

# High-Dimensional Brain: A Tool for Encoding and Rapid Learning of Memories by Single Neurons

Ivan Tyukin<sup>1,2</sup>  · Alexander N. Gorban<sup>1</sup> · Carlos Calvo<sup>3</sup> · Julia Makarova<sup>4,5</sup> · Valeri A. Makarov<sup>3,5</sup>

Received: 16 October 2017 / Accepted: 4 March 2018  
© The Author(s) 2018

**Abstract** Codifying memories is one of the fundamental problems of modern Neuroscience. The functional mechanisms behind this phenomenon remain largely unknown. Experimental evidence suggests that some of the memory functions are performed by stratified brain structures such as the hippocampus. In this particular case, single neurons in the CA1 region receive a highly multidimensional input from the CA3 area, which is a hub for information processing. We thus assess the implication of the abundance of neuronal signalling routes converging onto single cells on the information processing. We show that single neurons can selectively detect and learn arbitrary information items, given that they operate in high dimensions. The argument is based on stochastic separation theorems and the concentration of measure phenomena. We demonstrate that a simple enough functional neuronal model is capable of explaining: (i) the extreme selectivity of single neurons to the information content, (ii) simultaneous separation of several uncorrelated stimuli or informational items from a large set, and (iii) dynamic learning of new items by associating them with already “known” ones. These results constitute a basis for organization of complex memories in ensembles of single neurons. Moreover, they show that no a priori assumptions on the

---

✉ Ivan Tyukin  
I.Tyukin@le.ac.uk

<sup>1</sup> Department of Mathematics, University of Leicester, University Road, Leicester LE1 7RH, UK  
<sup>2</sup> Saint-Petersburg State Electrotechnical University, Prof. Popova Str. 5, Saint Petersburg, Russia  
<sup>3</sup> Instituto de Matemática Interdisciplinar, Faculty of Mathematics, Universidad Complutense de Madrid, Avda Complutense s/n, 28040 Madrid, Spain  
<sup>4</sup> Department of Translational Neuroscience, Cajal Institute, CSIC, Madrid, Spain  
<sup>5</sup> Lobachevsky State University of Nizhny Novgorod, Gagarin Ave. 23, Nizhny Novgorod, Russia 603950

structural organization of neuronal ensembles are necessary for explaining basic concepts of static and dynamic memories.

**Keywords** Neural memories · Single-neuron learning · Perceptron · Stochastic separation theorems

## 1 Introduction

The human brain is arguably among the most sophisticated and enigmatic nature creations. Over millions of years it has evolved to amass billions of neurons, featuring on average  $86 \times 10^9$  cells (Herculano-Houzel 2012). This remarkable figure is several orders of magnitude higher than that of the most mammals and several times larger than in primates (Herculano-Houzel 2011). While measuring roughly 2% of the body mass, the human brain consumes about 20% of the total energy (Clark and Sokoloff 1999).

The significant metabolic cost associated with a larger brain in humans, as opposed to mere body size—a path that great apes might have evolved (Herculano-Houzel 2011), must be justified by evolutionary advantages. Some of the benefits may be related to the development of a remarkably important social life in humans. This, in particular, requires extensive abilities in the formation of complex memories. Indirectly this hypothesis is supported by the significant difference among species in the number of neurons in the cortex (Herculano-Houzel 2009) and the hippocampus (Andersen et al. 2007). For example, in the CA1 area of the hippocampus there are  $0.39 \times 10^6$  pyramidal neurons in rats,  $1.3 \times 10^6$  in monkeys, and  $14 \times 10^6$  in humans.

Evolutionary implications in relation to cognitive functions have been widely discussed in the literature (see, e.g., Platek et al. 2007; Sherwood et al. 2012; Sousa et al. 2017). Recently, it has been shown that in humans new memories can be learnt very rapidly by supposedly individual neurons from a limited number of experiences (Ison et al. 2015). Moreover, some neurons can exhibit remarkable selectivity to complex stimuli, the evidence that has led to debates around the existence of the so-called “grand mother” and “concept” cells (Quiroga et al. 2005; Viskontas et al. 2009; Quiroga 2012), and their role as elements of a declarative memory. These findings suggest that not only the brain can learn rapidly but also it can respond selectively to “rare” individual stimuli. Moreover, experimental evidence indicates that such a cognitive functionality can be delivered by single neurons (Ison et al. 2015; Quiroga et al. 2005; Viskontas et al. 2009). The fundamental questions, hence, are: How is this possible? and What could be the underlying functional mechanisms?

Recent theoretical advances achieved within the Blue Brain Project show that the brain can operate in many dimensions (Reimann et al. 2017). It is claimed that the brain has structures operating in up to eleven dimensions. Groups of neurons can form the so-called *cliques*, i.e., networks of specially interconnected neurons that generate precise representations of geometric objects. Then the dimension grows with the number of neurons in the clique. Multidimensional representation of spatiotemporal information in the brain is also implied in the concept of generalized cognitive maps (see, e.g., Villacorta-Atienza et al. 2015; Calvo et al. 2016; Villacorta-

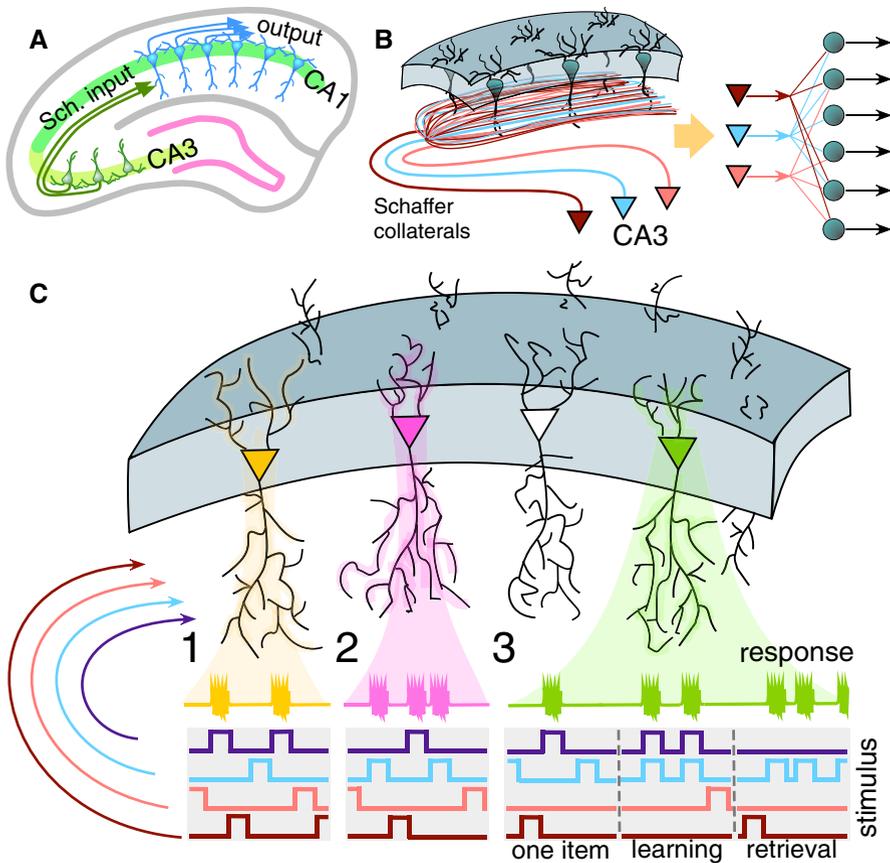
Atienza et al. 2017). Within this theory, spatiotemporal relations between objects in the environment are encoded as static (cognitive) maps and represented as elements of an  $n$ -dimensional space ( $n \gg 1$ ). The cognitive maps as information items can be learnt, classified, and retrieved on demand (Villacorta-Atienza and Makarov 2013). However, the questions concerning how the brain or individual neurons can distinguish among a huge number of different maps and select an appropriate one remain unknown.

In this work we propose that brain areas with a predominant laminar topology and abundant signalling routes simultaneously converging on individual cells (e.g., the hippocampus) are propitious for a high-dimensional processing and learning of complex information items. We show that a canonical neuronal model, the perceptron Rosenblatt (1962), in combination with a Hebbian-type of learning may provide answers to the above-mentioned fundamental questions. In particular, starting from stochastic separation theorems (Gorban and Tyukin 2017, 2018) we demonstrate that individual neurons gathering multidimensional stimuli through a sufficiently large number of synaptic inputs can exhibit extreme selectivity either to individual information items or to groups of items. Moreover, neurons are capable of associating and learning uncorrelated information items. Thus, a large number of signalling routes simultaneously converging on a large number of single cells, as it is widely observed in laminar brain structures, translates into a natural environment for rapid formation and maintenance of extensive memories. This is vital for social life and hence may constitute a significant evolutionary advantage, albeit, at the cost of high metabolic expenditure.

## 2 Fundamental Problems of Encoding Memories

Different brain structures, such as the hippocampus, have a pronounced laminar organization. For example, the CA1 region of the hippocampus is constituted by a palisade of morphologically similar pyramidal cells with their main axis in parallel and forming a monolayer (Fig. 1a). The major excitatory input to these neurons comes through Schaffer collaterals from the CA3 region (Amaral and Witter 1989; Ishizuka et al. 1990; Wittner et al. 2007), which is a hub routing information among many brain structures. Each CA3 pyramidal neuron sends an axon that bifurcates and leaves multiple collaterals in the CA1 with dominant parallel orientation (Fig. 1b). This topology allows multiple parallel axons conveying multidimensional “spatial” information from one area (CA3) simultaneously leave synaptic contacts on multiple neurons in another area (CA1). Thus, we have simultaneous convergence and divergence of the information content (Fig. 1b, right).

Experimental findings show that multiple CA1 pyramidal cells distributed in the rostro-caudal direction are activated near-synchronously by assemblies of simultaneously firing CA3 pyramidal cells (Ishizuka et al. 1990; Li et al. 1994; Benito et al. 2014). Thus, an ensemble of single neurons in the CA1 can receive simultaneously the same synaptic input (Fig. 1b, left). Since these neurons have different topology and functional connectivity (Finnerty and Jefferys 1993), their response to the same input can be different. Moreover, experimental *in vivo* results show that long-term poten-



**Fig. 1** General principles of encoding memories by single neurons in laminar structures. **a** Laminar organization of the CA3 and CA1 areas in the hippocampus facilitates multiple parallel synaptic contacts between neurons in these areas by means of Schaffer collaterals. **b** Axons from CA3 pyramidal neurons bifurcate and pass through the CA1 area in parallel (left panel) giving rise to the convergence–divergence of the information content (right panel). Multiple CA1 neurons receive multiple synaptic contacts from CA3 neurons. **c** Schematic representation of three memory encoding schemes. (1) Selectivity. A neuron (shown in yellow) receives inputs from multiple presynaptic cells that code different information items. It detects (responds to) only one stimulus (purple trace), whereas rejecting the others. (2) Clustering. Similar to 1, but now a neuron (shown in pink) detects a group of stimuli (purple and blue traces) and ignores the others. (3) Acquiring memories. A neuron (shown in green) learns dynamically a new memory item (blue trace) by associating it with a know one (purple trace) (Color figure online)

tiation can significantly increase the spike transfer rate in the CA3–CA1 pathway (Fernandez-Ruiz et al. 2012). This suggests that the efficiency of individual synaptic contacts can be increased selectively.

In this work we will follow conventional and rather general functional representation of signalling in the neuronal pathways. We assume that upon receiving an input, a neuron can either generate a response or remain silent. Forms of the neuronal responses as well as the definitions of synaptic inputs vary from one model to another. Therefore, here we adopt a rather general functional approach. Under a stimulus we understand

a number of excitations simultaneously (or within a short time window) arriving to a neuron through several axones and thus transmitting some “spatially coded” information items (Benito et al. 2016). If a neuron responds to a stimulus (e.g., generates output spikes or increases its firing rate), we then say that the neuron *detects* the informational content of the given stimulus.

We follow the standard machine learning assumptions (Vapnik and Chapelle 2000; Cucker and Smale 2002). The stimuli are generated in accordance with some distribution or a set of distributions (“Outer World Models”). All stimuli that a neuron may receive are samples from this distribution. The sampling itself may be a complicated process, and for simplicity we assume that all samples are identically and independently distributed (i.i.d.). Once a sample is generated, a stimuli sub-sample is independently selected for testing purposes. If more than one neuron is considered, we will assume that a rule (or a set of rules) is in place that determines how a neuron is selected from the set. The rules can be both deterministic and randomized. In the latter case we will specify this process.

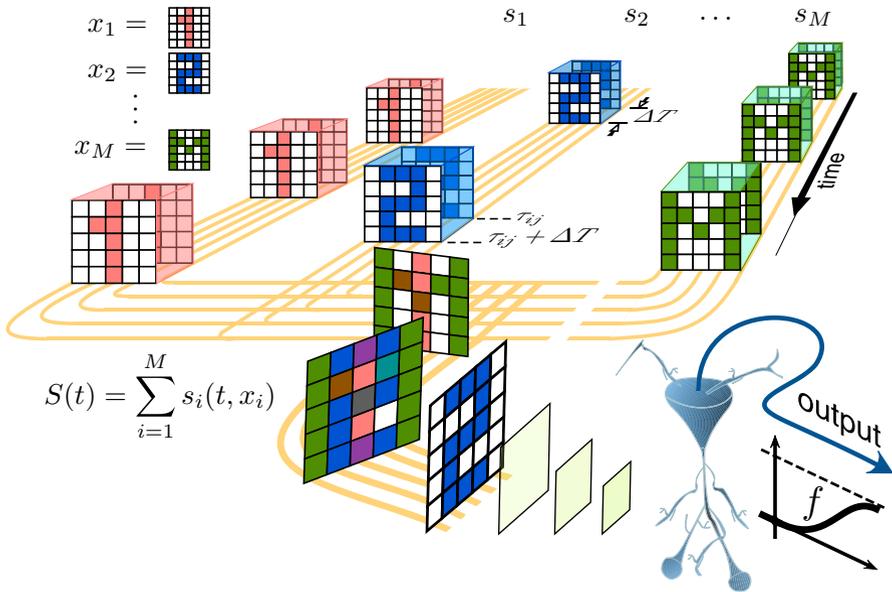
Let us now pose the following fundamental questions related to the information encoding and formation of memories by single neurons and their ensembles in laminated brain structures:

1. *Selectivity: detection of one stimulus from a set* (Fig. 1C.1) Pick an arbitrary stimulus from a reasonably large set such that a single neuron from a neuronal ensemble detects this stimulus. Then what is the probability that this neuron is stimulus-specific, i.e., it rejects all the other stimuli from the set?
2. *Clustering: detection of a group of stimuli from a set* (Fig. 1C.2) Within a set of stimuli we select a smaller subset, i.e., a group of stimuli. Then what is the probability that a neuron detecting all stimuli from this subset stays silent for all remaining stimuli in the set?
3. *Acquiring memories: learning new stimulus by associating it with one already known* (Fig. 1C.3) Let us consider two different stimuli  $s_1$  and  $s_2$  such that for  $t \leq t_0$  they do not overlap in time and a neuron detects  $s_1$ , but not  $s_2$ . In the next interval  $(t_0, t_1]$ ,  $t_1 > t_0$  the stimuli start to overlap in time (i.e., they stimulate the neuron together). For  $t > t_1$  the neuron receives only stimulus  $s_2$ . Then what is the probability that for some  $t_2 \geq t_1$  the neuron detects  $s_2$ ?

