

Representing *Where* along with *What* Information in a Model of a Cortical Patch

Yasser Roudi^{1,2*}, Alessandro Treves^{2,3}

1 Gatsby Computational Neuroscience Unit, UCL, United Kingdom, **2** Cognitive Neuroscience Sector, SISSA, Italy, **3** Centre for the Biology of Memory, NTNU, Norway

Abstract

Behaving in the real world requires flexibly combining and maintaining information about both continuous and discrete variables. In the visual domain, several lines of evidence show that neurons in some cortical networks can simultaneously represent information about the position and identity of objects, and maintain this combined representation when the object is no longer present. The underlying network mechanism for this combined representation is, however, unknown. In this paper, we approach this issue through a theoretical analysis of recurrent networks. We present a model of a cortical network that can retrieve information about the identity of objects from incomplete transient cues, while simultaneously representing their spatial position. Our results show that two factors are important in making this possible: A) a metric organisation of the recurrent connections, and B) a spatially localised change in the linear gain of neurons. Metric connectivity enables a localised retrieval of information about object identity, while gain modulation ensures localisation in the correct position. Importantly, we find that the amount of information that the network can retrieve and retain about identity is strongly affected by the amount of information it maintains about position. This balance can be controlled by global signals that change the neuronal gain. These results show that anatomical and physiological properties, which have long been known to characterise cortical networks, naturally endow them with the ability to maintain a conjunctive representation of the identity and location of objects.

Citation: Roudi Y, Treves A (2008) Representing *Where* along with *What* Information in a Model of a Cortical Patch. *PLoS Comput Biol* 4(3): e1000012. doi:10.1371/journal.pcbi.1000012

Editor: Karl J. Friston, University College London, United Kingdom

Received: August 24, 2007; **Accepted:** January 29, 2008; **Published:** March 21, 2008

Copyright: © 2008 Roudi Y, Treves A. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Financial support by Human Frontier Grant RGP0047/2004-C, National Institute of Mental Health Grant R01MH62447, and Gatsby Charitable Foundation.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: yasser@gatsby.ucl.ac.uk

Introduction

Visual object perception, which is often effortless despite partial occlusion or changes in view, shading, size, etc., has been associated to attractor dynamics in local cortical circuits [1–5]. A single pattern of neuronal activity would be associated with an object, and retrieved when an input cue engages the corresponding basin of attraction. This would lead to a distribution of activity over a cortical patch that can be read out by other areas and can persist even after the object is removed. Attractor dynamics can be realised in neuronal networks by Hebbian modifications of synaptic weights on the recurrent connections of a local population of cortical neurons [6]. The experimental observation of persistent activity in monkey prefrontal cortex (PFC) [7–9] and inferior temporal cortex (IT) [10–12] during memory related tasks supports the idea that attractor dynamics is involved in such tasks.

The above-mentioned paradigm is conceptually very successful in explaining how information about the identity of an object can be retrieved from noisy input and maintained in working memory, even when the input is transient. However, in day to day life, the identity of an object is hardly the only type of information that one needs to retrieve and maintain about it. If you look at a scene for a short time and then turn your head away, you will still remember details about what objects were present in the scene and where they were located. You can even do this if many of the objects in the scene were occluded. These abilities allow us to maintain a

coherent representation of our surrounding environment and are crucial for most real world visually guided behaviours. Visually guided behaviour often requires extracting information about identity of objects (*what* information) from noisy sensory input, and combining this *what* information with information about the position of objects (*where* information). It also requires maintaining this combined representation of position and identity of objects in working memory after the visual input is removed. The underlying neural mechanisms for these abilities are, however, unknown. In this paper, we analyse a network model of how this may be accomplished in the brain.

A great deal of experimental work has been focused on understanding this issue [13–18]. Single cell recordings from PFC during the delay period of a delay match to sample task show that neurons in this area can maintain information about the conjunction of position and identity [13,14]. Rao and colleagues [13] also found that some PFC neurons can change their selectivity from conveying *what* information to conveying *where* information when the type of information that is required by the task is changed. Selectivity for object-position pairs is further supported by the presence of retinotopically organised maps in PFC regions that are involved in identity working memory tasks [16]. Furthermore, a recent neuroimaging study by Sala and Courtney [17] shows that dorsal and ventral PFC can maintain an integrated representation of position *and* identity when it is relevant to the task, but represent position *or* identity when only one of them is

Author Summary

Forming a coherent picture of our surrounding environment requires combining visual information about the position of objects (*where* information) with information about their identity (*what* information). It also requires the ability to maintain this combined information for short periods of time after the stimulus is removed. Here, we propose a theoretical model of how this is accomplished in the brain, particularly when sensory input is incomplete, and missing *what* information should be supplied from what is stored in memory. The main idea is that local connectivity in cortical networks can allow the formation of localised states of activity. *Where* information can then be represented by the position of such “bumps”, and *what* information by the fine structure of the neuronal activity within them. We show that there is a difficulty with implementing this idea: noise and heterogeneity in connectivity cause bumps to drift, thereby losing *where* information. This problem can be solved by incorporating a localised increase in neuronal gain; this, however, interferes with retrieving *what* information and maintaining it in working memory. We quantify this interference via theoretical analysis of the model and show that, despite the interference, the proposed mechanism is an efficient one in retrieving *what* information while representing *where* information.

task relevant. Although most studies that address the issue of combining *what* and *where* information have focused on PFC, similar observations have been reported in IT. While some studies report a considerable position invariance in the response of IT neurons [19–21], this view has been challenged by others. More recent studies show that IT neurons can have small receptive fields and can convey detailed information about the position as well as the identity of objects [22,23]. Furthermore, it has been reported that the receptive fields of IT neurons are much smaller in natural scenes when compared to plain background and are closer to the fovea, thus conveying increased spatial information in their response [24]. Consistent with these properties, Hung et al [25] have shown how, using a simple linear classifier, spatial position can effectively be read off the response of IT neurons. Neuroimaging studies also show that temporal visual areas, just like V4 [26], can be involved in processing the spatial information of objects as well as their identity [15]. Although these neurophysiological studies have not directly assessed the ability of IT neurons to maintain combined *what* and *where* information after removing the stimulus, the possibility should be considered that, like PFC, delay activity in IT can also transmit *where* information in addition to *what* information. The degree to which the neural code in IT and PFC is committed to one versus the other most likely depends on task requirements, attention or learning [15,17,22,23,27,28].

In this paper, we study how a recurrent network can retrieve *what* information from noisy/transient input, while simultaneously representing *where* information. In the model that we present here, we consider a recurrent network embedded in a two dimensional tissue, and to each object associate a single discrete pattern of neuronal activity. These patterns do not have any spatial preference and are stored in the synaptic weights of the recurrent connections through Hebbian learning. We show that, when the connectivity between neurons is metrically organised (that is, nearby neurons are more likely to be connected than those far apart) [29–31], the network can retrieve these patterns in a spatially focused way by maintaining *localised retrieval states* (or

“retrieval bumps”), similar to what has been previously studied in one dimensional networks [32,33]. A localised retrieval state is a stable and localised pattern of activity which has a high correlation with one of the stored patterns, but low correlation with the others. The idea that we elaborate here is to use the position of the bump to represent the position of the object, while the distribution of activity inside the bump represents its identity (In this paper, when we say that a pattern of neuronal activity “represents” a variable, we mean that that variable can be decoded from the pattern of activity). In this way, ideally a *continuum* of firing patterns would represent the object in different positions. The difficulty in implementing this idea, as we show, is that the retrieval bumps cannot be localised at any target position on the surface of the network, but rather on a limited number of discrete positions. To resolve this problem we need to introduce some additional mechanisms. We show that it takes small modulations of neuronal gain to stabilise the bump on arbitrary positions on the tissue. The gain modulation can be provided or at least initiated by the cue that initiates the retrieval of the pattern, or it can be provided by other areas, e.g. in the form of attentional signals [34,35]. Importantly, stabilising the bump at a given position through such gain modulation affects the process of retrieving *what* information from stored representations. We quantify this effect and show that it can be negative, that is a trade-off between the representation of *what* information and *where* information, or it can be positive. Whether the effect is positive or negative depends on the average neuronal gain. When neuronal gain is high the effect is negative but it is positive when the gain is low. Moreover, when the cue is incomplete (that is when the stimulus is noisy or occluded) localising the cue in the gain modulated part of the network further helps identity retrieval. We finally discuss the possibility of retrieving multiple patterns, in the form of multiple bumps of activity. The distribution of activity inside each bump again reflects the identity of the corresponding object, and it can in principle be maintained in working memory while a serial attentional mechanism facilitates retrieval of another memory pattern at a different position.

Model

Firing rate description of the network. In our model of a cortical patch, we assume that the network responds to stimuli with attractor dynamics, thus autoassociatively retrieving activity patterns from memory. At any given time, the response of a neuron indexed i is represented by its firing rate $v_i \geq 0$, which is determined by the input h_i that it receives. We assume that the input to a neuron at time t is related to the firing rate of presynaptic neurons projecting to it through

$$h_i(t) = \sum_{j \neq i} J_{ij} v_j(t), \quad (1)$$

where J_{ij} is the weight of the connection from neuron j to neuron i . Given the input $h_i(t)$, the output firing rate at time $t+1$ is then determined through a threshold-linear gain function

$$v_i(t+1) = g_i F(h_i(t) - Th), \quad (2)$$

where g_i is the linear gain of neuron i , and Th is its threshold, such that $v_i = g_i F(h_i - Th) = 0$ if $h_i < Th$, and $v_i = g_i F(h_i - Th) = g_i (h_i - Th)$ if $h_i \geq Th$. Such rate based description of neurons can be derived from a more complicated model comprised of spiking neurons with conductance based synaptic connections, and it captures many essential features of the behaviour of the spiking model [36,37].

In this paper, we do not explicitly model inhibitory neurons. However, we model their effect on the dynamics of the network by adjusting the threshold of excitatory neurons, such that the mean activity of the network remains constant, i.e. at all times we have

$$\frac{1}{N} \sum_i v_i(t) = a, \quad (3)$$

where a is a constant. The rationale behind having a fixed mean activity at all times is that cortical networks are believed to operate in a balanced state [38–41]. In the balanced state, the inhibitory feedback to excitatory neurons is automatically adjusted such that the mean activity is roughly constant, regardless of whether the network is in the background state or engaged in memory retrieval [42,43]. In our network we achieve this constant mean activity state by adjusting the threshold: at each time step, we first compute the input to all neurons via Eq. (1). We then choose the threshold, Θ , such that when these inputs are passed through the gain function, Eq. (2), the constant mean activity condition, Eq. (3), is satisfied. In this case, the inhibitory feedback is effectively included in the threshold.

Stored memory patterns and synaptic weights. We assume that the network has “learnt”, that is, stored, p different patterns of activity, each representing an object. Each pattern of activity is represented by a vector

$$\eta^\mu = (\eta_1^\mu, \eta_2^\mu, \dots, \eta_N^\mu), \quad \mu = 1, \dots, p,$$

where $\eta_i^\mu = 1$ if neuron i is active in pattern μ (that is selective for the corresponding object) and $\eta_i^\mu = 0$ if it is not. In our model we assume that the stored patterns are generated from the binary distribution

$$\eta_i^\mu = \begin{cases} 1 & \text{with probability } a \\ 0 & \text{with probability } 1 - a, \end{cases} \quad (4)$$

independently for each unit and each pattern, and we have taken the mean activity of the patterns to be the same as the mean activity of the network fixed by the threshold (see Eq. (3)). These patterns are stored through modification of the synaptic weights J_{ij} such that the dynamical attractors of the network include these predefined patterns. One functional form for the synaptic weights which has been widely used in the literature [44,45] is a sparsely coded version of the “covariance Hebbian learning rule” [44,46]. This learning rule assumes a linear summation of contributions from the storage of each pattern

$$J_{ij} = \frac{\overline{\varpi}_{ij}}{Ca^2} \sum_{\mu=1}^p (\eta_i^\mu - a)(\eta_j^\mu - a), \quad (5)$$

where C is the average number of connections per neuron, $\overline{\varpi}_{ij} = 1$ if there is a connection between neurons i and j , and $\overline{\varpi}_{ij} = 0$ otherwise [47]. Based on this learning rule, the weight of the synapses that originate from a given neuron can be both negative and positive. This might seem in contradiction with the Dale’s law and our initial assertion that we would only model excitatory neurons. However, as elaborated in section “Synaptic weights that follow Dale’s law” (see Materials and Methods), one can think of adding a background weight to these values of J_{ij} , such that all weights become positive. The added background weight can then be included in the threshold, without further effects on the dynamics of the network.

Metrically organised versus randomly connected networks. In the model we present here, the way neurons are connected to each other plays a major role in determining whether the network can represent *what* and *where* information together. In general, we can distinguish between two types of neuronal connectivity. The first one is the case of a randomly connected network. In this case each $\overline{\varpi}_{ij}$ in Eq. (5) is set to zero or one with a probability that is independent of i and j , i.e.

$$\Pr(\overline{\varpi}_{ij} = 1) = C/N. \quad (6)$$

This is the case which has been considered in most previous analyses of autoassociative networks. The alternative is a network with *metric connectivity* in which the probability that a connection exists between two neurons depends on their distance i.e.

$$\Pr(\overline{\varpi}_{ij} = 1) = \varphi(|\mathbf{r}_i - \mathbf{r}_j|) \equiv \varphi_{ij}, \quad (7)$$

where \mathbf{r}_i and \mathbf{r}_j are position vectors of neurons i and j in the 2-dimensional (2D) cortical tissue and φ is a smooth function. Associative networks with metric connectivity have recently attracted attention [32,33,48–50] and will be used in the model that we present here.

Local overlap and self-consistent equations. To assess whether the activity of the network at a given time can reflect which object was presented to it in the past, we need to quantify the degree of correlation between the stored patterns and the activity of the network. We do this by measuring the *local overlaps*. The local overlap of pattern μ at position i is defined as [32,33]

$$m_i^\mu \equiv \frac{1}{Ca} \sum_j \overline{\varpi}_{ij} (\eta_j^\mu - a) v_j. \quad (8)$$

Intuitively, m_i^μ is the difference between the level of activity of the population of neurons selective for pattern μ in the neighbourhood of neuron i , and the average activity of all neurons in the neighbourhood of neuron i . The sum of m_i^μ over i , normalised with the number of neurons, is simply the dot product overlap (minus the mean activity, a) between the stored pattern μ and the activity of the network, and is denoted by m^μ ,

$$m^\mu = \frac{1}{N} \sum_i m_i^\mu = \frac{1}{Na} \sum_j \eta_j^\mu v_j - a. \quad (9)$$

This implies that if the sum of m_i^μ over i for one pattern, say $\mu = 1$, is large, while for the others, say $\mu \neq 1$, it hovers around zero, a simple dot product decoding can point at it as a retrieved pattern. In addition to reflecting which pattern is retrieved, the local overlap reflects, in its distribution over i , any emerging spatial structure in the distribution of activity in the network. Thus from the local overlap one can assess the encoding of both *what* and *where* information (see the following section, “Metric connectivity can produce localised states in 2D”, for an example of how we use this in practice).

