

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Research report

From sensorimotor learning to memory cells in prefrontal and temporal association cortex: A neurocomputational study of disembodiment

Friedemann Pulvermüller^{a,*} and Max Garagnani^{a,b,*}^a Brain Language Laboratory, Department of Philosophy, Freie Universität Berlin, Berlin, Germany^b School of Computing and Mathematics, Plymouth University, Plymouth, UK

ARTICLE INFO

Article history:

Received 2 May 2013

Reviewed 20 August 2013

Revised 20 November 2013

Accepted 8 February 2014

Action editor Stefano Cappa

Published online 11 March 2014

Keywords:

Cognitive theory

Embodied cognition

Hebbian learning

Neurobiology of working memory

Perception action circuit

ABSTRACT

Memory cells, the ultimate neurobiological substrates of working memory, remain active for several seconds and are most commonly found in prefrontal cortex and higher multisensory areas. However, if correlated activity in “embodied” sensorimotor systems underlies the formation of memory traces, why should memory cells emerge in areas distant from their antecedent activations in sensorimotor areas, thus leading to “disembodiment” (movement away from sensorimotor systems) of memory mechanisms? We modelled the formation of memory circuits in six-area neurocomputational architectures, implementing motor and sensory primary, secondary and higher association areas in frontotemporal cortices along with known between-area neuroanatomical connections. Sensorimotor learning driven by Hebbian neuroplasticity led to formation of cell assemblies distributed across the different areas of the network. These action-perception circuits (APCs) ignited fully when stimulated, thus providing a neural basis for long-term memory (LTM) of sensorimotor information linked by learning. Subsequent to ignition, activity vanished rapidly from APC neurons in sensorimotor areas but persisted in those in multimodal prefrontal and temporal areas. Such persistent activity provides a mechanism for working memory for actions, perceptions and symbols, including short-term phonological and semantic storage. Cell assembly ignition and “disembodied” working memory retreat of activity to multimodal areas are documented in the neurocomputational models’ activity dynamics, at the level of single cells, circuits, and cortical areas. Memory disembodiment is explained neuromechanistically by APC formation and structural neuroanatomical features of the model networks, especially the central role of multimodal prefrontal and temporal cortices in bridging between sensory and motor areas. These simulations answer the “where” question of cortical working memory in terms of distributed APCs and their inner structure, which is, in part, determined by neuroanatomical structure. As the neurocomputational model provides a mechanistic explanation of how memory-related “disembodied” neuronal activity emerges in “embodied” APCs, it may be key to solving aspects of the embodiment debate and eventually to a better understanding of cognitive brain functions.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

* Corresponding authors. Brain Language Laboratory, Department of Philosophy and Humanities, Freie Universität Berlin, 19145 Berlin, Germany.

E-mail addresses: friedemann.pulvermuller@fu-berlin.de (F. Pulvermüller), M.Garagnani@fu-berlin.de (M. Garagnani).

<http://dx.doi.org/10.1016/j.cortex.2014.02.015>

0010-9452/© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

When animals keep in mind the shape or colour of a stimulus and perform a concordant matching response after a delay of several seconds (delayed matching to sample task), some neurons fire at an enhanced level throughout the delay (Fig. 1, Constantinidis, Franowicz, & Goldman-Rakic, 2001; Fuster, 1995; Fuster & Alexander, 1971; Kojima & Goldman-Rakic, 1982; Romanski & Goldman-Rakic, 2002). Intriguingly, the persistent responses of these *memory cells* are frequently stimulus specific, thus being strong when, for example, red or cross-shaped stimuli have to be memorized, but being reduced for other items (e.g., Fuster & Jervey, 1981, 1982; Miyashita & Chang, 1988). Memory cell activity provides a neurobiological basis for working memory (Baddeley, 1992), in both animals and humans, for stimuli from different modalities (visual, auditory, speech) and their related concepts and motor patterns (Baddeley, 2003; D'Esposito, 2007; Fuster, 2003; Goldman-Rakic, 1995; Linden, 2007; Postle, 2006). Explanations for persistent memory activity have been offered in terms of reverberant firing (Verduzco-Flores, Bodner, Ermentrout, Fuster, & Zhou, 2009; Zipser, Kehoe, Littlewort, & Fuster, 1993) and/or cell-intrinsic properties, especially calcium-induced synaptic facilitation in prefrontal cortex (Fransen, Tahvildari, Egorov, Hasselmo, & Alonso, 2006; Loewenstein & Sompolinsky, 2003; Mongillo, Barak, & Tsodyks, 2008), within distributed memory networks. The former account postulates that reverberation of neuronal activity in strongly connected cell assemblies distributed over different cortical areas is the neuronal mechanism underlying working memory; the latter proposes changes in calcium levels consequent to stimulation and associated excitability modulation as the critical mechanisms. Although these views are not *a priori* incompatible with each other, we here provide new arguments in favour of a network memory account based on perception action circuits. Using a neurocomputational model mimicking cortical neuroanatomical structure we show that network memory

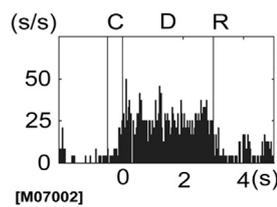


Fig. 1 – Memory cell activity recorded from the prefrontal cortex of a macaque monkey during a delayed-matching-to-sample task. This peri-stimulus time histogram (PSTH) plots firing rate (spikes per second, calculated for 30 msec time bins) against time. In this task, the monkey was required to make a saccade to a location where a cue had been presented. There were 8 possible predetermined locations (0°, 45°, 90°, ..., 315°). The cue (C) stimulus was presented for .5 sec (cue offset at time zero), followed by a fixation point (FP) presented at the centre of the monitor. After a 3 sec delay (D), the FP was removed, which gave the monkey the signal to respond (R)., adapted from Takeda & Funahashi, 2002).

mechanisms account for a feature of working memory left unexplained by previous models, namely the topography and distribution of memory cells in the cortex.

A most prominent feature of memory cell activity is that it cannot be seen in all cortical areas to the same degree. Strong and long-lasting persistent activity is reported primarily in anterior–inferior temporal areas for visual stimuli, in superior temporal areas for auditory stimuli including spoken words, in posterior parietal cortex for tactile stimuli, and in specific sections of dorsolateral and inferior prefrontal cortex for stimuli of all modalities (for review, see D'Esposito, 2007; Fuster, 2009; Romanski, 2004). While some memory cell activity has also been reported in primary visual and auditory as well as premotor cortex (Fuster, 1990; Lemus, Hernandez, & Romo, 2009; Romo, Hernandez, & Zainos, 2004; Serences, Ester, Vogel, & Awh, 2009; Super, Spekreijse, & Lamme, 2001), the percentage of memory cells in sensorimotor areas is typically lower than in prefrontal and anterior-temporal areas (Fuster, 1995, 2001; Linden, 2007). This local specificity of memory cell activity poses problems to both of the above accounts. Memory circuits resulting from correlated neural activity across sensorimotor areas (see Fuster, 1995; Hebb, 1949; Pulvermüller, 1999) can be seen as functionally uniform, and are, in this case, compatible with equal activity across their neurons. Therefore, in the context of learnt working memory circuits, the topographically specific distribution of memory cells across frontal, temporal and parietal cortex begs for an explanation. Equally, the cell-intrinsic account fails to explain local specificity because neurons and their excitability changes to stimulation are unlikely to be unique to one or a few areas. If anything, one may be tempted to argue that memory activity should be most pronounced where stimuli elicit strongest responses and consequent modulation of extracellular calcium concentrations would therefore be most pronounced. Hence, experiments using visual stimuli should elicit stronger memory cell activity in the directly stimulated primary sensory areas, but not in dorsolateral prefrontal cortex, where stimulus-elicited activation typically arrives after several additional synaptic steps. This prediction is in stark contrast with the experimental observation that primary cortices show some memory activity, but memory cells are by far more common in anterior-temporal and prefrontal cortex (see, for example, Fuster, 1995, 2001; Linden, 2007).

To trace area-specificity of memory cell activity, we use a model mimicking area structure and connectivity patterns across a range of cortices, including primary, secondary and higher-association areas (Garagnani & Pulvermüller, 2011; Garagnani, Wennekers, & Pulvermüller, 2007, 2008, 2009; Wennekers, Garagnani, & Pulvermüller, 2006), interconnected according to evidence revealed by neuroanatomical studies on long-range corticocortical fibres. Using Hebbian (Hebb, 1949) learning principles, we trained this model so that perception-action circuits (cell assemblies) emerged in it as a consequence of the repeated presentation of specific sensory input and motor output pattern pairs (D'Esposito, 2007; Fuster, 2003). In this model, the mechanism for associating sensory and motor patterns is the build-up, by way of correlation learning, of distributed cell assemblies spanning different

areas and linking sensory with motor information, as suggested by major brain models of working memory. These cell assemblies are considered the cortical mechanism underlying long-term memory (LTM), and their sustained activity the neural basis of working memory. This approach explains the emergence of memory traces as a consequence of stimulation of the senses and concomitant motor patterns and is applicable to the learning of basic perception-action contingencies (Fuster, 2003), the acquisition of phonological forms of spoken words (Braitenberg & Pulvermüller, 1992; Pulvermüller & Fadiga, 2010), as well as to the grounding of concepts and semantics in perception and action (Barsalou, 2008; Kiefer & Pulvermüller, 2012; Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012; Pulvermüller, 2005). As correlated sensory and motor activity is seen as the driving force of memory formation, this approach provides a mechanism for so-called “embodied” cognitive processes for perception, action, language and concepts (Kiefer & Pulvermüller, 2012). A crucial challenge for such an embodied neuronal model is to provide a mechanism for the transition from sensorimotor activation to memory trace formation, and from sensory stimulation to persistent working memory activations, which, as mentioned, are most typically seen in areas far removed from the primary cortical areas where stimulus information first arrives and motor responses are being triggered (Fuster, 2009). Below, we provide a mechanistic account for this transition from sensory/motor to higher cortices, from “embodied” to “disembodied” network activity. The explanation capitalizes on principles and features of neuroanatomical structure and function.

We should emphasize that structure and mechanisms implemented in the model are aimed at reflecting real cortical structural and functional features known from neurobiology: sparse and predominantly local connectivity within areas, lateral inhibition, patchy and topographic between-area links conforming with tracer and diffusion tensor and diffusion-weighted imaging (DTI, DWI) studies are among the well-documented anatomical features implemented (Amir, Harel, & Malach, 1993; Braitenberg & Schüz, 1998; Douglas & Martin, 2004; Gilbert & Wiesel, 1983; Rilling, 2014). Non-linear activity summation, neuronal adaptation and synaptic plasticity are among the neurophysiological features, with Hebbian learning mechanisms implemented according to the established synaptic plasticity phenomena of long-term potentiation and depression (LTP, LTD) (Artola & Singer, 1993; Malenka & Nicoll, 1999; Matthews, 2001).

2. Materials and methods

We used a mean field network model subdivided into banks or “areas” of artificial neurons with reciprocal connections between and within “areas”. The model was constructed so as to mimic a range of biologically realistic properties. We included the following features:

1. Area structure: six areas, modelling pre-specified sensorimotor and multimodal brain systems;
2. Between-area connectivity, constrained by specific neuroanatomical data and obeyed general neuroanatomical

principles in being sparse, random, initially weak and topographic;

3. Within-area connectivity, which was similarly sparse, random and initially weak, and exhibited a neighbourhood bias towards local links;
4. Local and area-specific global regulation mechanisms;
5. Synaptic modification by way of Hebb-type learning including both LTP and LTD;
6. Constant presence of uniform uncorrelated white noise during all phases of learning and retrieval in all parts of the network, and
7. Additional noise added to the stimulus patterns to mimic realistic noisy input conditions during retrieval.

These features are discussed in more detail below (see also Appendix A, Fig. 1, Discussion) and in previous publications (Garagnani et al., 2008, 2009; Garagnani & Pulvermüller, 2011; Garagnani & Pulvermüller, 2013).