These questions are in the core of a broad range of puzzling phenomena reported in Ison et al. (2015), Quiroga et al. (2005), Viskontas et al. (2009). In what follows we will show that, remarkably, these three non-trivial fundamental questions can be answered within a simple classical modeling framework, whereby a neuron is represented by a mere perceptron equipped with a Hebbian-type of learning.

### 3 Formal Statement of the Problem

In this section we specify the information content to be processed by neurons and define a mathematical model of a generic neuron equipped with synaptic plasticity. Before going any further, let us first introduce notational agreements used throughout the text. Given two vectors  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ , their inner product  $\langle \mathbf{x}, \mathbf{y} \rangle$  is:  $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$ . If  $\mathbf{x} \in \mathbb{R}^n$  then  $\|\mathbf{x}\|$  stands for the usual Euclidean norm of  $\mathbf{x}$ :  $\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$ . By



**Fig. 2** Codification of high-dimensional information by a neuron. Each of  $M$  stimuli comprises of the “spatial” information,  $\mathbf{x}_i \in \mathbb{R}^n$ , (e.g.,  $M$  images) conducted through  $n$  axons (in yellow) and the temporal part,  $c(t - \tau_{i,j})$ , reflecting the times of stimuli presentation. A neuron (in blue) receives the stimuli and generates responses determined by some transfer function  $f$  (Color figure online)

$B_n(1) = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| \leq 1\}$  we denote a unit  $n$ -ball centered at the origin;  $\mathcal{V}(\Xi)$  is the Lebesgue volume of  $\Xi \subset \mathbb{R}^n$ , and  $|\mathcal{M}|$  is the cardinality of a finite set  $\mathcal{M}$ . Symbol  $\mathcal{C}(\mathcal{D})$ ,  $\mathcal{D} \subseteq \mathbb{R}^m$  stands for the space of continuous real-valued functions on  $\mathcal{D}$ .

### 3.1 Information Content and Classes of Stimuli

We assume that a neuron receives and processes a large but finite set of different stimuli codifying different information items:

$$\mathcal{S} = \{s_i\}. \tag{1}$$

Figure 2 illustrates schematically the information flow. Each individual stimulus  $i$  is modeled by a function  $s : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ :

$$s(t, \mathbf{x}_i) = \mathbf{x}_i \sum_j c(t - \tau_{i,j}), \tag{2}$$

where  $\mathbf{x}_i \in \mathbb{R}^n \setminus \{0\}$  is the stimulus content codifying the information to be transmitted over  $n$  individual “axons”. An example of an information item could be an  $l \times k$  image (see Fig. 2). In this case the dimension of each information item is  $n = l \times k$ .

In Eq. (2) the function  $c(\cdot)$  defines the stimulus *context*, i.e., the time window when the stimulus arrives to the neuron. For the sake of simplicity we use a rectangular window:

$$c(t) = \begin{cases} 1, & \text{if } t \in [0, \Delta T] \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where  $\Delta T > 0$  is the window length. The time instants of the stimulus presentations,  $\tau_{i,j}$ , are ordered and satisfy:

$$\tau_{i,j+1} > \tau_{i,j} + \Delta T, \quad \forall j. \tag{4}$$

Different stimuli arriving to the neuron are added linearly on the neuronal membrane. Thus, the overall neuronal input  $S$  can be written as:

$$S(t) = \sum_{i,j} \mathbf{x}_i c(t - \tau_{i,j}). \tag{5}$$

We assume that the information content of stimuli (5) and (2), i.e., vectors  $\mathbf{x}_i$  are drawn i.i.d. from some distribution. For convenience, we partition all information items into two sets:

$$\mathcal{M} = \{\mathbf{x}_1, \dots, \mathbf{x}_M\}, \quad \mathcal{Y} = \{\mathbf{x}_{M+1}, \dots, \mathbf{x}_{M+m}\}, \tag{6}$$

where  $M$  is large but finite and  $m \geq 1$  is in general smaller than  $M$ . The set  $\mathcal{M}$  contains a *background* content for a given neuron, whereas the set  $\mathcal{Y}$  models the informational content *relevant* to the task at hand. In other words, to accomplish a static memory task the neuron should be able to detect all elements from  $\mathcal{Y}$  and to reject all elements from  $\mathcal{M}$ .

The sets  $\mathcal{M}$  and  $\mathcal{Y}$  give rise to the corresponding subsets of stimuli:

$$\begin{aligned} \mathcal{S}(\mathcal{M}) &= \{s_i \in \mathcal{S} \mid s_i(\cdot) = s(\cdot, \mathbf{x}_i), \mathbf{x}_i \in \mathcal{M}\}, \\ \mathcal{S}(\mathcal{Y}) &= \{s_i \in \mathcal{S} \mid s_i(\cdot) = s(\cdot, \mathbf{x}_i), \mathbf{x}_i \in \mathcal{Y}\}. \end{aligned} \tag{7}$$

### 3.2 Neuronal Model

To stay within functional description of the information processing let us consider the most basic class of model neurons, a perceptron (Rosenblatt 1962). A single neuron receives a stimulus  $s(t, \mathbf{x})$  through  $n$  synaptic inputs (Fig. 2) and its membrane potential,  $y \in \mathbb{R}$ , is given by

$$y(\mathbf{s}, \mathbf{w}) = \langle \mathbf{w}, \mathbf{s} \rangle, \tag{8}$$

where  $\mathbf{w} \in \mathbb{R}^n$  is a vector of the synaptic weights. The neuron generates a response,  $v \in \mathbb{R}$ , according to:

$$v(\mathbf{s}, \mathbf{w}, \theta) = f(y(\mathbf{s}, \mathbf{w}) - \theta), \tag{9}$$

where  $\theta \in \mathbb{R}$  is the “firing” threshold and  $f : \mathbb{R} \rightarrow \mathbb{R}$  is the transfer function (Fig. 2):  $f \in \mathcal{C}(\mathbb{R})$ ,  $f$  is locally Lipschitz,  $f(u) = 0$  for  $u \in (-\infty, 0]$ , and  $f(u) > 0$  for  $u \in (0, \infty)$ .

Model (8), (9) captures the summation of postsynaptic potentials and the threshold nature of the neuronal activation but disregards the specific dynamics accounted for in other more advanced models. Nevertheless, as we will show in Sect. 4, this phenomenological model is already sufficient to explain the fundamental properties of information processing discussed in Sect. 2.

### 3.3 Synaptic Plasticity

In addition to the basic neuronal response mechanism (Sect. 3.2), we also model the synaptic plasticity. The description adopted here relies on the neuronal firing rate and Hebbian learning. Such a learning rule implies that the dynamics of  $\mathbf{w}$  should depend on the product of the input signal,  $\mathbf{s}$ , and the neuronal output,  $v$ . We thus arrive to a modified classical Oja rule (Oja 1982):

$$\begin{aligned} \dot{\mathbf{w}} &= \alpha v(\mathbf{s}, \mathbf{w}, \theta) y(\mathbf{s}, \mathbf{w}) (\mathbf{s} - \mathbf{w} y(\mathbf{s}, \mathbf{w})), \\ \mathbf{w}(t_0) &= \mathbf{w}_0 \in \mathbb{R}^n, \quad \mathbf{w}_0 \neq 0, \end{aligned} \quad (10)$$

where  $\alpha > 0$  defines the relaxation time. The multiplicative term  $v$  in (10) ensures that plastic changes of  $\mathbf{w}$  occur only when an input stimulus evokes a nonzero neuronal response. The fact that  $\mathbf{w}_0 \neq 0$  reflects the assumption that synaptic connections have already been established, albeit their efficacy could be subjected to plastic changes. In addition to capturing general principle of the classical Hebbian rule, model (10) guarantees that synaptic weights  $\mathbf{w}$  are bounded in forward time (see “Appendix A”) and hence conforms with physiological plausibility.

## 4 Formation of Memories in High Dimensions

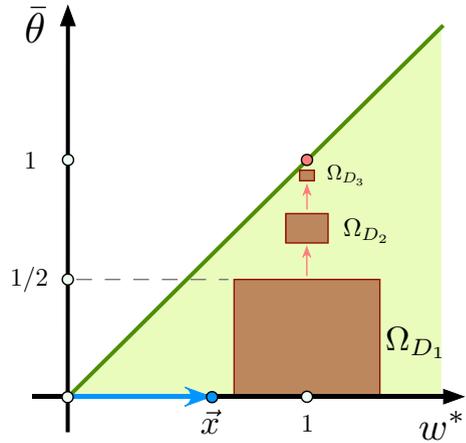
In Sect. 2 we formulated three fundamental problems of organization of memories in laminar brain structures. Let us now show how they can be treated given that pyramidal neurons operate in high dimensions.

To formalize the analysis let  $\mathcal{U}$  be a subset of the stimulus set  $\mathcal{S}$ . A neuron (8), (9) parameterized by  $(\mathbf{w}, \theta)$  partitions the set  $\mathcal{U}$  into the following subsets:

$$\begin{aligned} \text{Activated}(\mathcal{U}, (\mathbf{w}, \theta)) &= \{s_i \in \mathcal{U} \mid \exists t \geq t_0 : v(s_i(t), \mathbf{w}, \theta) > 0\}, \\ \text{Silent}(\mathcal{U}, (\mathbf{w}, \theta)) &= \{s_i \in \mathcal{U} \mid v(s_i(t), \mathbf{w}, \theta) = 0 \quad \forall t \geq t_0\}. \end{aligned} \quad (11)$$

The first set corresponds to the stimuli detected by the neuron, while the second one collects background stimuli.

**Fig. 3** Selection of neuronal parameters  $\bar{\theta} = \theta/\|\mathbf{x}_{M+1}\|$  and  $w^*$ , such that the neuron responds to the relevant information  $\mathbf{x}_{M+1}$ . Neurons corresponding to points within the green area detect the stimulus  $\mathbf{x}_{M+1}$ . Brown areas show projections of hypercylinders defined in Theorem 1 for  $D_1 = 0.3$ ,  $D_2 = 0.1$ ,  $D_3 = 0.03$  and  $\|\mathbf{x}_{M+1}\| = 0.6$  (Color figure online)



### 4.1 Extreme Selectivity of a Single Neuron to Single Stimuli

Consider the case when the set  $\mathcal{Y}$  in (6) contains only one element, i.e.,  $|\mathcal{Y}| = 1$ ,  $\mathcal{Y} = \{\mathbf{x}_{M+1}\}$ , whereas the set  $\mathcal{M}$  is allowed to be sufficiently large ( $|\mathcal{M}| = M \gg 1$ ). Let us also assume that the stimuli with different information content,  $s(\cdot, \mathbf{x}_i)$ , do not overlap in time, i.e., we present them to a neuron one by one.

For a given nonzero  $\mathbf{x}_{M+1} \in \mathcal{Y}$  and stimulus  $s(\cdot, \mathbf{x}_{M+1})$  such that it is not identically zero for  $t \geq t_0$  we can always construct a neuron which would generate a nonzero response to the stimulus  $s(\cdot, \mathbf{x}_{M+1})$  at some  $t \geq t_0$ . In other words,  $s(\cdot, \mathbf{x}_{M+1}) \in \text{Activated}(\mathcal{S}(\mathcal{Y}), (\mathbf{w}, \theta))$ . Mathematically such a neuron can be defined as follows. Let

$$\mathbf{w}^* = \frac{\mathbf{x}_{M+1}}{\|\mathbf{x}_{M+1}\|}. \tag{12}$$

Then the space from which the synaptic weights are chosen can be represented as a direct sum of the one-dimensional linear subspace  $L^\parallel(\mathbf{w}^*)$  spanned by  $\mathbf{w}^*$  and an  $(n - 1)$ -dimensional subspace  $L^\perp(\mathbf{w}^*)$  of  $\mathbb{R}^n$  that is orthogonal to  $\mathbf{w}^*$ . In this representation, if a neuron with the synaptic weight  $\mathbf{w}$  generates a nonzero response to  $s(\cdot, \mathbf{x}_{M+1})$ , then the coupling weight  $w^* = \langle \mathbf{w}, \mathbf{w}^* \rangle$  should satisfy the following condition (Fig. 3, green area):

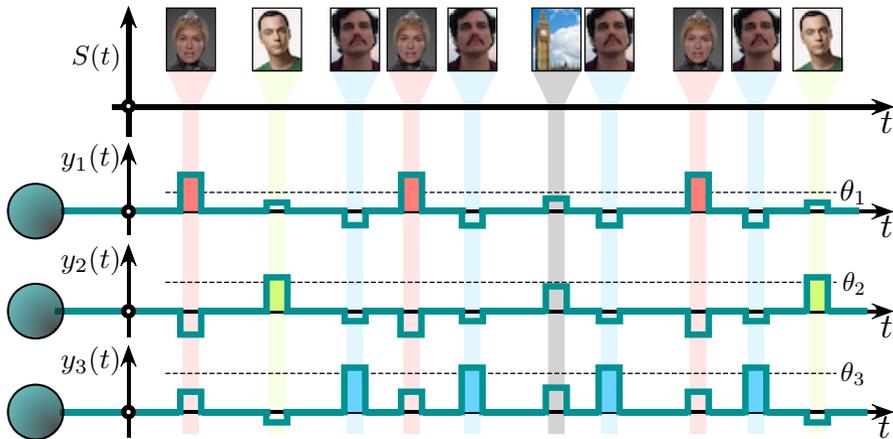
$$w^* > \frac{\theta}{\|\mathbf{x}_{M+1}\|}.$$

Indeed, such a choice is equivalent to

$$v(\mathbf{x}_{M+1}, \mathbf{w}, \theta) = f(w^*\|\mathbf{x}_{M+1}\| - \theta) > 0,$$

which in turn implies that  $v(s(t, \mathbf{x}_{M+1}), \mathbf{w}^*, \theta) > 0$  at some  $t$  and vice-versa.