The local overlap is also important for a more technical reason: it turns out to be a natural parameter for which one can easily write a self-consistent equation whose solutions describe the attractors of the network, as shown in our previous work [32,33] and summarised in section “Self-consistent equations” in the

Materials and Methods. By analysing such self-consistent equation one can determine whether pattern selective bumps of activity exist or not. In section “Self-consistent equations” (Materials and Methods) we show that, for our network, the self-consistent equations take the following form

$$m_i = \frac{1}{C} \sum_j \varphi_{ij} g_j \left\langle \left(\frac{\eta}{a} - 1 \right) \bar{F}_j \left(\left(\frac{\eta}{a} - 1 \right) m_j - \text{Th} \right) \right\rangle_{\eta}, \quad (10)$$

in which m_i is the local overlap with the retrieved pattern, $\langle \cdot \rangle_{\eta}$ represents an average over the distribution of $\{\eta\}$ (see Eq. (4)) and \bar{F}_j is defined in Eq. (24). Intuitively, \bar{F}_j is the gain function in Eq. (2), smoothed out with a Gaussian kernel whose variance is proportional to the storage load, defined as

$$\alpha = \frac{p}{C}; \quad (11)$$

see Eq. (29).

By solving Eq. (10) for m_i , one can find the steady-state distribution of activity in the network, when e.g. pattern μ is retrieved, through the following equation

$$v_i = g_i F \left(\left(\frac{\eta_i^{\mu}}{a} - 1 \right) m_i + z_i - \text{Th} \right), \quad (12)$$

where z_i is a zero mean Gaussian random variable whose variance is given in Eq. (29). The possibility of having retrieval bumps can be analysed through Eqs. (10) and (12) [32,33]. The upshot of such analysis is that when the probability of connection, φ_{ij} in Eq. (7), is such that a significant number of connections to each neuron comes from nearby neurons, the solution of Eq. (10) can depend on i and therefore the local average of v_i can be higher in some areas and lower in others. We will show examples of such localised retrieval states in the next section.

Results

Metric connectivity can produce localised states in 2D

In what follows, we show an example of a retrieval bump in a 2D recurrent network with metrically organised connectivity. We consider a two dimensional network containing $N = 4900$ neurons in total. The neurons are arranged on a regular lattice with 70 neurons on each side and distance l between neighbouring sites. The connections between neurons have a metric structure: a neuron in position \mathbf{r}_i is connected to a neuron in position \mathbf{r}_j with probability

$$\Pr(\varpi_{ij} = 1) = \frac{0.05Nl^2}{2\pi\sigma^2} \exp\left(-\frac{|\mathbf{r}_i - \mathbf{r}_j|^2}{2\sigma^2}\right). \quad (13)$$

In the simulations reported here the width of the connectivity, σ , is set to $7.5l$. Since l is the distance between two adjacent neurons, this means that the probability that two adjacent neurons are connected to each other is ~ 0.7 . Experimental data estimate this probability to be 0.5–0.8 [30]. The gain of all neurons in the simulations reported in this section is set to a background level $g = 0.5$.

At the beginning of the simulation a 15×15 square centred on the neuron in position (58,58) is chosen. The activity of neurons inside this square are initialised to their activity in the first stored pattern while the activity of other neurons are set to zero, that is in the beginning of simulation $v_i = \eta_i^1$ if \mathbf{r}_i is in the square and $v_i = 0$ if

\mathbf{r}_i is outside it. In this way at the beginning of the simulation the dot product overlap with the first pattern and the others have the following values

$$m^1(t=0) = \frac{1}{Na} \sum_i \eta_i^1 v_i - a \approx 0.037$$

$$m^{\mu \neq 1}(t=0) = \frac{1}{Na} \sum_i \eta_i^{\mu \neq 1} v_i - a \approx 0.$$

Fig. 1A shows the local overlap with the cued pattern ($\mu = 1$) at the beginning of a simulation. The local overlap (Eq. (8)) with the cued pattern after 200 synchronous updates is shown in Fig. 1B and the distribution of activity $\{v_i\}$, also after 200 time steps, is shown in Fig. 1C. We see that the activity of the network is concentrated on a part of the 2D network, and so is the local overlap. The important point is that this final pattern of activity has a high dot product overlap with the cued pattern but not with other stored patterns, i.e.

$$m^1(t=200) = \frac{1}{Na} \sum_i \eta_i^1 v_i - a \approx 0.8$$

$$m^{\mu \neq 1}(t=200) = \frac{1}{Na} \sum_i \eta_i^{\mu \neq 1} v_i - a \approx 0.$$

Thus by calculating these dot products, or equivalently calculating the sum of the local overlaps m_i^{μ} over i , in the end of the simulation we can say which pattern was presented, i.e. in this example the first pattern.

The spatial distribution of activity would have been different (Fig. 1D), if instead of the probability distribution in Eq. (13), we had used a uniform distribution

$$\Pr(\varpi_{ij} = 1) = C/N = 0.05.$$

In this case, too, by cueing one of the patterns, as we did for the metrically organised network, after 200 time steps, we have $m^1(t=200) \approx 0.8$ and $m^{\mu \neq 1}(t=200) \approx 0$, thus indicating retrieval of the pattern. The difference between the two connectivity models emerges, however, in the final distribution of activity. Whereas in Fig. 1D the activity is distributed uniformly across the network (at a gross spatial scale, since at a fine scale individual units are activated in relation to their selectivity for the cued pattern), in Fig. 1C the activity is localised over a portion of the 2D network. So, metric recurrent connections, as predicted by the mathematical analyses of attractor states and as confirmed by many other simulations, allow activity to stabilise in spatially modulated distributions.

Even though Fig. 1 shows the possibility of localised retrieval in the network with the Gaussian connectivity in Eq. (13), a critical observation is that in Fig. 1B the final local overlap is in a *different* position than the initial cue (Fig. 1A). The trajectory of the local overlap follows during the retrieval process is shown in Fig. 2. The green square shows the peak at the beginning of the simulation, before any updates take place (Fig. 1A), and the red circle shows the peak after 200 time steps (Fig. 1B). It is clear that, during retrieval, the “bump” of activity drifts away from its initial position. This raises the question addressed in this paper, of whether *where* information in the cue can be preserved by spatially modulated attractor states.

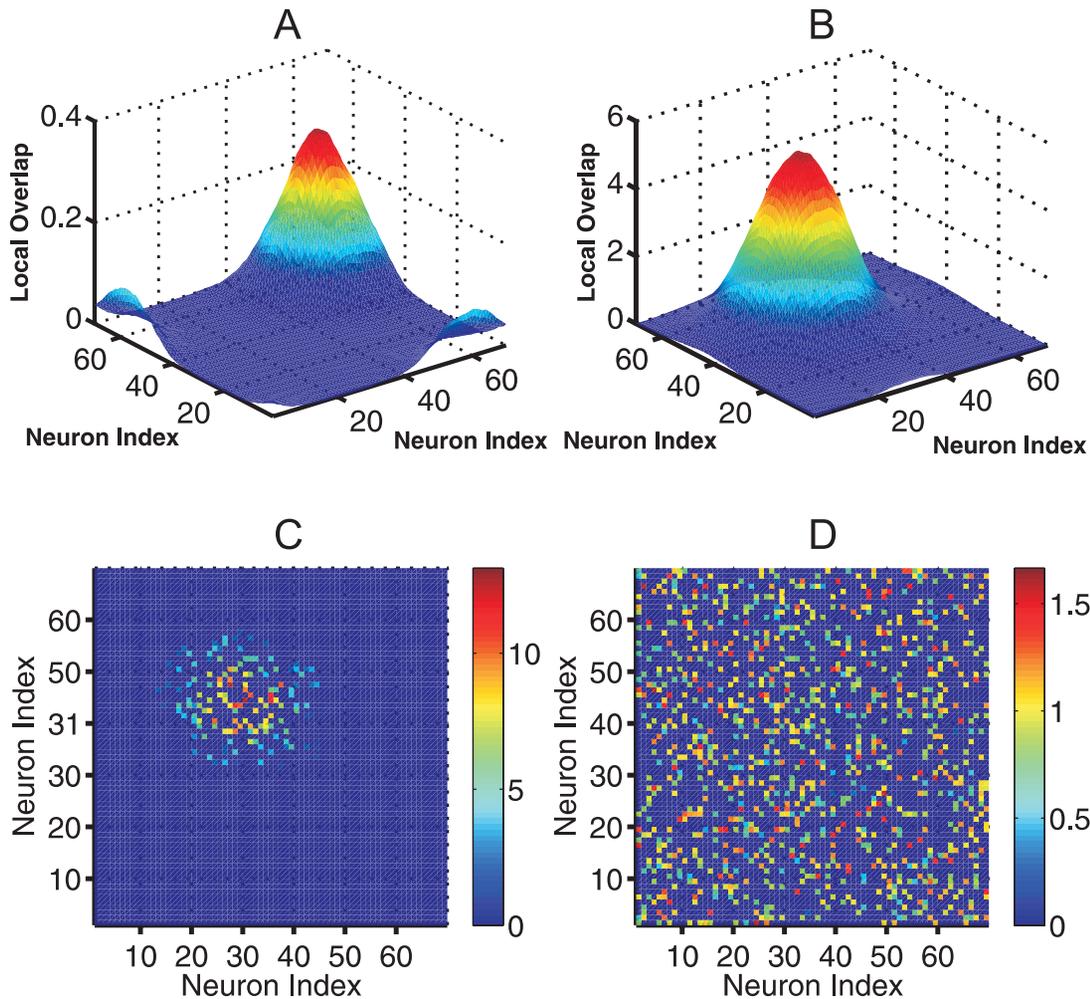


Figure 1. Short-range connectivity allows localised attractor states, but activity drifts away from the position of the cue. The network is comprised of $N=70 \times 70=4900$ neurons, each connected to $C=0.05N$ other neurons, chosen using the Gaussian distribution in Eq. (13). The gain of all neurons is fixed to a uniform background value $g=0.5$ and we have $p=5$ and $a=0.2$. (A) The local overlap with the cued memory pattern at the beginning of a simulation. In the beginning of the simulation a 15×15 square centred on the neuron in position (58,58) is chosen. The activity of neurons inside this square is set to their corresponding activities in the first pattern, and the rest of the neurons are silent. (B) The local overlap after 200 synchronous time steps. (C) The activity distribution across the network after 200 synchronous time steps. (D) The distribution of the activity of the neurons in an identical network operating with identical cue and parameters, except that recurrent connectivity is random, i.e. each neuron is connected to other neurons with probability $C/N=0.05$. Whereas the distribution in (C) is localised in space, in (D) the activity is distributed across the network. Note that in both cases (the network with metric connectivity and the one with random connectivity), the cued pattern has been successfully retrieved. This can be seen by measuring the overlap between the network activity and the cued pattern and comparing it with its overlap with non-cued ones (see the text): after 200 time steps, the overlaps, m^i (that is the sum of the local overlap m^i over i ; see Eq. (9)), are $m^1(t=200) \approx 0.8$ and $m^{i \neq 1}(t=200) \approx 0$, indicating the retrieval of the cued pattern. Periodic boundary conditions are assumed here and elsewhere in this paper.

doi:10.1371/journal.pcbi.1000012.g001

Metric connectivity alone is not sufficient to preserve *where* information

Can we code the position of an object by the position of the peak of the retrieval bump? The answer to this question depends on whether the retrieval process can end with the peak of the bump on the intended position. We first examine whether the position of the cue (which can be thought of as the position of an object in the retina) determines the positions of the retrieval bump. If the retrieval bump appears at the same position as (or is uniquely determined by) the centre of the cue, it is possible to read the activity of the network and simply decode both *what* information, that is, which cue has been presented (the pattern with the highest overlap with final activity), and, in addition, *where*

it has been presented: object position is simply coded by the position of the centre of the bump.

Continuous attractors are fragmented by superimposed memories

To examine the relation between the position of the initial cue and the final position of the retrieval bump, we ran simulations in which the position of the initial cue was systematically changed across the network and the distance between the position of the retrieval bump and the position of the initial cue was measured. In Fig. 3, we summarise the results from simulating a network of 70×70 neurons with the Gaussian connectivity pattern Eq. (13) with $\sigma = 7.5l$, as used in Fig. 1. At the beginning of each trial, the

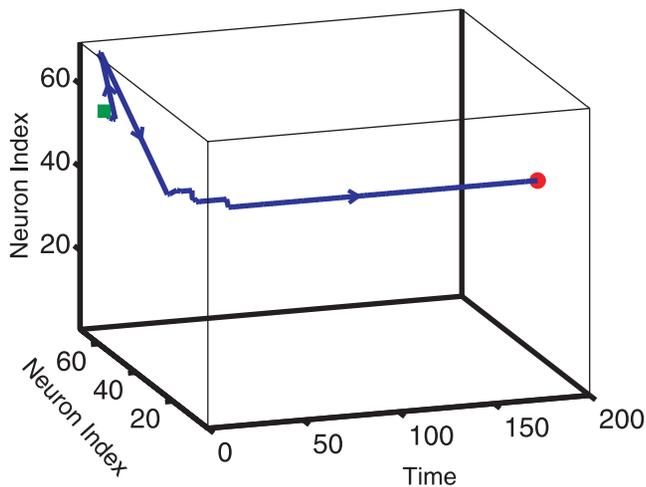


Figure 2. During retrieval, the peak of the local overlap drifts away from its initial position. This figure shows the trajectory that the peak of the local overlap follows during the simulation reported in Fig 1. The green square represented the position of the peak of the local overlap at the beginning of the simulation (Fig. 1A) and the red circle represents the position of the peak after 200 time steps (Fig. 1B). Over the first ~40 time steps the local overlaps drifts until it settles into an equilibrium state.

doi:10.1371/journal.pcbi.1000012.g002

first pattern was cued by initialising the activity of neurons in the following way: $v_i = \eta_i^1$, if neuron i was within a 15×15 square, whose centre was varied, across trials, over the entire network; while $v_i = 0$, if neuron i was outside the square. The activity of all neurons was then synchronously updated for 200 time steps and the local overlap with each pattern was monitored. Fig. 3A shows that the position of the bump at the end of each trial (red circles) does not match the peak of the local overlap with pattern 1 at the beginning of the trial (green squares). The bump drifts away from

its initial positions, and stabilises on one of, in this particular case, 4 final positions. This small number of final stable positions indicates that one cannot decode from the final position of the retrieval bump where the cue was located, at least not with high accuracy. In fact, by looking at the final position of the bump, one might say whether the initial position of the cue was among the 23 initial positions that converge to the upper left red circle or among the 10 initial positions that converge to the lower right red circle, but nothing more.