2.1. Model architecture

To investigate the neurobiological mechanisms that underlie the formation of memory traces in the cortex, we implemented a biologically grounded neural-network model (illustrated in Fig. 2) reproducing structural and functional properties of sensory, motor and association areas of the brain along with their long-distance connectivity (Fig. 2A). The model architecture used in previous simulations (Fig. 2, see Garagnani et al., 2007, 2008) replicated the structure of the left perisylvian cortex involved in storing correlations between articulatory–phonological patterns constituting spoken word forms and their corresponding auditory–phonological signals (Pulvermüller, 1999). This architecture used three main auditory areas, auditory core including primary auditory cortex (A1), auditory belt (AB), and parabelt areas (PB), and three inferior frontal areas, primary articulatory motor cortex (M1), inferior premotor (PM) and prefrontal cortex (PF, Fig. 2A). The model also realised some of the documented between-area connections (green-coloured arrows, Fig. 2C), especially the reciprocal links documented between adjacent areas (e.g., A1 and AB, AB and PB etc., Kaas & Hackett, 2000; Pandya, 1995; Pandya & Yeterian, 1985; Pulvermüller, 1992; Rauschecker & Tian, 2000; Young, Scannell, & Burns, 1995; Young, Scannell, Burns, & Blakemore, 1994). Furthermore, long-distance connections between inferior prefrontal and auditory parabelt areas in anterior, lateral and posterior superior temporal cortex were realised (Catani & Ffytche, 2005; Makris et al., 1999; Parker et al., 2005; Petrides & Pandya, 2002; Romanski et al., 1999). In addition to these previously implemented next-neighbour connections and links between association cortices (Fig. 2C), an increasing body of recent evidence indicates the existence of “jumping” links, skipping one intermediate area, especially between parabelt and premotor (Fig. 2D, Petrides & Pandya, 2009; Saur et al., 2008) and between belt and prefrontal areas (Kaas & Hackett, 2000; Rauschecker & Scott, 2009; Romanski et al., 1999). Furthermore, previous evidence had already indicated further “jumping” links between non-adjacent areas within superior-temporal and inferior frontal cortex (Deacon, 1992; Pandya & Yeterian, 1985; Young et al., 1994). To improve the neuroanatomical plausibility of

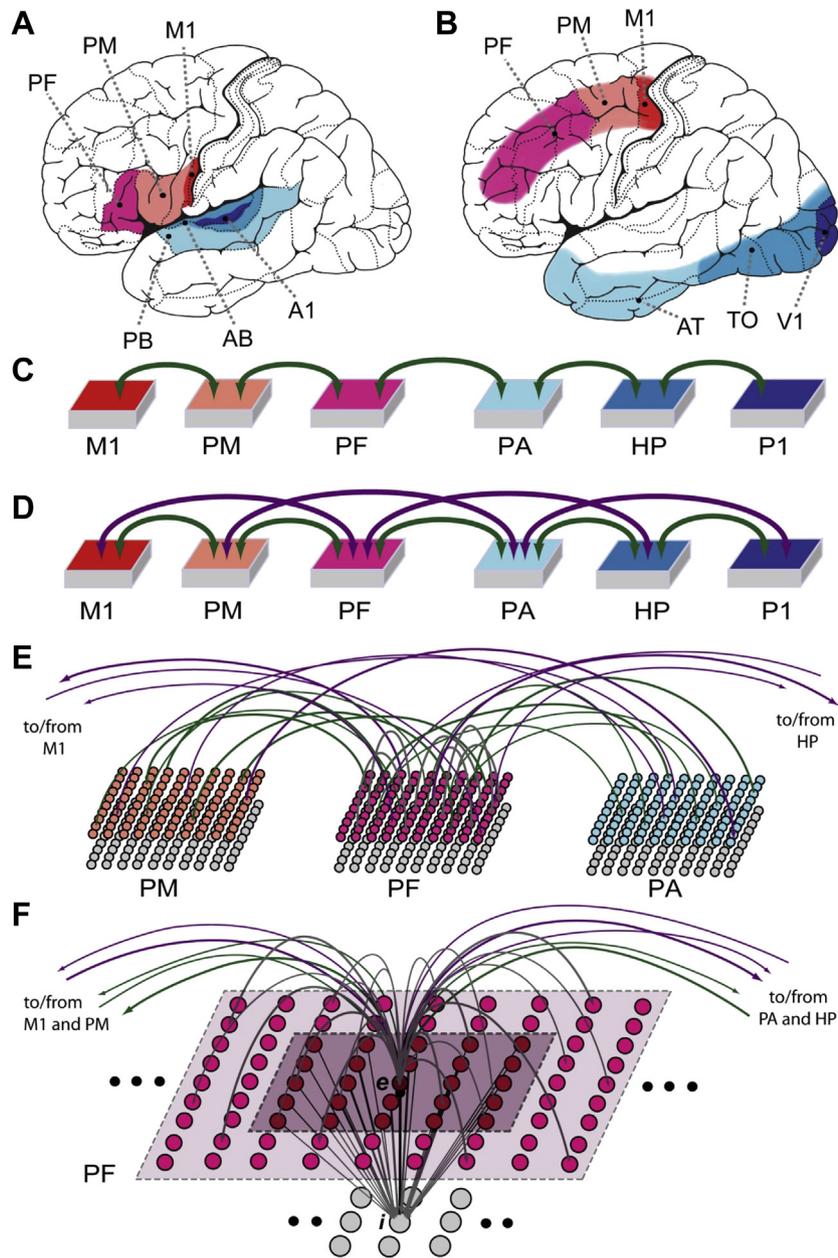


Fig. 2 – Brain areas, model architecture and connectivity. (A)–(B) Sets of cortical areas, which were imitated by the network’s area structure and long-distance connectivity. Sensory (different shades of blue) and motor (shades of red) areas relevant for learning the associations (A) between articulatory movements and the resultant sounds – that is, primary auditory cortex (AB), auditory belt (PB) and parabelt (PB), and inferior primary motor (M1), premotor (PM) and prefrontal (PF) cortex – and (B) between visual stimuli and hand motor actions – that is, primary visual (V1), temporo-occipital (TO) and anterior-temporal (AT) areas, as well as dorsolateral motor (M1), premotor (PM) and prefrontal cortex (PF). (C) Area macrostructure and between area connections of a previous network in which adjacent areas are connected reciprocally with each other and long-distance connections link prefrontal and temporal association areas in the “centre” of the network. (D) Macrostructure of the improved model motivated by neuroanatomical data (see text); in addition to next-neighbour between-area connections displayed in (C), this network also includes connections that skip one area (purple arrows). Architecture (D) was used for simulating sensorimotor learning and working memory. Model areas correspond to cortical areas in (A) and (B) as indicated by the colour code, in particular primary motor (M1), premotor (PM), prefrontal (PF), and primary perceptual (P1), higher perceptual (HP) and perceptual association (PA) areas. (E) Structure of network areas and connections within and between them. Each area is comprised of two layers of 25×25 units, made of excitatory (upper colour-filled circles) and inhibitory (lower grey-filled circles) cells. Each unit, or “node”, represents a cluster of pyramidal cells, and is implemented as a graded-response leaky integrator (Appendix A). Within- (grey arcs) and between- (green and purple arcs) area connections are not “all-to-all” but random, sparse and patchy, and topographically organised (not illustrated). (F) Microstructure of the connectivity of one single excitatory cell (labelled “e”). Local (lateral) inhibition is implemented by an underlying cell “i”

the network, all of these “jumping” connections (purple-coloured arrows, Fig. 2D) were included in the model used in the present study.

An important aspect of this six-area architecture (Fig. 2D) is that all mechanisms and connectivity features implemented in it are closely paralleled by properties found throughout the cortex. Because of this feature, although the network’s motor, sensory and higher areas were originally intended to model inferior frontal motor systems and superior temporal auditory areas (Fig. 2A), they can also simulate processes in other sets of sensory and motor areas which exhibit similar connection structure and function. In fact, a similar area structure can be distinguished in the ventral visual or “what” stream in inferior-occipital and -temporal cortex and in the dorsolateral motor system (Fig. 2B). These sets of areas are especially important for learning correlations between visual stimuli and hand-actions, for example for connecting information a monkey processes in a prototypical delayed-matching-to-sample experiment where visual stimuli must be responded to by concordant motor movements of the hand. The dorsolateral prefrontal cortex allows for the same distinction of motor, premotor and prefrontal areas as outlined for its inferior part (Pandya & Yeterian, 1985; Rizzolatti & Luppino, 2001). The hand/arm motor area and adjacent premotor and more rostral prefrontal regions are reciprocally connected (Arikuni, Watanabe, & Kubota, 1988; Dum & Strick, 2002, 2005; Lu, Preston, & Strick, 1994; Pandya & Yeterian, 1985; Rizzolatti & Luppino, 2001), with links also documented between prefrontal and primary motor areas (Guye et al., 2003; Young et al., 1995). Although the visual system has an extremely complex anatomy (Young et al., 1995), a focus on the inferior temporal “what” stream (Mishkin, Ungerleider, & Macko, 1983; Ungerleider & Haxby, 1994) also allows, with some generalisation, a distinction into primary visual, inferior temporal-occipital and inferior anterior-temporal areas (blue-shaded areas, Fig. 2B). These areas exhibit reciprocal next-neighbour-connections (Distler, Boussaoud, Desimone, & Ungerleider, 1993; Nakamura, Gattass, Desimone, & Ungerleider, 1993) and are also linked via the “jumping” link, by way of the inferior longitudinal fascicle, from V1 to anterior temporal regions (Catani, Jones, Donato, & Ffytche, 2003; Wakana, Jiang, Nagae-Poetscher, van Zijl, & Mori, 2004). Furthermore, neuroanatomical studies (Nakamura et al., 1993; Pandya & Barnes, 1987; Ungerleider, Gaffan, & Pelak, 1989; Webster, Bachevalier, & Ungerleider, 1994) and experiments inducing inactivation by local cooling in the macaque monkey (Bauer & Fuster, 1976, 1978; Chafee & Goldman-Rakic, 2000; Fuster, Bauer, & Jervey, 1981, 1985) suggest the presence of both long-distance and jumping links between (visual) inferior temporal and mid-dorsolateral prefrontal and premotor cortices analogous to those connecting (auditory) superior-temporal and inferior frontal areas described earlier.

As the neuroanatomical model structure (Fig. 2D) shows similarity with at least two large brain systems, those for

auditory–articulatory and for visual–manual association, we use a general terminology for its areas. On the sensory side, a primary perceptual, P1, a higher-perceptual, HP, and a perceptual association area or convergence zone (Damasio, 1989), PA, are complemented in frontal lobe by a primary motor, M1, premotor, PM, and prefrontal area, PF. The structure can also be interpreted as implementation of a crossing of these systems (e.g., visual-to-articulatory association) and other sensory-motor brain systems (for example haptic-motor, olfactory-motor) may be captured by this or a similar scheme. Our choice of target brain systems is motivated by pre-existing experiments and the resultant host of data from the typical visual delayed-matching-to-sample tasks (using finger or hand responses) and from verbal working memory tasks (implicating auditory input and the articulators as output system), where distinct patterns of brain responses have been reported (D’Esposito, 2007; Fuster, 2009; Postle, 2006).

Each model area included $25 \times 25 = 625$ excitatory and the same number of inhibitory neurons (Fig. 2E, for details, see caption). Following general principles of cortical anatomy in mammals, connections between and within areas were sparse (thus realising only a small fraction of all possible connections), patchy and topographic (Amir et al., 1993; Braitenberg & Schüz, 1998; Gilbert & Wiesel, 1983), and such that local connection probability fell off with distance (Braitenberg & Schüz, 1998). As a network correlate of further features of brain structure and function, local and area-specific inhibition mechanisms were also implemented (Fig. 2F, caption, Bibbig, Wennekens, & Palm, 1995; Palm, 1982; Wennekens et al., 2006). These inhibition mechanisms act as a means to regulate and control activity in the network and provide a correlate of attention (Braitenberg, 1978; Garagnani et al., 2008; Palm, 1982). Details of the neuron model and the Hebb-type learning mechanisms implemented, which covered both LTP and LTD, are summarised in Appendix A.

2.2. Simulations

Model simulations were carried out in three steps. After *learning* of sensorimotor patterns and build-up of cell assemblies, the resultant learnt LTM traces were precisely identified and their integrity qualitatively tested. Finally, as a critical step for this investigation, the *retrieval phase* scrutinised the dynamic activation process that followed sensory stimulation of memory traces, paying special attention to persisting working memory activity.

2.2.1. Learning

Initially, all synaptic links and weights were random, sparse and weak, mimicking well-known properties of cortical connectivity (Braitenberg & Schüz, 1998). Twelve different to-be-memorised sensorimotor patterns w were built and presented to the network; each pattern w included both a

(representing a cluster of inhibitory interneurons situated within the same cortical column) receiving excitatory input from all cells situated within a local (5×5) neighbourhood (dark-coloured area) and projecting back to e , inhibiting it. Within-area sparse excitatory links (in grey) to and from e are limited to a (19×19) neighbourhood (light-coloured area); between-area excitatory projections (green and purple arcs) are topographic and target 19×19 neighbourhoods in other areas (not depicted).

“sensory” and a “motor” component (17 specific cells in P1 and another 17 in M1, equalling 2.72% of the neurons in each respective 25×25 area). Each component was built by choosing 17 cells at random within an area. By “a stimulus pattern w was presented”, we mean that its 2×17 cells were activated. A Hebbian learning rule (Appendix A) was used to re-adjust synaptic weights throughout training. Due to sensorimotor activation, neural activity was present in specific neurons in P1 and M1, which partly activated further neural elements connected to these stimulated ones. Correlated activity led to synaptic strengthening so that, eventually, sensorimotor stimulation led to increasingly stronger activation spreading to specific neuron sets throughout the network, which finally led to formation of a distributed LTM trace, a cell assembly (for detailed description and analyses of cell assembly formation in this type of network, see Garagnani et al., 2008, 2009). At the onset of each “learning trial”, areas P1 and M1 were simultaneously stimulated with one sensorimotor pattern for 2 times steps. After each stimulation, network activity was allowed to return to a predefined baseline value, so that, when the next stimulus pair appeared, any possibly contaminating activity related to the previous input was minimal or absent. Because the amount of activity induced in the network differed between patterns and changed over learning, learning trials were followed by different inter-stimulus intervals (ISIs). During these ISIs the only input to the network was uniform white noise, simulating the spontaneous baseline neuronal firing observed in real neuronal cells. Note that all parts of the network were subjected to the same amount of noise. The trial-to-trial presentation sequence of the different patterns was random. The training stopped after each pattern had been presented 3000 times. Hebbian learning was effective throughout learning trials, during stimulus presentation and ISIs.