Once a neuron that detects relevant information item, i.e.,  $\mathbf{x}_{M+1}$ , is specified we can proceed with assessing its selectivity properties.



**Fig. 4** Example of selective neuronal responses to stimulation with different (30 × 38)-pixels images (only first few stimulus are shown in the time line). Each neuron responds to its own (relevant) stimulus only and rejects the other (background) stimuli

**Definition 1** (*Neuronal Selectivity*) We say that a *neuron is selective to the information content*  $\mathcal{Y}$  iff it detects the relevant stimuli from the set  $S(\mathcal{Y})$  and ignores all the others from the set  $S(\mathcal{M})$ .

The notion of selectivity, as stated in Definition 1, could be relaxed to account for partial detection and rejection of information content from  $\mathcal{Y}$  and  $\mathcal{M}$ , respectively. This naturally gives rise to various levels of neuronal selectivity determined, for instance, by the proportion of elements from  $\mathcal{M}$  that correspond to stimuli that have been rejected. As we will see below, different admissible pairs  $(\mathbf{w}, \theta)$  (Fig. 3) produce different selectivity levels. The closer to the bisector, the higher the selectivity. One can pick an arbitrary firing threshold  $\theta \geq 0$  and select the synaptic efficiency at  $t = t_0$  as:

$$\mathbf{w}(t_0) = \frac{\theta + \epsilon}{\|\mathbf{x}_{M+1}\|} \mathbf{w}^* + \mathbf{w}^\perp, \quad \epsilon > 0, \quad \mathbf{w}^\perp \in L^\perp. \tag{13}$$

It can be shown (see “Appendix A”) that if the stimulus  $s(\cdot, \mathbf{x}_{M+1})$  is persistent over time and  $\mathbf{w}(t_0)$  satisfies (13) then synaptic efficiency  $\mathbf{w}(t, \mathbf{w}_0)$  converges asymptotically (as  $t \rightarrow \infty$ ) to:

$$\mathbf{w}_\infty = \begin{cases} \mathbf{w}^*, & \text{if } \theta < \|\mathbf{x}_{M+1}\| \\ \frac{\theta}{\|\mathbf{x}_{M+1}\|} \mathbf{w}^* + \mathbf{w}_\infty^\perp, & \text{if } \theta \geq \|\mathbf{x}_{M+1}\|, \end{cases} \tag{14}$$

where  $\mathbf{w}_\infty^\perp$  is an element of  $L^\perp$ .

Figure 4 shows typical responses of neurons parameterized by different pairs  $(\mathbf{w}, \theta)$  and subjected to stimulation by different information items  $\mathbf{x}_i$ . Here  $\mathbf{x}_i$  correspond to (30 × 38)-pixels color images (i.e.,  $\mathbf{x}_i \in \mathbb{R}^{3420}$ ). Firing thresholds  $\theta$  have been chosen at random, and weights  $\mathbf{w}$  have been set in accordance with (13) with the first three

images serving as the relevant information items for the three corresponding neurons. No plastic changes in  $\mathbf{w}$  were allowed. The neurons detect their own (relevant) stimuli, as expected. Moreover, they do not respond to the stimulation by other background information items (4 out of  $10^3$  images are shown in Fig. 4). Thus, the neurons indeed exhibit high stimulus selectivity.

The following theorem provides theoretical justification for these observations.

**Theorem 1** *Let elements of the sets  $\mathcal{M}$  and  $\mathcal{Y}$  be i.i.d. random vectors drawn from the equidistribution in  $B_n(1)$ . Consider the sets of stimuli  $\mathcal{S}(\mathcal{M})$  and  $\mathcal{S}(\mathcal{Y})$  specified by (7). Let  $(\mathbf{w}, \theta)$  be the neuron parameters such that*

$$s_{M+1} \in \text{Activated}(\mathcal{S}(\mathcal{Y}), (\mathbf{w}, \theta)) \text{ and } 0 < \theta < \|\mathbf{w}\|.$$

Then:

1. *The probability that the neuron is silent for all background stimuli  $s_i \in \mathcal{S}(\mathcal{M})$  is bounded from below by:*

$$\begin{aligned}
 P(s_i \in \text{Silent}(\mathcal{S}(\mathcal{M}), (\mathbf{w}, \theta)) \forall s_i \in \mathcal{S}(\mathcal{M}) \mid \mathbf{w}, \theta) &\geq \\
 &\geq \left[ 1 - \frac{1}{2} \left( 1 - \frac{\theta^2}{\|\mathbf{w}\|^2} \right)^{\frac{n}{2}} \right]^M.
 \end{aligned}
 \tag{15}$$

2. *There is a family of sets parametrized by  $D$  ( $0 < D < \min\{\frac{1}{2}, \|\mathbf{x}_{M+1}\|\}$ ):*

$$\Omega_D = \left\{ (\mathbf{w}, \theta) \mid \|\mathbf{w} - \mathbf{w}^*\| < D, D \leq \|\mathbf{x}_{M+1}\| - \theta \leq 2D \right\},
 \tag{16}$$

where  $\mathbf{w}^* = \mathbf{x}_{M+1} / \|\mathbf{x}_{M+1}\|$ , such that  $s_{M+1} \in \text{Activated}(\mathcal{S}(\mathcal{Y}), (\mathbf{w}, \theta))$ , for  $(\mathbf{w}, \theta) \in \Omega_D$  and

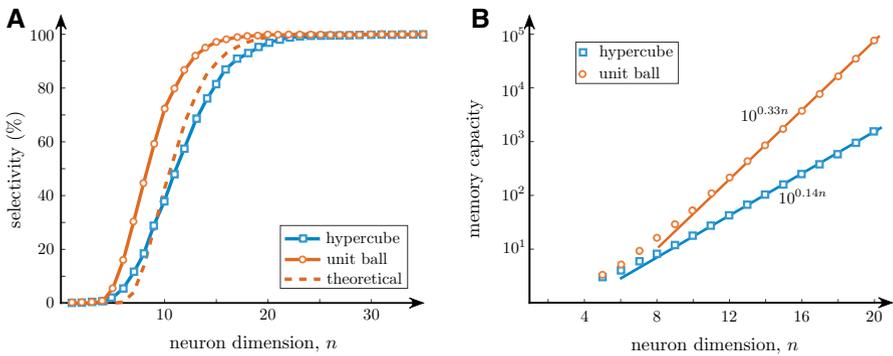
$$\begin{aligned}
 P(s_i \in \text{Silent}(\mathcal{S}(\mathcal{M}), (\mathbf{w}, \theta)) \forall s_i \in \mathcal{S}(\mathcal{M}) \mid \forall (\mathbf{w}, \theta) \in \Omega_D) &\geq \\
 &\geq \max_{\varepsilon \in (0, 1-2D)} (1 - (1 - \varepsilon)^n) \left[ 1 - \frac{1}{2} \rho(\varepsilon, D)^{\frac{n}{2}} \right]^M
 \end{aligned}
 \tag{17}$$

where

$$\rho(\varepsilon, D) = 1 - \left( \frac{1 - \varepsilon - 2D}{1 + D} \right)^2.$$

The proof is provided in ‘‘Appendix B’’.

*Remark 1* For an admissible fixed  $D > 0$ , the volume  $\mathcal{V}(\Omega_D) > 0$ . Therefore, the estimate provided by Theorem 1 is robust to small perturbations of  $(\mathbf{w}, \theta)$ , and slight fluctuations of neuronal characteristics are not expected to affect neuronal functionality.



**Fig. 5** Extreme selectivity to stimuli and memory capacity of single neurons. **a** Stimulus selectivity vs the neuron dimension. The selectivity index steeply increases for  $n \in [10, 20]$ . For  $n > 20$  practically all neurons become selective to a set of  $10^3$  random stimuli. **b** Memory capacity with reliability 0.95 of a neuronal ensemble versus the neuron dimension. For both types of stimuli the memory capacity grows exponentially (straight lines show regressions)

*Remark 2* Theorem 1 (part 2) specifies a non-iterative procedure for constructing sets of selective neurons. Such neurons detect given stimuli and reject the others, with high probability. Figure 3 (in brown) shows examples of three projections of the hypercylinders (16) ensuring robust selective stimulus detection. The smaller is the cylinder, the higher is the selectivity.

To illustrate Theorem 1 numerically we fixed the neuronal dimensionality parameter  $n$  and generated two random sets of information items comprising of  $10^3$  elements each, i.e.,  $\{\mathbf{x}_i\}_{i=1}^{10^3}$ . One set was sampled from the equidistribution in a unit ball  $B_n(1)$  centered at the origin (i.e.,  $\|\mathbf{x}_i\|_2 \leq 1$ ), and the other from the equidistribution in the hypercube  $\|\mathbf{x}_i\|_\infty \leq 1$  (a product distribution). For each set of informational items, a neuronal ensemble of  $10^3$  single neurons parameterized by  $(\mathbf{w}_i, \theta_i)$  was created. Each neuron was assigned fixed firing threshold  $\theta_i = 0.5$ ,  $i = 1, \dots, 10^3$ , whereas the synaptic efficiencies were set as  $\mathbf{w}_i = (\theta_i + \epsilon)\mathbf{x}_i / \|\mathbf{x}_i\|$ ,  $\epsilon = 0.05$ . For these neuronal ensembles and their corresponding stimuli sets we evaluated output of each neuron and assessed the neuronal selectivity (see Def. 1). The procedure was repeated 10 times. This was followed by evaluation of the frequencies of selective neurons in the pool for each  $n$ .

Figure 5a shows frequencies of selective neurons in an ensemble, for  $10^3$  stimuli taken from: i) a unit ball (red), ii) a hypercube (blue), and iii) the estimate provided by Theorem 1 (dashed). For  $n$  small ( $n < 6$ ) neurons exhibit no selectivity, i.e., they confuse different stimuli and generate nonspecific responses. As expected, when neuronal dimensionality,  $n$ , increases, the neuronal selectivity increases rapidly; and at around  $n = 20$  it approaches 100%.

### 4.2 Extreme Selectivity of a Single Neuron and Ensemble Memory Capacity

The property of a neuron to respond selectively to a single element from a large set of stimuli can be related to the notion of *memory capacity* of a neuronal ensemble comprising of a set of selective neurons.

Recall that in the framework of associative memory (Hopfield 1982), for each informational item (pattern)  $\mathbf{x}_i$  from the set  $\mathcal{M}$  there is a vicinity  $\mathcal{V}_i$  associated with  $\mathbf{x}_i$  and corresponding to all admissible perturbations of  $\mathbf{x}_i$ . Suppose that for each  $\mathbf{x}_i$  there is a neuron in the ensemble that is activated for all stimuli with informational content  $\mathbf{x}$  in  $\mathcal{V}_i$  and is silent for all other stimuli, i.e., for stimuli with  $\mathbf{x}$  in  $\cup_{j \neq i} \mathcal{V}_j$ . The maximal size of the set  $\mathcal{M}$  for which this property holds will be referred to as the (*absolute*) *memory capacity* of the ensemble (cf. Hopfield 1982; Barrett et al. 2004; Leung et al. 1995).

This conventional mechanistic definition of memory capacity, however, is too restrictive to account for variability and uncertainty that biological neuronal ensembles and systems are to deal with. Indeed, informational items themselves may bear a degree of uncertainty resulting in that  $\mathcal{V}_i \cap \mathcal{V}_j \neq \emptyset$  for some  $j, i, i \neq j$ . Furthermore, errors in memory retrievals are known to occur in classical artificial associative memory models too (see, e.g., Hopfield 1982; Amit et al. 1985; Leung et al. 1995). To be able to formally quantify such errors in relation to the number of informational items an ensemble is to store, we extend the classical notion as follows.

Suppose that for each  $\mathbf{x}_i$  there is a neuron in the ensemble that is activated for all stimuli with informational content  $\mathbf{x} \in \mathcal{V}_i$  and, with probability  $\phi$ , is silent for all stimuli with  $\mathbf{x} \in \mathcal{V}_j, j \neq i$ . The maximal size of the set  $\mathcal{M}$  for which this property holds will be referred to as the *memory capacity with reliability*  $\phi$  of the ensemble.

Assuming that  $\mathcal{V}_i$  are sufficiently small, an estimate of the memory capacity with reliability  $\phi$  of a neuronal ensemble follows from Theorem 1.

**Corollary 1** *Let elements of the sets  $\mathcal{M}$  and  $\mathcal{Y}$  be i.i.d. random vectors drawn from the equidistribution in  $B_n(1)$ . Consider the set of stimuli  $\mathcal{S}(\mathcal{M})$  as defined in (7). Then for a given fixed  $\phi \in (0, 1)$  the maximal size  $\overline{M}$  of the stimuli set  $\mathcal{S}(\mathcal{M})$  for which the following holds*

$$P(s_i \in \text{Silent}(\mathcal{S}(\mathcal{M}), (\mathbf{w}, \theta)) \forall s_i \in \mathcal{S}(\mathcal{M}) \mid \mathbf{w}, \theta) \geq \phi$$

*grows at least exponentially with the neuronal dimension  $n$ :*

$$\overline{M} > -\ln(\phi) (2e^{\alpha n} - 1), \text{ where } \alpha = \ln \left[ \frac{\|\mathbf{w}\|}{\sqrt{\|\mathbf{w}\|^2 - \theta^2}} \right] > 0. \tag{18}$$

The proof is given in ‘‘Appendix C’’.