Relation to continuous attractors without stored memory

The small number of final stable positions of the bump resembles what has been noticed before in recurrent networks with distance dependent weights between neurons but without stored memory patterns. In such models the synaptic weight between two neurons is generally taken to be excitatory at short distances while inhibitory at long distances [51–58]. The distance between two neurons in these models can be anatomical distance, or distance, in the feature space, between the features that the neurons are selective for. Models of this type have been used to conceptualise how local networks of orientation selective neurons in visual cortex [52], head-direction neurons [53], location selective neurons in prefrontal cortex [54] and hippocampal neurons [57,58] can maintain selectivity after the external input has been removed. Studies on rate based models [51–53] as well as networks with spiking neurons [54–56] show that, under very mild conditions, the stable activity profile of these networks is of the form of a localised “bump”. If the network is strictly homogeneous, the bump can potentially exist anywhere on the network, and it can be smoothly moved from one position to the other. Any small inhomogeneity in the network, however, fractures the continuum of solutions, which therefore represents an ideal limit case, and stable bumps are allowed only at a number of discrete positions [53,57,59–61]. Coming back to the associative memory network with metric connectivity, it is clear that inhomogeneity is an unavoidable part of its overlaid memory structure. Synaptic

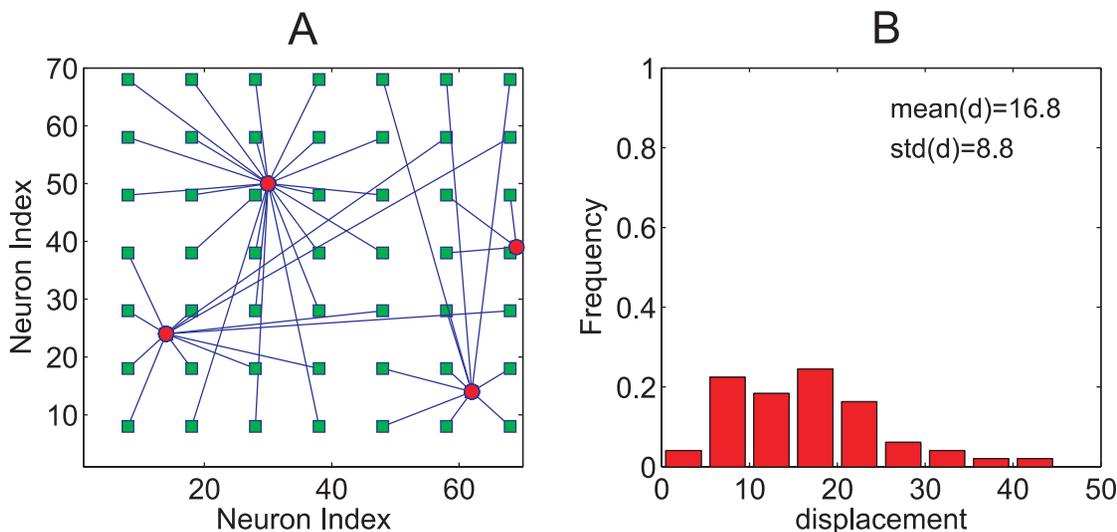


Figure 3. Steady-state activity bumps are only weakly correlated with cue position, in the constant gain model. This graph illustrates how the position of the cue affects the position of the bump after retrieval is completed. Cue position (the centre of a 15×15 square window) was chosen on a 7×7 grid with grid spacing 10, whose vertices are shown in (A) as green squares. The red circles show the position of the peak of the local overlap, as it drifted away from the linked green square, over 200 time iterations of the simulated retrieval process. All runs in this example resulted in successful retrieval, i.e. the overlap between the final activity of the network and the cued pattern was higher than with other patterns. (B) The distribution of the distances between the centre of the cue and the centre of the local overlap after 200 time steps. In the legend, d is the distance averaged across all trials. The neuronal gain in this network is $g = 0.5$ and the number of stored patterns is $p = 5$.

doi:10.1371/journal.pcbi.1000012.g003

weights are required to be different from each other in such a network, to support the retrieval of memory patterns, a situation where a neuron can be active while its nearest neighbour is inactive. As a result, a retrieval bump in our model cannot be maintained at any arbitrary position on the network.

Scaling of the number of stable positions

Even though the final position of the bump cannot accurately tell where the cue was initiated, it may still be able to code for a large number of positions in a network with realistic size. This happens if the number of final stable positions increases with the size of the network. To examine this relation, we scaled up the simulated network. The result of such scaling analysis is reported in Fig. 4, which shows the number of final positions resulting with different network sizes, while keeping the number of connections and the width of connectivity constant. One sees a roughly linear increase in the number of stable bump positions.

The approximately linear scaling of the number of final positions with network size indicates that a large number of positions can be represented in realistically large networks, but not *any* arbitrary position: with our regular 2D lattices and our parameters, the number of stable bump positions is about one thousand times smaller than the number of lattice nodes. Furthermore, the few stable positions of the retrieval bump are different for different patterns (data not shown). This makes the

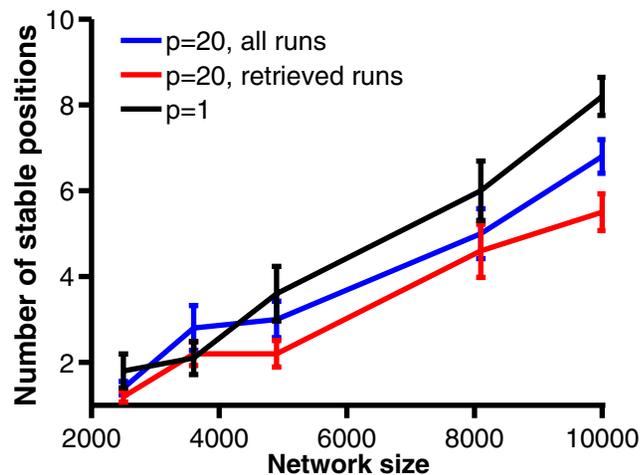


Figure 4. Scaling of the number of stable positions of the bump with the size of the network. The scaling is done for $p=1$ (black line) and $p=20$ (red and blue lines). For $p=20$, the blue line represents the number of stable positions when both successful and unsuccessful runs are averaged and the red line represents this number when only the successful ones are counted. For each network, of size N , one of the hundred positions on a 10×10 regular lattice was chosen in each trial. The activity of neurons inside a square that covers $0.04 \times N$ neurons and whose centre was on the chosen position, was set to their activity in the stored pattern. For $p=1$ the activity of neurons outside this square were set to zero. For $p=20$, a fraction of 10% of neurons outside the square also had their activity set to their activity in the first pattern, while the rest has zero activity level. The difference between the ways we initialised the retrieval for $p=1$ and $p=20$ was introduced to ensure a higher quality of initial cue for $p=20$; this would be needed for having a reasonable number of successful runs. The width of the connectivity was set to 7.5 and the average number of connections per neuron was always 245. The number of final positions of the bump of activity after 200 time steps was then counted. We then plotted the average and standard error of this number over the hundred positions of the initial cue and 10 realizations of the network and of the stored pattern, versus N .

doi:10.1371/journal.pcbi.1000012.g004

representation of position dependent on object identity and thus hard to decode. We ask, therefore, whether it is indeed possible to stabilise bumps at any arbitrary position. This is discussed in the following sections.

Gain modulation and bump localisation

In this section we show that the bump of activity can be stabilised on an arbitrary position on the network if neurons around that position have a slightly higher linear gain than the rest of the neurons. This increase in the linear gain applies to all neurons in that area in the same manner, whether they are selective for the cued pattern or not; that is, it is not pattern selective and solely reflects object position. This *local gain modulation* can be triggered by an attentional mechanism that modulates the responsiveness of neurons in the part of the network which corresponds to the position of the object. It could also be produced by the pattern itself: when the cue to initialise retrieval is given to the network, the mean activity of the part of the network that receives the cue would be higher than the rest of network. This could trigger changes in the gain of the neurons that may last for several seconds [62–64]. We leave discussing the sources of the gain modulation to the Discussion section and first answer the following questions. Can such localised gain modulation stabilise the bump at any desired position and, if so, how strong should it be? How does localised gain modulation affect pattern retrieval?

Modulating single neuron gain can stabilise the bump

Suppose that a non-pattern-selective signal changes the gain of those neurons which correspond to the position of the object in the visual scene. The effect of such gain modulation is shown in Fig. 5.

In the simulations of Fig. 5, the activity of $\approx 4.6\%$ of the neurons, randomly distributed across the network, are initially set to their activity in the first pattern, while the rest are silent (note that the quality of the cue is then the same as what we used in the simulations of Fig. 3, but now the cue is not localised). The localised gain modulation is incorporated into the simulations by first choosing, at each trial, a square box at a different position over the network. The linear gain of neurons inside the square is then increased by a factor of β relative to that of the other neurons in the network. The position of the centre of the high gain square box is in fact chosen in exactly the same way as we chose the centre of the cue in Fig. 3, i.e. at the nodes of a regular lattice, shown as green squares in Fig. 5A and Fig. 5C.

The result of such change in the spatial distribution of the gain is evident for $\beta=1.5$ (Fig. 5A, 5B) and even more for $\beta=3$ (Fig. 5C, 5D). Even though the pattern-selective cue does not contain spatial information, a spatially selective increase in the linear gain of the neurons in a restricted region of the network helps localising the bump in that region. Notably, as shown in Fig. 5D, the distance that the peak in the local overlap drifts from the initial focus is minimal, particularly for successful trials (red circles) (d), whereas averaging across unsuccessful runs (black circles) (d^*) substantially increases the drift, as if jumping to the wrong basin of attraction in the space of patterns facilitates similar jumps in physical space.

It should be noted that while in Fig. 3 the localised cue had been removed after initialising the activity, in the results shown in Fig. 5 the change in gain is maintained throughout the simulation. It is true that keeping the localised cue would have helped localising the bump at the right position, without gain modulation, but the essential difference between the two mechanisms should be appreciated: the change in gain is independent of the memory pattern to be retrieved and could thus be produced by a mere spatial signal, with little specific information content besides spatial

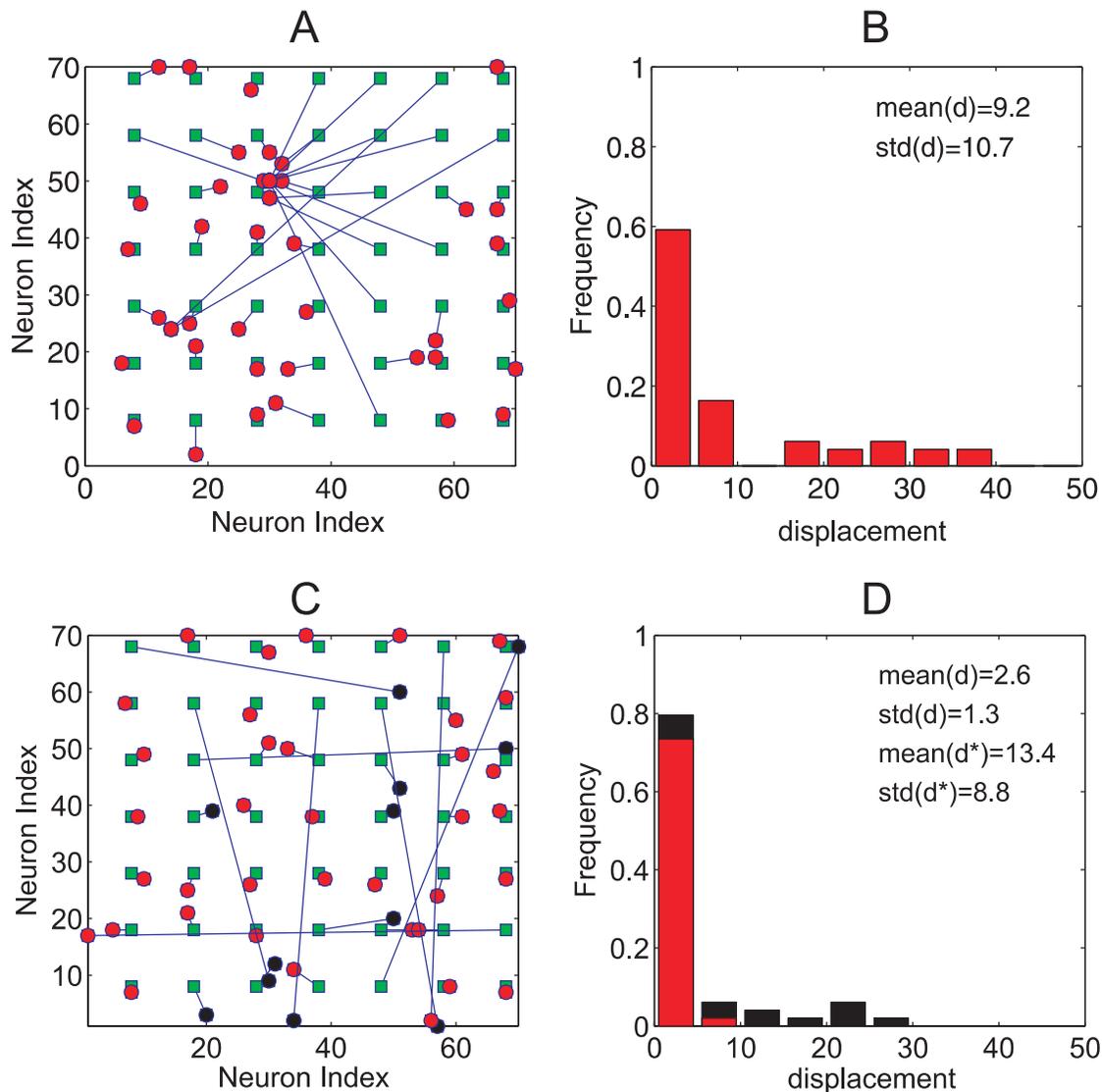


Figure 5. Local gain modulation largely fixes the position of the bump. The panels summarise the results of simulations conducted as for Fig. 3 except for two factors. First, the pattern-selective cue is not localised, i.e. the local overlap at the beginning of each simulation is a uniform function across the network; hence, the distribution of the peak of the local overlap in the beginning of the simulation is not shown: there are no significant peaks. Second, neurons inside the 15×15 square centred around each green square in (A) have a gain factor g which is 1.5 times larger than the rest of the network, and in (C), 3 times larger. (B) and (D) report the distributions of distances between the centre of the gain modulated square and the peak of the final local overlap corresponding to simulations in (A) and (C), respectively. Red circles in (A) and (C), and red portions of the bars in (B) and (D) correspond to successful runs (defined as runs in which the overlap with the cued pattern, after 200 time steps, is higher than the overlap with any other pattern) and black circles and black portions of the bars represent unsuccessful ones. $mean(d)$ and $std(d)$ are the mean and the standard deviation of the distances averaged over successful runs and $mean(d^*)$ and $std(d^*)$ averaged over unsuccessful runs. doi:10.1371/journal.pcbi.1000012.g005

position itself. The pattern-selective cue, instead, can be thought to commit the informational resources (e.g., the channel capacity [65]) of the ventral visual form processing stream, and it makes sense to hypothesise that it should be removed as soon as possible, to make room for the analysis of other objects by the same pathway.