2.2.2. Cell assembly definition

As mentioned, formation of cell assemblies, that is, sets of strongly and reciprocally linked distributed LTM circuits associating the paired sensory (in P1) and motor (in M1) pattern parts with each other, was observed. The pattern neurons in the “periphery” of the network architecture (areas P1 and M1, Fig. 2D) were thus linked with each other by way of cells in the more “central” areas (HP, PA, PF, PM), where sensory and motor information converges (Damasio, 1989). After the 3000 learning steps, 10 out of 12 patterns with subparts in all six areas of the network could be successfully retrieved, i.e., stimulus-specific ignition and spreading activation across all areas followed sensory stimulation (i.e., presentation of only the P1 pattern part).¹ Subsequent analyses were based on these 10 successfully learnt circuits. To identify the neurons forming each of the 10 cell assemblies across the different network areas, activity of all 3750 excitatory network cells was monitored. An excitatory cell was considered a member of a given cell assembly if and only if its output (firing rate) reached the .5 threshold (recall that a cell’s output is a continuous value between 0 and 1) within 15 time steps upon

¹ In Discussion below, we relate this limited retrieval performance of the network to features that make the network more biologically realistic (noise, random patchy connectivity etc.).

presentation of a (noisy²) sensory pattern. This threshold was chosen on the basis of simulation results obtained with the present and previous networks (Garagnani et al., 2008, 2009). Following standard definitions in the literature on auto-associative memories (see, e.g., Braitenberg, 1978; Palm, 1990), only excitatory cells were considered to be part of an assembly.³ Average cell assembly size was 210 cells, with the following average numbers of neurons in each of the areas: P1: 50; HP: 39; PA: 37; PF: 36; PM: 36; M1: 12.

2.2.3. Analysis of area, cell assembly and cell dynamics

Activity dynamics in the network was studied following sensory stimulation of P1 with learnt pattern parts, to which random noise had been added. To ascertain that noisy stimulation led to cell assembly activation, stimulus duration was set to 5 simulation time-steps. To obtain results with good signal-to-noise ratio, each of the stimuli was presented 12 times, in 12 separate “trials”, and activation time courses were averaged over “trials” for evaluation of results. For each of the 12 trials, we recorded the membrane potential of all assembly cells within the network and the total activity in each of the six cortical areas (i.e., sum of the firing rates of all cells within an area) during the 5 time-steps preceding stimulus onset and 180 following off-set. Each trial was thought to imitate the stimulation and memory maintenance part of trials of a delayed matching-to-sample (DMTS) or verbal working memory experiment, which typically last for seconds (see Fig. 1). To investigate area dynamics, average activity (total sum of firing rates of all cells averaged over trials) within each area was plotted against time. The specific membrane potential dynamics of a range of representative individual cell-assembly neurons was also plotted, to give an impression of the typical single cell response produced by the simulations. As data from real single-unit recordings are typically measured in number-of-spikes and displayed as peri-stimulus time histograms (PSTHs, Fig. 1), to compare simulated results with experimental data we also produced, for a small selection of typical cells monitored, examples of simulated spiking behaviour. This was achieved by transforming the cells’ membrane potential $y = V(x,t)$ (defined by Eq. (A.1), Appendix A) into a “spiking probability” value $PSpike(y)$. More precisely, $PSpike(y)$ was obtained by rescaling (by a constant k) the graded output $O(y)$ of a cell (representing the average firing rate of a cluster of neurons) and adding a spontaneous (baseline) firing rate probability η to it:

$$PSpike(y) = k \cdot O(y) + \eta \quad (1)$$

During the testing phase, the output of a cell $O(y)$ was computed according to the sigmoid function (2):

² During testing, area P1 was stimulated with learnt sensory pattern parts to which random noise had been added (on each trial, cells not belonging to the original pattern might be “switched on” with 5% probability).

³ As each excitatory cell in the network had its own local “twin” inhibitory cell (cf. Fig. 2F) and as connections between excitatory cell and twin inhibitors were not modified by learning, inclusion of the latter in the count would double the number of cell assembly neurons but otherwise leave the results unchanged.

$$O(y) = \frac{1}{1 + e^{-2\beta(y-\varphi)}} \quad (2)$$

where y is the cell's membrane potential, β a parameter determining the slope of the sigmoid function at the inversion point, and φ is the threshold (i.e., the sigmoid inversion point). We used $\beta = 1.5$, $\varphi = 3.5$, $k = .4$ and $\eta = .1$, thereby adjusting minimal and maximal spike rates to neurobiologically plausible values. In particular, when the model-cell membrane y approaches its resting potential (which in the model is 0) and, thus, $O(y) \rightarrow 0$, the probability of a spike occurring at any one simulation time-step approaches η (i.e., ~ 5 spikes/sec) on average. On the other hand, when $y \gg \varphi$ (i.e., $O(y) \rightarrow 1$) and the cell is firing at its maximum rate, the spiking probability $p_{\text{Spike}}(y) \rightarrow k + \eta$ (i.e., ~ 25 spikes/sec). These values are close to experimental data obtained from real single-cell recordings in awake, behaving monkeys (see, for example, Fig. 1, and Table 1 in Shinomoto et al., 2009). The resulting signal-to-noise ratio ($k:\eta = 4:1$) is also comparable with that observed in single-unit recordings from dorsolateral prefrontal (Funahashi, 2006; Funahashi, Bruce, & Goldman-Rakic, 1989; Fuster, Bauer, & Jervey, 1982) and inferotemporal (Fuster & Jervey, 1981; Naya, Sakai, & Miyashita, 1996) cortices of the monkey brain during delayed matching-to-sample tasks.

2.3. Statistical analysis

As preliminary data indicated a clear dissociation of activation patterns between an *early* phase (from 7 to 14 time-steps post stimulus onset), a *middle* (from 30 to 60 time-steps), and a *late* (from 90 to 120 time-steps) interval, statistical analyses focused on these time ranges. For statistical analysis, the number of assembly cells exceeding pre-defined membrane potential thresholds was contrasted between areas and time intervals. These thresholds were set to the values of 0, 10 and

Table 1 – Statistical results of three-way ANOVAs performed on the number of active cell assembly neurons determined at low (0), medium (10) and high (20) cut-off threshold at different time Intervals, in central versus peripheral and anterior versus posterior areas. Corresponding means and standard errors are presented in Fig. 5.

Threshold	Effect	df	F	$p <$
$\theta = 0$	Interval	2, 18	110.37	.00001
	Posterior–Anterior	1, 9	8.21	.05
	Centrality	2, 18	136.63	.00001
	Interval \times Posterior–Anterior	2, 18	114.78	.00001
	Interval \times Centrality	4, 36	112.38	.00001
	Posterior–Anterior \times Centrality	2, 18	10.43	.001
	Interval \times Posterior–Ant. \times Centrality	4, 36	169.22	.0001
$\theta = 10$	Interval	1, 9	108.49	.00001
	Centrality	2, 18	126.65	.00001
	Interval \times Centrality	2, 18	53.00	.00001
$\theta = 20$	Interval	1, 9	42.64	.005
	Centrality	2, 18	29.42	.00001
	Interval \times Centrality	2, 18	32.56	.00001

20 to cover the entire range of weaker and stronger activations. As before, results for each pattern were obtained by averaging 12 “trials” of its presentation. Statistics were performed over the 10 different patterns learnt by one network.

For statistical analysis, the data from the six areas were first entered as one six-level factor “Area”. Because a significant effect of this factor is ambiguous and could be due either to a processing difference between anterior (frontal) and posterior (perceptual) areas or to a difference between primary, secondary and higher areas, or to both of these differences, a second set of analysis was performed. In this case, the six-level Area factor was regrouped into two factors, the two-level factor “Anterior–Posterior” or distinguishing anterior/frontal from posterior/sensory areas or systems, and the three-level factor “Centrality”, which distinguished, within each systems, between primary (P1, M1), higher (HP, PM) and association areas (PA, PF). Note that, in this case, any general difference between frontal and perceptual area function would emerge as a main effect of Anterior–Posterior, any difference along the Centrality gradient, between primary and gradually “higher” and more multimodal areas, would surface as a main effect of the three-level factor, and a non-additive contribution of both factors would be manifest in a significant interaction. Two- and three-way repeated-measures Analyses of Variance (ANOVAs) were run with the two factors Interval (3 levels; *early vs middle vs late*), and Area (six levels for the six areas) and with three factors, Interval, Posterior–Anterior (2 levels; *posterior, anterior*), and Centrality (3 levels; *primary, secondary, central*). Separate ANOVAs were carried out for each of the three thresholds. As no cells exceeded the $\theta = 10$ threshold in the early interval, the Interval factor included only two levels at that threshold and above. Additional 3-way ANOVAs were also run (for $\theta = 10$ and $\theta = 20$ only) to investigate more specifically the “central” regions of the network (higher sensory/motor, sensory association/prefrontal), with factors Interval (*middle, late*), Posterior–Anterior (*posterior, anterior*), and Centrality (*secondary, central*). These analyses were performed on both raw and normalized data, obtained by dividing each area's data point (number of cells) by the maximum value across the six areas (separately for each stimulus, time interval, and threshold θ). Normalisation was performed to check whether any interactions between area and time variables were mere multiplicative scaling effects (possibly due to a larger quantity of assembly cells in one area generally) or rather whether they persisted after removing scaling differences between areas (i.e., multiplicative area main effects), thus revealing true differences in area-specific cell dynamics.

3. Results

Upon stimulation, activation was first present in area P1, where it peaked approximately 5 time steps after stimulation onset. Peaks in higher perceptual areas emerged slightly later (time-step 20) shortly followed by prefrontal and premotor areas (21–25). P1's activity showed a second, smaller, peak (23–28), which emerged together with activity in M1. Whereas persistent activity was (present but) at a low level in the periphery (P1 and M1; total area firing rate < 15), more “central”

areas, that is, higher perceptual/motor and association cortex, showed strong lasting activation (firing rate > 20), which dropped to lower levels towards the end of the observation period. Comparing activity levels between areas, association areas PA and PF seemed to show strongest memory activation towards the end of the observation period. Fig. 3 illustrates the network response to stimulation of area P1 with one of the learnt sensorimotor patterns. The other nine learnt patterns produced qualitatively similar results.

Fig. 4 plots examples of single cell responses taken from the different areas during stimulation of P1 with the same pattern used to produce the data plotted in Fig. 3. As the total number of cell assembly neurons in given areas varied between 10 and 50 (see Methods), one randomly chosen cell was plotted for every seven cell assembly neurons to give an impression of the range of responses obtained. To facilitate comparison with existing neurophysiological data (as in Fig. 1), for each area the action potential frequency of one of these cells is plotted in a PSTH. Single cell plots confirm predominance of long-lasting memory activity in association areas PF and PA throughout the observation period, and more short-lived memory activity in secondary areas PM and HP. Examples of the typical, more or less long-lasting responses in higher and association areas are illustrated in the PSTHs. Primary areas P1 and M1 did not produce large numbers of cells exhibiting long-lasting sustained activity, although a small number of memory cells (2 in P1 and 3 in M1) could also be identified in the sample selected. However, P1 showed very consistent stimulus-elicited early activation in the majority of its cell assembly neurons.

Results consistent with the observations from Fig. 4 were revealed by statistical tests performed on data obtained over all patterns and cell assemblies. Fig. 5 further quantifies and summarizes these results on the area-specificity of dynamic activation patterns by illustrating the outcomes of statistical analyses. The analyses took advantage of activation in all 6 areas and all 10 cell assemblies elicited by presentations of 10 different sensory pattern parts (see Methods). Each single response was determined by averaging over 12 replications of the same pattern part presentation. The two-way ANOVAs showed significant interactions of the factors Interval \times Area regardless of which activation threshold was chosen, thus confirming that responses differed between areas and time intervals ($\theta = 0$: $F(10,90) = 130.85$, $p < .0001$; $\theta = 10$: $F(5,45) = 21.57$, $p < .0001$; $\theta = 20$: $F(5,45) = 17.05$, $p < .0001$). For the three-way ANOVAs, slightly different results emerged for the different thresholds. When taking into account all active neurons ($\theta = 0$), a significant interaction of all three factors, Interval \times Posterior–Anterior \times Centrality, resulted, which was due to general enhancement of the number of active neurons in the central layers compared with activity in peripheral areas, with the only exception of particularly strong activity in the early interval in primary perceptual area ($F(4,36) = 169.22$, $p < .0001$; Fig. 5, left-bottom panel), which by far surpassed that of all other areas and time intervals ($p < .0001$). Memory activity thresholded at $\theta = 10$ demonstrated a significant interaction of the factors Interval \times Centrality ($F(2, 18) = 53.0$, $p < .00001$), without any effect of the Posterior–Anterior variable, thus suggesting that responses in frontal and posterior cortex were similar to each

other but the distance from primary cortex led to a change in the number of memory-active neurons: There were significantly larger numbers of active cells in the four central (PA, PF, HP, PM) areas compared with the peripheral ones (P1, M1), along with a reduction of activity with time in the four central areas only. Interestingly, this interaction was still significant when the test was repeated on the raw and normalised data from the four central areas only ($F(1,9) = 24.55$, $p < .001$), indicating that the number of cells reaching criterion dropped more quickly (i.e., cells lost activation more rapidly) in the secondary (HP, PM) than in the association (PA, PF) areas (see Fig. 5, middle diagram). This means that long-lasting memory-activity predominates in these association areas. Finally, the most strongly active cells that remained after application of a threshold criterion of $\theta = 20$ confirmed another Interval \times Centrality interaction (Fig. 5, top-right); however, this interaction did not survive normalisation in the analysis of the four central areas only. Instead, main effects of Interval (middle vs late) and Centrality (secondary vs central) confirmed that cells responding very strongly to the stimulus were generally more numerous during the middle than late time windows ($F(1,9) = 22.76$, $p < .01$), and more frequent in the central (PA, PF) than in the secondary (HP, PM) areas ($F(1,9) = 55.90$, $p < .0001$).

These results show that the most active memory cells were found predominantly in prefrontal and higher association cortices, and that this strong activity was long-lasting almost exclusively at these loci. Table 1 lists all significant main effects and interactions.

