \times \sum_{j=1}^{N_E} w_{ji}^{NMDA} s_j^{NMDA}(t) \quad (5)$$

$$I_{GABA}(t) = g_{GABA}(V(t)-V_I)\sum_{j=1}^{N_I} w_{ji}^{GABA} s_j^{GABA}(t) \quad (6)$$

where $V_E = 0$ mV, $V_I = -70$ mV, w_j are the synaptic weights, s_j's the fractions of open channels for the different receptors and g's the synaptic conductances for the different channels. The NMDA synaptic current depends on the membrane potential and the extracellular concentration of Magnesium ($[Mg^{++}] = 1$ mM (Jahr & Stevens, 1990)). The values for the synaptic conductances for excitatory neurons are $g_{AMPA,ext} = 2.08$ nS, $g_{AMPA,rec} = 0.104$ nS, $g_{NMDA} = 0.327$ nS and $g_{GABA} = 1.25$ nS; and for inhibitory neurons $g_{AMPA,ext} = 1.62$ nS, $g_{AMPA,rec} = 0.081$ nS, $g_{NMDA} = 0.258$ nS and $g_{GABA} = 0.973$ nS. These values are obtained from the ones used by Brunel and Wang (2001) by correcting for the different numbers of neurons. The conductances were calculated so that in an unstructured network the excitatory neurons have a spontaneous spiking rate of 3 Hz and the inhibitory neurons a spontaneous rate of 9 Hz. The fractions of open channels are described by:

$$\frac{ds_j^{AMPA,ext}(t)}{dt} = -\frac{s_j^{AMPA,ext}(t)}{\tau_{AMPA}} + \sum_k \delta(t-t_j^k) \quad (7)$$

$$\frac{ds_j^{AMPA,rec}(t)}{dt} = -\frac{s_j^{AMPA,rec}(t)}{\tau_{AMPA}} + \sum_k \delta(t-t_j^k) \quad (8)$$

$$\frac{ds_j^{NMDA}(t)}{dt} = -\frac{s_j^{NMDA}(t)}{\tau_{NMDA,decay}} + \alpha x_j(t)(1-s_j^{NMDA}(t)) \quad (9)$$

$$\frac{dx_j(t)}{dt} = -\frac{x_j(t)}{\tau_{NMDA,rise}} + \sum_k \delta(t-t_j^k) \quad (10)$$

$$\frac{ds_j^{GABA}(t)}{dt} = -\frac{s_j^{GABA}(t)}{\tau_{GABA}} + \sum_k \delta(t-t_j^k), \quad (11)$$

where $\tau_{NMDA,decay} = 100$ ms is the decay time for NMDA synapses, $\tau_{AMPA} = 2$ ms for AMPA synapses (Hestrin, Sah, & Nicoll, 1990; Spruston, Jonas, & Sakmann, 1995) and $\tau_{GABA} = 10$ ms for GABA synapses (Salin & Prince, 1996; Xiang, Huguenard, & Prince, 1998); $\tau_{NMDA,rise} = 2$ ms is the rise time for NMDA synapses (the rise times for AMPA and GABA are neglected because they are typically very short) and $\alpha = 0.5\ \text{ms}^{-1}$. The sums over k represent a sum over spikes formulated as δ-Peaks $\delta(t)$ emitted by presynaptic neuron j at time t_j^k.

The equations were integrated numerically using a second order Runge-Kutta method with step size 0.02 ms. The Mersenne Twister algorithm was used as random number generator for the external Poisson spike trains.