Even though increasing the gain in a spatially restricted part of the network stabilised the final bump there, there is a disadvantage with this strategy: by using such non-uniform gain, the number of successful runs decreases. Remember that the quality of the cue used in Fig. 5 is the same as the one in Fig. 3, however, there were no unsuccessful runs in Fig. 3 and Fig. 5A, whereas there are 12 unsuccessful runs in Fig. 5C (shown by black circles): better

preservation of spatial information (higher gain modulation) is accompanied by, in this example, a higher number of unsuccessful runs. This suggests that preservation of spatial information through gain modulation affects the retrieval of the pattern. In Fig. 5 the effect is negative, an interference, but as we show below it can also be a positive effect. In the following sections, we quantify this interaction using information theory and demonstrate efficient ways to minimise the negative interaction.

The relation between I_{what} and I_{where} : An information theoretic analysis

In order to quantify the interaction between *what* and *where* information, we use Shannon information theory. We estimate the

amount of information that the activity of the network, after retrieval, represents about *what* and *where*. We do this for different degrees of gain modulation, levels of the average gain, number of stored patterns and also different ways of presenting the cue. This provides us with a quantitative picture of the relation between *what* and *where* information.

We denote by I_{what} and I_{where} the amount of information about *what* and *where*, respectively. To compute I_{what} we look at network activity after 200 time steps and compute its overlap with all stored patterns (Eq. (9)). The pattern with the highest overlap is considered as retrieved and I_{what} measures how much information knowing this retrieved pattern gives us about which pattern was presented. I_{where} , on the other hand, is the information between the position of the bump of activity after retrieval and the centre of the gain modulated area (we put $I_{where}=0$ when there is no gain modulation; see section “Continuous attractors are fragmented by superimposed memories”). For exact definitions and details about how we compute I_{what} and I_{where} from the simulations see section “Mutual information measures” in the Materials and Methods.

To start with, we consider a network (with the architecture used before) that has stored p patterns and assume that in the beginning of the simulations a cue similar to one of the patterns is presented (the exact cue presentation is described in the three Conditions below). All neurons have a background gain of g . During recall, either the gain of all neurons is kept equal to g , which is the case of uniform gain, or the localised gain modulation mechanism is turned on. In the latter case the gain of the neurons inside a 15×15 square whose centre is on one of 49 preassigned positions on the network is boosted to βg . Different values of β are chosen in different simulations. In each run, one of the patterns is chosen as a cue and one of the 49 positions is chosen as the centre of the high gain region. As in the previous sections, the centre of the squares surrounding the high gain region is chosen from one of the 49 nodes of a 7×7 regular lattice covering the entire 2D network. Each pattern and each of the 49 positions for the high gain region is used exactly once.

High gain regime

We first calculate I_{what} and I_{where} for a network with the global gain chosen to be $g=0.5$. We do this for the case of uniform gain (all neurons have the same gain, thus equal to the background gain g), three degrees of gain modulation, with $\beta=1.5, 2$ and 3 , and three values of $p=5, 10$ and 15 . We consider three alternative ways in which the cue can be presented to the network. These cueing conditions and the resulting $I_{what}-I_{where}$ relation are described below.

Condition 1: Complete cue. In this condition, in the beginning of the simulation, the activity of all neurons in the network is set to their activity in the pattern to be retrieved. Fig. 6A shows how I_{where} covaries with I_{what} for three values of $p=5$ (full line), $p=10$ (dashed line) and $p=15$ (dashed-dotted line) when the initial cue was complete. On each curve different points correspond to different degrees of gain modulation. There is a clear trade-off between I_{what} and I_{where} . The maximum amount of *what* information and the minimum amount of *where* information occurs when there is no gain modulation. As we increase the degree of localised gain modulation, I_{what} decreases while I_{where} increases. Moreover, the value of p for which maximum I_{what} can be retrieved depends on the degree of gain modulation and thus on I_{where} . In other words, for any given I_{where} there is a critical p for which I_{what} is maximised. In the absence of gain modulation, $I_{where}=0$ and the maximum of I_{what} is obtained for a certain storage load $10 \leq p \leq 15$, whereas, as the degree of gain modulation

increases, the number of patterns resulting in maximal identity information decreases.

In this condition, we had provided the network with a perfect cue. It is important to see whether a similar trade-off occurs in the more interesting condition in which the cue is incomplete, so that the network has not merely to maintain but also to actually retrieve identity information.

Condition 2: Non-localised incomplete cue. In the second cueing condition, the cue that is given to the network in the beginning of the simulation is incomplete, and it does not have any spatial information. In this case, at the beginning of the simulation, the activity of 4.6% of randomly chosen neurons are set to their activity in the pattern to be retrieved, and the rest at zero. Using such imperfect cue serves to assess the degree to which the network is able to maintain I_{where} and retrieve I_{what} from a degraded version of the stimulus. Results are shown in Fig. 6B. The same pattern of trade-off between I_{what} and I_{where} seen with full cues is also evident here. The main difference is that now the maximum value of I_{what} is decreased relative to the first condition, which is a direct consequence of the cue being incomplete. Again, as in the previous condition, increasing the degree of gain modulation, which favours I_{where} at the expense of I_{what} , also favours networks with low storage load p .

Condition 3: Incomplete and localised cue. In the first two conditions the cue did not have any spatial information by itself and the spatial information was provided entirely through gain modulation. In the third cueing condition, the quality of the cue is the same as condition 2, but now the cue is itself localised. On each run, a 15×15 square is considered whose centre is on one of the 49 pre-assigned positions on the network. In this condition, the square with higher gain coincides with the localisation of the cue. The activity of neurons in the square is set to their activity in the pattern to be retrieved, and the rest to zero. Note that the quality of the cue is the same as condition 2, but now these neurons are not randomly chosen, but are rather localised in a certain region of space.

The analysis of simulations in this Condition 3 is shown in Fig. 6C. A result evident by comparing Fig. 6C with Fig. 6B is that both the degree of preservation of spatial information and the number of successful runs have increased in Condition 3 compared to Condition 2. For instance when $p=10$ and $\beta=2$ both I_{where} and I_{what} are near their maximum values. By its effectiveness in retrieving I_{what} while maintaining I_{where} condition 3 suggests the advantage of modulating neuronal gain through the change of single neuron properties triggered by the cue itself. As the activity level around neurons which are receiving the cue is increased, this may induce an increase in the excitability of these neurons, which lasts for some time (see section “Possible sources of gain modulation” in the Discussion).

The dependence of the $I_{what}-I_{where}$ curve in all three panels of Fig. 6C on p indicates that the optimal storage load depends on the conditions prevailing at retrieval, and in particular on the degree of gain modulation.

Low gain regime

In the previous section, the background gain was $g=0.5$. Without gain modulation, the network could reach high I_{what} values, sometimes retrieving all stored patterns, even from a very small initial cue. With gain modulation, I_{where} increased but I_{what} decreased. Here, we show that when the background gain is low, the interaction can be reversed, that is, gain modulation can actually increase both I_{what} and I_{where} . We set the background gain to $g=0.25$. As shown in Fig. 7, for the case of complete cue (as in Condition 1 above) even without gain modulation I_{what} is very

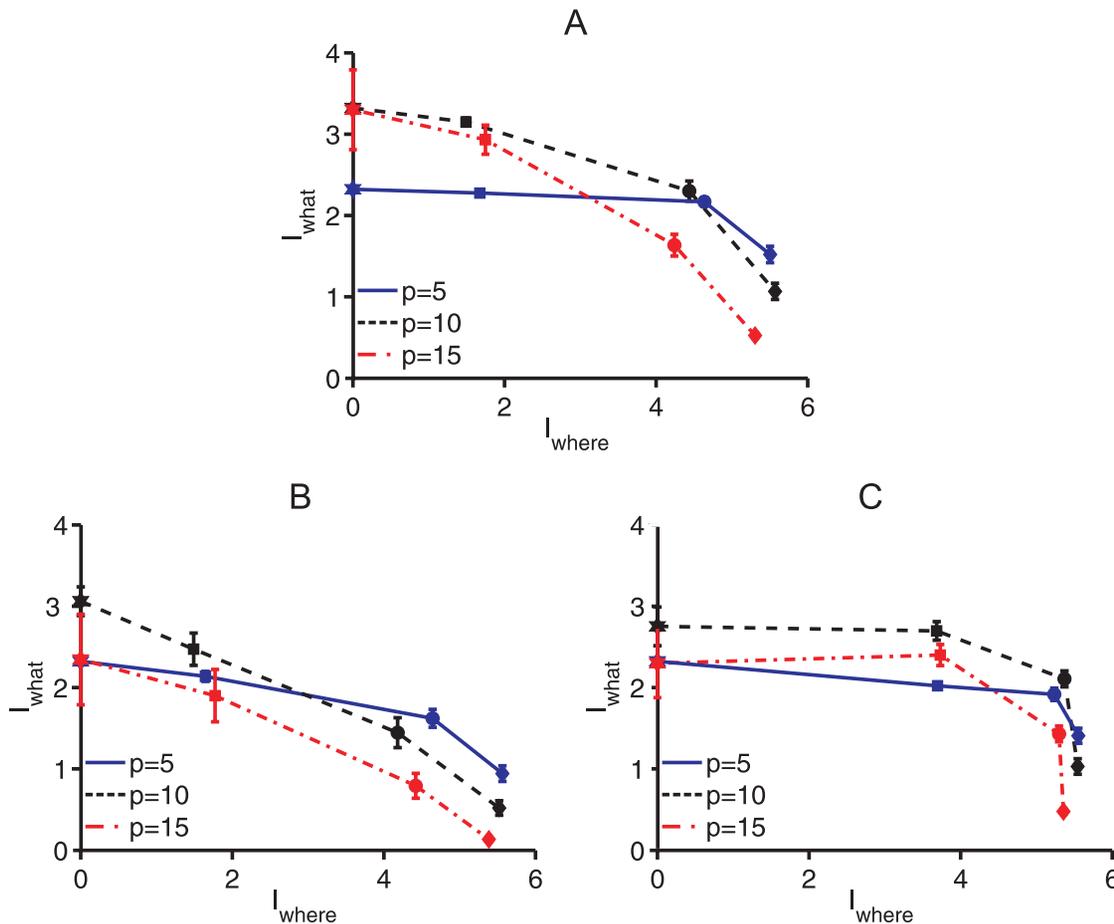


Figure 6. The trade-off between *what* and *where* information for different cueing conditions, different numbers of stored patterns and different levels of gain modulation, as discussed in the text. (A) Condition 1: complete cue. (B) Condition 2: 4.6% of the neurons are randomly chosen and their activity is initialised to their activity in the cued pattern, while the rest has zero activity. (C) Condition 3: similar to Condition 2 but the neurons that receive the cue are spatially co-localised with gain modulation. For each cueing condition, the trade-off is shown for $p=5$ (full line), $p=10$ (dashed line) and $p=15$ (dashed-dotted line). For each value of p , three different degrees of localised gain modulation, $\beta=1.5$ (square), $\beta=2$ (circle) and $\beta=3$ (diamond), were tested in addition to the uniform gain of $g=0.5$ (star). doi:10.1371/journal.pcbi.1000012.g006

small. When incorporating a gain modulation mechanism, however, the amount of *what* information maintained by the network increases, together with the amount of *where* information. In section “Low gain regime versus high gain regime” (see Materials and Methods), we discuss why the relation between I_{what} and I_{where} is different in the low gain and high gain regimes. Intuitively, the reason is as follows. Successful retrieval occurs only when the gain of the neurons that support the retrieved pattern is between a minimum g_{min} and a maximum g_{max} . In the low gain regime, the level of background gain is below g_{min} and the network cannot retrieve the patterns. When the gain is increased in part of the network, then it may enter the range $[g_{\text{min}}, g_{\text{max}}]$, allowing for retrieval to occur. At the same time, since that region has a higher gain, the retrieval bump does not drift away. When the background gain is high, instead, gain modulation stabilises the bump in the gain modulated area. This is accompanied, however, by a decrease in the size of the bump. The reason is that the higher neuronal gain increases the firing rate of neurons inside the bump (the peak of the bump is higher) and, to comply with the constant mean activity condition (Eq. (3)), this increase in the peak activity is accompanied by a decrease in the spatial extent of the bump. Therefore, fewer connections are involved in retrieving the pattern and I_{what} decreases. As expected from this argument, increasing β

too much even in the low gain regime should decrease I_{what} . This can be seen in Fig. 7 for $\beta=5$ and $p=10$.

Multiple bumps

When a retrieval bump is localised on a particular position, one can in principle use the rest of the network to retrieve other patterns, in the form of additional bumps of activity. If they can coexist with the first bump, the network would then be able to represent the position and identity of multiple objects simultaneously, without encountering the problem of binding. In random networks with no metric connectivity nor localised retrieval, retrieving multiple patterns is indeed possible, at very low storage loads [1,45,66]; in these networks, however, it is not possible to represent the position of the objects, which has to be represented elsewhere. If the *what* and *where* of multiple objects are represented in different networks, a binding problem arises. The localised retrieval process described here does not suffer from this problem. It is then important to assess the conditions which make it possible to stabilise (at least) two retrieval bumps simultaneously.

Assume that a pattern is retrieved and, using localised gain modulation, the bump of activity is stabilised on a desired position. A second cue may then be presented to the network at another position. Even though most of the connections to each neuron in

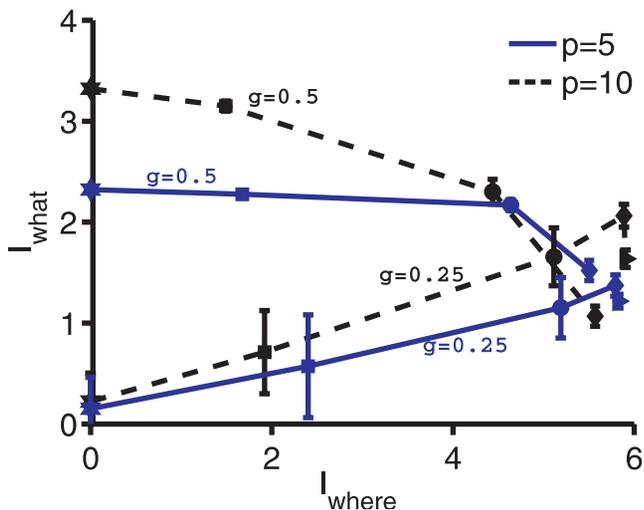


Figure 7. The relation between I_{where} and I_{what} in the network with low gain ($g=0.25$), with $p=5$ and $p=10$. When there is no gain modulation, the network fails to retrieve any of the stored patterns (star). When localised gain modulation is incorporated, both I_{what} and I_{where} increase. $\beta=1.5$ (square), $\beta=2$ (circle), $\beta=3$ (diamond) and $\beta=5$ (right pointing triangle). The corresponding relation between I_{where} and I_{what} for $g=0.5$ is replotted here from Fig. 5A for comparison. doi:10.1371/journal.pcbi.1000012.g007

the network come from nearby neurons, the second pattern would still affect the first retrieval bump, because of the global inhibition in the simplest version of our model, as inhibition is taken to regulate a common threshold, such that the mean activity of the network is constant (Eq. (3)). This introduces interactions between distal neurons, which are not directly connected by excitatory synapses, and such interactions are generally disruptive. A simple way to reduce such interaction is to assume that when the local mean activity in part of the network exceeds some limit value, the threshold is raised but only locally, regardless of the activity of neurons outside that region. The local threshold may also be regulated downward, to facilitate the emergence of a retrieved pattern in a region which would otherwise be kept at too low a mean activity level. With such additional provisions, multiple bumps can be formed and stabilised, as shown in the example in Fig. 8.