4. Discussion

A neurocomputational model mimicking the neuroanatomical area structure of sensory, motor and adjacent multimodal brain systems, the known connectivity within and between these areas along with biologically-inspired physiological and learning principles was used to model and explain the topographical specificity of perception processes and working memory in the human brain. No *a priori* assumptions were incorporated in the model other than 1) knowledge about connectivity structure within and between cortical areas, and 2) neurophysiological principles, as manifest in mechanisms of neuronal activation, deactivation and synaptic plasticity. The model successfully replicated both the early predominance of sensory-evoked activity in primary perceptual cortices and, crucially, the local specificity of persistent working memory activity primarily seen in areas distant from the loci of sensorimotor activation. Persistent memory-related activity was strong in these “central” areas of the network – in higher perceptual and premotor, as well as in prefrontal and perceptual association areas, the “higher” multimodal convergence zones of the model – but was rare, although occasionally observed, in primary cortices, both on the perceptual and motor sides. Critically, strong long-lasting “memory” activity was present almost exclusively in areas that are, in terms of connectivity and localisation in the brain and in the model, distant from sensory input and motor output, namely in the prefrontal and higher-perceptual temporal areas of the network. Converging results emerged from

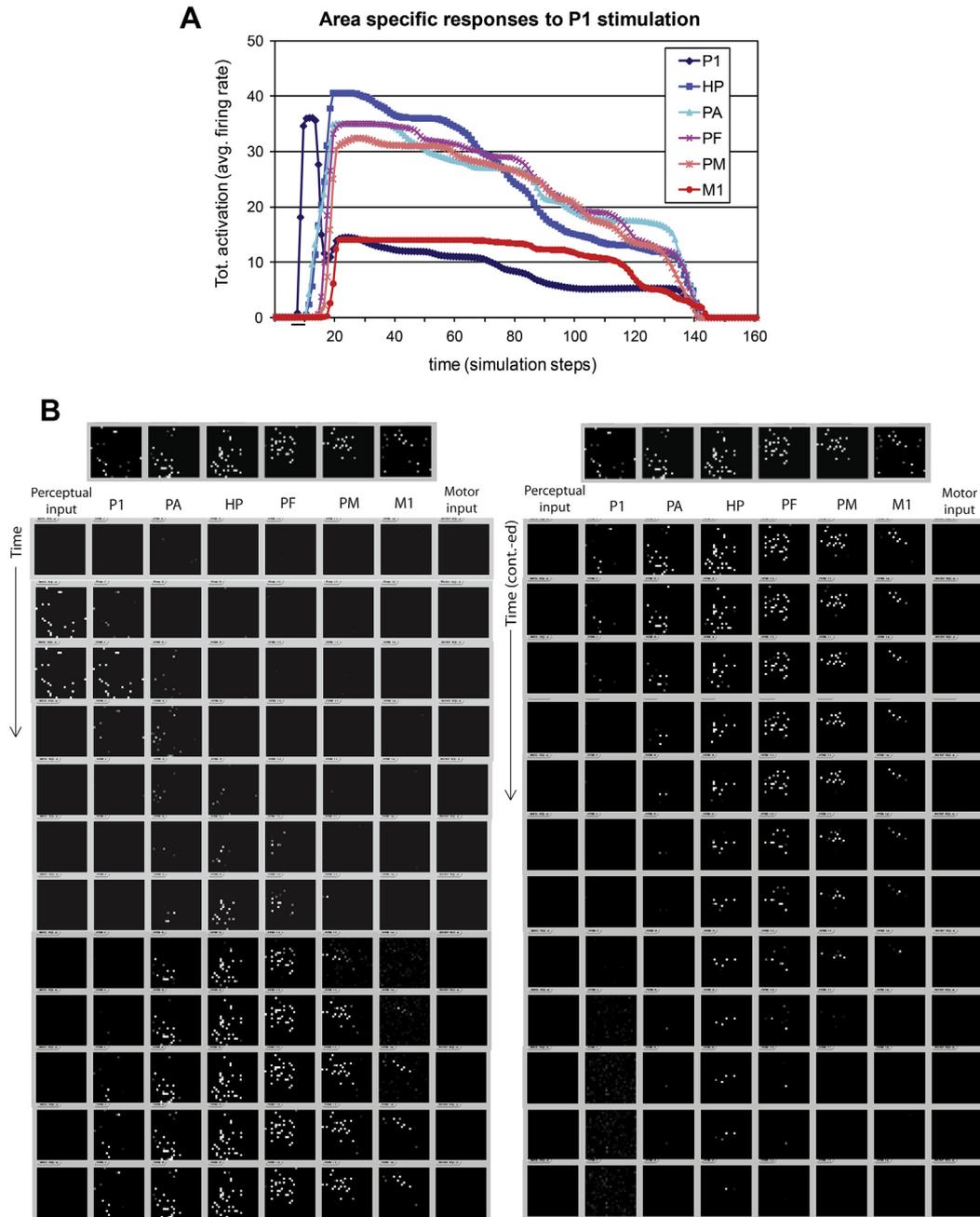


Fig. 3 – Memory dynamics in the “areas” of the neurocomputational model. (A) Example of a cell assembly, which is being stimulated in area A1 (time step 6–11), then ignites (time steps 10–22), subsequently exhibits reverberant working memory activity in all “higher” areas (HP, PA, PF, PM) and finally deactivates (time steps 130–140). Area P1 was stimulated by a previously learnt sensory pattern. For each area, the averaged cumulative within-area firing rate (summed over all cells of the area and averaged over 12 repeated presentations), is plotted against time. Stimulus onset and offset are indicated by the small horizontal segment. Note the stronger sustained response of the four central model areas (HP, PA, PF, PM) compared with that of the two peripheral (sensory and motor) areas (P1, M1). In area P1, an early response (peaking at around 12 time-steps), driven by the presence of the stimulus in the sensory input, predominates, whereas in other areas activity tends to peak slightly later and fall off gradually. **(B)** Activation in the six areas of the network to stimulation of one previously learnt cell assembly. Areas are shown from left to right; the leftmost columns show stimulation to A1; white dots indicate active neurons. Temporal dynamics of one single cell assembly stimulation (top left), ignition (middle and bottom left), subsequent reverberation (top right) and deactivation (bottom right). To facilitate comparison, the structure of the cell assembly is shown (twice) at the top left and right. Note that most cell assembly neurons are active during ignition and activity is maintained longest in the “central” areas PF and HP.

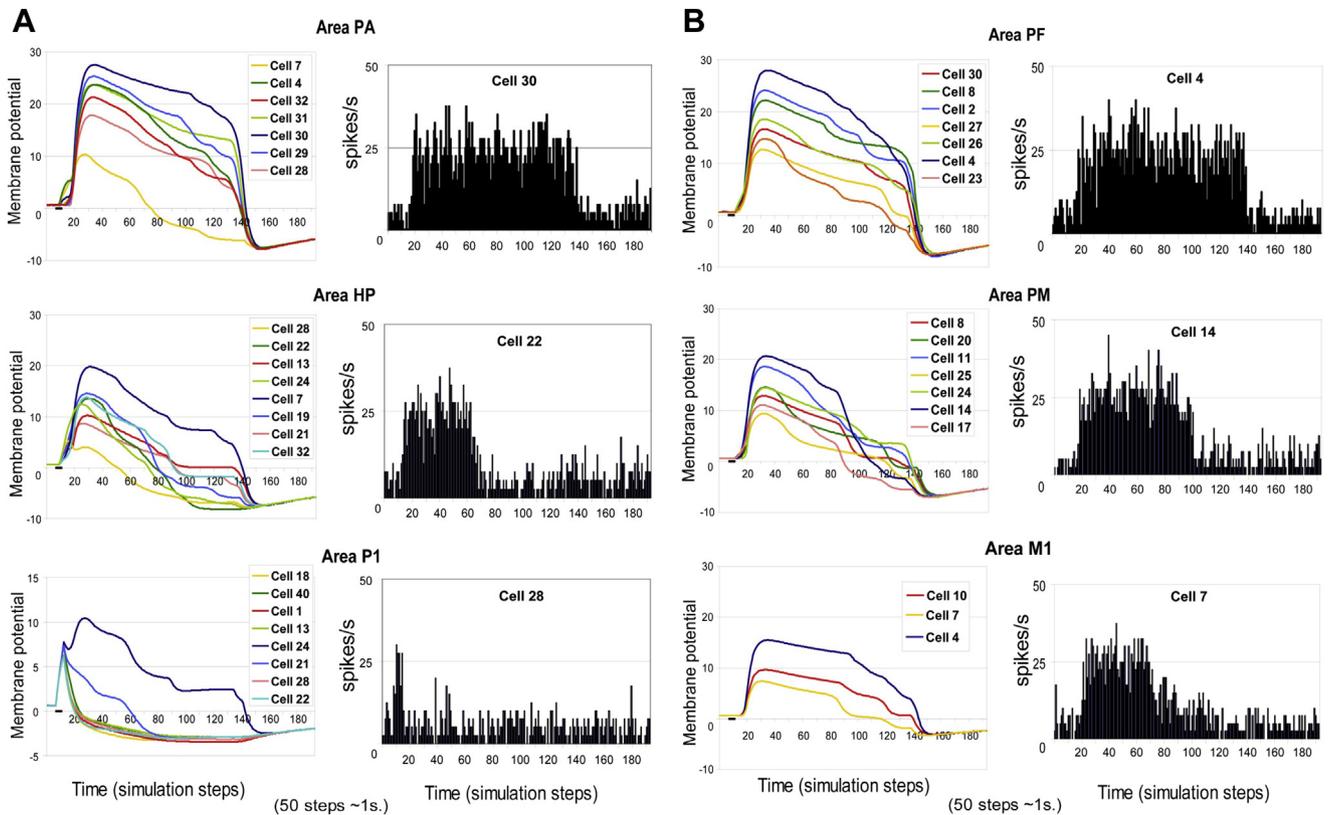


Fig. 4 – Simulated single-cell responses. Figure part (A) presents cell responses from the three sensory areas, and Figure part (B) for the three motor areas. All responses are averaged over 12 trials following stimulation with the same sensory pattern. Areas central to the network architecture, i.e., PF and PA, are at the top, followed by successively more “peripheral” areas, i.e., PM and HP, and finally M1 and P1 at the bottom. Left panels: the average membrane potentials of cells responding strongly to P1 stimulation with a given learned pattern are plotted against time. Right panels: The peristimulus time histograms, or PSTHs, are plotted for one representative example cell from the corresponding panels on the left. Firing rates are computed using Eq. (1) in Methods. Note transient activation in P1 and longer-lasting activation in all other areas. Note furthermore, that longest-lasting strongest neuron responses are seen in “central” areas PF and PA. Some cells in “peripheral” areas of the network, P1 and M1, also exhibit moderate sustained responses, but these are rare.

simulated area-responses, cell assembly dynamics and single cell responses. These features replicate important results from single cell neurophysiology and large-scale neuroimaging (D’Esposito, 2007; Fuster, 2009), as we discuss in more detail below. Our results may help explain why neuronal activity reflecting working memory is so frequently seen in multimodal higher association cortices albeit the correlated patterns of cortical activation driving memory formation are present in primary areas.

4.1. The cortical topography of active memory is explained by neuroanatomical connectivity structure

Between-area connectivity of the model mimicked neuroanatomical connectivity between cortical areas as revealed by tracer and DTI and DWI studies (for details, see Methods). A key observation is that the connectivity structures of visuo-motor and auditory-motor systems show important parallels. Therefore, the same network architecture was used to simulate these systems and the results of corresponding memory experiments. Results from two specific experimental

domains were addressed: typical visual DMTS tasks – as they are carried out in animal experiments – and verbal working memory processes for spoken language. In both cases, posterior sensory (visual vs auditory) and frontal areas with similar between-area connectivity structure (Fig. 1) are relevant. In a typical visual DMTS task, a visual stimulus has to be memorised in view of a specific action (typically a button press) to be performed after a delay. In verbal working memory experiments, a spoken meaningless pseudoword or meaningful word, or series of such items, have to be kept in mind for later verbal–articulatory reproduction. In both cases, sensory and motor systems carry input patterns that had previously been associated with each other through learning. In both cases, stimulation is to primary sensory cortex. In both cases, however, experimental research documents that memory-related brain activity is present predominantly in prefrontal cortex regardless of stimulus type, and in sensory association cortex, especially in anterior temporal lobe in visual DMTS tasks and posterior superior and middle temporal cortex in verbal working memory, and to a lesser degree in primary sensory areas (D’Esposito, 2007; Fuster, 2009; Linden,

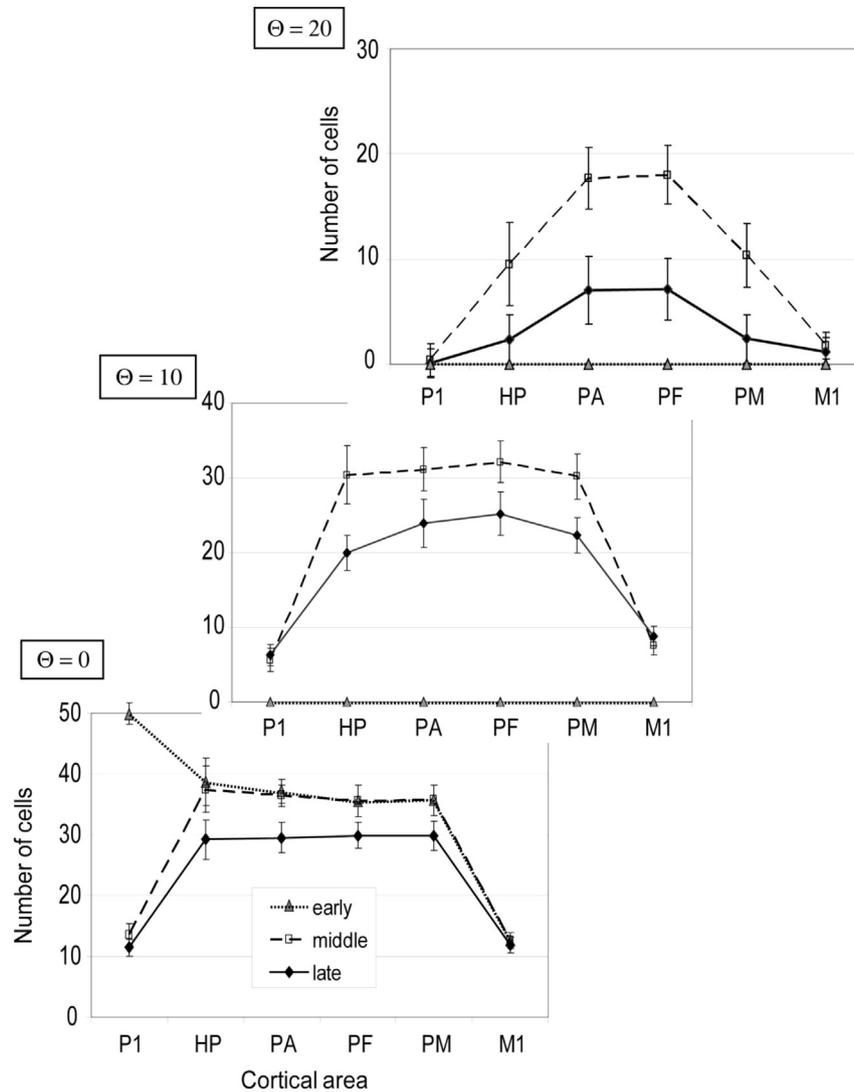


Fig. 5 – Cortical distribution of cells and activation profiles. For each model area, the number of active cells is plotted for different time intervals after stimulation, an early (time steps 7–14; triangles), middle (30–60; squares), and late one (90–120; diamonds). Results are plotted for three different activation criteria, applying thresholds to each cell’s average membrane potential, for $\theta = 0$ (bottom-left), $\theta = 10$ (central), and $\theta = 20$ (top-right). Means are plotted for ten different previously learnt patterns. Error bars give standard errors of the mean (SE). Note the predominance of early activation in P1 (bottom-left panel, left triangle) and predominance of strong long-lasting activity in central association areas, in PF and PA (top-right panel, diamonds).