Figure 5b illustrates how the memory capacity with reliability  $\phi$  grows with neuronal dimension  $n$ . For each neuronal dimension  $n$  we generated i.i.d. samples  $\mathcal{M}$  with  $|\mathcal{M}| = M$  from the equidistribution in  $B_n(1)$  and the  $n$ -cube  $[-1, 1]^n$ . For each sample, we defined neuronal ensembles comprising of  $M$  neurons with synaptic weights  $\mathbf{w}_i = \mathbf{x}_i / \|\mathbf{x}_i\|$  and thresholds  $\theta_i = 0.5$ , and calculated the proportion of neurons in the

ensemble that are activated by each stimulus. If the proportion was smaller than 0.05 of the total number of neurons, we incremented the value of  $M$ , generated a new sample  $\mathcal{M}$  with increased cardinality  $M$ , and repeated the experiment. The values of  $M$  corresponding to samples at which the process stopped have been recorded and retained. These constituted empirical estimates of the maximal number of stimuli for which the proportion of neurons responding to a single stimulus is at most  $0.05 = 1 - \phi$ . Figure 5b shows empirical means of such numbers for the unit ball and in the hypercube. As follows from these observations, memory capacity grows exponentially with the neuron dimension in both cases. Such a fast growth can easily cover quite exigent memory necessities.

### 4.3 Selectivity of a Single Neuron to Multiple Stimuli

To organize memories, the ability to associate different information items is essential (Fig. 1C2). To determine if such associations are feasible at the level of single neurons we assess neuronal selectivity to multiple stimuli. In particular, we consider the set  $\mathcal{Y}$  [Eq. (6)] containing  $m > 1$  random vectors:  $\mathcal{Y} = \{\mathbf{x}_{M+1}, \dots, \mathbf{x}_{M+m}\}$ . As in Sect. 4.1, here we assume that all stimuli do not overlap in time and arrive to the neuron separately. The question of interest is: Can we find a neuron [i.e., parameters  $(\mathbf{w}, \theta)$ ], such that it would generate a nonzero response to all  $\mathbf{s}_i \in \mathcal{S}(\mathcal{Y})$  and, with high enough probability, would be silent to all  $\mathbf{s}_i \in \mathcal{S}(\mathcal{M})$ ?

Below we will show that this is indeed possible, provided that the neuronal dimensionality,  $n$ , is large enough. Moreover, the separation can be achieved by a neuron with the vector of synaptic weights,  $\mathbf{w} = \mathbf{w}^*$ , closely aligned with the mean vector of the stimulus set  $\mathcal{Y}$ :

$$\bar{\mathbf{x}} = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_{M+i}, \quad \mathbf{w}^* = \frac{\bar{\mathbf{x}}}{\|\bar{\mathbf{x}}\|}. \quad (19)$$

This vector points to the center of the group to be separated from the set  $\mathcal{M}$ . In low dimensions, e.g., when  $n = 2$ , such functionality appears to be extremely unlikely. However, high-dimensional neurons can accomplish this task with probability close to one. Formal statement of this property is provided in Theorem 2.

**Theorem 2** *Let elements of the sets  $\mathcal{M}$  and  $\mathcal{Y}$  be i.i.d. random vectors drawn from the equidistribution in  $B_n(1)$ . Consider the sets of stimuli  $\mathcal{S}(\mathcal{M})$  and  $\mathcal{S}(\mathcal{Y})$  specified by (7) and let  $D, \varepsilon, \delta \in (0, 1)$  be chosen such that*

$$\theta^* = \frac{(1 - \varepsilon)^3 - \delta(m - 1)}{\sqrt{m(1 - \varepsilon)[1 - \varepsilon + \delta(m - 1)]}} \in (D, 1). \quad (20)$$

Let  $\mathbf{w}^* = \bar{\mathbf{x}}/\|\bar{\mathbf{x}}\|$  and consider the set:

$$\Omega_D = \left\{ (\mathbf{w}, \theta) \mid \|\mathbf{w} - \mathbf{w}^*\| < D, \theta \in (0, \theta^* - D) \right\}.$$

Then

$$P\left([s_i \in \text{Activated}(\mathcal{S}(\mathcal{Y}), \mathbf{w}, \theta) \forall s_i \in \mathcal{S}(\mathcal{Y})] \& [s_i \in \text{Silent}(\mathcal{S}(\mathcal{M}), \mathbf{w}, \theta) \forall s_i \in \mathcal{S}(\mathcal{M})] \mid (\mathbf{w}, \theta) \in \Omega_D\right) \geq p(\varepsilon, \delta, D, m), \tag{21}$$

where

$$p(\varepsilon, \delta, D, m) = (1 - (1 - \varepsilon)^n)^m \prod_{d=1}^{m-1} \left(1 - d \left(1 - \delta^2\right)^{\frac{n}{2}}\right) \left[1 - \frac{1}{2} \Delta^{\frac{n}{2}}\right]^M, \\ \Delta = 1 - \frac{\theta^2}{(1 + D)^2}.$$

The proof is provided in ‘‘Appendix D’’. The theorem admits the following corollary.

**Corollary 2** *Suppose that the conditions of Theorem 2 hold. Let  $\theta^* > 2D$  and consider the set:*

$$\Omega_D^* = \left\{(\mathbf{w}, \theta) \mid \|\mathbf{w} - \mathbf{w}^*\| < D, \theta \in [\theta^* - 2D, \theta^* - D]\right\}.$$

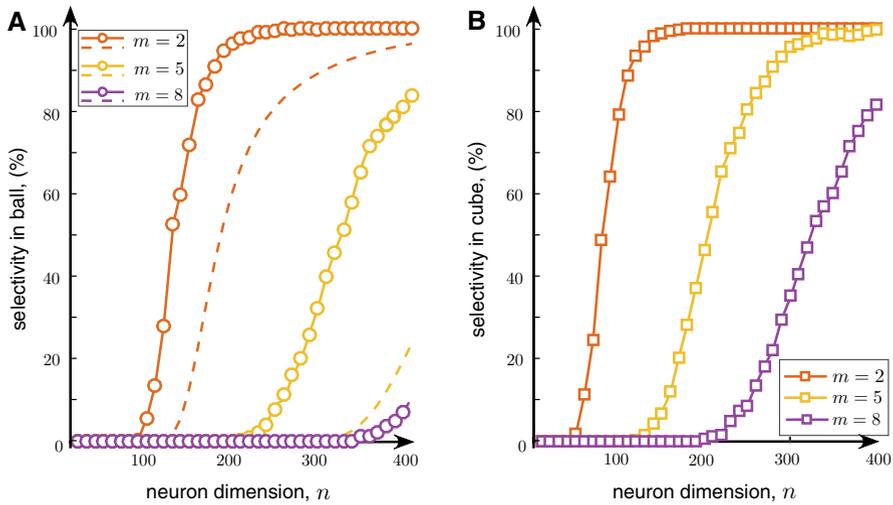
Then

$$P\left([s_i \in \text{Activated}(\mathcal{S}(\mathcal{Y}), \mathbf{w}, \theta) \forall s_i \in \mathcal{S}(\mathcal{Y})] \& [s_i \in \text{Silent}(\mathcal{S}(\mathcal{M}), \mathbf{w}, \theta) \forall s_i \in \mathcal{S}(\mathcal{M})] \mid (\mathbf{w}, \theta) \in \Omega_D^*\right) \geq (1 - (1 - \varepsilon)^n)^m \prod_{d=1}^{m-1} \left(1 - d \left(1 - \delta^2\right)^{\frac{n}{2}}\right) \left[1 - \frac{1}{2} \Delta^{\frac{n}{2}}\right]^M, \tag{22} \\ \Delta = 1 - \left(\frac{\theta^* - 2D}{1 + D}\right)^2.$$

*Remark 3* Estimates (21), (22) hold for all feasible values of  $\varepsilon$  and  $\delta$ . Maximizing the r.h.s of (21), (22) over feasible domain of  $\varepsilon, \delta$  provides lower-bound ‘‘optimistic’’ estimates of the neuron performance.

*Remark 4* The term  $\theta^*$  in Theorem 2 and Corollary 2 is an upper bound for the firing threshold  $\theta$ . The larger is the value of  $\theta$ , the higher is the neuronal selectivity to multiple stimuli. The value of  $\theta^*$ , however, decays with the number of stimuli  $m$ .

The extent to which the decay mentioned in Remark 4 affects neuronal selectivity to a group of stimuli depends largely on the neuronal dimension,  $n$ . Note also that the probability of neuronal selective response to multiple stimuli, as provided by Theorem 2, can be much larger if elements of the set  $\mathcal{Y}$  are spatially close to each other or positively correlated (Tyukin et al. 2017) (see also Lemma 4 in ‘‘Appendix F’’).



**Fig. 6** Selectivity of a single neuron to multiple stimuli. **a** Corresponds to the case when the informational content vectors,  $\mathbf{x}_i$ , are sampled from the equidistribution in the unit ball  $B_n(1)$ , and **b** corresponds to the equidistribution in the  $n$ -cube centered in the origin. In both cases the neuronal selectivity approaches 100% when the dimension  $n$  grows. In **(a)** dashed curves show the estimates provided by Theorem 2. Parameter values:  $\varepsilon = 0.01$ ,  $D = 0.001$ ,  $\delta = (1 - \varepsilon)/2(m - 1)$ ,  $\theta = \theta^* - D$

*Remark 5* Similarly to the case considered in Corollary 1, the maximal size of the stimuli set  $\mathcal{S}(\mathcal{M})$  for which selective response is ensured, with some fixed probability, grows exponentially with dimension  $n$ . Indeed, denoting  $\phi = (1 - z)^M$ , letting  $z = 1/2\Delta^{n/2}$  (with  $\Delta$  defined in Theorem 2) and invoking (34), (35) from the proof of Corollary 1, we observe that

$$\overline{M} > -\ln(\phi)(z^{-1} - 1) = -\ln(\phi)(2e^{\beta n} - 1), \quad \beta = \ln \frac{1 + D}{\sqrt{(1 + D)^2 - \theta^2}}.$$

Thus, for  $M = |\mathcal{S}(\mathcal{M})| \leq \overline{M}$ , the r.h.s. of (21) is bounded from below by

$$(1 - (1 - \varepsilon)^n)^m \prod_{d=1}^{m-1} \left(1 - d \left(1 - \delta^2\right)^{\frac{n}{2}}\right) \phi.$$

Similar estimate can be provided for the case considered in Corollary 2.

To illustrate Theorem 2 we conducted several numerical experiments. For each  $n$  we generated  $M = 10^3$  of background information items  $\mathbf{x}_i$  (the set  $\mathcal{M}$ ) and  $m = 2, 5, 8$  relevant vectors (the sets  $\mathcal{Y}$ ). In the first group of experiments all  $M + m$  i.i.d. random vectors were chosen from the equidistribution in  $B_n(1)$ . Neuronal parameters were set in accordance with Theorem 2 (i.e., Eqs. 19–21). Figure 6a illustrates the results.

Similarly to the case of neuronal selectivity to a single item (Fig. 5a), we observe a steep growth of the selectivity index with the neuronal dimension. The sharp increase

occurs, however, at significantly higher dimensions. The number of random and uncorrelated stimuli,  $m$ , to which a neuron should be able to respond selectively is fundamentally linked to the neuron dimensionality. For example, the probability that a neuron is selective to  $m = 5$  random stimuli becomes sufficiently high only at  $n > 400$ . This contrasts sharply with  $n = 120$  for  $m = 2$ .

Our numerical experiments also show that the firing threshold specified in Theorem 2 for arbitrarily chosen fixed values of  $\delta$  and  $\varepsilon$  is not optimal in the sense of providing the best possible probability estimates. Playing with  $\theta$  one can observe that the values of  $n$  at which neuronal selectivity to multiple stimuli starts to emerge are in fact significantly lower than those predicted by Eq. (22). This is not surprising. First, since estimate (22) holds for all admissible values of  $\delta$  and  $\varepsilon$ , it should also hold for the maximizer of  $p(\varepsilon, \delta, D, m)$ . Second, the estimate is conservative in the sense that it is based on conservative estimates of the volume of spherical cups  $\mathcal{C}_n$  (see, e.g., proof of Theorem 1). Deriving more accurate numerical expressions for the latter is possible, although at the expense of simplicity.

To demonstrate that dependence of the selectivity index on the firing threshold is likely to hold qualitatively for broader classes of distributions from which the sets  $\mathcal{M}$  and  $\mathcal{Y}$  are drawn, we repeated the simulation for the equidistribution in an  $n$ -cube centered at the origin. In this case, Theorem 2 does not formally apply. Yet, an equivalent statement can still be produced (cf. Gorban and Tyukin 2017). In these experiments synaptic weights were set to  $\mathbf{w} = \bar{\mathbf{x}}/\|\bar{\mathbf{x}}\|$  and  $\theta = 0.5\|\bar{\mathbf{x}}\|$ . The results are shown in Fig. 6b. The neuron’s performance in the cube is markedly better than that of in  $B_n(1)$ . Interestingly, this is somewhat contrary to expectations that might have been induced by our earlier experiments (shown in Fig. 5) in which neuronal selectivity to a single stimulus was more pronounced for  $B_n(1)$ .

Overall, these results suggest that single neurons can indeed separate random uncorrelated information items from a large set of background items with probability close to one. This gives rise to a possibility for a neuron to respond selectively to various arbitrary uncorrelated information items simultaneously. The latter property provides a natural mechanism for accurate and precise grouping of stimuli in single neurons.

#### 4.4 Dynamic Memory: Learning New Information Items by Association

In the previous sections we dealt with a static model of neuronal functions, i.e., when the synaptic efficiency  $\mathbf{w}$  either did not change at all or the changes were negligibly small over large intervals of stimuli presentation. In the presence of synaptic plasticity (10), the latter case corresponds to  $0 \leq \alpha \ll 1$  in (10). In this section we explicitly account for the time evolution of the synaptic efficiency,  $\mathbf{w}(t, \mathbf{w}_0)$  [Eq. (10)]. As we will see below, this may give rise to dynamic memories in single neurons.

