Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

Research Report

Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain

Edmund T. Rolls[a,*], *Gustavo Deco*[b,c]

[a]*Oxford Centre for Computational Neuroscience, Oxford, UK*
[b]*Universitat Pompeu Fabra, Theoretical and Computational Neuroscience, Roc Boronat 138, 08018 Barcelona, Spain*
[c]*Institucio Catalana de Recerca i Estudis Avancats (ICREA), Spain*

ARTICLE INFO

Article history:
Accepted 8 September 2014

Keywords:
Syntax
Stochastic neurodynamics
Short-term memory
Attractor network
Semantics
Language

ABSTRACT

Neural principles that provide a foundation for memory, perception, and decision-making include place coding with sparse distributed representations, associative synaptic modification, and attractor networks in which the storage capacity is in the order of the number of associatively modifiable recurrent synapses on any one neuron. Based on those and further principles of cortical computation, hypotheses are explored in which syntax is encoded in the cortex using sparse distributed place coding. Each cortical module 2–3 mm in diameter is proposed to be formed of a local attractor neuronal network with a capacity in the order of 10,000 words (e.g. subjects, verbs or objects depending on the module). Such a system may form a deep language-of-thought layer. For the information to be communicated to other people, the modules in which the neurons are firing which encode the syntactic role, as well as which neurons are firing to specify the words, must be communicated. It is proposed that one solution to this (used in English) is temporal order encoding, for example subject–verb–object. It is shown with integrate-and-fire simulations that this order encoding could be implemented by weakly forward-coupled subject–verb–object modules. A related system can decode a temporal sequence. This approach based on known principles of cortical computation needs to be extended to investigate further whether it could form a biological foundation for the implementation of language in the brain.

This article is part of a Special Issue entitled SI: Brain and Memory.

1. Introduction

Previous research on networks in the cortex involved in memory and perception is described concisely in Sections 2–5, and is used to suggest principles that might contribute to advances in understanding one of the major unsolved problems in brain research, how the syntax needed for language might be implemented in cortical networks. Those principles

[*]Corresponding author.
E-mail address: Edmund.Rolls@oxcns.org (E.T. Rolls).
URL: http://www.oxcns.org (E.T. Rolls).

http://dx.doi.org/10.1016/j.brainres.2014.09.021
0006-8993/© 2014 Elsevier B.V. All rights reserved.

Please cite this article as: Rolls, E.T., Deco, G., Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain. Brain Research (2014), http://dx.doi.org/10.1016/j.brainres.2014.09.021

are then explored and investigated using attractor network models and simulations of cortical function. This paper is thus intended to address the topic of this special issue of the Journal on the Brain and Memory, but focusses on a major new perspective, how language may be implemented in the brain, using principles based on our understanding of the operation of cortical attractor networks that implement memory, perceptual, short-term memory, attentional, and decision-making functions in the brain (Rolls, 2008).

2. A quantitative theory of the implementation of episodic memory in the brain

David Marr pioneered quantitative approaches to understanding how the hippocampus operates in memory. Marr (1971) showed how a network with recurrent collaterals could complete a memory using a partial retrieval cue, and how sparse representations could increase the number of memories stored. Early work of Gardner-Medwin (1976) showed how progressive recall could operate in a network of binary neurons with binary synapses. The analysis of these autoassociation or attractor networks was developed by Kohonen (1977), Kohonen et al. (1981), and Hopfield (1982), and the value of sparse representations was quantified by Treves and Rolls (1991). Marr (1971) did not specify the functions of the dentate granule cells vs the CA3 cells vs the CA1 cells (which were addressed by Rolls, 1987; Rolls, 1989a,b,c and by Treves and Rolls, 1992, 1994), nor how retrieval to the neocortex of hippocampal memories could be produced, for which a quantitative theory was developed by Treves and Rolls (1994).

Rolls (1987) produced a theory of the hippocampus in which the CA3 neurons operated as an autoassociation memory to store episodic memories including object and place memories, and the dentate granule cells operated as a preprocessing stage for this by performing pattern separation so that the mossy fibres could act to set up different representations for each memory to be stored in the CA3 cells. The architecture showing the connections of the system is shown in Fig. 1. McNaughton and Morris (1987) at about the same time suggested that the CA3 network might be an autoassociation network, and that the mossy fibre to CA3 connections might implement ‘detonator’ synapses. The concepts that the dentate acted as a competitive network to perform pattern separation, and that the mossy fibers act as a randomising system to contribute to pattern separation in