A.2. Calcium-dependent spike frequency adaptation mechanism

A specific implementation of the spike-frequency adaptation mechanism using Ca^{++}-activated K^+ hyper-polarizing currents (Liu & Wang, 2001) is described next, and was used by Deco and Rolls (2005). We assume that the intrinsic gating of K^+ After-Hyper-Polarizing current (I_{AHP}) is fast, and therefore its slow activation is due to the kinetics of the cytoplasmic Ca^{2+} concentration. This can be introduced in the model by adding an extra current term in the integrate-and-fire model, i.e. by adding I_{AHP} on the right hand of Eq. (12), which describes the evolution of the subthreshold membrane potential $V(t)$ of each neuron:

$$C_m\frac{dV(t)}{dt} = -g_m(V(t)-V_L)-I_{syn}(t) \quad (12)$$

where $I_{syn}(t)$ 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. The extra current term that is introduced into this equation is as follows:

$$I_{AHP} = -g_{AHP}[Ca^{2+}](V(t)-V_K) \quad (13)$$

where V_K is the reversal potential of the potassium channel. Further, each action potential generates a small amount (α) of calcium influx, so that I_{AHP} is incremented accordingly. Between spikes the $[Ca^{2+}]$ dynamics is modelled as a leaky integrator with a decay constant τ_{Ca}. Hence, the calcium dynamics can be described by following system of equations:

$$\frac{d[Ca^{2+}]}{dt} = -\frac{[Ca^{2+}]}{\tau_{Ca}} \quad (14)$$

If $V(t) = \theta$, then $[Ca^{2+}] = [Ca^{2+}] + \alpha$ and $V = V_{reset}$, and these are coupled to the equations of the neural dynamics provided here and elsewhere (Rolls, 2016; Rolls & Deco, 2010). The $[Ca^{2+}]$ is initially set to be 0 μM, $\tau_{Ca} = 300$ ms, $\alpha = 0.002$, $V_K = -80$ mV and $g_{AHP} = 0$–40 nS. $g_{AHP} = 40$ nS simulates the effect of high levels of acetylcholine produced alertness and attention, and $g_{AHP} = 0$ nS simulates the effect of low levels of acetylcholine in normal aging (Rolls & Deco, 2015b).

A.3. The model parameters used in the simulations

The fixed parameters of the model are shown in Table 1, 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 1
Parameters used for each module in the integrate-and-fire simulations.

N_E	6400
N_I	1600
r	0.1
w_+	2.1
w_I	1.0
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}

References

Abbott, L. F., Varela, J. A., Sen, K., & Nelson, S. B. (1997). Synaptic depression and cortical gain control. *Science, 275*, 220–224.

Abeles, A. (1991). *Corticonics.* New York: Cambridge University Press.

Amit, D. J. (1989). *Modeling brain function. The world of attractor neural networks.* Cambridge: Cambridge University Press.

Battaglia, F. P., & Treves, A. (1998). Attractor neural networks storing multiple space representations: A model for hippocampal place fields. *Physical Review E, 58*, 7738–7753.

Bauer, R., & Dow, B. M. (1991). Local and global principles of striate cortical organization: An advanced model. *Biological Cybernetics, 64*(6), 477–483.

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature, 495*(7441), 327–332.

Braitenberg, V., & Schütz, A. (1991). *Anatomy of the cortex.* Berlin: Springer Verlag.

Brown, D. A., Gähwiler, B. H., Griffith, W. H., & Halliwell, J. V. (1990). Membrane currents in hippocampal neurons. *Progress in Brain Research, 83*, 141–160.

Brunel, N., & 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.

Burton, M. J., Rolls, E. T., & 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.

Caspers, J., Palomero-Gallagher, N., Caspers, S., Schleicher, A., Amunts, K., & Zilles, K. (2015). Receptor architecture of visual areas in the face and word-form recognition region of the posterior fusiform gyrus. *Brain Structure and Function, 220*(1), 205–219.

Cerasti, E., & Treves, A. (2013). The spatial representations acquired in CA3 by self-organizing recurrent connections. *Frontiers in Cellular Neuroscience, 7*, 112.

Deco, G., & Rolls, E. T. (2005). Sequential memory: A putative neural and synaptic dynamical mechanism. *Journal of Cognitive Neuroscience, 17*, 294–307.

Deco, G., & Rolls, E. T. (2006). A neurophysiological model of decision-making and Weber's law. *European Journal of Neuroscience, 24*, 901–916.

Deco, G., Rolls, E. T., Albantakis, L., & Romo, R. (2013). Brain mechanisms for perceptual and reward-related decision-making. *Progress in Neurobiology, 103*, 194–213.