2007; Postle, 2006). Our model reproduces this between-area shift and the resultant local dissociation between early sensory-evoked and later working memory processes. Results on differential area activation in early perceptual and subsequent memory processes were bolstered statistically using ANOVAs. Statistics confirmed three important points: (1) that early strong activation occurred in modality-specific sensory systems, (2) that subsequent persistent activity emerged predominantly in “higher” or “central” areas of the networks (including premotor, prefrontal, higher perceptual and perceptual association cortex) distant from the primary sensory input and motor output areas, and, most importantly, (3) that strongest and most long-lasting memory responses were specific to multimodal areas, in the model equivalents of prefrontal cortex and temporal association cortex. These

results hold true for different sensorimotor systems, thus indicating a degree of generality of the “upward shift” of memory activity away from modality-specific primary and towards multimodal “higher” areas.

After learning, stimulation with a learnt perceptual pattern elicited activity in all areas of the network, with gradual differences between areas. Fig. 3A shows that around time steps 5–15, sensory activation in P1 dominated the network response. Note that, although one may want to characterise this early activation as purely “perceptual”, it cannot be explained by sensory stimulation alone. In fact, even activity in P1 persists a few time steps after external stimulation ceased, providing evidence of memory maintenance. However, this perceptually-evoked activity is short-lived and degrades towards time step 15 (blue curve in Fig. 3A).

Interestingly, at time steps 15–20, activation now peaks in all areas of the network near-simultaneously (bottom left of Fig. 3B). This is the almost simultaneous “explosion-like” cell assembly activation process sometimes called *ignition* in the literature on brain theory (Braitenberg, 1978; Dehaene & Changeux, 2011; Friston, Breakspear, & Deco, 2012; Pulvermüller, 1999; Wennekers et al., 2006). Ignition can be related to LTM retrieval and, in our present simulations, reflects the rapid spreading of activation within a given learnt cell assembly that links together specific sensory and motor patterns. Therefore, these assemblies can be characterised as *action-perception circuits* (APCs) and can be considered the basis of LTM (Braitenberg & Schüz, 1998; Fuster, 1995; Hebb, 1949; Pulvermüller & Fadiga, 2010). Ignition-related activation is strong in the “central” areas and weak in the primary cortices. Upon ignition, activity vanishes gradually, but tends to last longest in prefrontal and higher perceptual areas, the multimodal convergence zones of the model (bottom right of Fig. 3B). These dynamics of perceptual activation, ignition and reverberation provide a putative neuronal model of the cognitive processes of perception, recognition and LTM access, and persistent working memory of/for familiar stimuli for which individuals have corresponding actions in their learnt action-schema repertoire. In a loose sense, all of the processes that occur after external stimulation has ceased can be considered to reflect active memory, that is, the functional consequences of activity maintenance in circuits of neurons, which emerged spontaneously via long-term changes in synaptic weights brought about by correlation learning. Persistent active memory is seen in most auto-associative and attractor networks (e.g., Hinton & Shallice, 1991; Verduzco-Flores et al., 2009).⁴

The explanation of the results of our simulations of working memory dynamics must rely on the between-area anatomical connectivity structure of the model network. Other explanations are difficult to maintain, as within-area connectivity structure and all physiological properties did not change across areas. In other words, specific “areas” of our model corresponded to real cortical areas in the human cortex insofar as they were given a similar inter-area connection structure, and therefore showed similar connectivity to those areas that received/produced input/output, as real cortical areas. The long-distance between-area links incorporated in this model were justified in light of tracer studies in monkeys and DTI work in human subjects (see Methods, Section 4.3 below, and, for example, Catani et al., 2003; Kaas & Hackett, 2000; Makris & Pandya, 2009; Pandya & Yeterian, 1985; Petrides & Pandya, 2009; Rizzolatti & Luppino, 2001; Wakana et al., 2004; Young et al., 1995). According to these data, the motor and sensory cortices are not directly linked by way of strong fibre bundles, but there are massive connections between adjacent areas, second-next areas, and additional links between frontal and temporal association cortices. In addition, recent work has shown links from prefrontal to auditory belt and inferior temporal areas, and between a range of superior and temporal pole areas

with premotor cortex. All of these well-established links were realised in the present model. These connections result, in a cumulative manner, in a high number of long-distance afferent and efferent connections in the prefrontal cortex and the temporal association areas (note the four arrow heads present in each of these areas in Fig. 2D), slightly less cortico-cortical links to/from premotor and higher perceptual cortices (three arrow heads), and a relatively low number of incoming and/outgoing connections in primary cortices. It is this connection structure that influenced the inner structure of the LTM circuits formed and therefore their functional dynamic activation and maintenance of activity. Our simulations therefore suggest that the most richly connected areas provide the connectivity basis for the most strongly and most persistently active memory cells. Because of the many synaptic links converging onto, and diverging from, prefrontal and temporal association cortices, cell assembly neurons located there are likely to be reciprocally and strongly connected to a high number of other assembly cells, and, therefore, to excite (as well as receive feedback from) such fellow circuit members during cell assembly activation and activity maintenance. As a result, activity within these neurons is most persistent. These cells represent the inner “core” of the cell assemblies, whereas neurons in primary (and intermediate) areas are more likely to become part of the assembly “halo” (Braitenberg, 1978), which is characterised by smaller numbers of (and therefore overall weaker) links to other cell assembly member-neurons. After sensory stimulation and consequent cell assembly ignition, activity gradually vanishes from the circuit’s neurons, first in the halo and then, gradually, also from neurons towards the more central circuit parts. As a result, memory activity gradually “retreats” to the inner core, which most heavily draws on prefrontal and higher perceptual areas. This mechanism of *memory retreat to cell assembly cores* together with neuroanatomical inter-area connectivity explain why memory cells are so common in higher multimodal association cortex and much less so in primary areas.

Based on the analogy proposed here between the inter-area connectivity structure of auditory–articulatory and visual–hand motor cortical systems, the explanation of memory cell topography in cortex, which we lay out above, can be applied to two different sensorimotor brain systems sometimes classified as being part of the ventral and dorsal stream, respectively. Only the ventral “what” stream of visual processing was simulated in the visual working memory model, assuming that for other memory processes capitalising on the location of visual stimuli in space, other more dorsal (“where” stream) areas in the parietal lobe might be relevant (Jeannerod, Arbib, Rizzolatti, & Sakata, 1995; Mishkin et al., 1983). In the speech-language domain, two fibre bundles connect posterior and anterior perisylvian regions, the dorsal arcuate fascicle and the ventral internal capsule (Makris & Pandya, 2009; Petrides & Pandya, 2009; Rilling et al., 2008; Saur et al., 2008). The arcuate in part carries “ventral” connections between *middle temporal* gyrus and prefrontal cortex (Glasser & Rilling, 2008), and the capsule has even been attributed a primary role in the ventral stream (Saur et al., 2008). Likewise, both of these connection highways link anterior, lateral and posterior *superior temporal* cortex with inferior

⁴ Note that, in the present network structure, activity ceases immediately if a naïve network is stimulated before any learning has taken place (see also Section 4.2 for further discussion).

prefrontal and premotor areas as part of the “dorsal” speech-language processing stream (Kelly et al., 2010; Petrides & Pandya, 2009; Rilling et al., 2008; Saur et al., 2008). Our model does not explicitly distinguish between arcuatus and capsula connections as part of “dorsal” superior temporal-to-inferior frontal connectivity, although such differentiation might lead to a worthwhile elaboration of the model in future. Crucially, the processes of memory retreat emerging in the present simulations are equally applicable to the higher convergence zones of the ventral-stream’s anterior inferior temporal areas (Fuster, 1995) and to those of the dorsal-stream’s anterior, lateral and posterior superior temporal areas of the auditory belt and parabelt (Kaas & Hackett, 2000; Romanski et al., 1999) and their ventral and dorsal frontocentral counterparts. Considering the anatomical differences between visual and auditory systems, the move away from early sensory cortex therefore results in an anterior-temporal shift in the ventral visual stream (Fuster, 1995; Fuster & Jervey, 1981), but, in contrast, in the auditory system, in a move away from primary cortex in anterior (towards temporal pole), lateral (to middle temporal sulcus and gyrus) and posterior (to temporoparietal junction) directions, which may be modulated by task and stimulus materials (see, for example, Acheson, Hamidi, Binder, & Postle, 2011). Future theoretical, neurocomputational and neuroimaging studies are necessary to elaborate, evaluate, and eventually explain such more fine-grained predictions and implications.

As mentioned, our neurocomputational study targets both long-term structural and short-term working memory. With this model, we assume that LTM mechanisms, especially the structural synaptic changes that underlie the formation of strongly connected distributed neuronal circuits, provide the basis for working memory. Nevertheless, this functional relationship between LTM and working memory should not lead one to equalise the two (see also Fuster, 1995). As our simulations show, working memory activity gradually retreats to “central” core parts of the LTM circuit, so that, although LTM provides, in a sense, the underpinning for longer-lasting working memory, the two may partly dissociate (Linden, 2007). The network implements similar memory processes in frontal and posterior areas, thus suggesting a representational rather than modulatory role of prefrontal cortex in working memory. Note that there is a dispute about this issue in the working memory literature. Some data suggested memory impairment following temporary prefrontal lesion (for example, Bauer & Fuster, 1976; Fuster, 1995; Fuster & Bauer, 1974; Owen, Sahakian, Semple, Polkey, & Robbins, 1995; Wheeler, Stuss, & Tulving, 1995), whereas other works report still intact memory after focal unilateral prefrontal damage, thus speaking in favour of a role of this region in modulating and controlling working memory (for example, Petrides, 2000). A possible integration of these results allows for a representational role of prefrontal cortex but admits that widely distributed prefrontal areas, possibly in both cortical hemispheres, can contribute, so that only large bilateral prefrontal lesions may cause a deficit (see D’Esposito, Cooney, Gazzaley, Gibbs, & Postle, 2006). Although addressing this issue was outside the scope of this work, it is possible to use the present model to simulate cortical lesions and help shed more light on the role of prefrontal cortex in working memory and attention.

4.2. Biological features and network performance

As mentioned in the Introduction and Methods sections above, our model aims at biological realism in several respects. Its complex area structure replicates six areas of pre-specified sensorimotor brain systems along with the links between them. The specific between-area connection structure shown in Fig. 1D, obtained from the evaluation of neuroanatomical studies in humans and macaques (see Methods), is replicated, along with the more general established properties of long-distance cortico-cortical projections, which are known to be sparse, random, initially weak and topographic. Real cortical within-area connectivity is mimicked in the model insofar as it emphasises local links, and, once again, is sparse, random and initially weak. The addition of local inhibition, by means of which each neuron activates a “twin” local feedback-inhibitor cell when active, and, in addition, of a more global area-specific regulation mechanism represent further network features inspired by cortical anatomy and physiology. On the functional side, a biologically-motivated Hebb-type rule for synaptic plasticity and learning realising both LTP and LTD was implemented. In addition, the constant presence of uncorrelated white noise in all parts of the network and, furthermore, of noise overlaying any sensory input during retrieval are further features that make the simulations more comparable with real brain activity and real life perceptual input.