As before, we will deal with two sets of stimuli, the relevant one,  $\mathcal{S}(\mathcal{Y})$ , and the background one,  $\mathcal{S}(\mathcal{M})$ . We will consider two time epochs: (i) Learning phase and (ii) Retrieval phase. Within the learning phase we assume that all stimuli from the set  $\mathcal{S}(\mathcal{Y})$  arrive to a neuron completely synchronized, i.e.,:

$$\tau_{M+1,j} = \tau_{M+2,j} = \dots = \tau_{M+m,j}, \quad \forall j. \tag{23}$$

Such a synchronization could be interpreted as a mechanism for associating or grouping different uncorrelated information items for the purposes of memorizing them at a later stage.

The dynamics of the synaptic weights for  $t \geq t_0$  is given Eq. (10) with the input signal  $s$  replaced with:

$$\bar{s}(t) = \sum_{i=1}^m s_{M+i}(t). \tag{24}$$

Let  $\mathbf{w}_0 = \mathbf{w}(t_0)$  and  $\theta$  satisfy the following condition:

$$\begin{aligned} \exists s_k \in \mathcal{S}(\mathcal{Y}) \text{ such that } s_k \in \text{Activated}(\mathcal{S}(\mathcal{Y}), \mathbf{w}_0, \theta) \\ s_i \in \text{Silent}(\mathcal{S}, \mathbf{w}_0, \theta) \text{ for all } s_i \in \mathcal{S} \setminus \{s_k\}. \end{aligned} \tag{25}$$

Thus, at  $t = t_0$  only one information item is “known” to the neuron. All other relevant items from the set  $\mathcal{Y}$  are “new” in the sense that the neuron rejects them at  $t = t_0$ . Theorem 1 specifies the sets of neuronal parameters  $\mathbf{w}_0, \theta$  for which condition (25) holds with probability close to one if  $n$  is large enough.

The question is: What is the probability that, during the learning phase the synaptic weights  $\mathbf{w}(t, \mathbf{w}_0)$  evolve in time so that the neuron becomes responsive to all  $s_i \in \mathcal{S}(\mathcal{Y})$  while remaining silent to all  $s_i \in \mathcal{S}(\mathcal{M})$  (Fig. 1C.3)? In other words, the neuron learns new items and recognizes them in the retrieval phase. The following theorem provides an answer to this question.

**Theorem 3** *Let elements of the sets  $\mathcal{M}$  and  $\mathcal{Y}$  be i.i.d. random vectors drawn from the equidistribution in  $B_n(1)$ . Consider the sets of stimuli  $\mathcal{S}(\mathcal{M})$  and  $\mathcal{S}(\mathcal{Y})$  specified by (7). Let (23) hold, the dynamics of neuronal synaptic weights satisfy (10), (24), and  $(\mathbf{w}_0, \theta)$  be chosen such that condition (25) is satisfied. Pick  $\varepsilon, \delta \in (0, 1)$  such that*

$$(1 - \varepsilon)^3 > \delta(m - 1).$$

Moreover, suppose that

1. There exist  $L, \kappa > 0$  such that

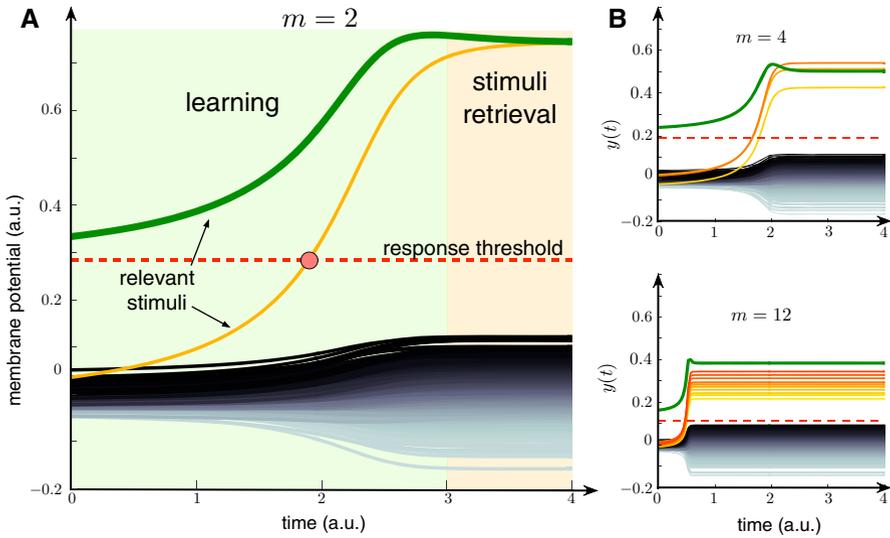
$$\int_t^{t+L} v(\bar{s}(\tau), \mathbf{w}(\tau, \mathbf{w}_0), \theta) \langle \bar{s}(\tau), \mathbf{w}(\tau, \mathbf{w}_0) \rangle^2 d\tau > \kappa, \quad \forall t \geq t_0.$$

2. The firing threshold,  $\theta$ , satisfies

$$0 < \theta < \frac{(1 - \varepsilon)^3 - \delta(m - 1)}{\sqrt{m(1 - \varepsilon)[(1 - \varepsilon) + \delta(m - 1)]}} = \theta^*.$$

Then for, any  $0 < D \leq \theta^* - \theta$ , there is  $t_1(D) > t_0$  such that

$$\begin{aligned} P([\mathcal{S}(\mathcal{Y}) \in \text{Activated}(\mathcal{S}, \mathbf{w}(t, \mathbf{w}_0), \theta)] \ \& \ [\mathcal{S}(\mathcal{M}) \in \text{Silent}(\mathcal{S}, \mathbf{w}(t, \mathbf{w}_0), \theta)]) \geq \\ (1 - (1 - \varepsilon)^n)^m \prod_{d=1}^{m-1} \left( 1 - d \left( 1 - \delta^2 \right)^{\frac{n}{2}} \right) \left[ 1 - \frac{1}{2} \left( 1 - \frac{\theta^2}{(1 + D)^2} \right)^{\frac{n}{2}} \right]^M \end{aligned}$$



**Fig. 7** Dynamic memory: Learning new information items by association. **a** Example of the dynamic association of a known stimulus (neuron’s response to the known stimulus is shown by green curve) and a new one (neuron’s response shown by orange curve). Two relevant stimuli out of 502 are learnt by the neuron. At  $t \approx 2$  (red circle) the orange curve crosses the threshold (red dashed line) and stays above it for  $t > 2$ . Thus the neuron detects the corresponding stimulus for  $t > 2$ . **b** Same as in A but for  $m = 4$  and  $m = 12$ . Parameter values:  $\varepsilon = 0.01$ ,  $D = 0.001$ ,  $\delta = (1 - \varepsilon)^3/2(m - 1)$ ,  $\alpha = 1$ ,  $M = 500$ ,  $\theta = \theta^* - D$ ,  $n = 400$  (Color figure online)

for all  $t \geq t_1(D)$ .

The proof is provided in “Appendix E”.

Figure 7 illustrates the theorem numerically. First we assumed that the relevant set  $\mathcal{Y}$  consists of  $m = 2$  items. One of them is considered as “known” to the neuron (Fig. 7a, green). Its informational content,  $\mathbf{x}_{M+1}$ , satisfies the condition  $\langle \mathbf{w}_0, \mathbf{x}_{M+1} \rangle > \theta$ , i.e., this stimulus evokes membrane potential above the threshold at  $t = t_0$ . Consequently, the neuron detects this stimulus selectively as described in Sect. 4.1. For the second relevant stimulus (Fig. 7a, orange), however, we have  $\langle \mathbf{w}_0, \mathbf{x}_{M+2} \rangle < \theta$ . Therefore, the neuron cannot detect such a stimulus alone. The background stimuli from the set  $\mathcal{S}(\mathcal{M})$  are also sub-threshold (Fig. 7a, back curves).

During the learning phase, the neuron receives  $M = 500$  background and  $m = 2$  relevant stimuli. The relevant stimuli from the set  $\mathcal{S}(\mathcal{Y})$  appear simultaneously, i.e., they are temporarily associated. The synaptic efficiency changes during the learning phase by action of the relevant stimuli. Therefore, the membrane potential,  $y(t) = \langle \mathbf{w}(t), \mathbf{w}_0 \rangle, \bar{s}(t)$ , progressively increases when the relevant stimuli arrive (Fig. 7a, green area). These neuronal adjustments give rise to a new functionality.

At some time instant (marked by red circle in Fig. 7a) the neuron becomes responsive to the new relevant stimulus (Fig. 7a, orange), which is synchronized with the “known” one. Note that all other background stimuli that show no temporal associativity remain below the threshold (Fig. 7a, black traces). Thus, after a transient period, the neuron

learns new stimulus. Once the learning is over, the neuron detects selectively either of the two relevant stimuli.

The procedure just described can be used to associate together more than two relevant stimuli. Figure 7b shows examples for  $m = 4$  and  $m = 12$ . In both cases the neuron was able to learn all relevant stimuli, while rejecting all background ones. We observed, however, that increasing the number of uncorrelated information items to be learnt, i.e., the value of  $m$ , reduces the gap between firing thresholds and the membrane potentials evoked by background stimuli. In other words, the neuron does detect the assigned group of new stimuli, but with lower accuracy. This behavior is consistent with the theoretical bound on  $\theta$  prescribed in the statement of Theorem 3.

## 5 Discussion

Theorems 1–3 and our numerical simulations demonstrate that the extreme neuronal selectivity to single and multiple stimuli, and the capability to learn uncorrelated stimuli observed in a range of empirical studies Quiroga et al. (2005), Viskontas et al. (2009), Ison et al. (2015) can be explained by simple functional mechanisms implemented in single neurons. The following basic phenomenological properties have been used to arrive to this conclusion: (i) the dimensionality  $n$  of the information content and neurons is sufficiently large, (ii) a perceptron neuronal model, Eq. (9), is an adequate representation of the neuronal response to stimuli, and (iii) plasticity of the synaptic efficiency is governed by Hebbian rule (10). A crucial consequence of our study is that no a priori assumptions on the structural organization of neuronal ensembles are necessary for explaining basic concepts of static and dynamic memories.

Our approach does not take into account more advanced neuronal behaviors reproduced by, e.g., models of spike-timing-dependent plasticity (Markram et al. 1997) and firing threshold adaptation (Fontaine et al. 2014). Nevertheless, our model captures essential properties of neuronal dynamics and as such is generic enough for the purpose of functional description of memories.

Firing threshold adaption, as reported in Fontaine et al. (2014), steers firing activity of a stimulated neuron to a homeostatic state. In this state, the value of the threshold is just large/small enough to maintain reasonable firing rate without over/under-excitation. In our model, such a mechanism could be achieved by setting the value of  $\theta$  sufficiently close to the highest feasible values specified in Theorems 1 and 2.

In addition to rather general model of neuronal behavior, another major theoretical assumption of our work was the presumption that stimuli informational content is drawn from an equidistribution in a unit ball  $B_n(1)$ . This assumption, however, can be relaxed, and results of Theorems 1–3 generalized to product measures. Key ingredients of such generalizations are provided in Gorban and Tyukin (2017), and their practical feasibility is illustrated by numerical simulations with information items randomly drawn from a hypercube (Figs. 5, 6, 7).

Our theoretical and numerical analysis revealed an interesting hierarchy of cognitive functionality implementable at the level of single neurons. We have shown that cognitive functionality develops with the dimensionality or connectivity parameter  $n$  of single neurons. This reveals explicit relationships between levels of the neural

connectivity in living organisms and different cognitive behaviors such organisms can exhibit (cf. Lobov et al. 2017). As we can see from Theorems 1, 2 and Figs. 5 and 6, the ability to form static memories increases monotonically with  $n$ . The increase in cognitive functionality, however, occurs in steps.

For  $n$  small ( $n \in [1, 10]$ ), neuronal selectivity to a single stimulus does not form. It emerges rapidly when the dimension parameter  $n$  exceeds some critical value, around  $n = 10 \div 20$  (see Fig. 5a). This constitutes the first critical transition. Single neurons become selective to single information items. The second critical transition occurs at significantly larger dimensions, around  $n = 100\text{--}400$  (see Fig. 6). At this second stage the neuronal selectivity to multiple *uncorrelated* stimuli develops. The ability to respond selectively to a given set of multiple uncorrelated information items is apparently crucial for rapid learning “by temporal association” in such neuronal systems. This learning ability as well as formation of dynamic memories are justified by Theorem 3 and illustrated in Fig. 7.

In the core of our mathematical arguments are the concentration of measure phenomena exemplified in Gorban et al. (2016), Gorban and Tyukin (2018) and stochastic separation theorems (Gorban and Tyukin 2017; Gorban et al. 2016). Some of these results, which have been central in the proofs of Theorem 2 and 3, namely, the statements that random i.i.d. vectors from equidistributions in  $B_n(1)$  and product measures are almost orthogonal with probability close to one, are tightly related to the notion of effective dimensionality of spaces based on  $\epsilon$ -*quasiorthogonality* introduced in Hecht-Nielsen (1994), Kainen and Kurkova (1993). In these works the authors demonstrated that in high dimensions there exist exponentially large sets of quasiorthogonal vectors. Gorban et al. (2016), however, as well as in our current work (see Lemma 3) we demonstrated that not only such sets exist, but also that they are typical.

Finally, we note that the number of multiple stimuli that can be selectively detected by single neurons is not extraordinarily large. In fact, as we have shown in Figs. 6 and 7, memorizing 8 information items at the level of single neurons requires more than 400 connections. This suggests that not only new memories are naturally packed *in quanta*, but also that there is a limit on this number that is associated with the cost of implementation of such a functionality. This cost is the number of individual functional synapses. Balancing the costs in living beings is of course a subject of selection and evolution. Nevertheless, as our study has shown, there is a clear functional gain that these costs may be paid for.

## 6 Conclusion

In this work we analyzed the striking consequences of the abundance of signalling routes for functionality of neural systems. We demonstrated that complex cognitive functionality derived from extreme selectivity to external stimuli and rapid learning of new memories at the level of single neurons can be explained by the presence of multiple signalling routes and simple physiological mechanisms. At the basic level, these mechanisms can be reduced to a mere perceptron-like behavior of neurons in response to stimulation and a Hebbian-type learning governing changes of the synaptic efficiency.