Elston, G. N., Benavides-Piccione, R., Elston, A., Zietsch, B., Defelipe, J., Manger, P., ... Kaas, J. H. (2006). Specializations of the granular prefrontal cortex of primates: Implications for cognitive processing. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology, 288*, 26–35.

Fuhrmann, G., Markram, H., & Tsodyks, M. (2002). Spike frequency adaptation and neocortical rhythms. *Journal of Neurophysiology, 88*, 761–770.

Goldman-Rakic, P. (1996). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society London B, 351*, 1445–1453.

Harris, K. D., & Shepherd, G. M. (2015). The neocortical circuit: Themes and variations. *Nature Neuroscience, 18*, 170–181.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation.* Wokingham, UK: Addison Wesley.

Hestrin, S., Sah, P., & Nicoll, R. (1990). Mechanisms generating the time course of dual component excitatory synaptic currents recorded in hippocampal slices. *Neuron, 5*, 247–253.

Holmgren, C., Harkany, T., Svennenfors, B., & Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *The Journal of Physiology, 551*(Pt 1), 139–153.

Hooks, B. M., Hires, S. A., Zhang, Y. X., Huber, D., Petreanu, L., Svoboda, K., & Shepherd, G. M. (2011). Laminar analysis of excitatory local circuits in vibrissal motor and sensory cortical areas. *PLoS Biology, 9*(1), e1000572.

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.

Jahr, C., & Stevens, C. (1990). Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. *Journal of Neuroscience, 10*, 3178–3182.

Koch, K. W., & Fuster, J. M. (1989). Unit activity in monkey parietal cortex related to haptic perception and temporary memory. *Experimental Brain Research, 76*, 292–306.

Kropff, E., & Treves, A. (2008). The emergence of grid cells: Intelligent design or just adaptation? *Hippocampus, 18*, 1256–1269.