Discussion

Behaviour requires processing and integrating different types of information, from various sources and modalities, into a coherent picture of the world. Within the visual domain, a specific question is how the brain can analyse the identity of objects, which has to be extracted from raw visual input, while maintaining information about their position, directly present in the input.

Previous theoretical work on the representation of objects in neural circuits has been mainly focused on two issues [2,5,67–69]: how the hierarchy of visual cortical areas builds representations that are invariant with respect to changes in position, view, etc. of objects, and how this may be accomplished while still preserving information about the relative position of features within objects, to enable object recognition. Among these studies, Olshausen et al [2] and Parga and Rolls [3] also considered how attractor dynamics can be used to retrieve *what* information from stored invariant representations. However, this body of work did not address how an activity pattern that carries information about both *what* and *where* can be produced when *what* information is

retrieved from memory. They also did not consider how this combined *what* and *where* representation can be maintained in working memory, after the visual stimulation has subsided. Retrieving information about object identity from memory, as well as maintaining this information in working memory, has been associated to attractor dynamics in local cortical networks. The most straightforward extension of the attractor idea, to store attractors associated to *what-where* pairs is, however, infeasible due to the extremely large storage capacity that it would require (see the following section “Comparison with storing attractors associated to object-position pairs”).

Our model sheds light onto this issue of how to combine the representation of *what* and *where*, by showing that a recurrent network can retrieve stored memories about objects from incomplete transient cues, while maintaining information about their positions. It can account for the *what-where* delay activity observed in monkey electrophysiology [13,14] and it can combine *what* and *where* information in a flexible manner as has been reported in experimental studies [13,17,28]. In our model this flexibility is expressed in the fact that by changing the level of background gain and localised gain modulation, one can control the levels of *what* and *where* information that the network retains. When the network is operating in the low gain regime, turning on the localised gain modulation increases both *what* and *where* information, whereas in the high gain regime *what* information decreases and *where* information increases. Behavioural experiments show a pattern of interaction between *what* and *where* information similar to this latter case [70,71]. It is interesting to note that it has been recently reported [72] that single IT neurons, when they show high selectivity (i.e. they respond vigorously to only a few images in a large sample) also tend to show less position tolerance, suggestive of their ability to convey more *where* information. This could be interpreted either as the more selective neurons contributing less *what* information to the population response, or even as implying a different trade-off at the single neuron level from the one we propose to prevail at the population level.

The localised retrieval process described here also offers the possibility of retrieving multiple objects while maintaining their position, without facing a binding problem [73]. The metric excitatory connectivity avoids interference effects mediated by excitatory connections, while inhibitory mechanisms should be such that two activity patterns retrieved at different locations do not destroy each other once they are formed. One such mechanism is briefly described in section “Multiple bumps”. The crucial questions about the coexistence of multiple bumps are of course still open: how does it depend on the parameters of the model, and in particular on its detailed dynamics? how many bumps can simultaneously coexist in a network of a given size? how does the ability to support multiple bumps changes the storage capacity? These questions require further investigations. In the context of networks with spatially dependent weights without stored memory, it has been shown that oscillatory weights can support multiple bumps of activity [74,75]. The possibility of supporting multiple retrieval bumps using more complicated connectivity schemes remains open in our model.

Comparison with storing attractors associated to object-position pairs

In our model, independent attractors are set up in a local cortical network only for object identity, as position invariant representations; but they can be accessed in a spatially focused mode, leading to position dependent activity. Associating a single representation to an object, which is then modulated by position, is

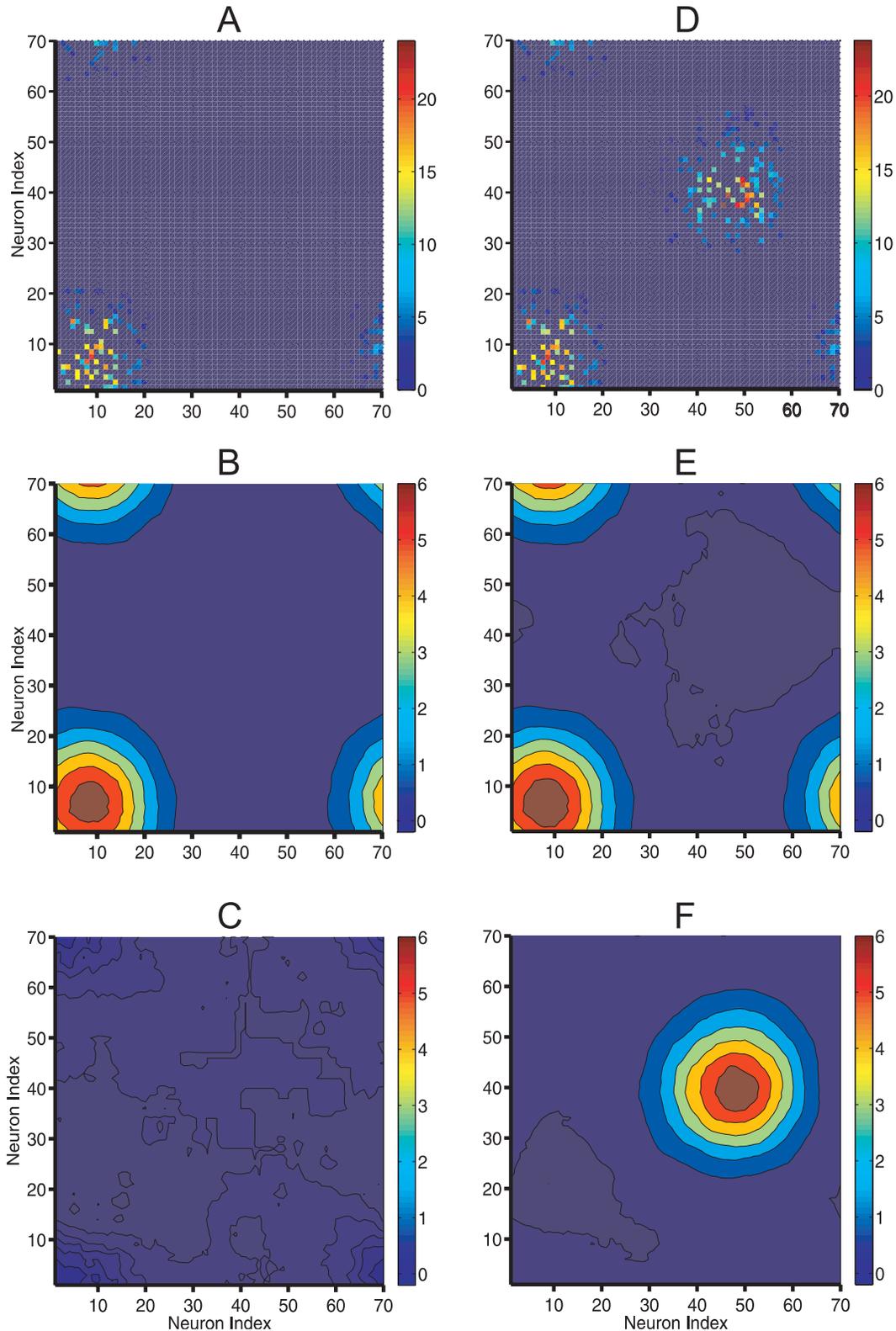


Figure 8. Formation of multiple bumps. In the beginning of the simulation, pattern 1 is presented to all units in a 15×15 square whose lower left corner is at node (1,1); that is, for each neuron i inside the square the activity is set to η_i^1 , and for those outside to zero. While simulating the network, background threshold values are set to regulate the mean network activity to a fixed level equal to 0.2. If mean activity inside a 30×30 square centred on the cue centre exceeds 1.0, the threshold of neurons inside this square will be regulated to keep its mean activity equal to 1.08, and neurons outside it will be assigned a high threshold. In the second phase of the simulation, in the right column, a second pattern is also presented to all units in a 15×15 square whose lower left corner is at node (36,36), accompanied by a local threshold decrease to facilitate the pattern “holding on”. The threshold is then regulated in the same way as the first pattern. (A) The distribution of activity, (B) the local overlap with the first pattern (cued in the beginning of the first phase) and (C) the local overlap with the second pattern (cued in the beginning of the second phase), all 100 time steps after the presentation of the first pattern. (D), (E), and (F) are the same quantities as (A), (B) and (C) but 100 time steps after the presentation of the second pattern.
doi:10.1371/journal.pcbi.1000012.g008

a particular case of what in cognitive neuroscience parlance is sometimes referred to as type (e.g. *table*) and token (particular instance of a type: e.g. a table in a particular position) [76,77]. In the language of our model, the type is the original pattern of activity associated to an object and the token is the bumpy pattern that is localised in a particular position. An alternative mechanism is to store attractors associated to object-position pairs, that is storing a neural activity pattern for each token [78]. In this way, when a particular object is presented in a particular position, the attractor corresponding to the object-position pair would be activated, and could remain activated even after the object has been removed from the scene. The problem is that models which hypothesise distinct, discrete attractors for each possible object-position combination would certainly violate any conceivable storage capacity limit, because of the infinitely large number of possible positions of an object [79]. Furthermore, there is a major difference between the nature of *what* and *where* information, which makes attractors associated to object-position pairs unlikely: as opposed to *what* information, to which the brain can contribute from the information that it has previously stored, the brain does not usually *retrieve* positional information from memory, but rather has to *maintain* it as well as it can. Thus, it would seem rather implausible that the brain uses its storage capacity, arguably its most precious resource [29], to store something that it does not have to retrieve.

The difference between these two mechanisms is directly reflected in the storage capacity required for object-position attractors, in order to represent the same amount of information as the mechanisms studied here does through spatial modulation. Representing 6 bits of I_{where} and 2 bits of I_{what} (corresponding to the black diamonds in Fig. 7) would require the storage of $2^{(6+2)} = 256$ distinct object-position attractors. This is ca. 2.5 times beyond the number of attractors that a randomly connected network, with the same number of connections per neuron and the same mean activity level as what we used, could store [32]. This exorbitant requirement is due to effectively committing storage space separately to each pair, instead of using the physical arrangement of neurons in the tissue to represent I_{where} . Analytical results valid in the limit of large networks and optimal storage further support this conclusion, as we show in section “Comparison with other models” in the Materials and Methods. There, we also show that the difference in the efficiency of the two models will be even more pronounced for larger networks. There is, of course, a price to pay: the addition of a gain modulation mechanism to stabilise the position of the bump. In what follows, we discuss the possible physiological substrates of this gain modulation.

Possible sources of gain modulation

In our model, localised gain modulation is crucial for maintaining *where* information as *what* information is being retrieved, and for maintaining both *what* and *where* information after the retrieval process is completed. When an object is presented as a stimulus, a signal should trigger an increase in the gain of neurons in an appropriate part of the network. Such higher gain should then be maintained by the same or a distinct mechanism during retrieval and thereafter, when the object is not present anymore but information about it has to be used (e.g. during the delay period of a delay-match-to-sample task).

What mechanisms can trigger the neuronal gain? *In vivo* studies show that increasing the activity of a local cortical network increases the gain of its neurons [80–82]. Therefore, any mechanism that increases the mean activity of a part of the network could be used for triggering the gain modulation. One such source of increase in the activity is the cue itself. This requires

that the pattern selective cue retains some spatial information; a scenario which we have shown to be particularly effective in minimising the trade-off between *what* and *where* information (see Fig. 6C). Although this mechanism would be effective in this sense, it is doubtful whether it could be the only source of gain modulation in high level visual cortices. This is because experimental studies show that the position of the peak of the activity in visual cortical areas during visual stimulation is strongly correlated with the categorical properties of the stimulus and exhibits a weaker level of retinotopy [83–85] (see also the following section “Storing patterns with spatial preference”). The situation may be different in more advanced cortical areas, such as PFC, in which such categorical maps have not been reported.

Another possible source for increasing the gain is attentional signals. In this case the increase in the activity level required for gain modulation is induced by the attentional signal and the position of the bump corresponds to the position of the attentional spotlight. There are several reasons that make attention a likely source of activity localisation through gain modulation. fMRI studies on human subjects show that the retinotopic representation of the position of an attended object in visual cortices show increased activity [86–94]. Evidence from monkey neurophysiology also supports the idea that the attentional spotlight increases the gain of neurons inside the spotlight [95–100]. Furthermore, many studies in cognitive neuropsychology suggest that spatial, focal attention is critical to allow the binding of *what* and *where* information [77,101], referred to as type and token information respectively [76]. Finally, a recent neuroimaging study shows that attention strongly enhances retinotopic representation in object selective visual areas, thus supporting the idea that attentional gain modulation is important for combined representation of *what* and *where* [102]. Although, these studies point to attentional signals as a strong candidate for initiating the gain modulation, a contribution may still be given by a weakly retinotopic initial cue. Further experimental work is required to disentangle the relative effect of the initial cue and attention on triggering the gain modulation.

Once the increase in the gain of neurons in the right part of the network is triggered, it should be maintained during retrieval. Although the same mechanisms that initiated gain modulation can keep the gain high during retrieval, a promising mechanism for maintaining high level of gain, particularly after the stimulus is removed, is single neuron memory. Several studies show that the recent history of spiking increases the responsiveness of neurons, and that this increase can last for several seconds, thus exhibiting a form of single neuron short-term memory [62–64]. Assuming that such single neuron short-term memory mechanisms are responsible for the higher gain of neurons inside the bump, global signals that turn them on or off can strongly affect the level of *what* and *where* information that the network represents in its activity.