The biologically inspired features of the model make it less efficient – from an “engineering” perspective – than recurrent or attractor networks that prioritize functionality at the expense of biological realism. For example, full all-to-all connectivity between layers and within memory layers as it is implemented in standard recurrent networks (e.g., Elman, 1991) makes it easy to associate (and recall) the different components of a stimulus pattern. In contrast, the more biologically realistic sparse connectivity implemented in the present work makes it harder for the specific neurons co-activated by a given pattern to link up with each other. This is because, in cases where direct links between pattern neurons are missing, indirect links need to be established, which requires interlinking additional neurons and therefore additional training. Networks with only one (fully connected) “hidden” layer (area) can efficiently store patterns and might seem sufficient to simulate memory (e.g., Rogers et al., 2004); however, it is biologically unrealistic to define a-priori one cortical area as the only site of memory, as auto-associative within-area connections characterise all parts of cortex, primary areas included (Braitenberg & Schüz, 1998). Thus, using a network structure with several auto-associative “memory layers” is more realistic when the phenomena of interest are known to engage a range of cortical areas. However, the more complex area structure comes with the need to modify more, and more indirect, connections, which increase the risk of retrieval errors. Likewise, the backpropagation learning rule applied in many neural network simulations is optimised for efficient learning across an entire network, whereas the Hebb-type rule that captures important aspects of neocortical plasticity requires strengthening of connections in the periphery of the network before synaptic changes may occur in

its centre; simulation time and learning efficiency, therefore, are far from being optimized here. Finally, the addition of perceptual and brain-generated noise decreases the signal-to-noise-ratio, rendering learning and retrieval more error-prone. In sum, our approach trades effectiveness for biological realism: whilst the former aspect was of no interest here, the latter plays a vital role in providing a novel explanation of area-specific cortical dynamics of working memory and other specific predictions on cognitive brain processes (see also Garagnani et al., 2008; Garagnani & Pulvermüller, 2011; Garagnani & Pulvermüller, 2013).

Given the biologically inspired nature of our models, it should not come as a surprise that the six-area network performs less well than standard neural networks with regard to storage and retrieval. In this context, it needs to be mentioned that only 10 out of the 12 patterns were learnt to criterion within the predefined learning phase. Given the network's 3750 neurons, this may seem as poor performance, as 375 neurons per stored pattern is far below standard associative memories' capacity (e.g., Palm, 1980). Nevertheless, one should bear in mind that vocabularies of ca. 40,000 words (Pinker, 1994) are normally not fully and errorlessly stored by the human cortex, which includes ca. $2 \cdot 10^{10}$ neurons (Pakkenberg & Gundersen, 1997), thus resulting in some 500,000 neurons per stored pattern.⁵ Therefore, after all, the "inefficient" network's performance appears not to be substantially below that of real brains, and the high neuron-per-pattern ratio may eventually turn out to be an additional biologically realistic feature. We should add, however, that the network's seemingly low memory capacity is not a necessary feature of this architecture; for example, we were able to successfully store up to 30 distributed patterns (each with neurons in each of the six areas) in such a model (ca. 100 neurons per pattern). While further work on optimising network's memory storage capacity and retrieval is certainly possible, this was not the aim of the present investigation.

As any model, the present implementation had to make simplifying assumptions. For example, to keep computational time manageable, we limited the overall network size (3750 neurons) and adopted a mean field approach; moreover, during the learning stage, the cells' transformation function was replaced by a computationally less costly version. While developing a complementary spiking neurons model with a larger number of cells might be desirable in the future, we should note that mean field models correctly describe the average behaviour of more realistic (but computationally more demanding) spiking networks (see, for example, Deco et al., 2013), and provide the sufficient level of complexity and realism required for the phenomena of interest here.

Although the connection structure of the network was motivated by neuroanatomical research, we note again that this structure implies abstracting away from and omitting some neuroanatomical detail (see *Methods*); therefore,

⁵ Note that these relationships change with different assumptions. Assuming that each artificial "neuron" of our model corresponds to a local neuronal cluster of ca. one thousand real cortical neurons, the network's (real-) neuron-per-pattern ratio becomes $3.75 \cdot 10^5$

advancing the model by including further brain-structural detail represents a fruitful perspective for the future. A further obvious limitation of our work lies in the limited set of areas (six in the present study); modelling additional potentially relevant areas (for example, in parietal cortex) may help us understand whether/how these sites may also play a role in working memory. In this sense, the "brain-part simulations" we offer do not exhaustively reveal the range of areas relevant for working memory. Extending the network would also lead to the addition of connections to and from the currently simulated areas, so that even the primary areas might increase their number of input and output connections. However, the *difference in connectivity between relevant primary, secondary and multimodal areas*, which is so essential for our results and explanation, seems manifest even when looking at large connection matrices obtained for whole brains. For example, one set of connectivity matrices (Sporns & Zwi, 2004) shows relatively low numbers of connections of primary perceptual areas V1 and M1 (BA 4) compared with the higher perceptual areas MT or the dorsolateral prefrontal area (BA 46). Therefore, it seems that the relatively richer connectivity that multimodal "convergence" areas exhibit when compared to primary ones, which is immanent to the present network structure, may indeed be a common feature of (smaller and larger) real anatomical networks. Furthermore, the selection of areas included in our model was made so as to allow for the most direct pathways between sensory and motor information (3 synaptic steps from P1 to M1), and therefore the addition of further areas with more indirect sensorimotor links (>3 steps) may not substantially change the results, because more indirect links are less likely to make important contributions to sensorimotor assembly formation.

4.3. Relation to pre-existing neurobiologically-inspired computational work

The present results build upon and extend a range of previous simulations of memory circuits which already implemented features of cortical anatomy and function (e.g., Bussey & Saksida, 2002; Deco & Rolls, 2005; Elman, 1990; Farah & McClelland, 1991; Knoblauch & Palm, 2002a, 2002b; Palm, 1987; Patterson, Nestor, & Rogers, 2007; Sommer & Wennekers, 2001; Verduzco-Flores et al., 2009; Willshaw & Buckingham, 1990; Willwacher, 1976; Zipser et al., 1993). We focus here on the question of *why* memory processes arise predominantly in specific areas of cortex, and answer this question on the basis of intrinsic neuroanatomical structure.⁶ This new neuroanatomically-grounded explanation for the specificity of the cortical distribution of memory cells shows that the network memory account (Fuster, 1997) can address a major question previously left unanswered. The present proposal, namely, that the topography of memory cells results from cortico-cortical connectivity structure, is not incompatible with other models of working memory physiology, for example at the single cell level in terms of extracellular calcium

⁶ Needless to say, there is a range of still open questions about the brain basis of working memory which our model does not address.

concentration (Mongillo et al., 2008), but points to the need for such accounts to explain why neurons in some areas show stronger and more long-lasting memory activity than those in other areas. Independently of the possibility to generate alternative accounts, the fact that the network memory theory can explain the local dissociation of perceptual and working memory processes using “embodied” APC formation and corticocortical connectivity structure increases the explanatory power of this account and hence further strengthens it.

We simulated memory cells that activate and subsequently lose activity with time. Sometimes, the neuron behaviour was almost bistable, but typically there was activity loss with time. Activity loss was sometimes faster, especially in perceptual areas, sometimes short-lasting, typically in secondary areas (premotor and secondary perceptual), and frequently long-lasting and slowly decaying, especially in “higher” prefrontal and perceptual association cortex where sensory and motor activity converged. The range of memory cell dynamics observed experimentally is even wider than the types documented in our simulations, also including cells that gain activity with time and neurons showing lasting inhibition (for an overview, see, Fuster, 1995). We did not focus on such additional cell responses but would like to remark that, for example, tonically inhibited neurons below the level of spontaneous activity can be seen in the cells selected from area P1 (Fig. 4A, bottom). Further cell dynamics may emerge when scrutinising not only excitatory cells that belong to cell assemblies – on which our current investigation focuses – but, in addition, inhibitory and excitatory cells adjacent to cell assembly neurons (which receive strong local inhibitory input through inhibitory cells and potentially some additional excitatory input) as well. Furthermore, use of a more complex neuron model with detailed implementation of cellular or synaptic mechanisms may lead to networks exhibiting an even richer set of cell activation dynamics. Because the replication of the full range of experimentally observed firing behaviours was not the focus of the present study, it is important to note that a recent simulation study using analogous attractor networks of spiking neurons spanning different cortical regions showed that “the multiple pattern types exhibited by cells in working memory networks are inherent in networks with dynamic synapses, and that the variability and firing statistics in such networks with distributed architectures agree with that observed in the cortex” (Verduzco-Flores et al., 2009).

There is a growing literature on neurocomputational models of language and cognition, including concepts and their relationships (for review, see Stramandinoli, Marocco, & Cangelosi, 2012; Wennekers et al., 2006; Wermter et al., 2009). As mentioned, most of these models use standard architectures such as variants of the perceptron (e.g., Farah & McClelland, 1991) or simple recurrent networks (Elman, 1991), which are not easily linked to concrete neural structures, let alone specific cortical areas and their patterns of connectivity. More recent approaches have used a wider array of area-like “layers”, which show some parallelism to brain structures. For example, Rogers and colleagues (Rogers et al., 2004) implement separate layers for the processing of visual features, perceptual, functional, linguistic and encyclopedic knowledge, and semantics, which they liken to occipital visual, perisylvian language, and anterior-temporal conceptual

cortex. However, as already mentioned, recurrent connections, which yield memory activity, are only implemented in their semantic (anterior-temporal) layer, so that the predefined network architecture determines the locus of memory. This model does not provide, or attempt at, a neuro-mechanistic explanation of the question of why the anterior-temporal lobe becomes so important for semantic memory, but, instead, implements this *a priori*. In contrast, our present simulations explain why the anterior temporal lobe (along with other higher convergence zones) becomes a hub of semantic memory activation and storage. Similarly, Plaut presents a brain-based model of optic aphasia (Plaut, 2002), which includes perceptual and motor areas directly linking into a semantic convergence layer. This model is used for lesion studies but could, with appropriate extension and inclusion of additional areas, also be applied, similar to our own, to target area-specific contributions to working memory processes. When comparing our own model to previous ones, main differences lie in the richer selection of areas linking perception to motor regions and the greater emphasis on neuroanatomical and neurophysiological features (see Section 4.2).

4.4. Embodiment and disembodiment: towards a mechanistic integration of cognitive theories

Over and above its contribution to an explanation of the cortical topography of memory cell activation, the proposed mechanistic neurocomputational framework may help solving the current dispute in cognitive science between proponents of the embodied grounding perspective on symbolic, conceptual and semantic processing (Barsalou, 1999; Glenberg, 1997; Lakoff & Johnson, 1999; Pulvermüller, 1999) and the contrarian cognitivist position attributing these processes to a disembodied and “amodal” system (Bedny & Caramazza, 2011; Fodor, 1983; Mahon & Caramazza, 2008). Proponents of embodiment have accumulated evidence for the activation of sensorimotor systems in symbol and concept processing and for a causal role of sensorimotor systems for speech perception and semantic understanding (Barsalou, 2008; Kiefer & Pulvermüller, 2012; Meteyard et al., 2012; Moseley et al., 2013; Pulvermüller & Fadiga, 2010). For example, support comes from the activation of specific sensory or motor brain regions to words with specific meaning or articulatory structure, and from cognitive-symbolic defects arising from lesions in these same sensorimotor areas (e.g., in motor cortex to hand-related words or auditory areas to sound-related words, Kemmerer, Rudrauf, Manzel, & Tranel, 2012; Kiefer & Pulvermüller, 2012; Neiningner & Pulvermüller, 2003; Trumpp, Kliese, Hoening, Haarmeier, & Kiefer, 2013). The disembodied perspective views symbolic processes to be located in an “amodal system”, which is probably best identified with multimodal cortical areas. This view is bolstered by the activation and importance of brain areas with seemingly general semantic and conceptual roles, especially in semantic memory, including prefrontal, high-parietal and anterior-temporal cortex (Binder & Desai, 2011; Bookheimer, 2002; Patterson et al., 2007). This apparent discrepancy resulted in a major dispute between the embodied and disembodied schools, where the most important symbolic and semantic processes and representations are attributed to “higher”

multimodal areas by one camp and to sensory and motor fields by the other. Of special relevance is the fact that it had been suggested that a role of multimodal (or “amodal”) areas in symbol or concept processing is inconsistent with the embodied action-perception perspective (Bedny & Caramazza, 2011). We here show that this is not correct: (1) Learning of action perception patterns leads to the formation of LTM circuits that include sensory and motor neurons along with additional neural elements in convergence zones bridging between the two. (2) The functional dynamics of these circuits are such that, after their activation, activity persists most reliably and most extensively in the multimodal areas. This investigation shows that the action perception perspective rooted in brain structure and function integrates and explains both embodied activation of sensorimotor systems in perception and memory access as well as disembodiment and retreat of activation to multimodal areas during working memory. Therefore, by integrating and explaining both embodied and disembodied processes in one model, for the specific domain of action perception learning and memory, the present neurocomputational work may contribute to a solution of the embodiment debate: Perception and comprehension processes driving LTM circuit formation draw upon “peripheral” sensorimotor areas (M1, P1), whereas working memory activity gradually retreats to the “central” or core parts of the circuits (PF, PA).

As mentioned, the present approach models the emergence of memory traces as a consequence of stimulation of the senses and concomitant motor patterns. This mechanism can be applied to learning and memory in a range of cognitive domains. It certainly applies to basic relationships between sensory input and motor output (e.g., the button press a monkey performs to a given colour stimulus in DMTS experiments, Fuster, 2003), as well as to that between acoustic and articulatory phonological schemas of spoken words in language acquisition (Pulvermüller & Fadiga, 2010), and equally to the action- and perception-related knowledge about concepts – e.g., that a hammer is characterized by typical visual shape, manual action and goal (Barsalou, 2008; Cappa & Pulvermüller, 2012; Kiefer & Pulvermüller, 2012; Meteyard et al., 2012; Pulvermüller, 2005). In spite of these parallelisms between sensorimotor, symbolic and conceptual learning insofar as sensorimotor association appears as a crucial component, conceptual and semantic processing certainly also engages additional mechanisms not addressed in the present study (Kiefer & Pulvermüller, 2012; Meteyard et al., 2012). For example the question about the embodiment or disembodiment of abstract concepts (see Barsalou & Wiemer-Hastings, 2005; Casasanto, 2009; Pulvermüller, 2013) is left untouched by the present simulations. Therefore, although the mechanistic model accounts for the differential relevance of embodied sensorimotor and disembodied multimodal systems to perception/comprehension and memory processes, it clearly does not cover all facets of the embodiment debate.