The observed phenomenon is robust. Remarkably, a simple generic model offers a clear-cut mathematical explanation of a wealth of empirical evidence related to *in vivo* recordings of “Grandmother” cells, “concept” cells, and rapid learning at the level of individual neurons (Quiroga et al. 2005; Viskontas et al. 2009; Ison et al. 2015). The results can also shed light on the question why Hebbian learning may give rise to neuronal selectivity in prefrontal cortex (Lindsay et al. 2017) and explain why adding single neurons to deep layers of artificial neural networks is an efficient way to acquire novel information while preserving previously trained data representations (Draelos et al. 2016).

Finding simple laws explaining complex behaviors has always been the driver of progress in Mathematical Biology and Neuroscience. Numerous examples of such simple laws can be found in the literature (see, e.g., Roberts et al. 2014; Jurica et al. 2013; Gorban et al. 2016; Perlovsky 2006). Our results not only provide a simple explanation of the reported empirical evidence but also suggest that such a behavior might be inherent to neuronal systems and hence organisms that operate with high-dimensional informational content. In such systems, complex cognitive functionality at the level of elementary units, i.e., single neurons, occurs naturally. The higher the dimensionality, the stronger the effect. In particular, we have shown that the memory capacity in ensembles of single neurons grows exponentially with the neuronal dimension. Therefore, from the evolutionary point of view, accommodating large number of signalling routes converging onto single neurons is advantageous despite the increased metabolic costs.

The considered class of neuronal models, being generic, is of course a simplification. It does not capture spontaneous firing, signal propagation in dendritic trees, and many other physiologically relevant features of real neurons. Moreover, in our theoretical assessments we assumed that the informational content processed by neurons is sampled from an equidistribution in a unit ball. The results, however, can already be generalized to product measure distributions (see, e.g., Gorban and Tyukin 2017). Generalizing the findings to models offering better physiological realism is the focus of our future works.

**Acknowledgements** This work has been supported by Innovate UK Grants KTP009890 and KTP010522, by the Spanish Ministry of Economy and Competitiveness under Grant FIS2014-57090-P, the Russian Federation Ministry of Education state assignment (No. 8.2080.2017/4.6), “Initiative scientific project” of the main part of the state plan of the Ministry of Education and Science of Russian Federation (Task No. 2.6553.2017/BCH Basic Part), and by the Russian Science Foundation Project 15-12-10018 (numerical assessment and results). Alexander N. Gorban was supported by the Ministry of Education and Science of Russian Federation (Project No. 14.Y26.31.0022).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### A Dynamics of Coupling Weights

The following results demonstrate that the neuronal model provided in Sect. 3 is well-posed.

**Lemma 1** Consider (9), (10) with the function  $s(\cdot, \mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$  defined as in (2). Then (1) solutions  $\mathbf{w}(\cdot, \mathbf{w}_0)$  of (10) are defined for all  $t \geq t_0$  and are unique and bounded in forward time.

If, in addition,  $\theta \geq 0$  and there exist numbers  $L, \delta > 0$  such that:

$$\int_t^{t+L} v(s(\tau, \mathbf{x}), \mathbf{w}(\tau, \mathbf{w}_0), \theta) (s(\tau, \mathbf{x}), \mathbf{w}(\tau, \mathbf{w}_0))^2 d\tau > \delta, \quad \forall t \geq t_0, \tag{26}$$

then

(2)  $\mathbf{x} / \|\mathbf{x}\|$  is an attractor, that is:

$$\lim_{t \rightarrow \infty} \mathbf{w}(t, \mathbf{w}_0) = \frac{\mathbf{x}}{\|\mathbf{x}\|}. \tag{27}$$

*Proof of Lemma 1* (1) The right-hand side of (10) is continuous in  $\mathbf{w}$  and piece-wise continuous in  $t$  with finite number of discontinuities of the first kind in any finite interval containing  $t_0$ , independently on the values of  $\mathbf{w}$ . Hence, in accordance with Peano Theorem, solutions of (10) are defined on some non-empty interval containing  $t_0$ . Let  $\mathcal{T}$  be the maximal interval of this solution’s definition (to the right of  $t_0$ ). Since the right-hand side of (10) is locally Lipschitz in  $\mathbf{w}$ , the solution  $\mathbf{w}(\cdot, \mathbf{w}_0)$  is uniquely defined on  $\mathcal{T}$ .

To show that  $\mathcal{T} = [t_0, \infty)$  consider

$$V(\mathbf{w}) = 1 - \|\mathbf{w}\|^2.$$

In the interval  $\mathcal{T}$  we have:

$$\dot{V} = -2\alpha v y^2 V.$$

Given that  $v y^2 \geq 0$ , the above expression implies that

$$|1 - \|\mathbf{w}_0\|^2| \geq |1 - \|\mathbf{w}(t, \mathbf{w}_0)\|^2| \geq \|\mathbf{w}(t, \mathbf{w}_0)\|^2 - 1.$$

Consequently,

$$\|\mathbf{w}(t, \mathbf{w}_0)\| \leq \left(1 + |1 - \|\mathbf{w}_0\|^2|\right)^{\frac{1}{2}} \tag{28}$$

for all  $t \geq t_0, t \in \mathcal{T}$ . Let  $t_1$  be an arbitrary point in the interval  $\mathcal{T}$ . Recall that the right-hand side of (10) is continuous and locally Lipschitz with respect to  $\mathbf{w}$  (uniformly in  $t$ ). Thus (28) implies existence of some  $\Delta(\mathbf{w}_0, \mathbf{x}) > 0$ , independent on  $t_1$ , such that the solution  $\mathbf{w}(\cdot, \mathbf{w}_0)$  is defined on the interval  $[t_0, t_1 + \Delta(\mathbf{w}_0, \mathbf{x})]$ .

Given that  $t_1$  was chosen arbitrarily in  $\mathcal{T}$ , we can conclude that  $\mathcal{T} = [t_0, \infty)$  (cf. Theorem 3.3 Khalil 2002).

(2) For the sake of convenience, we denote

$$p(t) = v(s(t, \mathbf{x}), \mathbf{w}(t, \mathbf{w}_0), \theta) \langle s(t, \mathbf{x}), \mathbf{w}(t, \mathbf{w}_0) \rangle^2.$$

Condition (26) assures that both  $\mathbf{x} \neq 0, \mathbf{w}_0 \neq 0$ . Moreover, since  $V(\mathbf{w}(t, \mathbf{w}_0))$  is defined for all  $t \geq t_0$ , we can conclude that

$$|V(t)| = \left| V_0 e^{-2\alpha \int_{t_0}^t p(\tau) d\tau} \right| \leq |V_0| e^{-2\alpha \delta \left[ \frac{t-t_0}{L} \right]}.$$

Hence

$$\lim_{t \rightarrow \infty} \|\mathbf{w}(t, \mathbf{w}_0)\| = 1. \tag{29}$$

Consider:

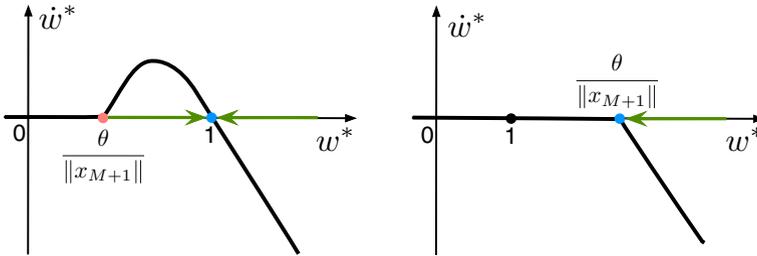
$$\begin{aligned} \mathbf{w}(t, \mathbf{w}_0) &= e^{-\alpha \int_{t_0}^t p(\tau) d\tau} \mathbf{w}_0 \\ &+ \alpha \left[ \int_{t_0}^t e^{-\alpha \int_{\tau}^t p(s) ds} v(s(\tau, \mathbf{x}), \mathbf{w}(\tau, \mathbf{w}_0), \theta) \langle s(\tau, \mathbf{x}), \mathbf{w}(\tau, \mathbf{w}_0) \rangle \right. \\ &\left. \sum_j c(\tau - \tau_j) d\tau \right] \mathbf{x}. \end{aligned}$$

Observe that the first term decays exponentially to 0, whereas the second term is proportional to  $\mathbf{x}$ . Moreover, since  $\theta \geq 0$ , the term  $v(s(\tau, \mathbf{x}), \mathbf{w}(\tau, \mathbf{w}_0), \theta) \langle s(\tau, \mathbf{x}), \mathbf{w}(\tau, \mathbf{w}_0) \rangle \geq 0$  for all  $\tau \geq t_0$ . Hence the coefficient in front of  $\mathbf{x}$  is non-negative. This, combined with (29), implies that (27) holds.  $\square$

Note that Lemma 1 applies to stimuli classes that are broader than the one defined by (2), (3). The results hold, e.g., for the functions  $c(\cdot)$  in (2) that are non-negative, piece-wise continuous, and bounded. On the other hand, to determine convergence and asymptotic properties of  $\mathbf{w}(\cdot, \mathbf{w}_0)$  for  $t \geq t_0$  (part 2 of the lemma) one needs to check that condition (26) holds. A drawback of this condition is that it requires availability of signals  $v(s(t, \mathbf{x}), \mathbf{w}(t, \mathbf{w}_0), \theta), \langle s(t, \mathbf{x}), \mathbf{w}(t, \mathbf{w}_0) \rangle$  for all  $t \geq t_0$ .

For  $c(\cdot)$  specified by (2) this latter condition can be drastically simplified. To see this, let us get a somewhat deeper geometrical insight into the dynamics of  $\mathbf{w}$  governed by (10). In order to bring the discussion in line with the question of neuronal selectivity, consider the stimuli sets (6), (7) with  $\mathcal{Y} = \{\mathbf{x}_{M+1}\}$ , and suppose that stimuli  $s(\cdot, \mathbf{x}_i), i = 1, \dots, M$  do not evoke any neuronal responses, i.e.,  $v(s(\cdot, \mathbf{x}_i), \mathbf{w}(\cdot, \mathbf{w}_0), \theta) = 0$  for all  $i = 1, \dots, M$ . Hence no changes in  $\mathbf{w}$  occur if the stimulus  $s$  in (10) is any of  $s(\cdot, \mathbf{x}_i), i = 1, \dots, M$ .

Consider system (10) with  $s(\cdot, \mathbf{x}_{M+1})$ . The variable  $\mathbf{w}$  may change only over those intervals of  $t$  when  $s(\cdot, \mathbf{x}_{M+1}) \neq 0$ . Between these intervals  $\mathbf{w}(t, \mathbf{w}_0)$  is constant. Let the stimulus be persistent in the sense that for any  $t' \geq t_0$  there is a  $t''$  such that  $s(t'', \mathbf{x}_{M+1}) \neq 0$ . Thus, without loss of generality and for the purposes of assessing



**Fig. 8** Sketch of the dynamics of  $w^*$ . Thick black curve shows the r.h.s. of (30) as a function of  $w^*$  for two cases:  $\theta < \|x_{M+1}\|$  (left) and  $\theta > \|x_{M+1}\|$  (right). Blue (red) dots correspond to stable (unstable) equilibria. Green arrows mark trajectories. In the first case (left)  $w^*$  tends to 1, whereas in the second (right) it goes asymptotically to  $\theta/\|x_{M+1}\|$  (Color figure online)

asymptotic behavior of  $w(t, w_0)$  at  $t \rightarrow \infty$  variable  $s(t, x_{M+1})$  in (8)–(10) may be replaced with  $x_{M+1}$ .

Recall that  $w(t, w_0)$  can be represented as a sum

$$w(t, w_0) = w^*(t, w_0)w^* + w^\perp(t, w_0), \quad w^*(t, w_0) = \langle w(t, w_0), w^* \rangle,$$

where  $w^*$  is defined in (12) and  $w^\perp \in L^\perp$ . In this representation,

$$\begin{aligned} \dot{w} &= \dot{w}^* w^* + \dot{w}^\perp = \alpha f(\langle x_{M+1}, w^* w^* + w^\perp \rangle - \theta) \langle x_{M+1}, w^* w^* + w^\perp \rangle (x_{M+1} - \langle x_{M+1}, w^* w^* + w^\perp \rangle [w^* w^* + w^\perp]) \\ &= \left[ \alpha f(w^* \|x_{M+1}\| - \theta) \|x_{M+1}\|^2 (1 - w^{*2}) \right] w^* w^* - \left[ \alpha f(w^* \|x_{M+1}\| - \theta) \|x_{M+1}\|^2 w^{*2} \right] w^\perp \end{aligned}$$

or, equivalently,

$$\dot{w}^* = \alpha \|x_{M+1}\|^2 f(w^* \|x_{M+1}\| - \theta) (1 - w^{*2}) w^* \tag{30}$$

$$\dot{w}^\perp = - \left[ \alpha f(w^* \|x_{M+1}\| - \theta) \|x_{M+1}\|^2 w^{*2} \right] w^\perp. \tag{31}$$

Obviously,  $L^\parallel$ ,  $L^\perp$ , and the set

$$\mathcal{W}(x_{M+1}, \theta) = \{(w^*, w^\perp), w^* \in \mathbb{R}, w^\perp \in L^\perp \mid w^* \|x_{M+1}\| - \theta \leq 0\}$$

are invariant with respect to (10). Let  $x_{M+1} \neq 0$ ,  $\theta \geq 0$ , and  $w_0 \notin \mathcal{W}(x_{M+1}, \theta)$ . Then two non-trivial alternatives (Fig. 8) are possible:

**A:** If  $\theta < \|x_{M+1}\|$  then  $w^*(t, w_0) \rightarrow 1$  and, according to (31),  $w^\perp(t, w_0) \rightarrow 0$  as  $t \rightarrow \infty$ . Thus,

$$\lim_{t \rightarrow \infty} w(t) = \frac{x_{M+1}}{\|x_{M+1}\|} = w^*.$$

**B:** If  $\theta \geq \|\mathbf{x}_{M+1}\|$  then  $w^*(t, \mathbf{w}_0) \rightarrow \theta/\|\mathbf{x}_{M+1}\|$  as  $t \rightarrow \infty$ . There is no guarantee, however, that  $\mathbf{w}^\perp(t, \mathbf{w}_0)$  converges to the origin asymptotically. Thus, there is a  $\mathbf{w}_\infty^\perp \in L^\perp$ :

$$\lim_{t \rightarrow \infty} \mathbf{w}(t) = \frac{\theta}{\|\mathbf{x}_{M+1}\|} \mathbf{w}^* + \mathbf{w}_\infty^\perp.$$

The above result can now be formalized as

**Lemma 2** Consider (9), (10) with the function  $s(\cdot, \mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^n$  defined as in (2). Let  $\theta \geq 0$  and  $\langle \mathbf{w}_0, \mathbf{x} \rangle > \theta$ . Furthermore, let the stimulus  $s(\cdot, \mathbf{x})$  be persistent in the sense that for any  $t' \geq t_0$  there is a  $t'' > t'$  such that  $s(t'', \mathbf{x}) \neq 0$ . Then the following alternatives hold:

- 1) If  $\theta < \|\mathbf{x}\|$  then  $\lim_{t \rightarrow \infty} \mathbf{w}(t, \mathbf{w}_0) = \mathbf{x}/\|\mathbf{x}\|$ .
- 2) If  $\theta \geq \|\mathbf{x}\|$  then  $\lim_{t \rightarrow \infty} \langle \mathbf{w}(t, \mathbf{w}_0), \mathbf{x}/\|\mathbf{x}\| \rangle = \theta/\|\mathbf{x}\|$ .