As discussed above, the attentional signal may trigger the increase in neuronal gain and maintain it elevated for some time. After the attentional signal is removed, the increase in neuronal gain can be maintained by single neuron short term memory mechanisms. Attention can then be directed to another object, while *what* and *where* information about the first object is still decodable from neuronal activity. How long this information survives depends on how long the short term increase of the gain can be maintained by single neuron mechanisms. Understanding such mechanisms and comparing their time scale with behavioural times for maintaining combined *what* and *where* information, as well as pharmacologically interfering with them, one can test whether our model is relevant to real visual perception.

One of the roles of attention is to bias the competition for limited processing resources in favour of the object that it is acting

on [103,104]. Therefore, if the localised gain modulation that is needed in our model for combining *what* and *where* is induced by attention, it should be able to do the same. This is verified by computer simulations as shown in Fig. 9. Two localised partial cues, corresponding to two different objects, are simultaneously given to a network. When the neuronal gain is uniform, the object with the larger cue will be retrieved, while the other one will be suppressed. However, if the neuronal gain in the area that receives the smaller cue is sufficiently large, the competition will be biased in favour of it. Interestingly, the level of gain modulation that is required to bias the competition towards the object with the small cue depends on the width of the connectivity, σ . Increasing the width of the neuronal connectivity increases the minimum level of gain modulation that is required for biasing the competition. This emphasises the role of local connectivity.

Storing patterns with spatial preference

In the model presented here, the units are taken to be arranged on a retinotopic patch of cortex, corresponding to at least a portion of visual field, but we assumed patterns of activity to be

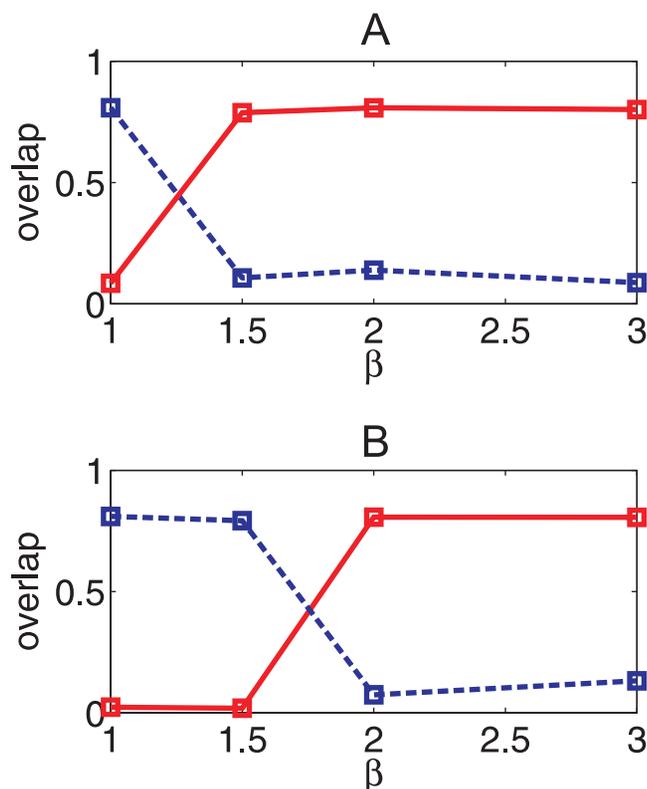


Figure 9. Two patterns, pattern one and two, are cued at two different positions at the beginning of the simulation. The cue corresponding to pattern one is given by setting the activity of neurons inside a 15×15 square box whose lower left corner is on neuron (1,1) to their activity in pattern one. The cue corresponding to pattern two is given by setting the activity of neurons inside a 20×20 square box whose lower left corner is on neuron (33,35) to their activity in pattern two. Plotted are the final dot product overlap (minus the mean activity; Eq. (9)) with the first stored pattern (full line) and the second stored pattern (dashed line) for two values of the connectivity width (A) $\sigma = 7.5$ and (B) $\sigma = 10$. Since the initial overlap with the second pattern is larger, without gain modulation it wins the competition, and it will be retrieved, as shown by the final dot product overlap with the two patterns. However, localised gain modulation biases the competition in favour of the first pattern.

doi:10.1371/journal.pcbi.1000012.g009

generated from a spatially uniform distribution (see Eq. (4)). A more realistic model, however, should allow for the storage of spatially organised patterns [105]. This is important since, in the case of high level visual cortical areas, the overall position of intense neural activity during visual stimulation is strongly correlated with object identity or category. Regions in the visual cortex have been located that are preferentially selective for faces [106–109], pictures of scenes [110,111] and buildings [112], and complex object features [113]. This strong categorical map may coexist with a retinotopic map. The details of this combined organisation are far from clear, however, particularly insofar as it is expressed in the putative attractor states, after the stimulus is removed (e.g. during delay periods), which is the situation relevant to our study. During visual stimulation, and when attention is not a main factor, some studies suggest that there is a weak retinotopy, with only a peripheral versus central bias and no angular representation [83–85]. Others, on the other hand, report the existence of multiple precise retinotopic maps in the same regions [114–116], although still much weaker than the level of retinotopy in primary visual areas [117]. As mentioned in the previous section, such retinotopic maps could be enhanced by attention [102].

To include the coexistence of categorical and retinotopic maps in the model presented here, one might consider two limit cases, which roughly correspond to these two views. In the first case, category specificity and weak retinotopy coexist at the same spatial scale; one should then assume, in a refined model, that patterns are generated from multiple distributions, each of them corresponding to one category of objects, and patterns drawn from each have higher activity at a *preferred* position on the network. In this case, when there is no gain modulation the peaks of the retrieved patterns cluster depending on which distribution they came from. The peaks will also be more weakly correlated with the position of the cue compared to the case of spatially uniform patterns that we have discussed. With attentional gain modulation, one expects to see a clearer retinotopic map. This is in fact consistent with the abovementioned finding that attentional gain modulation enhances the retinotopic representation in advanced visual areas [102]. In the second limit case, retinotopy is expressed in object selective visual areas at a finer scale than category specificity, in which case one should allow for the present model to be simply multiplexed, to include one array on a distinct cortical patch for each object category.

Further work is required, especially in view of many intermediate possibilities, to assess, for example, how much more gain modulation would be needed in order to stabilise a bump of activity away from its preferred position, and how this would affect retrieval.

Laminated networks

The ability to represent *what* and *where* information in the same network has also been proposed to be crucial to understand the functional significance of the differentiation among cortical layers [118]. Whereas most network models used to study attractor dynamics in associative memory do not consider cortical lamination, the core hypothesis of the proposal is that layer IV units, by virtue of their distinct connectivity, may privilege the representation of position information. Furthermore, through less adaptive spiking activity they may influence the dynamics of pyramidal units in the superficial layers only *after* these have engaged the attractor basin that leads to retrieve object identity. The differentiation was shown to be advantageous, in the model, through computer simulations, conducted with external inputs maintained active. In this regime no assessment was possible of

whether genuine dynamical attractors had indeed been formed during memory storage, that will drive network dynamics in the absence of the cue. While the present work clarifies the conditions allowing a single layer network to represent *what* and *where* information, how they could be realized in a network with differentiated cortical layers remains to be explored.

Combining other variables

In discussing *what* and *where* information, we have made explicit reference, here, to object identity and position in the visual field. *Where* information could however be any feature that is mapped in the gross topography of the cortical sheet, such as frequency in the auditory system [119], and in relation to which there is no meaning to using attractor dynamics in order to refine the afferent signal with what is stored in memory. In fact, this mapping need not even be topographically organised: the crucial factor is the existence of a map (topographic or not) [120], that is produced as a result of the dependence of ϖ_{ij} in Eq. (5) on i and j , and that is independent of the stored patterns. *Where* information would ideally be expressed by a continuous attractor and thus maintained e.g. as delay activity, except that continuity at a fine scale is disrupted by the storage of *what* memories. *What* information could instead be any feature that could benefit from attractor dynamics, because of its uneven statistical distribution, which makes some interpretation of the afferent signal more likely than others.

Materials and Methods

Synaptic weights that follow Dale's law

If synaptic weights are produced by Eq. (5), the weights of the connections that originate from a given neuron can be both negative and positive. This is against Dale's law and against our assertion that all neurons in the model network are excitatory. In this section, we show how the model described in sections "Firing rate description of the network" and "Stored memory patterns and synaptic weights" (see Model) can be conceptually derived from a more realistic formulation, in which all synaptic weights are positive. Let us first consider a network in which the firing rate of neuron i at time $t+1$ is determined by

$$v_i(t+1) = g_i F(H_i(t) - \text{Th}_i - I_i), \quad (14)$$

in which Th_i is the threshold of neuron i , I_i is its inhibitory input, and

$$H_i(t) = \sum_{j \neq i} W_{ij} v_j(t). \quad (15)$$

The synaptic weights, W_{ij} , in this network take the following form

$$W_{ij} = \frac{\varpi_{ij}}{Ca^2} \left(J_{back} + \sum_{\mu=1}^p (\eta_i^\mu - a)(\eta_j^\mu - a) \right), \quad (16)$$

where J_{back} is the background weight, $\varpi_{ij}=1$ if there is a connection from neuron j to neuron i and $\varpi_{ij}=0$ otherwise, and C is the average number of connections per neurons. For sufficiently large J_{back} , the resulting synaptic weights in Eq. (16) will be all positive. We can now show that a network with uniform threshold, as assumed in Eq. (2), and synaptic weights of the form Eq. (5), has equivalent dynamics as described by Eqs. (14) and (16), when an additional condition is satisfied.

Combining Eq. (14) with Eqs. (15) and (16), the firing rate of neuron i can be written in terms of the firing rate of the other neurons as

$$v_i(t+1) = g_i F \left(\sum_j J_{ij} v_j(t) + \frac{J_{back}}{Ca^2} \sum_j \varpi_{ij} v_j(t) - \text{Th}_i - I_i \right), \quad (17)$$

in which J_{ij} is the weight of the connection from neuron j to neuron i according to the prescription Eq. (5). The assumption we now make is that the inhibitory feedback reacts in such a way that for each neuron, the last three terms in the parenthesis in Eq. (17) together become equal to a uniform effective threshold, Th . This effective threshold is simply chosen such that Eq. (3) holds. In this way, Eq. (17) reduces to

$$v_i(t+1) = g_i F \left(\sum_j J_{ij} v_j(t) - \text{Th} \right), \quad (18)$$

which is the same as Eq. (2).

Self-consistent equations

In this section we briefly describe how the self-consistent equation for the local overlap with the retrieved pattern (Eq. (10)) can be derived. We refer the reader to [32,33] for more details.

To start with, we assume, without loss of generality, that the first pattern ($\mu=1$) is retrieved and therefore $m_i^1 \gg m_i^\mu$ for $\mu \neq 1$. Using Eqs. (1), (5) and (8), we then write the input to neuron i as

$$h_i = \left(\frac{\eta_i^1}{a} - 1 \right) m_i^1 + \sum_{\mu \neq 1, j} \frac{\varpi_{ij}}{C} \left(\frac{\eta_i^\mu}{a} - 1 \right) \left(\frac{\eta_j^\mu}{a} - 1 \right) v_j. \quad (19)$$

Defining z_i as

$$z_i = \sum_{\mu \neq 1, j} \frac{\varpi_{ij}}{C} \left(\frac{\eta_i^\mu}{a} - 1 \right) \left(\frac{\eta_j^\mu}{a} - 1 \right) v_j, \quad (20)$$

and combining Eq. (19) and Eq. (2), the activity of neuron i can be written as

$$v_i = g_i F \left(\left(\frac{\eta_i^1}{a} - 1 \right) m_i^1 + z_i - \text{Th} \right). \quad (21)$$

Inserting Defining v_i from Eq. (21) into Eq. (8) we arrive at the following self-consistent equation for $m_i = m_i^1$

$$m_i = \frac{1}{C} \sum_j \varpi_{ij} g_j \left(\frac{\eta_j^1}{a} - 1 \right) F \left(\left(\frac{\eta_j^1}{a} - 1 \right) m_j + z_j - \text{Th} \right). \quad (22)$$

Averaging the right hand side of Eq. (22) over the distribution of z_j , η , and the connectivity pattern, yields the following equation (which is the same as Eq. (10))

$$m_i = \frac{1}{C} \sum_j \varphi_{ij} g_j \left\langle \left(\frac{\eta}{a} - 1 \right) \bar{F}_j \left(\left(\frac{\eta}{a} - 1 \right) m_j - \text{Th} \right) \right\rangle_\eta, \quad (23)$$

where $\langle \rangle_\eta$ stands for averaging over the distribution of η , φ_{ij} is the probability of connection (Eq. (7)), and \bar{F}_j is the gain function, F ,

averaged over the distribution of z_j

$$\bar{F}_j^k(h) \equiv \int dz \Pr_j(z_j) F^k(h+z_j). \quad (24)$$

We now find the distribution of z_j , which we denote by $z_j \Pr_i(z_i)$. To do this we note that if the first pattern is retrieved, v_{js} , on the right hand side of Eq. (20) will be independent from each other and from η^μ for $\mu \neq 1$. The assumption of independence is strictly correct when the network is highly diluted, that is when the number of presynaptic neurons shared by any two postsynaptic neurons is small [121,122]. When the network is *not* highly diluted, the calculation will be more involved, but yields qualitatively the same results [32,123]. Thus, for the sake of simplicity, we assume that the assumption of independence holds; for a complete derivation we refer the reader to the aforementioned references. With this independence assumption, the right hand side of Eq. (20) will be a sum of independent random variables, and therefore, $\Pr_i(z_i)$ will be a Gaussian distribution. In the following we show that the mean of this Gaussian distribution is zero and also find a self-consistent equation for its variance.

Noting that

$$\left\langle \left(\frac{\eta_j^\mu}{a} - 1 \right) \right\rangle_\eta = 0, \quad (25)$$

$$\left\langle \left(\frac{\eta_j^\mu}{a} - 1 \right)^2 \right\rangle_\eta = \frac{1}{a} - 1 \equiv T_0, \quad (26)$$

and using Eq. (21), we get the following equations

$$\langle z_i \rangle_{\eta, \varpi_{ij}} = 0, \quad (27)$$

$$\delta_i^2 \equiv \langle z_i^2 \rangle_{\eta, \varpi_{ij}} = \frac{\alpha T_0^2}{C} \sum_j \varphi_{ij} g_j^2 \left\langle F^2 \left(\left(\frac{\eta}{a} - 1 \right) m_j + z_j - \text{Th} \right) \right\rangle_\eta, \quad (28)$$

where $\langle \rangle_{\eta, \varpi_{ij}}$ indicates averaging over the distribution of η and ϖ_{ij} , and $\langle \rangle_\eta$ indicates averaging over η . From Eq. (27), we see that the mean of $\Pr_i(z_i)$ is zero. In order to find the variance of $\Pr_i(z_i)$, we should average both sides of Eq. (28) over the distribution of z_j . This is because, in the limit of large N and large C , this variance is expected not to depend on the exact realisation of any z_j in the right hand side of Eq. (28), but only on its statistical distribution. Performing this average yields the following equation for the variance that we denote by ρ_i^2

$$\rho_i^2 \equiv \langle \delta_i^2 \rangle_{\{z_j\}} = \frac{\alpha T_0^2}{C} \sum_j \varphi_{ij} g_j^2 \left\langle \bar{F}_j^2 \left(\left(\frac{\eta}{a} - 1 \right) m_j - \text{Th} \right) \right\rangle_\eta. \quad (29)$$

Equations (23) and (29) form a closed set of equations whose solutions determine the steady states of the system. Finding m_i and ρ_i that satisfy these equations, we can find the activity of neurons in the steady states by plugging them in Eq. (21). In the case of a randomly connected network, that is when φ_{ij} is independent of i and j , and g_i are also the same for all neurons, the solution of Eqs. (23) and (29) will be of the form $m_i = m$ and $\rho_i = \rho$. In this case the only spatial dependence of the steady state activities, Eq. (21),

will come from the dependence of η_i^1 on i and since they are generated identically for each i , the probability that a neuron is active in the steady state will be uniform over the network. Spatially localised retrieval can be observed when φ_{ij} depends on the distance between i and j .