5. Conclusions

Distributed APCs in neural networks incorporating features of cortical neuroanatomy can explain one specific aspect of

working memory processes in the human and monkey brain, namely its cortical topography. Memory cell activity is rarely present in primary cortices, pronounced in higher secondary perceptual and premotor areas, and especially strong and long-lasting in higher association cortex, where multimodal information from sensory and motor domains massively converges. The mechanistic cause and explanation of this topography of working memory lie in correlation learning and the between-area neuroanatomical connectivity structure of the cortex, both of which were incorporated in the neurocomputational architecture applied here. As we argue in the discussion, the dispute between the proponents of embodiment and disembodiment of symbolic and conceptual processing may, in part, be reconciled by the observation that, although APCs are the basis of sensorimotor and semantic learning, the working memory processes supported by these APCs show “memory retreat” of activation to the circuits’ own multimodal “cores” in prefrontal and perceptual association areas.

Acknowledgements

Supported by the Freie Universität Berlin, Deutsche Forschungsgemeinschaft (Excellence Cluster *Languages of Emotion*), and the EPSRC and BBSRC, UK (project on *Brain-inspired architecture for brain embodied language*, BABEL (grant no. EP/J004561/1)). The authors declare no conflict of interest.

Appendix A

Each cell or “node” of the network represents a cortical column of approximately .25 mm² size (Hubel, 1995), containing $\sim 2.5 \cdot 10^4$ neurons⁷ (Braitenberg & Schüz, 1998). The state of each cell x is uniquely defined by its membrane potential $V(x,t)$, representing the average of the sum of all excitatory and inhibitory post-synaptic potentials acting upon neural pool (cluster) x at time t . The membrane potential $V(x,t)$ at time t of a model cell x with membrane time-constant τ is governed by the equation:

$$\tau \cdot \frac{dV(x,t)}{dt} = -V(x,t) + V_{in}(x,t) \quad (\text{A.1})$$

where $V_{in}(x,t)$ is the total input to x (sum of all excitatory and inhibitory synaptic inputs to cell x at time t ; inhibitory synapses are given a negative sign).

The output of an excitatory cell x at time t is defined as follows:

$$O(x,t) = \begin{cases} 0 & \text{if } V(x,t) \leq \varphi \\ (V(x,t) - \varphi) & \text{if } 0 < (V(x,t) + \varphi) \leq 1 \\ 1 & \text{otherwise} \end{cases} \quad (\text{A.2})$$

$O(x,t)$ represents the average (graded) firing rate (number of action potentials per time unit) of cluster x at time t ; it is a

⁷ These figures are meant to provide only an estimate of the grain of the model; as noted previously Hubel D. *Eye, brain, and vision*. New York: Scientific American Library, 1995, the size of a macrocolumn (or “module”) varies substantially between cortical layers (ranging from 0.1 mm² in layer 4C to 4 mm² in layer 3) and cortical areas (*ibid.*, p. 130).

piecewise-linear sigmoid function of the cell's membrane potential $V(x,t)$, clipped into the range $[0, 1]$ and with slope 1 between the lower and upper thresholds φ and $\varphi + 1$. The output $O(x,t)$ of an inhibitory cell is 0 if $V(x,t) < 0$, and $V(x,t)$ otherwise. The value of φ varies in time, as explained below. Thus, a clipped version of the transfer function (maximum output = 1) was used to calculate the excitatory unit output, whereas the inhibitory units' output was not clipped. This was done in an attempt to provide powerful local inhibition and regulation preventing instability of the network. These features were constant across all "areas" of the network and therefore cannot underlie any local functional differences between areas.

The piecewise-linear function $O(x,t)$ described above was used – during the training phase of the simulations – as an approximation of a continuous sigmoidal function. This choice was motivated by practical considerations: calculating the value of a piecewise linear function is much less computationally expensive than computing with non-linear transformation functions; even with the linear variant, computation time for learning in each individual network was several days. Therefore, to keep computation effort manageable, the piecewise-linear approximation was used during learning. However, as the aim of the testing phase was to generate network responses comparable with real experimental data, a more realistic, continuous version of the transformation function was used during testing. To this end, $O(x,t)$ was re-defined as the following sigmoid (cf. also Eq. (2) in the main text):

$$O(x, t) = \frac{1}{1 + e^{-2\beta(V(x,t) - \varphi)}} \quad (\text{A.2.1})$$

where $\beta/2$ is the slope at the inversion point, and φ (the threshold) determines the inversion point of the sigmoid. The transformation function for inhibitory cells remained unchanged.

Neuronal adaptation was realised (in excitatory cells only) by allowing the threshold φ in Eqs. (A.2) and (A.2.1) to be cell-specific and vary in time. More precisely:

$$\varphi(x, t) = \alpha \cdot \omega(x, t) \quad (\text{A.3})$$

where $\omega(x,t)$ is the time-average of the cell's recent output and α is the "adaptation strength" (see below for parameter values used in the simulations).

For any excitatory cell x , the approximate time-average $\omega(x,t)$ of its output $O(x,t)$ is estimated by integrating Eq. (A.4) below, assuming initial average $\omega(x,0) = 0$:

$$\tau_A \cdot \frac{d\omega(x, t)}{dt} = -\omega(x, t) + O(x, t) \quad (\text{A.4})$$

The low-pass dynamics of the cells [Eq. (A.1), (A.2) and (A.4)] are integrated using the Euler scheme with step size Δt , where $\Delta t = .5$ (in arbitrary units of time). Other parameter values are reported below.

The learning rule used to simulate synaptic plasticity is based on the Artola–Bröcher–Singer model of LTP/LTD (Artola & Singer, 1993). In the implementation, we discretized the continuous range of possible synaptic efficacy changes into two possible levels, $+\Delta w$ and $-\Delta w$ (with $\Delta w \ll 1$ and fixed). We defined as "active" any link from a cell x such that

the output $O(x,t)$ of cell x at time t is larger than θ_{pre} , where $\theta_{pre} \in [0,1]$ is an arbitrary threshold representing the minimum level of pre-synaptic activity required for LTP (or LTD) to occur. Thus, given any two cells x and y linked with weight $w_t(x,y)$, the new weight $w_{t+1}(x,y)$ is calculated as follows:

$$w_{t+1}(x, y) = \begin{cases} w_t(x, y) + \Delta w & \text{if } O(x, t) \geq \theta_{pre} \text{ and } V(y, t) \geq \theta_+ \\ w_t(x, y) - \Delta w & \text{if } O(x, t) \geq \theta_{pre} \text{ and } \theta_- \leq V(y, t) < \theta_+ \\ w_t(x, y) - \Delta w & \text{if } O(x, t) < \theta_{pre} \text{ and } V(y, t) \geq \theta_+ \\ w_t(x, y) & \text{otherwise} \end{cases} \quad (\text{A.5})$$

Parameter values used for the simulations are:

Eq. (A.1)	Excitatory cells: $\tau = 2.5$ (in simulation time-steps); Inhibitory cells: $\tau = 5$ (in simulation time-steps);
Eq. (A.2.1)	$\beta = 1.5$, $\varphi = 3.5$
Eq. (A.3)	Adaptation: $\alpha = .026$;
Eq. (A.4)	Time constant for computing gliding-average of cell activity: $\tau_A = 15$ (in simulation time-steps);
Eq. (A.5)	Post-synaptic potential thresholds for LTP/LTD: $\theta_- = .15$, $\theta_+ = .25$; Pre-synaptic output activity required for synaptic change: $\theta_{pre} = .05$; Learning rate: $\Delta w = .0005$.

REFERENCES

- Acheson, D. J., Hamidi, M., Binder, J. R., & Postle, B. R. (2011). A common neural substrate for language production and verbal working memory. *Journal of Cognitive Neuroscience*, 23, 1358–1367.
- Amir, Y., Harel, M., & Malach, M. (1993). Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *Journal of Comparative Neurology*, 334, 19–46.
- Arikuni, T., Watanabe, K., & Kubota, K. (1988). Connections of area 8 with area 6 in the brain of the macaque monkey. *Journal of Comparative Neurology*, 277, 21–40.
- Artola, A., & Singer, W. (1993). Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends in Neurosciences*, 16, 480–487.
- Baddeley, A. (1992). Working memory. *Science*, 255, 556–559.
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, 4, 829–839.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–609. discussion 610–660.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.
- Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher, & R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129–163). New York: Cambridge University Press.
- Bauer, R. H., & Fuster, J. M. (1976). Delayed-matching and delayed-response deficit from cooling dorsolateral prefrontal cortex in monkeys. *Journal of Comparative and Physiological Psychology*, 90, 293–302.
- Bauer, R. H., & Fuster, J. M. (1978). Effects of d-amphetamine and prefrontal cortical cooling on delayed matching-to-sample behavior. *Pharmacology Biochemistry and Behavior*, 8, 243–249.
- Bedny, M., & Caramazza, A. (2011). Perception, action, and word meanings in the human brain: the case from action verbs. *Annals of the New York Academy of Sciences*, 1224, 81–95.

- Bibbig, A., Wennekers, T., & Palm, G. (1995). A neural network model of the cortico-hippocampal interplay and the representation of contexts. *Behavioural Brain Research*, 66, 169–175.
- Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, 15, 527–536.
- Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, 25, 151–188.
- Braitenberg, V. (1978). Cell assemblies in the cerebral cortex. In R. Heim, & G. Palm (Eds.), *Lecture notes in biomathematics: Vol. 21. Theoretical approaches to complex systems* (pp. 171–188). Berlin: Springer.
- Braitenberg, V., & Pulvermüller, F. (1992). Entwurf einer neurologischen theorie der sprache. *Naturwissenschaften*, 79, 103–117.
- Braitenberg, V., & Schüz, A. (1998). *Cortex: Statistics and geometry of neuronal connectivity*. Berlin: Springer.
- Bussey, T. J., & Saksida, L. M. (2002). The organization of visual object representations: a connectionist model of effects of lesions in perirhinal cortex. *European Journal of Neuroscience*, 15, 355–364.
- Cappa, S. F., & Pulvermüller, F. (2012). Language and the motor system. *Cortex*, 48, 785–787.
- Casasanto, D. (2009). Embodiment of abstract concepts: good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, 138, 351–367.
- Catani, M., & Ffytche, D. H. (2005). The rises and falls of disconnection syndromes. *Brain*, 128, 2224–2239.
- Catani, M., Jones, D. K., Donato, R., & Ffytche, D. H. (2003). Occipito-temporal connections in the human brain. *Brain*, 126, 2093–2107.
- Chafee, M. V., & Goldman-Rakic, P. S. (2000). Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. *Journal of Neurophysiology*, 83, 1550–1566.
- Constantinidis, C., Franowicz, M. N., & Goldman-Rakic, P. S. (2001). The sensory nature of mnemonic representation in the primate prefrontal cortex. *Nature Neuroscience*, 4, 311–316.
- D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 362, 761–772.
- D'Esposito, M., Cooney, J. W., Gazzaley, A., Gibbs, S. E., & Postle, B. R. (2006). Is the prefrontal cortex necessary for delay task performance? Evidence from lesion and fMRI data. *Journal of the International Neuropsychological Society*, 12, 248–260.
- Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition*, 33, 25–62.
- Deacon, T. W. (1992). Cortical connections of the inferior arcuate sulcus cortex in the macaque brain. *Brain Research*, 573, 8–26.
- Deco, G., Ponce-Alvarez, A., Mantini, D., Romani, G. L., Hagmann, P., & Corbetta, M. (2013). Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. *Journal of Neuroscience*, 33, 11239–11252.
- Deco, G., & Rolls, E. T. (2005). Attention, short-term memory, and action selection: a unifying theory. *Progress in Neurobiology*, 76, 236–256.
- Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70, 200–227.
- Distler, C., Boussaoud, D., Desimone, R., & Ungerleider, L. G. (1993). Cortical connections of inferior temporal area TEO in macaque monkeys. *Journal of Comparative Neurology*, 334, 125–150.
- Douglas, R. J., & Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annual Review of Neuroscience*, 27, 419–451.
- Dum, R. P., & Strick, P. L. (2002). Motor areas in the frontal lobe of the primate. *Physiology & Behavior*, 77, 677–682.
- Dum, R. P., & Strick, P. L. (2005). Frontal lobe inputs to the digit representations of the motor areas on the lateral surface of the hemisphere. *Journal of Neuroscience*, 25, 1375–1386.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, 195–225.
- Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 120, 339–357.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Fransen, E., Tahvildari, B., Egorov, A. V., Hasselmo, M. E., & Alonso, A. A. (2006). Mechanism of graded persistent cellular activity of entorhinal cortex layer v neurons. *Neuron*, 49, 735–746.
- Friston, K., Breakspear, M., & Deco, G. (2012). Perception and self-organized instability. *Frontiers in Computational Neuroscience*, 6, 44.
- Funahashi, S. (2006). Prefrontal cortex and working memory processes. *Neuroscience*, 139, 251–261.
- Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61, 331–349.
- Fuster, J. M. (1990). Inferotemporal units in selective visual attention and short-term memory. *Journal of Neurophysiology*, 64, 681–697.
- Fuster, J. M. (1995). *Memory in the cerebral cortex. An empirical approach to neural networks in the human and nonhuman primate*. Cambridge, MA: MIT Press.
- Fuster, J. M. (1997). Network memory. *Trends in Neurosciences*, 20, 451–459.
- Fuster, J. M. (2001). The prefrontal cortex—an update: time is of the essence. *Neuron*, 30, 319–333.
- Fuster, J. M. (2003). *Cortex and mind: Unifying cognition*. Oxford: Oxford University Press.
- Fuster, J. M. (2009). Cortex and memory: emergence of a new paradigm. *Journal of Cognitive Neuroscience*, 21, 2047–2072.
- Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science*, 173, 652–654.
- Fuster, J. M., & Bauer, R. H. (1974). Visual short-term memory deficit from hypothermia of frontal cortex. *Brain Research*, 81, 393–400.
- Fuster, J. M., Bauer, R. H., & Jervey, J. P. (1981). Effects of cooling inferotemporal cortex on performance of visual memory tasks. *Experimental Neurology*, 71, 398–409.
- Fuster, J. M., Bauer, R. H., & Jervey, J. P. (1982). Cellular discharge in the dorsolateral prefrontal cortex of the monkey in cognitive tasks. *Experimental Neurology*, 77, 679–694.
- Fuster, J. M., Bauer, R. H., & Jervey, J. P. (1985). Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain Research*, 330, 299–307.
- Fuster, J. M., & Jervey, J. P. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science*, 212, 952–955.
- Fuster, J. M., & Jervey, J. P. (1982). Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *Journal of Neuroscience*, 2, 361–375.
- Garagnani, M., & Pulvermüller, F. (2011). From sounds to words: a neurocomputational model of adaptation, inhibition and memory processes in auditory change detection. *NeuroImage*, 54, 170–181.