Lanthorn, T., Storm, J., & 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.
Lefort, S., Tomm, C., Floyd Sarria, J. C., & Petersen, C. C. (2009). The excitatory neuronal network of the c2 barrel column in mouse primary somatosensory cortex. *Neuron, 61*, 301–316.
Liu, Y., & Wang, X.-J. (2001). Spike-frequency adaptation of a generalized leaky integrate-and-fire model neuron. *Journal of Computational Neuroscience, 10*, 25–45.
Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., ... Kennedy, H. (2014). Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology, 522*, 225–259.
Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., ... Schurmann, F. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell, 163*(2), 456–492.
Marlin, T. E. (2000). *Process control: Designing processes and control systems for dynamic performance.* Boston: McGraw-Hill <http://pc-textbook.mcmaster.ca/> .
Mason, A., & 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.
McCormick, D., Connors, B., Lighthall, J., & Prince, D. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons in the neocortex. *Journal of Neurophysiology, 54*, 782–806.
Montagnini, A., & Treves, A. (2003). The evolution of mammalian cortex, from lamination to arealization. *Brain Research Bulletin, 60*(4), 387–393.
Morari, M., & Zafiriou, E. (1989). *Robust process control.* Englewood Cliffs, NJ: Prentice Hall <http://control.ee.ethz.ch/info/people/morari/robust_process_control.pdf> .
Mora, F., Rolls, E. T., & 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.
Nicoll, R. A. (1988). The coupling of neurotransmitter receptors to ion channels in the brain. *Science, 241*, 545–551.
Pandya, D. N., Seltzer, B., Petrides, M., & Cipolloni, P. B. (2015). *Cerebral cortex: Architecture, connections, and the dual origin concept.* Oxford: Oxford University Press.
Renart, A., Parga, N., & Rolls, E. T. (1999a). Associative memory properties of multiple cortical modules. *Network, 10*, 237–255.
Renart, A., Parga, N., & Rolls, E. T. (1999b). Backprojections in the cerebral cortex: Implications for memory storage. *Neural Computation, 11*, 1349–1388.
Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research, 179*, 3–20.
Rolls, E. T. (2008). *Memory, attention, and decision-making. A unifying computational neuroscience approach.* Oxford: Oxford University Press.
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). 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. (2014). *Emotion and decision-making explained.* Oxford: Oxford University Press.
Rolls, E. T. (2016). *Cerebral cortex: Principles of operation.* Oxford: Oxford University Press.
Rolls, E. T. (2017). A scientific theory of ars memoriae: Spatial view cells in a continuous attractor network with linked items. *Hippocampus, 27*, 570–579.
Rolls, E. T., Burton, M. J., & Mora, F. (1976). Hypothalamic neuronal responses associated with the sight of food. *Brain Research, 111*, 53–66.
Rolls, E. T., & Deco, G. (2002). *Computational neuroscience of vision.* Oxford: Oxford University Press.
Rolls, E. T., & Deco, G. (2010). *The noisy brain: Stochastic dynamics as a principle of brain function.* Oxford: Oxford University Press.
Rolls, E. T., & 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., & 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., Grabenhorst, F., & Deco, G. (2010). Choice, difficulty, and confidence in the brain. *Neuroimage, 53*, 694–706.
Rolls, E. T., Sanghera, M. K., & 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., Stringer, S. M., & Trappenberg, T. P. (2002). A unified model of spatial and episodic memory. *Proceedings of The Royal Society B, 269*, 1087–1093.
Rolls, E. T., & Treves, A. (1998). *Neural networks and brain function.* Oxford: Oxford University Press.
Rolls, E. T., & Treves, A. (2011). The neuronal encoding of information in the brain. *Progress in Neurobiology, 95*, 448–490.
Roudi, Y., & Treves, A. (2008). Representing where along with what information in a model of a cortical patch. *PLoS Computational Biology, 4*(3), e1000012.
Sah, P. (1996). Ca^{2+}-activated K^+ currents in neurones: Types, physiological roles and modulation. *Trends in Neuroscience, 19*, 150–154.
Sah, P., & Faber, E. S. (2002). Channels underlying neuronal calcium-activated potassium currents. *Progress in Neurobiology, 66*, 345–353.
Salin, P., & Prince, D. (1996). Spontaneous GABA-A receptor mediated inhibitory currents in adult rat somatosensory cortex. *Journal of Neurophysiology, 75*, 1573–1588.
Spruston, N., Jonas, P., & Sakmann, B. (1995). Dendritic glutamate receptor channel in rat hippocampal CA3 and CA1 pyramidal neurons. *Journal of Physiology, 482*, 325–352.
Stella, F., & Treves, A. (2015). The self-organization of grid cells in 3D. *eLife, 4*, e05913.
Szymanski, F. D., Garcia-Lazaro, J. A., & Schnupp, J. W. (2009). Current source density profiles of stimulus-specific adaptation in rat auditory cortex. *Journal of Neurophysiology, 102*(3), 1483–1490.
Tovee, M. J., Rolls, E. T., Treves, A., & Bellis, R. P. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *Journal of Neurophysiology, 70*, 640–654.
Treves, A. (1993). Mean-field analysis of neuronal spike dynamics. *Network, 4*, 259–284.
Treves, A. (2003). Computational constraints that may have favoured the lamination of sensory cortex. *Journal of Computational Neuroscience, 14*, 271–282.
Treves, A., & Rolls, E. T. (1994). A computational analysis of the role of the hippocampus in memory. *Hippocampus, 4*, 374–391.
Tuckwell, H. (1988). *Introduction to theoretical neurobiology.* Cambridge: Cambridge University Press.
Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron, 36*, 955–968.
Wilson, F. A. W., O'Scalaidhe, S. P., & Goldman-Rakic, P. (1994). 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, F. A. W., & 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., & Rolls, E. T. (1990b). Neuronal responses related to reinforcement in the primate basal forebrain. *Brain Research, 509*, 213–231.
Wilson, F. A. W., & 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.
Wiskott, L., & Sejnowski, T. J. (2002). Slow feature analysis: Unsupervised learning of invariances. *Neural Computation, 14*, 715–770.
Xiang, Z., Huguenard, J., & Prince, D. (1998). GABA-A receptor mediated currents in interneurons and pyramidal cells of rat visual cortex. *Journal of Physiology, 506*, 715–730.