Mutual information measures

In this section we show how we compute *what* and *where* information, I_{what} and I_{where} , from simulations. We estimate the amount of *what* information, I_{what} , from the frequency of successful retrieval runs. To see how, let us assume that we cue pattern μ_c . Then after some time we look at the pattern of activity of the network, compute its dot product overlap with all stored patterns (Eq. (9)) and find that pattern μ_r , say, has been “retrieved” in that particular run, i.e., it has the highest overlap with the activity of the network. We denote the probability of retrieving pattern μ_r given that we have cued pattern μ_c by $\Pr(\mu_r|\mu_c)$. Estimating this probability from the simulations, we can compute the information that the pattern of activity gives us about which pattern was presented as

$$I_{\text{what}} = \sum_{\mu_c, \mu_r} \Pr(\mu_r|\mu_c) \Pr(\mu_c) \log \left(\frac{\Pr(\mu_r|\mu_c)}{\Pr(\mu_r)} \right), \quad (30)$$

where $\Pr(\mu_c)$ is the probability of cueing pattern μ_c and

$$\Pr(\mu_r) = \sum_{\mu_c} \Pr(\mu_r|\mu_c) \Pr(\mu_c). \quad (31)$$

In the simulations all patterns are presented an equal number of times, therefore,

$$\Pr(\mu_c) = p^{-1}. \quad (32)$$

We denote the fraction of successful runs (when $\mu_c = \mu_r$) that we measure from the simulations by f , that is

$$\Pr(\mu_c|\mu_c) = f. \quad (33)$$

Since in unsuccessful runs (when $\mu_c \neq \mu_r$), all patterns, except for μ_c are *a priori* equally likely to be retrieved, we have

$$\Pr(\mu_r|\mu_c) = \frac{(1-f)}{p-1} \quad \mu_c \neq \mu_r. \quad (34)$$

Using Eqs. (32)–(34) in Eq. (30), we can thus write for fixed degree of gain modulation, fixed background gain, and fixed number of patterns,

$$I_{\text{what}} = \log_2(p) + f \log_2(f) + (1-f) \log_2 \left(\frac{1-f}{p-1} \right). \quad (35)$$

Note that the above is, strictly speaking, only a measure of the information implicit in the selection among the p patterns operated by attractor dynamics; under certain conditions, however, it can also serve as an indicator of the total information available in the firing pattern itself [124].

I_{where} is the mutual information between the peak of the local overlap after 200 time steps and the centre of the gain modulated area (or the centre of the cue when there is no gain modulation). To estimate *where* information, I_{where} , we first measure the distance between the peak of the final overlap of the successful runs and the centre of the gain modulation, for each cued pattern. Then we make a histogram of these distances and calculate the fraction of runs which fall in any of the 10 distance bins chosen to be $b_1 = [0,5], b_2 = [5,10], \dots, b_{10} = [45,50]$. In this way we have the conditional probability, $\Pr(k | x)$, of having the peak of the activity in the k th distance bin, given that the peak was initially at position x on the lattice. With N neurons on each side of the lattice, we have $\Pr(x) = 1/N^2$, and we can write I_{where} as

$$\begin{aligned} I_{where} &= \sum_{x,k} \Pr(k|x) \Pr(x) \log_2 \left(\frac{\Pr(k|x)}{\Pr(k)} \right) \\ &= \sum_{k=1}^{10} \Pr_k \log_2 \left(\frac{\Pr_k N^2}{(2k-1)(5 \times 5)\pi} \right) \\ &= \log \left(\frac{N^2}{25\pi} \right) + \sum_{k=1}^{10} \Pr_k \log_2 \left(\frac{\Pr_k}{2k-1} \right), \end{aligned} \quad (36)$$

in which we have used the fact that $\Pr(k | x)$ does not explicitly depend on x and we can simply denote it by \Pr_k . Similarly to what we do for I_{what} , we have also assumed that for any such ring between the circles of radius $5k$ and $5(k-1)$, centred on the gain modulation square, the final bump can be anywhere, with equal probability, on the ring. In this expression the factor $2k-1$ accounts for the fact that the area covered by the k th bin is $2k-1$ times the area of the first bin, and hence its *a priori* probability is $2k-1$ times higher. The first term in Eq. (36), is the maximum information value, $I_{where} \approx 6$ bits, in this approximation, i.e., the logarithm (in base 2) of the ratio between the ‘‘area’’ of the network (4900) and that of the smallest bin ($5 \times 5 \times \pi$), and is achieved when all successful runs end up with a bump at $d \leq 5$ from its intended position.

Low gain regime versus high gain regime

In this section we discuss why in the low gain regime, gain modulation aids retrieval of the patterns whereas in the high gain regime it has a negative effect. We start from the self-consistent equations, Eqs. (23) and (29).

Assume that the steady state of the network is a bump of activity over a part of the network with single neuron gain βg , whereas the rest of the network is silent with gain g . Furthermore, assume that m_i and ρ_i that satisfy Eqs. (23) and (29) are nonzero inside the bump and zero elsewhere. Consider that inside the bump $m_i = m$ and $\rho_i = \rho$, where m and ρ can be regarded, just for simplicity, to be roughly constant. Then from Eqs. (23) and (29) we have:

$$m = \beta g \int Dx \left\langle \left(\frac{\eta}{\alpha} - 1 \right) F \left(\left(\frac{\eta}{\alpha} - 1 \right) m + \rho x - \text{Th} \right) \right\rangle_{\eta} \quad (37a)$$

$$\rho^2 = \alpha T_0^2 \beta^2 g^2 \int Dx \left\langle F^2 \left(\left(\frac{\eta}{\alpha} - 1 \right) m + \rho x - \text{Th} \right) \right\rangle_{\eta}, \quad (37b)$$

where $\alpha = \beta/C$ is the storage load and

$$Dx \equiv \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2\pi} \right). \quad (38)$$

Eqs. (37) are of the form of mean-field equations of a recurrent network with non-metric connections [79,125] (assuming uniform

values for m_i and ρ_i inside the bump and zero outside is equivalent to assuming that the part of the network, over which the bump is formed, is behaving as an independent network). For each value of α , Eqs. (37) have non-zero solution for m , and thus the network can retrieve the stored patterns, if and only if $g_{min}(\alpha) < \beta g < g_{max}(\alpha)$, where $g_{max}(\alpha)$ and $g_{min}(\alpha)$ are functions of α . The effect of background gain g can now be readily seen. When $g < g_{min}(\alpha)$ retrieval does not happen without gain modulation. With gain modulation, however, the neuronal gain of the part of the network that is gain modulated will be boosted by a factor of β and for large enough β , the neuronal gain will be in the regime that supports retrieval i.e. $g_{min}(\alpha) < \beta g < g_{max}(\alpha)$. When the background gain g is high, βg can exceed $g_{max}(\alpha)$, thus retrieval will not be successful.

Comparison with other models

In this section, we discuss why it is more efficient to spatially modulate attractor states associated to objects, than to store distinct attractors for different positions of each object.

Under optimal conditions, the number of attractors that an associative memory with C connections per neuron, but without metric connectivity, can retrieve is

$$p_{max} = kC, \quad (39)$$

where k is a constant that is primarily determined by the sparsity of the stored patterns [79]. Metric connectivity, which enables localised retrieval, decreases k by a moderate factor $\gamma_1 \approx 3-4$ [32]. Localised gain modulation, that stabilises the bump at an arbitrary position, decreases k again by another factor, γ_2 , that for the parameters and network size we used turns out to be $\gamma_2 \approx 4$. This is actually an overestimation of the decrease in storage capacity due to localised gain modulation, for realistic size networks. This is because when we calculate the mean of the right hand side of Eq. (22) over the distribution of connectivity patterns and η to get Eq. (23), we ignore the fluctuations around this mean, that behave as $O(1/\sqrt{C})$. These fluctuations are what break the translational symmetry of the self-consistent equation, Eq. (23), and make the bump favour a few positions over the others, and are compensated for by the localised gain modulation. As a result, less gain modulation is required for stabilising the bump when there are more connections per neuron. However, even with this estimate for γ_2 , the process described here results in a moderate reduction in storage capacity

$$p_{max} = \frac{k}{\gamma_1 \gamma_2} C. \quad (40)$$

The spatial modulation described here can represent positional information with a resolution $\ell = l\sqrt{N/N_p}$, where l is the lattice spacing and N_p is the number of distinct position that can be resolved-in a large network, $N_p \sim O(N)$ (see Fig. 5).

On the other hand, the naive storage of distinct, unrelated attractors for each object position pair decreases the number of objects, whose identity could be retrieved, to

$$p_{max} = kC/N_p, \quad (41)$$

illustrating the wasteful use of memory resources for positional information, which in itself requires no memory.

An alternative arrangement might be to associate attractors to objects, but allow each attractor to be a continuous 2D manifold, different for each object, so that position can be represented by the position of a bump of activity on such attractor manifold, unrelated to the position of the active neurons in the tissue. This arrangement corresponds to the multiple spatial charts model of Samsonovich and McNaughton [58], introduced to account for the ability of rodents to track their own position in multiple spatial environments, by coding it as a group of coactive hippocampal place cells, which comprise a bump on a chart corresponding to each environment. Instead of assigning distinct charts to distinct spatial contexts, such as a square recording box rather than a circular one, one could well assign distinct charts to distinct objects, each of which would then have its own “private” continuous or quasi-continuous attractor, unrelated to the 2D arrangement of neurons in the tissue. The mathematical analysis of the multiple charts model [126] reveals that a network can store a number of charts equivalent to the number of attractors in a standard associative network of the same connectivity, reduced by a factor N_b , which is the number of place cell ensembles, uncorrelated with each other, required to “tile” a chart. In the simplest version of the model, each neuron shows a single place field in each environment (at a different spatial position in each chart) covering a fraction a of the total area of the environment. Then $N_b \approx (1/a)$ and, although the number of positions that can be represented accurately can be larger than N_b , still $a \ll 1$ for the network to be able to resolve position in space. Therefore, adapting the hippocampal model would also yield a

lower capacity

$$p_{max} \simeq kaC, \quad (42)$$

because of the cost of creating a separate “virtual” space for each object. Simply utilising the position of neurons in the tissue to represent physical position for all objects, and reserving memory resources for object identity, provides the most efficient solution to combine *what* and *where* information. Note instead that in the hippocampus, to the extent that it utilises coactivity patterns to discriminate between different spatial contexts [127], the position of neurons in the tissue cannot be used to code for position in real space, and in fact place field position in the chart is found to be unrelated to cell position in the tissue [128]. It is also worth mentioning that the same problem that we encountered for stabilising the bump at an arbitrary position will also appear in models that associate a distinct chart to each object [57]. Therefore, an extra mechanism will be required in this case, too, and the real p_{max} will be smaller than p_{max} in Eq. (42) by a factor similar to γ_2 in our model.

Acknowledgments

Discussions with Bharathi Jagadeesh, Peter Latham, and Peter Dayan are gratefully acknowledged, as well as helpful suggestions by the anonymous reviewers.

Author Contributions

Conceived and designed the study: YR AT. Performed the experiments/simulations/mathematical derivations: YR. Analyzed the data: YR. Wrote the paper: YR AT.

References

- Amit DJ (1989) Modeling brain function. Cambridge: Cambridge University Press.
- Olshausen BA, Anderson CH, van Essen DC (1993) A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J Neurosci* 13: 4700–4719.
- Parga N, Rolls ET (1998) Transform invariant recognition by association in a recurrent network. *Neural Computation* 10: 1507–1525.
- Bartlett MS, Sejnowski TJ (1998) Learning viewpoint-invariant face representations from visual experience in an attractor network. *Network: Comp Neur Sys* 9: 399–417.
- Deco G, Rolls ET (2002) Computational Neuroscience of Vision. Oxford: Oxford University Press.
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Nat Acad Sci U S A* 79: 2554–2558.
- Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. *Science* 173: 652–654.
- Kubota K, Niki H (1971) Prefrontal cortical unit activity and delayed alternation performance in monkeys. *J Neurophysiol* 34: 337–347.
- Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci* 16: 5154–5167.
- Miyashita Y, Chang HS (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature* 331: 68–70.
- Miyashita Y (1988) Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* 335: 817–820.
- Sakai K, Miyashita Y (1991) Neural organization for the long-term memory of paired associates. *Nature* 354: 152–155.
- Rao SC, Rainer G, Miller EK (1997) Integration of what and where in the primate prefrontal cortex. *Science* 276: 821–824.
- Rainer G, Asaad WF, Miller EK (1998) Memory fields of neurons in the primate prefrontal cortex. *Proc Nat Acad Sci USA* 95: 15008–15013.
- Postle BR, D’Esposito M (1999) What-then-where in visual working memory: An event-related fMRI study. *J Cogn Neurosci* 11: 585–597.
- Hagler DJ, Sereno MI (2006) Spatial maps in frontal and prefrontal cortex. *Neuroimage* 29: 567–577.
- Sala JB, Courtney SM (2007) Binding of what and where during working memory maintenance. *Cortex* 43: 5–21.
- Baddeley AD (2007) Working memory, thought and action. Oxford: Oxford University Press.
- Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4: 2051–2062.
- Tovee MJ, Rolls ET, Azzopardi P (1994) Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert monkey. *J Neurophysiol* 72: 1049–1060.
- Leuschow A, Miller E, Desimone R (1994) Inferior temporal mechanisms for invariant object recognition. *Cereb Cortex* 5: 523–531.
- Op de Beeck H, Vogels R (2000) Spatial sensitivity of macaque inferior temporal neurons. *J Comp Neurol* 426: 505–518.
- DiCarlo JJ, Maunsell JH (2003) Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J Neurophysiol* 89: 3264–3278.
- Rolls ET, Aggelopoulos NC, Zheng F (2003) The receptive fields of inferior temporal cortex neurons in natural scenes. *J Neurosci* 23: 339–348.
- Hung CP, Kreiman GK, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. *Science* 310: 863–866.
- Hansen KA, Kay KN, Gallant J (2007) Topographic organization in and near human visual area v4. *J Neurosci* 27: 11896–11911.
- Schiller PH (1993) The effects of V4 and middle temporal (MT) area lesions on visual performance in the rhesus monkey. *Visual Neuroscience* 10: 717–746.
- Duncan J (2001) An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience* 2: 820–829.
- Braitenberg V, Schuz A (1998) Cortex: statistics and geometry of neuronal connectivity. Berlin: Springer.
- Hellwig B (2000) A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. *Biol Cyber* 82: 111–121.
- Holmgren C, Svennensfors B, Zilberter Y (2003) Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *J Physiol* 551.
- Roudi Y, Treves A (2004) An associative network with spatially organized connectivity. *J Stat Mech: Theo and Exp* 1: P070102.
- Roudi Y, Treves A (2006) Localized activity profiles and storage capacity of rate-based associative networks. *Physical Review E* 73: 061904/1–9.
- Armstrong KM, Fitzgerald JK, Moore T (2006) Changes in visual receptive fields with microstimulation of frontal cortex. *Neuron* 50: 791–798.
- Hamker FH, Zirnsaka M (2006) V4 receptive field dynamics as predicted by a systems-level model of visual attention using feedback from the frontal eye field. *Neural Networks* 19: 1371–1382.
- Treves A (1993) Mean-field analysis of neuronal spike dynamics. *Network* 4: 259–284.
- Shriki O, Hansel D, Sompolinsky H (2003) Rate models for conductance-based cortical neuronal networks. *Neural Comput* 15: 1809–1841.