Note that alternative 1) in Lemma 2 is equivalent to the second statement of Lemma 1. Alternative 2) corresponds to the case when condition (26) of Lemma 1 is not satisfied.  $\square$

## B Proof of Theorem 1

1. Let us first assume that  $\|\mathbf{w}\| = 1$ . Notice that the condition

$$\langle \mathbf{w}, \mathbf{x}_i \rangle \leq \theta \quad \forall \mathbf{x}_i \in \mathcal{M}, \quad (32)$$

assures that  $v = 0$  and hence  $s_i \in \text{Silent}(\mathcal{S}(\mathcal{M}), (\mathbf{w}, \theta)) \quad \forall s_i \in \mathcal{S}(\mathcal{M})$ .

In this case the neuron is silent for all stimuli except  $s_{M+1}$  that does evoke a response by construction. Therefore, it is sufficient to estimate the probability that (32) holds.

Let  $\mathcal{C}_n(\mathbf{w}, \theta)$  be the spherical cap:

$$\mathcal{C}_n(\mathbf{w}, \theta) = \{\mathbf{x} \in B_n(1) \mid \langle \mathbf{w}, \mathbf{x} \rangle > \theta\}.$$

Then the ratio of volumes  $\mathcal{V}(\mathcal{C}_n(\mathbf{w}, \theta))/\mathcal{V}(B_n(1))$  is the probability that a random vector  $\mathbf{x}_i \in \mathcal{C}_n(\mathbf{w}, \theta)$ . Observe that

$$\frac{\mathcal{V}(\mathcal{C}_n(\mathbf{w}, \theta))}{\mathcal{V}(B_n(1))} \leq \frac{1}{2}(1 - \theta^2)^{\frac{n}{2}}.$$

Thus, the probability that all  $\mathbf{x}_i \in \mathcal{M}$  are outside the cap  $\mathcal{C}_n(\mathbf{w}, \theta)$  is bounded from below:

$$P = \left[ 1 - \frac{\mathcal{V}(\mathcal{C}_n(\mathbf{w}, \theta))}{\mathcal{V}(B_n(1))} \right]^M \geq \left[ 1 - \frac{1}{2}(1 - \theta^2)^{\frac{n}{2}} \right]^M, \quad (33)$$

which is equivalent to (15), given that  $\|\mathbf{w}\| = 1$ .

Let  $\|\mathbf{w}\| \neq 1$ . Noticing that, for  $\|\mathbf{w}\| > 0$

$$\langle \mathbf{w}, \mathbf{x}_i \rangle \leq \theta \quad \forall \mathbf{x}_i \in \mathcal{M} \Leftrightarrow \langle \mathbf{w}/\|\mathbf{w}\|, \mathbf{x}_i \rangle \leq \theta/\|\mathbf{w}\| \quad \forall \mathbf{x}_i \in \mathcal{M},$$

and substituting  $\theta/\|\mathbf{w}\|$  in place of  $\theta$  in (33) results in (15).

- Let us show that for  $(\mathbf{w}, \theta) \in \Omega_D$  the neuron detects the relevant stimulus  $\mathbf{s}_{M+1}$ , i.e.,  $v > 0$ . Using (16) we observe that

$$\begin{aligned} \langle \mathbf{w}, \mathbf{x}_{M+1} \rangle - \theta &= \langle \mathbf{w} - \mathbf{w}^*, \mathbf{x}_{M+1} \rangle + \|\mathbf{x}_{M+1}\| - \theta \geq \langle \mathbf{w} - \mathbf{w}^*, \mathbf{x}_{M+1} \rangle + D \geq \\ &\geq -\|\mathbf{w} - \mathbf{w}^*\| \|\mathbf{x}_{M+1}\| + D > D(1 - \|\mathbf{x}_{M+1}\|) \geq 0, \end{aligned}$$

implying that  $\mathbf{s}_{M+1} \in \text{Activated}(\mathcal{Y}, (\mathbf{w}, \theta))$ .

Let us evaluate the probability that the neuron rejects all background stimuli for all  $(\mathbf{w}, \theta) \in \Omega_D$ . According to (16) the following holds:

$$\frac{\theta}{\|\mathbf{w}\|} \geq \frac{\|\mathbf{x}_{M+1}\| - 2D}{1 + D}, \quad \forall (\mathbf{w}, \theta) \in \Omega_D.$$

Moreover,  $\|\mathbf{x}_{M+1}\| \geq 1 - \varepsilon$  with probability  $p = 1 - (1 - \varepsilon)^n$ . Therefore, with probability larger or equal to  $p$ , the ratio  $\frac{\theta}{\|\mathbf{w}\|}$  is bounded from below as:

$$\frac{\theta}{\|\mathbf{w}\|} \geq \frac{1 - \varepsilon - 2D}{1 + D}.$$

Finally, since the value of  $\varepsilon$  can be chosen arbitrarily in the interval  $(0, 1 - 2D)$  and taking into account that the right-hand side of (33) is a monotone and increasing function with respect to  $\theta$  in the interval  $[0, 1]$ , estimate (17) immediately follows from (33) and (15).

□

### C Proof of Corollary 1

Consider (15) and denote

$$z = \frac{1}{2} \left[ 1 - \frac{\theta^2}{\|\mathbf{w}\|^2} \right]^{\frac{n}{2}}, \quad \phi = (1 - z)^{\overline{M}}. \tag{34}$$

According to (34),  $(1 - z)^M \geq \phi$  for all  $0 < M \leq \overline{M}$ . Given that  $z \in (0, 1)$ , from Eq. (34) we get  $\ln(\phi) = \overline{M} \ln(1 - z)$ . Recall that  $\ln(1 - z) > -z/(1 - z), \forall z \in (0, 1)$ . Thus, we can conclude that

$$\overline{M} > -\ln(\phi) \frac{1 - z}{z} = -\ln(\phi)(z^{-1} - 1) = -\ln(\phi)(2e^{an} - 1), \tag{35}$$

where  $\alpha$  is given by (18). Thus, according to (35), for  $0 < M \leq -\ln(\phi) (2e^{an} - 1) < \overline{M}$  the following holds

$$P(s_i \in \text{Silent}(\mathcal{S}(\mathcal{M}), (\mathbf{w}, \theta)) \forall s_i \in \mathcal{S}(\mathcal{M}) \mid \mathbf{w}, \theta) \geq \phi.$$

□

## D Proof of Theorem 2

The proof of the Theorem is essentially contained in Lemmas 3 and 4 (Sect. F). Consider the set  $\mathcal{Y}$ . With probability  $(1 - (1 - \varepsilon)^n)^m$ , all elements  $\mathbf{x}_i \in \mathcal{Y}$  satisfy the condition  $\|\mathbf{x}_i\| \geq 1 - \varepsilon$ . Hence, using Lemma 3 we have that the following inequality

$$|\langle \mathbf{x}_i, \mathbf{x}_j \rangle| \leq \frac{\delta}{1 - \varepsilon}, \quad \forall \mathbf{x}_i, \mathbf{x}_j \in \mathcal{Y}, \quad i \neq j$$

holds with probability

$$p_0 \geq (1 - (1 - \varepsilon)^n)^m \prod_{d=1}^{m-1} \left(1 - d(1 - \delta^2)^{\frac{n}{2}}\right).$$

This implies that, with probability  $p_0$ , the following conditions are met

$$\|\mathbf{x}_i\| \geq 1 - \varepsilon, \quad -\frac{(m-1)\delta}{1 - \varepsilon} \leq \sum_{j=1, j \neq i}^m \langle \mathbf{x}_i, \mathbf{x}_j \rangle \leq \frac{(m-1)\delta}{1 - \varepsilon}, \quad \forall \mathbf{x}_i \in \mathcal{Y}.$$

Consider  $\ell(\mathbf{x}) = \langle \mathbf{w}^*, \mathbf{x} \rangle - \theta^* + D$ . Invoking Lemma 4 and setting  $\beta_1 = \delta/(1 - \varepsilon)$ ,  $\beta_2 = -\delta/(1 - \varepsilon)$ , we can conclude that, with probability  $p_0$ ,

$$\ell(\mathbf{x}) \geq D, \quad \forall \mathbf{x} \in \mathcal{Y}.$$

In fact, we can conclude that with probability  $p_0$

$$\begin{aligned} \ell_0(\mathbf{x}) &= \langle \mathbf{w}, \mathbf{x} \rangle - \theta = \ell(\mathbf{x}) + \langle \mathbf{w} - \mathbf{w}^*, \mathbf{x} \rangle - \theta \\ &+ (\theta^* - D) > 0, \quad \forall (\mathbf{w}, \theta) \in \Omega_D, \quad \mathbf{x} \in \mathcal{Y}. \end{aligned}$$

Thus, the probability that  $\ell_0(\mathbf{x}) > 0$  for all  $\mathbf{x} \in \mathcal{Y}$  and that  $\ell_0(\mathbf{x}) \leq 0$  for all  $\mathbf{x} \in \mathcal{M}$  is bounded from below by

$$(1 - (1 - \varepsilon)^n)^m \prod_{d=1}^{m-1} \left(1 - d(1 - \delta^2)^{\frac{n}{2}}\right) \left[1 - \frac{1}{2} \left(1 - \frac{\theta^2}{\|\mathbf{w}\|^2}\right)^{\frac{n}{2}}\right]^M.$$

Noticing that  $\|\mathbf{w}\| \leq 1 + D$ , we can conclude that (21) holds. □

### E Proof of Theorem 3

According to Lemma 1, solutions  $\mathbf{w}(t, \mathbf{w}_0)$  are defined for all  $t \geq t_0$ . Moreover, condition 1 of the theorem and Lemma 1 imply that

$$\lim_{t \rightarrow \infty} \mathbf{w}(t, \mathbf{w}_0) = \frac{\sum_{i=1}^m \mathbf{x}_{M+i}}{\|\sum_{i=1}^m \mathbf{x}_{M+i}\|} = \bar{\mathbf{x}} / \|\bar{\mathbf{x}}\| = \mathbf{w}^*. \tag{36}$$

Let  $D > 0$  be chosen so that

$$0 < \theta + D \leq \theta^*.$$

Given that  $0 < \theta < \theta^*$ , such  $D$ s always exist. Equation (36) implies that there is a  $t_1(D) > t_0$  such that

$$\|\mathbf{w}(t, \mathbf{w}_0) - \mathbf{w}^*\| < D, \theta \in (0, \theta^* - D] \forall t \geq t_1(D). \tag{37}$$

The theorem now follows immediately from Theorem 2. □

### F Auxiliary Results

**Lemma 3** (cf. Gorban et al. 2016) *Let  $\mathcal{Y} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$  be a set of  $k$  i.i.d. random vectors from the equidistribution in the unit ball  $B_n(1)$ . Let  $\delta, r \in (0, 1)$ , and suppose that  $\|\mathbf{x}_i\| \geq r$ , for all  $i \in \{1, \dots, k\}$ .*

*Then the probability that the elements of  $\mathcal{Y}$  are pair-wise  $\delta/r$ -orthogonal, that is*

$$|\cos(\angle(\mathbf{x}_i, \mathbf{x}_j))| \leq \frac{\delta}{r} \text{ for all } i \neq j, i, j \in \{1, \dots, k\},$$

*is bonded from below as*

$$\begin{aligned} \mathcal{P} \left( |\cos(\angle(\mathbf{x}_i, \mathbf{x}_j))| \leq \frac{\delta}{r} \forall i, j \in \{1, \dots, k\}, i \neq j \mid \|\mathbf{x}_i\| \geq r, 1 \leq i \leq k \right) \\ \geq \prod_{d=1}^{k-1} \left( 1 - d \left( 1 - \delta^2 \right)^{\frac{n}{2}} \right). \end{aligned}$$

*Proof of Lemma 3* Let  $\mathbf{x}_i, i = 1, \dots, k$  be random vectors satisfying conditions of the lemma. Let  $E_\delta(\mathbf{x}_i)$  be the delta-thickening of the largest equator of  $B_n(1)$  that is orthogonal to  $\mathbf{x}_i$ . There is only one such equator, and it is uniquely determined by  $\mathbf{x}_i$ .