- Garagnani, M., & Pulvermüller, F. (2013). Neuronal correlates of decisions to speak and act: spontaneous emergence and dynamic topographies in a computational model of frontal and temporal areas. *Brain and Language*, 127(1), 75–85.
- Garagnani, M., Wennekers, T., & Pulvermüller, F. (2007). A neuronal model of the language cortex. *Neurocomputing*, 70, 1914–1919.
- Garagnani, M., Wennekers, T., & Pulvermüller, F. (2008). A neuroanatomically-grounded Hebbian learning model of attention-language interactions in the human brain. *European Journal of Neuroscience*, 27, 492–513.
- Garagnani, M., Wennekers, T., & Pulvermüller, F. (2009). Recruitment and consolidation of cell assemblies for words by way of Hebbian learning and competition in a multi-layer neural network. *Cognitive Computation*, 1, 160–176.
- Gilbert, C. D., & Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *Journal of Neuroscience*, 3, 1116–1133.
- Glasser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cerebral Cortex*, 18, 2471–2482.
- Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences*, 20, 1–19. discussion 19–55.
- Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron*, 14, 477–485.
- Guye, M., Parker, G. J., Symms, M., Boulby, P., Wheeler-Kingshott, C. A., Salek-Haddadi, A., et al. (2003). Combined functional MRI and tractography to demonstrate the connectivity of the human primary motor cortex in vivo. *NeuroImage*, 19, 1349–1360.
- Hebb, D. O. (1949). *The organization of behavior. A neuropsychological theory*. New York: John Wiley.
- Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: investigation of acquired dyslexia. *Psychological Review*, 98, 74–95.
- Hubel, D. (1995). *Eye, brain, and vision*. New York: Scientific American Library.
- Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends in Neurosciences*, 18, 314–320.
- Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 11793–11799.
- Kelly, C., Uddin, L. Q., Shehzad, Z., Margulies, D. S., Castellanos, F. X., Milham, M. P., et al. (2010). Broca's region: linking human brain functional connectivity data and non-human primate tracing anatomy studies. *European Journal of Neuroscience*, 32, 383–398.
- Kemmerer, D., Rudrauf, D., Manzel, K., & Tranel, D. (2012). Behavioural patterns and lesion sites associated with impaired processing of lexical and conceptual knowledge of action. *Cortex*, 48, 826–848.
- Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex*, 48, 805–825.
- Knoblauch, A., & Palm, G. (2002a). Scene segmentation by spike synchronization in reciprocally connected visual areas. I. Local effects of cortical feedback. *Biological Cybernetics*, 87, 151–167.
- Knoblauch, A., & Palm, G. (2002b). Scene segmentation by spike synchronization in reciprocally connected visual areas. II. Global assemblies and synchronization on larger space and time scales. *Biological Cybernetics*, 87, 168–184.
- Kojima, S., & Goldman-Rakic, P. S. (1982). Delay-related activity of prefrontal neurons in rhesus monkeys performing delayed response. *Brain Research*, 248, 43–49.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.
- Lemus, L., Hernandez, A., & Romo, R. (2009). Neural encoding of auditory discrimination in ventral premotor cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 14640–14645.
- Linden, D. E. (2007). The working memory networks of the human brain. *Neuroscientist*, 13, 257–267.
- Loewenstein, Y., & Sompolinsky, H. (2003). Temporal integration by calcium dynamics in a model neuron. *Nature Neuroscience*, 6, 961–967.
- Lu, M. T., Preston, J. B., & Strick, P. L. (1994). Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. *Journal of Comparative Neurology*, 341, 375–392.
- Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology (Paris)*, 102, 59–70.
- Makris, N., Meyer, J. W., Bates, J. F., Yeterian, E. H., Kennedy, D. N., & Caviness, V. S. (1999). MRI-based topographic parcellation of human cerebral white matter and nuclei ii. Rationale and applications with systematics of cerebral connectivity. *NeuroImage*, 9, 18–45.
- Makris, N., & Pandya, D. N. (2009). The extreme capsule in humans and rethinking of the language circuitry. *Brain Structure and Function*, 213, 343–358.
- Malenka, R. C., & Nicoll, R. A. (1999). Long-term potentiation—a decade of progress? *Science*, 285, 1870–1874.
- Matthews, G. G. (2001). *Cellular physiology of nerve and muscle*. Malden, MA: Blackwell.
- Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex*, 48, 788–804.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, 6, 414–417.
- Miyashita, Y., & Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, 331, 68–70.
- Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, 319, 1543–1546.
- Moseley, R. L., Mohr, B., Lombardo, M. V., Baron-Cohen, S., Hauk, O., & Pulvermüller, F. (2013). Brain and behavioural correlates of action semantic deficits in autism. *Frontiers in Human Neuroscience*, 7.
- Nakamura, H., Gattass, R., Desimone, R., & Ungerleider, L. G. (1993). The modular organization of projections from areas v1 and v2 to areas v4 and TEO in macaques. *Journal of Neuroscience*, 13, 3681–3691.
- Naya, Y., Sakai, K., & Miyashita, Y. (1996). Activity of primate inferotemporal neurons related to a sought target in pair-association task. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 2664–2669.
- Neininger, B., & Pulvermüller, F. (2003). Word-category specific deficits after lesions in the right hemisphere. *Neuropsychologia*, 41, 53–70.
- Owen, A. M., Sahakian, B. J., Semple, J., Polkey, C. E., & Robbins, T. W. (1995). Visuo-spatial short-term recognition memory and learning after temporal lobe excisions, frontal lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia*, 33, 1–24.
- Pakkenberg, B., & Gundersen, H. J. G. (1997). Neocortical neuron number in humans: effect of sex and age. *Journal of Comparative Neurology*, 384, 312–320.
- Palm, G. (1980). On associative memory. *Biological Cybernetics*, 36, 19–31.
- Palm, G. (1982). *Neural assemblies*. Berlin: Springer.
- Palm, G. (1987). Associative memory and threshold control in neural networks. In J. L. Casti, & A. Karlqvist (Eds.), *Real brains, artificial minds* (pp. 165–179). New York: North-Holland.

- Palm, G. (1990). Cell assemblies as a guideline for brain research. *Concepts in Neuroscience*, 1, 133–147.
- Pandya, D. N. (1995). Anatomy of the auditory cortex. *Revue Neurologique (Paris)*, 151, 486–494.
- Pandya, D. N., & Barnes, C. L. (1987). Architecture and connections of the frontal lobe. In E. Perecman (Ed.), *The frontal lobes revisited* (pp. 41–72). New York: The IRBN Press.
- Pandya, D. N., & Yeterian, E. H. (1985). Architecture and connections of cortical association areas. In A. Peters, & E. G. Jones (Eds.), *Association and auditory cortices: Vol. 4. Cerebral cortex* (pp. 3–61). London: Plenum Press.
- Parker, G. J., Luzzi, S., Alexander, D. C., Wheeler-Kingshott, C. A., Ciccarelli, O., & Lambon Ralph, M. A. (2005). Lateralization of ventral and dorsal auditory-language pathways in the human brain. *NeuroImage*, 24, 656–666.
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8, 976–987.
- Petrides, M. (2000). Dissociable roles of mid-dorsolateral prefrontal and anterior inferotemporal cortex in visual working memory. *Journal of Neuroscience*, 20, 7496–7503.
- Petrides, M., & Pandya, D. N. (2002). Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *European Journal of Neuroscience*, 16, 291–310.
- Petrides, M., & Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biology*, 7, e1000170.
- Pinker, S. (1994). *The language instinct. How the mind creates language*. New York: Harper Collins Publishers.
- Plaut, D. C. (2002). Graded modality-specific specialisation in semantics: a computational account of optic aphasia. *Cognitive Neuropsychology*, 19, 603–639.
- Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, 139, 23–38.
- Pulvermüller, F. (1992). Constituents of a neurological theory of language. *Concepts in Neuroscience*, 3, 157–200.
- Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22, 253–336.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6, 576–582.
- Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, 17, 458–470.
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11, 351–360.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12, 718–724.
- Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 11800–11806.
- Rilling, J. K. (2014). Comparative primate neuroimaging: insights into human brain evolution. *Trends in Cognitive Sciences*, 18(1), 46–55.
- Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., et al. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience*, 11, 426–428.
- Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron*, 31, 889–901.
- Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychological Review*, 111, 205–235.
- Romanski, L. M. (2004). Domain specificity in the primate prefrontal cortex. *Cognitive, Affective, & Behavioral Neuroscience*, 4, 421–429.
- Romanski, L. M., & Goldman-Rakic, P. S. (2002). An auditory domain in primate prefrontal cortex. *Nature Neuroscience*, 5, 15–16.
- Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*, 2, 1131–1136.
- Romo, R., Hernandez, A., & Zainos, A. (2004). Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron*, 41, 165–173.
- Saur, D., Kreher, B. W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M. S., et al. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 18035–18040.
- Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, 20, 207–214.
- Shinomoto, S., Kim, H., Shimokawa, T., Matsuno, N., Funahashi, S., Shima, K., et al. (2009). Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS Computational Biology*, 5, e1000433.
- Sommer, F. T., & Wennekers, T. (2001). Associative memory in networks of spiking neurons. *Neural Networks*, 14, 825–834.
- Sporns, O., & Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics*, 2, 145–162.
- Stramandinoli, F., Marocco, D., & Cangelosi, A. (2012). The grounding of higher order concepts in action and language: a cognitive robotics model. *Neural Networks*, 32, 165–173.
- Super, H., Spekrijse, H., & Lamme, V. A. (2001). A neural correlate of working memory in the monkey primary visual cortex. *Science*, 293, 120–124.
- Takeda, K., & Funahashi, S. (2002). Prefrontal task-related activity representing visual cue location or saccade direction in spatial working memory tasks. *Journal of Neurophysiology*, 87, 567–588.
- Trumpp, N. M., Kliese, D., Hoening, K., Haarmeier, T., & Kiefer, M. (2013). Losing the sound of concepts: damage to auditory association cortex impairs the processing of sound-related concepts. *Cortex*, 49, 474–486.
- Ungerleider, L. G., Gaffan, D., & Pelak, V. S. (1989). Projections from inferior temporal cortex to prefrontal cortex via the uncinate fascicle in rhesus monkeys. *Experimental Brain Research*, 76, 473–484.
- Ungerleider, L. G., & Haxby, J. V. (1994). ‘What’ and ‘where’ in the human brain. *Current Opinion in Neurobiology*, 4, 157–165.
- Verduzco-Flores, S., Bodner, M., Ermentrout, B., Fuster, J. M., & Zhou, Y. (2009). Working memory cells' behavior may be explained by cross-regional networks with synaptic facilitation. *PLoS One*, 4, e6399.
- Wakana, S., Jiang, H., Nagae-Poetscher, L. M., van Zijl, P. C., & Mori, S. (2004). Fiber tract-based atlas of human white matter anatomy. *Radiology*, 230, 77–87.
- Webster, M. J., Bachevalier, J., & Ungerleider, L. G. (1994). Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cerebral Cortex*, 4, 470–483.
- Wennekers, T., Garagnani, M., & Pulvermüller, F. (2006). Language models based on Hebbian cell assemblies. *Journal of Physiology (Paris)*, 100, 16–30.
- Wermter, S., Page, M., Knowles, M., Gallese, V., Pulvermüller, F., & Taylor, J. (2009). Multimodal communication in animals, humans and robots: an introduction to perspectives in brain-inspired informatics. *Neural Networks*, 22, 111–115.

- Wheeler, M. A., Stuss, D. T., & Tulving, E. (1995). Frontal lobe damage produces episodic memory impairment. *Journal of the International Neuropsychological Society*, 1, 525–536.
- Willshaw, D. J., & Buckingham, J. T. (1990). An assessment of Marr's theory of the hippocampus as a temporary memory store. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 329, 205–215.
- Willwacher, G. (1976). Fähigkeiten eines assoziativen speichersystems im vergleich zu gehirnfunktionen. *Biological Cybernetics*, 24, 181–198.
- Young, M. P., Scannell, J. W., & Burns, G. (1995). *The analysis of cortical connectivity*. Heidelberg: Springer.
- Young, M. P., Scannell, J. W., Burns, G., & Blakemore, C. (1994). Analysis of connectivity: neural systems in the cerebral cortex. *Review in Neuroscience*, 5, 227–249.
- Zipser, D., Kehoe, B., Littlewort, G., & Fuster, J. (1993). A spiking network model of short-term active memory. *Journal of Neuroscience*, 13, 3406–3420.