38. Shu Y, Hasenstaub A, McCormick DA (1993) Turning on and off recurrent balanced cortical activity. *Nature* 423: 288–293.
39. Haider B, Duque A, Hasenstaub AR, McCormick DA (2006) Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J Neurosci* 26: 4535–4545.
40. van Vreeswijk C, Sompolinsky H (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274: 1724–1726.
41. van Vreeswijk C, Sompolinsky H (1998) Chaotic balanced state in a model of cortical circuits. *Neural Comput* 10: 1321–1371.
42. Renart A, Moreno-Bote R, Wang XJ, Parga N (2006) Mean-driven and fluctuation-driven persistent activity in recurrent networks. *Neural Comp* 19: 1–46.
43. Roudi Y, Latham P (2007) A balanced memory network. *PLoS Comp Biol* 3: e141.
44. Tsodyks MV, Feigman MV (1988) The enhanced storage capacity in neural networks with low activity level. *Europhys Lett* 6: 101–106.
45. Buhmann J, Divko R, Schulten K (1989) Associative memory with high information content. *Phys Rev A* 39: 2689–2692.
46. Sejnowski T (1977) Storing covariance with nonlinearly interacting neurons. *J Math Biol* 4: 303–321.
47. Sompolinsky H (1986) Neural networks with nonlinear synapses and a static noise. *Phys Rev A* 34: 2571–2574.
48. Anishchenko A, Treves A (2007) Autoassociative memory retrieval and spontaneous activity bumps in small-world networks of integrate-and-fire neurons. *J Physiol (Paris)* 100: 225–236.
49. Koroutchev K, Korutcheva E (2005) Conditions for the emergence of spatial asymmetric states in attractor neural network. *Cent Eur J Phys* 3: 409–419.
50. Koroutchev K, Korutcheva E (2006) Bump formation in a binary attractor neural network. *Phys Rev E* 73: 027107/1–11.
51. Amari S (1977) Dynamics of pattern formation in lateral-inhibition type neural fields. *Biolog Cybernet* 27: 77–87.
52. Ben-Yishai R, Bar-Or RL, Sompolinsky H (1995) Theory of orientation tuning in visual cortex. *Proc Natl Acad Sci U S A* 92: 3844–3848.
53. Zhang K (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *J Neurosci* 16: 2112–2126.
54. Compte A, Brunel N, Goldman-Rakic P, Wang X (2000) Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb Cortex* 10: 910–923.
55. Laing CR, Chow CC (2001) Stationary bumps in networks of spiking neurons. *Neural Comp* 13: 1473–1494.
56. Chow CC, Coombes S (2006) Existence and wandering of bumps in a spiking neural network model. *SIAM J Appl Dynamical Syst* 5: 552–574.
57. Tsodyks MV, Sejnowski TJ (1995) Associative memory and hippocampal place cells. *Intl J Neural Systems* 6: 81–86.
58. Samsonovich A, McNaughton BL (1997) Path integration and cognitive mapping in a continuous attractor neural network model. *J Neuroscience* 17: 5900–5920.
59. Stringer SM, Trappenberg TP, Rolls ET, de Araujo IE (2002) Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells. *Network* 13: 217–242.
60. Renart A, Song P, Wang XJ (2003) Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron* 38: 473–482.
61. Hamaguchi K, Hatchett JPL, Okada M (2006) Analytic solution of neural network with disordered lateral inhibition. *Phys Rev E* 73: 051104/1–7.
62. Turrigiano GG, Marder E, Abbott LF (1996) Cellular short-term memory from a slow potassium conductance. *J Neurophys* 75: 963–966.
63. Marder E, Abbott LF, Turrigiano GG, Liu Z, Golowasch J (1996) Memory from the dynamics of intrinsic membrane currents. *Proc Natl Acad Sci USA* 93: 13481–13486.
64. Daoual G, Debanne D (2003) Long-term plasticity of intrinsic excitability: Learning rules and mechanisms. *Learning Memory* 10: 456–465.
65. McCarley JS, Mounds JR, Krame AF (2007) Spatially mediated capacity limits in attentive visual perception. *Acta Psychol (Amst)* 126: 98–119.
66. Roudi Y, Treves A (2003) Disappearance of spurious states in analog associative memories. *Phys Rev E* 67: 041906.
67. Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. *Nat Neurosci* 2: 1019–1025.
68. Riesenhuber M, Poggio T (1999) Are cortical models really bound by the ‘binding problem’? *Neuron* 24: 87–93.
69. Riesenhuber M, Poggio T (2000) Models of object recognition. *Nat Neurosci* 3: 1199–1204.
70. Akdal G, Hodgson TL, Hill AC, Mannan S, Kennard C (2002) Visual object memory and memory-guided saccades rely on shared mental representations. *Exp Brain Res* 143: 509–514.
71. Finke K, Bublak P, Neugebauer U, Zihl J (2005) Combined processing of what and where information within the visuospatial scratchpad. *Eur J Cogn Psych* 17: 1–22.
72. Zoccolan D, Kouh M, Poggio M, DiCarlo JJ (2007) Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *J Neurosci* 27: 12292–12307.
73. Aggelopoulos NC, Rolls ET (2005) Natural scene perception: inferior temporal cortex neurons encode the positions of different objects in the scene. *Eur J Neurosci* 22: 2903–2916.
74. Zemel RS, Pillow JW (2000) Encoding multiple orientations in a recurrent network. *Neurocomputing* 32–33: 609–616.
75. Laing CR, Troy WC, Gutkin BS, Ermentrout GB (2002) Multiple bumps in a neuronal model of working memory. *SIAM Journal of Applied Mathematics* 63: 62–97.
76. Kanwisher N (1987) Repetition blindness: Type recognition without token individuation. *Cognition* 27: 117–143.
77. Kanwisher N (1991) Repetition blindness and illusory conjunctions: Errors in binding visual types with visual tokens. *J Exp Psych: Human Perception Performance* 17: 404–421.
78. Deco G, Rolls ET, Horwitz B (2004) What and where in visual working memory: a computational neurodynamical perspective for integrating fmri and single-neuron data. *J Cogn Neurosci* 16: 683–701.
79. Treves A, Rolls ET (1991) What determines the capacity of autoassociative memories in the brain? *Network: Comp Neural Syst* 2: 371–397.
80. Arieli A, Sterkin A, Grinvald A, Aertsen A (1996) Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science* 273: 1868–1871.
81. Azouz R, Gray CM (1999) Cellular mechanisms contributing to response variability of cortical neurons in vivo. *J Neurosci* 19: 2209–2223.
82. Haider B, Duque A, Hasenstaub AR, Yu Y, McCormick DA (2007) Enhancement of visual responsiveness by spontaneous local network activity in vivo. *J Neurophysiol* 97: 4186–4202.
83. Grill-Spector K, Kushnir T, Hendler T, Edelman S, Itzhak Y, et al. (1998) A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp* 6: 316–328.
84. Levy I, Hasson U, Avidan G, Hendler TRM (2001) Center-periphery organization of human object areas. *Nat Neurosci* 4: 533–539.
85. Malach R, Levy I, Hasson U (2002) The topography of high-order human object areas. *Trends Cogn Sci* 6: 176–184.
86. Kastner S, De Weerd P, Desimone R, Ungerleider L (1998) Mechanisms of directed attention in the human extrastriate cortex as revealed by functional mri. *Science* 282: 108–111.
87. Tootell R, Hadjikhani N, Hall E, Marrett S, Vanduffel W, et al. (1998) The retinotopy of visual spatial attention. *Neuron* 21: 1409–1422.
88. Brefczynski J, DeYoe E (1999) A physiological correlate of the “spotlight” of visual attention. *Nat Neurosci* 2: 370–374.
89. Gandhi S, Heeger D, Boynton G (1999) Spatial attention affects brain activity in human primary visual cortex. *Proc Natl Acad Sci U S A* 96: 3314–3319.
90. Martinez A, Arillo-Vento L, Sereno M, Frank L, Buxton R, et al. (1999) Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat Neurosci* 2: 364–369.
91. Somers D, Dale A, Seiffert A, Tootell R (1999) Functional mri reveals spatial specific attentional modulation in human primary visual cortex. *Proc Natl Acad Sci U S A* 96: 1663–1668.
92. Muller N, Bartelt O, Donner T, Villringer A, Brandt S (2003) A physiological correlate of the zoom lens of visual attention. *J Neurosci* 23: 3561–3565.
93. McMains S, Somers D (2004) Multiple spotlights of attentional selection in human visual cortex. *Neuron* 42: 677–686.
94. Sylvester CM, Shulman GL, Jack AI, Corbetta M (2007) Asymmetry of anticipatory activity in visual cortex predicts the locus of attention and perception. *J Neurosci* 27: 14424–14433.
95. Connor CE, Gallant JL, Preddie DC, van Essen DC (1996) Responses in area v4 depend on the spatial relationship between stimulus and attention. *J Neurophysiol* 75: 1306–1308.
96. Connor CE, Gallant JL, Preddie DC, van Essen DC (1997) Spatial attention effects in macaque area v4. *J Neurosci* 17: 3201–3214.
97. Treue S, Martinez-Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399: 575–579.
98. McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *J Neurosci* 19: 431–441.
99. McAdams CJ, Maunsell JH (2000) Attention to both space and feature modulates neuronal responses in macaque area v4. *J Neurophysiol* 83: 1751–1755.
100. Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. *Annu Rev Neurosci* 27: 611–647.
101. Treisman A, Schmidt H (1982) Illusory conjunctions in the perception of objects. *Cognitive Psychology* 14: 107–141.
102. Saygin AP, Sereno MI (2007) Attention and retinotopic maps in human cortex. *Cerebral Cortex*. In press.
103. Yeshurun Y, Carrasco M (1998) Attention improves or impairs visual performance by enhancing spatial resolution. *Nature* 396: 72–75.
104. Kastner S, Ungerleider L (2000) Mechanisms of visual attention in the human cortex. *Ann Rev Neurosci* 23: 315–341.
105. Dorotheyev EA, Rotundo G, Tirozzi B (1995) Energy landscape of neural networks storing spatially correlated patterns. *J Phys A: Math Gen*. pp 3733–3741.
106. Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17: 4302–4311.
107. Kanwisher N, Yovel G (2006) The fusiform face area: A cortical region specialized for the perception of faces. *Phil Trans Roy Soc Lond B*. pp 2109–2128.

108. Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science* 311: 670–674.
109. Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque cerebral cortex. *Nat Neurosci* 6: 989–995.
110. Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392: 598–601.
111. Epstein R, Harris A, Stanley D, Kanwisher N (1999) The parahippocampal place area: recognition, navigation, or encoding? *Neuron* 23: 115–125.
112. Aguirre GK, Zarahn E, Esposito M (1998) An area within human ventral cortex sensitive to building stimuli: evidence and implications. *Neuron* 21: 373–383.
113. Tanaka K (1996) Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19: 109–139.
114. Brewer AA, Liu J, Wade AR, Wandell BA (2005) Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nat Neurosci* 8: 1102–1109.
115. Larsson J, Heeger DJ (2006) Two retinotopic visual areas in human lateral occipital cortex. *J Neurosci* 26: 13128–13142.
116. Wandell BA, Brewer AA, Dougherty RF (2005) Visual field map clusters in human cortex. *Philos Trans R Soc Lond B Biol Sci* 360: 693–707.
117. Hemond C, Kanwisher NG, Op de Beeck HP (2007) A preference for contralateral stimuli in human object- and face-selective cortex. *PLoS ONE* 2: e574.
118. Treves A (2003) Computational constraints that may have favoured the lamination of sensory cortex. *J Comput Neurosci* 12: 271–282.
119. Fritz JB, Elhilal M, Davida SV, Shamma SA (2007) Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in a1? *Hearing Research* 229: 186–203.
120. Bressloff PC (2005) Pattern formation in visual cortex. In: Chow C, Gutkin B, Hansel D, Meunier C, Dalibard J, eds (2005) *Methods and models in neurophysics*. Amsterdam: Elsevier.
121. Derrida B, Gardner E, Zippelius A (1987) An exactly solvable asymmetric neural network model. *Europhys Lett* 4: 167–173.
122. Evans MR (1989) Random dilution in a neural network for biased patterns. *J Phys A: Math Gen* 22: 2103–2118.
123. Treves A (1991) Dilution and sparse coding in threshold-linear nets. *J Phys A: Math Gen* 24: 327.
124. Samengo I, Treves A (2001) Representational capacity of a set of independent neurons. *Phys Rev E* 63: 11910–11924.
125. Treves A (1990) Graded-response neurons and information encoding. *Phys Rev A* 42: 2418–2430.
126. Battaglia FP, Treves A (1998) Attractor neural networks storing multiple space representations: A model for hippocampal place fields. *Phys Rev E* 58: 7738–7753.
127. Fyhn M, Hafting T, Treves A, Moser E, Moser MB (2007) Hippocampal remapping and grid realignment in entorhinal cortex. *Nature* 446: 190–194.
128. Redish AD, Battaglia FP, Chawla MK, Ekstrom AD, Gerrard JL, et al. (2001) Hippocampal pyramidal cells located near each other anatomically do not show related spatial firing correlates. *J Neurosci* 21: 1–6.