Consider the following probabilities:

$$\begin{aligned} &P(\mathbf{x}_2 \in E_\delta(\mathbf{x}_1)) \\ &P([\mathbf{x}_3 \in E_\delta(\mathbf{x}_2)] \& [\mathbf{x}_3 \in E_\delta(\mathbf{x}_1)]) \\ &P([\mathbf{x}_4 \in E_\delta(\mathbf{x}_3)] \& [\mathbf{x}_4 \in E_\delta(\mathbf{x}_2)] \& [\mathbf{x}_4 \in E_\delta(\mathbf{x}_1)]) \\ &\dots \\ &P([\mathbf{x}_k \in E_\delta(\mathbf{x}_{k-1})] \& \dots \& [\mathbf{x}_k \in E_\delta(\mathbf{x}_1)]). \end{aligned}$$

Pick  $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{Y}, i \neq j$ . Recall that, for any random events  $A_1, \dots, A_k$ , the probability

$$P(A_1 \& A_2 \& \dots \& A_k) \geq 1 - \sum_{i=1}^k (1 - P(A_i)). \quad (38)$$

According to (38), the probability that  $\mathbf{x}_i \in E_\delta(\mathbf{x}_j)$  is bounded from below by  $1 - (1 - \delta^2)^{\frac{n}{2}}$  (cf. Gorban et al. 2016, Proposition 3; see also Fig. 1 in Gorban et al. 2016 for illustration). Then

$$\begin{aligned} &P(\mathbf{x}_2 \in E_\delta(\mathbf{x}_1)) \geq 1 - (1 - \delta^2)^{\frac{n}{2}} \\ &P([\mathbf{x}_3 \in E_\delta(\mathbf{x}_2)] \& [\mathbf{x}_3 \in E_\delta(\mathbf{x}_1)]) \geq 1 - 2(1 - \delta^2)^{\frac{n}{2}} \\ &P([\mathbf{x}_4 \in E_\delta(\mathbf{x}_3)] \& [\mathbf{x}_4 \in E_\delta(\mathbf{x}_2)] \& [\mathbf{x}_4 \in E_\delta(\mathbf{x}_1)]) \geq 1 - 3(1 - \delta^2)^{\frac{n}{2}} \quad (39) \\ &\dots \\ &P([\mathbf{x}_k \in E_\delta(\mathbf{x}_{k-1})] \& \dots \& [\mathbf{x}_k \in E_\delta(\mathbf{x}_1)]) \geq 1 - (k - 1)(1 - \delta^2)^{\frac{n}{2}}. \end{aligned}$$

The fact that  $\mathbf{x}_i \in E_\delta(\mathbf{x}_j)$  combined with the condition that  $\|\mathbf{x}_i\| \geq r, \|\mathbf{x}_j\| \geq r$  imply:

$$|\cos(\angle(\mathbf{x}_i, \mathbf{x}_j))| \leq \frac{\delta}{r}.$$

Finally, given that  $\mathbf{x}_1, \dots, \mathbf{x}_k$  are drawn independently and that the distribution is rotationally invariant, the probability that all vectors in  $\mathcal{Y}$  are pair-wise orthogonal is the product of all probabilities in the left-hand side of (39). Thus the statement follows.  $\square$

**Lemma 4** Let  $\mathcal{Y} = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$  be a finite set from  $B_n(1)$ . Let  $\|\mathbf{x}_i\| \geq 1 - \varepsilon$ ,  $\varepsilon \in (0, 1)$  for all  $\mathbf{x}_i \in \mathcal{Y}$ , and  $\beta_1, \beta_2 \in \mathbb{R}$  be such that the following condition holds:

$$\beta_2(m - 1) \leq \sum_{j \in \{1, \dots, m\}, j \neq i} \langle \mathbf{x}_i, \mathbf{x}_j \rangle \leq \beta_1(m - 1) \text{ for all } i = 1, \dots, m. \quad (40)$$

Consider

$$\ell(\mathbf{x}) = \left\langle \frac{\bar{\mathbf{y}}}{\|\bar{\mathbf{y}}\|}, \mathbf{x} \right\rangle - \frac{1}{\sqrt{m}} \left( \frac{(1 - \varepsilon)^2 + \beta_2(m - 1)}{\sqrt{1 + (m - 1)\beta_1}} \right), \quad \bar{\mathbf{y}} = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i,$$

and suppose that parameters  $\beta_1, \beta_2$  satisfy:

$$(1 - \varepsilon)^2 + \beta_2(m - 1) > 0, \quad 1 + (m - 1)\beta_1 > 0.$$

Then

$$\ell(\mathbf{x}_i) \geq 0 \text{ for all } \mathbf{x}_i \in \mathcal{Y}. \tag{41}$$

*Proof of Lemma 4* Consider the set  $\mathcal{Y}$ . According to the lemma assumptions,  $\|\mathbf{x}_i\| \geq 1 - \varepsilon$  for some given  $\varepsilon \in (0, 1)$  and all  $i = 1, \dots, m$ . Consider now the mean vector  $\bar{\mathbf{y}}$

$$\bar{\mathbf{y}} = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i,$$

and evaluate the following inner products

$$\left\langle \frac{\bar{\mathbf{y}}}{\|\bar{\mathbf{y}}\|}, \mathbf{x}_i \right\rangle = \frac{1}{m\|\bar{\mathbf{y}}\|} \left( \|\mathbf{x}_i\|^2 + \sum_{j \in \{1, \dots, m\}, j \neq i} \langle \mathbf{x}_i, \mathbf{x}_j \rangle \right), \quad i = 1, \dots, m.$$

According to assumption (40), the following holds

$$\left\langle \frac{\bar{\mathbf{y}}}{\|\bar{\mathbf{y}}\|}, \mathbf{x}_i \right\rangle \geq \frac{1}{m\|\bar{\mathbf{y}}\|} \left( (1 - \varepsilon)^2 + \beta_2(m - 1) \right),$$

and, respectively,

$$\frac{1}{m} (1 + (m - 1)\beta_1) \geq \langle \bar{\mathbf{y}}, \bar{\mathbf{y}} \rangle = \|\bar{\mathbf{y}}\|^2 \geq \frac{1}{m} \left( (1 - \varepsilon)^2 + \beta_2(m - 1) \right)$$

Let  $(1 - \varepsilon)^2 + \beta_2(m - 1) > 0$  and  $1 + \beta_1(m - 1) > 0$ . It is clear that for  $\ell$ , as defined by (41), the following holds for all  $i = 1, \dots, m$ :  $\ell(\mathbf{x}_i) \geq 0$ . □

## References

Andersen P, Morris R, Amaral D, Bliss T, O’Keefe J (eds) (2007) The hippocampus book. Oxford University Press, Oxford

Amaral DG, Witter MP (1989) The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience* 31:571–591

Amit DJ, Gutfreund H, Sompolinsky H (1985) Storing infinite numbers of patterns in a spin-glass model of neural networks. *Phys Rev Lett* 55:1530–1533

- Barrett Lisa Feldman, Tugade Michele M, Engle Randall W (2004) Individual differences in working memory capacity and dual-process theories of the mind. *Psychol Bull* 130(4):553
- Benito N, Fernandez-Ruiz A, Makarov VA, Makarova J, Korovaichuk A, Herreras O (2014) Spatial modules of coherent activity in pathway-specific lfps in the hippocampus reflect topology and different modes of presynaptic synchronization. *Cereb Cortex* 11(7):1738–1752
- Benito N, Martin-Vazquez G, Makarova J, Makarov VA, Herreras O (2016) The right hippocampus leads the bilateral integration of gamma-parsed lateralized information. *eLife* 5:e16658. <https://doi.org/10.7554/eLife.16658>
- Calvo C, Villacorta-Atienza JA, Mironov VI, Gallego V, Makarov VA (2016) Waves in isotropic totalistic cellular automata: application to real-time robot navigation. *Adv Complex Syst* 19(4):1650012–18
- Clark DD, Sokoloff L (1999) Circulation and energy metabolism of the brain. In: Siegel GJ, Agranoff BW, Albers RW, Fisher SK, Uhler MD (eds) *Basic neurochemistry: molecular, cellular and medical aspects*. Lippincott, Philadelphia, pp 637–670
- Cucker F, Smale S (2002) On the mathematical foundations of learning. *Bull Am Math Soc* 39(1):1–49
- Draeos TJ, Miner NE, Lamb CC, Vineyard CM, Carlson KD, James CD, Aimone JB (2016) Neurogenesis deep learning. arXiv preprint [arXiv:1612.03770](https://arxiv.org/abs/1612.03770)
- Fernandez-Ruiz A, Makarov VA, Herreras O (2012) Sustained increase of spontaneous input and spike transfer in the ca3-ca1 pathway following long term potentiation in vivo. *Front Neural Circuits* 6:71
- Finnerty CT, Jefferys JGR (1993) Functional connectivity from ca3 to the ipsilateral and contralateral ca1 in the rat dorsal hippocampus. *Neuroscience* 56(1):101
- Fontaine B, Peña JL, Brette R (2014) Spike-threshold adaptation predicted by membrane potential dynamics in vivo. *PLoS Comput Biol* 10(4):e1003560
- Gorban AN, Tyukin IY, Romanenko I (2016) The blessing of dimensionality: Separation theorems in the thermodynamic limit. *IFAC-PapersOnLine* 49(24):64–69, 2016. 2th IFAC Workshop on Thermodynamic Foundations for a Mathematical Systems Theory TFMST 2016
- Gorban AN, Tyukin IY (2018) Blessing of dimensionality: mathematical foundations of the statistical physics of data. *Philosophical Trans R Soc A*. <https://doi.org/10.1098/rsta.2017.0237>
- Gorban AN, Tyukin IY (2017) Stochastic separation theorems. *Neural Netw* 94:255–259
- Gorban AN, Tyukin IY, Prokhorov DV, Sofeikov KI (2016) Approximation with random bases: pro et contra. *Inf Sci* 364–365:129–145
- Gorban AN, Tyukina TA, Smirnova EV, Pokidysheva LI (2016) Evolution of adaptation mechanisms: adaptation energy, stress, and oscillating death. *J Theor Biol* 405:127–139
- Hecht-Nielsen R (1994) Context vectors: general-purpose approximate meaning representations self-organized from raw data. In: Zurada J, Marks R, Robinson C (eds) *Computational intelligence: imitating life*. IEEE Press, London
- Herculano-Houzel S (2009) The human brain in numbers: a linearly scaled-up primate brain. *Front Hum Neurosci* 3:31
- Herculano-Houzel S (2011) Gorilla and orangutan brains conform to the primate cellular scaling rules: implications for human evolution. *Brain Behav Evol* 77:33–44
- Herculano-Houzel S (2012) The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proc Nat Acad Sci* 109:10661–10668
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Nat Acad Sci* 79(8):2554–2558
- Ishizuka N, Weber J, Amaral DG (1990) Organization of intrahippocampal projections originating from ca3 pyramidal cells in the rat. *J Comp Neurol* 295(580–623):580
- Ison MJ, Quiroga R Quian, Fried I (2015) Rapid encoding of new memories by individual neurons in the human brain. *Neuron* 87(1):220–230
- Jurica P, Gepshtein S, Tyukin I, van Leeuwen C (2013) Sensory optimization by stochastic tuning. *Psychol Rev* 120(4):798–816
- Kainen PC, Kurkova V (1993) Quasiorthogonal dimension of euclidian spaces. *Appl Math Lett* 6(3):7–10
- Khalil H (2002) *Nonlinear systems*, 3rd edn. Prentice Hall, Upper Saddle River
- Leung Chi-Sing, Chan Lai-Wan, Lai Edmund (1995) Stability, capacity, and statistical dynamics of second-order bidirectional associative memory. *IEEE Trans Syst Man Cybernet* 25(10):1414–1424
- Li XG, Somogyi P, Ylinen A, Buzsaki G (1994) The hippocampal ca3 network: an in vivo intracellular labeling study. *J Comp Neurol* 339:181–208
- Lindsay GW, Rigotti M, Warden MR, Miller EK, Fusi S (2017) Hebbian learning in a random network captures selectivity properties of prefrontal cortex. *bioRxiv*, p 133025

- Lobov SA, Zhuravlev MO, Makarov VA, Kazantsev VB (2017) Noise enhanced signaling in stdp driven spiking-neuron network. *Math Model Nat Phenom* 12(4):109–124
- Markram H, Lubke J, Frotscher M, Sakmann B (1997) Regulation of synaptic efficacy by coincidence of postsynaptic apss and epsps. *Science* 275(5297):213–215
- Oja E (1982) A simplified neuron model as a principal component analyzer. *J Mathe Biol* 15:267–273
- Perlovsky LI (2006) Toward physics of the mind: concepts, emotions, consciousness, and symbols. *Phys Life Rev* 3(1):23–55
- Platek M, Keenan JP, Shackelford T K (2007) *Evolutionary cognitive neuroscience*. MIT Press, Cambridge
- Quiroga R Quian (2012) Concept cells: the building blocks of declarative memory functions. *Nat Rev Neurosci* 13(8):587–597
- Quiroga R Quian, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435(7045):1102–1107
- Reimann MW, Nolte M, Scolamiero M, Turner K, Perin R, Chindemi G, Dlotko P, Levi R, Hess K, Markram H (2017) Cliques of neurons bound into cavities provide a missing link between structure and function. *Front Comput Neurosci* 11:48
- Roberts A, Conte D, Hull M, Merrison-Hort R, al Azad AK, Buhl E, Borisyuk R, Soffe SR (2014) Can simple rules control development of a pioneer vertebrate neuronal network generating behavior? *J Neurosci* 34(2):608–621
- Rosenblatt F (1962) *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Spartan Books, Sparta
- Sherwood CC, Bauernfeind AL, Bianchi S, Raghanti MA, Hof PR (2012) Human brain evolution writ large and small. *Prog Brain Res* 195:237–254
- Sousa AM, Meyer KA, Santpere G, Gulden FO, Sestan N (2017) Evolution of the human nervous system function, structure, and development. *Cell* 170(2):226–247
- Tyukin IY, Gorban AN, Sofeikov K, Romanenko I (2017) Knowledge transfer between artificial intelligence systems. arXiv preprint [arXiv:1709.01547](https://arxiv.org/abs/1709.01547)
- Vapnik V, Chappelle O (2000) Bounds on error expectation for support vector machines. *Neural Comput* 12(9):2013–2036
- Villacorta-Atienza JA, Makarov VA (2013) Neural network architecture for cognitive navigation in dynamic environments. *IEEE Trans Neural Netw Learn Syst* 24(12):2075–2087
- Villacorta-Atienza JA, Calvo C, Makarov VA (2015) Prediction-for-compactness: navigation in social environments using generalized cognitive maps. *Biol Cybernet* 109(3):307–320
- Villacorta-Atienza JA, Calvo C, Lobov S, Makarov VA (2017) Limb movement in dynamic situations based on generalized cognitive maps. *Math Model Nat Phenom* 12(4):15–29
- Viskontas IV, Quiroga R Quian, Fried I (2009) Human medial temporal lobe neurons respond preferentially to personally relevant images. *Proc Nat Acad Sci* 106(50):21329–21334
- Wittner L, Henze DA, Zaborszky L, Buzsaki G (2007) Three-dimensional reconstruction of the axon arbor of a ca3 pyramidal cell recorded and filled in vivo. *Brain Struct Funct* 212(1):75–83