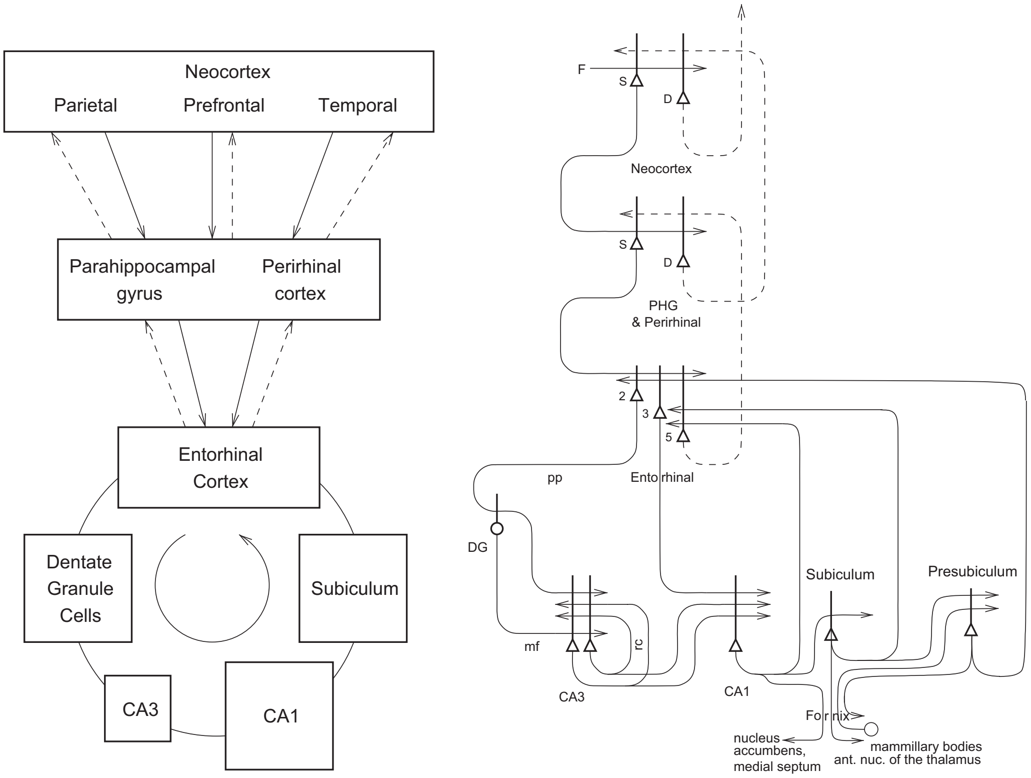

Fig. 1 – Forward connections (solid lines) from areas of cerebral association neocortex via the parahippocampal gyrus and perirhinal cortex, and entorhinal cortex, to the hippocampus; and backprojections (dashed lines) via the hippocampal CA1 pyramidal cells, subiculum, and parahippocampal gyrus to the neocortex. There is great convergence in the forward connections down to the single network implemented in the CA3 pyramidal cells; and great divergence again in the backprojections. Left: block diagram. Right: more detailed representation of some of the principal excitatory neurons in the pathways. Abbreviations – D, Deep pyramidal cells; DG, Dentate Granule cells; F, Forward inputs to areas of the association cortex from preceding cortical areas in the hierarchy; mf, mossy fibres; PHG, parahippocampal gyrus and perirhinal cortex; pp, perforant path; rc, recurrent collaterals of the CA3 hippocampal pyramidal cells; S, Superficial pyramidal cells. 2: pyramidal cells in layer 2 of the entorhinal cortex. 3: pyramidal cells in layer 3 of the entorhinal cortex. The thick lines above the cell bodies represent the dendrites.

Please cite this article as: Rolls, E.T., Deco, G., Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain. Brain Research (2014), http://dx.doi.org/10.1016/j.brainres.2014.09.021

CA3 was proposed by Rolls (1989b). The concepts that the diluted mossy fibre connectivity might implement selection of a new random set of CA3 cells for each new memory, and that a direct perforant path input to CA3 is needed to initiate retrieval, were analyzed quantitatively by Treves and Rolls (1992). Rolls (1987) suggested that the CA1 cells operate as a recoder for the information recalled from the CA3 cells to a partial memory cue, so that the recalled information would be represented more efficiently to enable recall, via the backprojection synapses, of activity in the neocortical areas similar to that which had been present during the original episode. This theory was developed further (Rolls, 1989a,b,c,d; Rolls, 1990a,b), including further details about how the backprojections could operate (Rolls, 1989b, 1989c), and how the dentate granule cells could operate as a competitive network (Rolls, 1989a). Quantitative aspects of the theory were then developed with A. Treves, who brought the expertise of theoretical physics applied previously mainly to understand the properties of fully connected attractor networks with binary neurons (Amit, 1989; Hopfield, 1982) to bear on the much more diluted connectivity of the recurrent collateral connections found in real biological networks (e.g. 2% between CA3 pyramidal cells in the rat), in networks of neurons with graded (continuously variable) firing rates, graded synaptic strengths, and sparse representations in which only a small proportion of the neurons is active at any one time, as is found in the hippocampus (Treves, 1990; Treves and Rolls, 1991). These developments in understanding quantitatively the operation of more biologically relevant recurrent networks with modifiable synapses were applied quantitatively to the CA3 region (Treves and Rolls, 1991), and to the issue of why there are separate mossy fibre and perforant path inputs to the CA3 cells of the hippocampus (Treves and Rolls, 1992). This whole model of the hippocampus was described in more detail, and a quantitative treatment of the theory of recall by backprojection pathways in the brain was provided by Treves and Rolls (1994) and tested by Rolls (1995).

Further developments of this theory, which remains the only quantitative theory of information storage for episodic memory in the hippocampus, and its quantitative recall back to neocortex, have been described (Rolls, 2008, 2010; Kesner and Rolls, 2014). These developments include the role of different parts of the hippocampal system in pattern separation and pattern completion (Rolls, 2013b; Cerasti and Treves, 2010; Stella et al., 2013); the utility of the diluted connectivity between cortical neurons provided by recurrent collaterals in ensuring that the memory capacity of the attractor network is not compromised (Rolls, 2012a); and the way in which time encoding neurons (MacDonald et al., 2011) in CA1 could be used to implement a temporal sequence memory for objects and odors (Kesner and Rolls, 2014). The theory is richly supported by empirical tests of the contributions of different subregions of the hippocampus (Rolls and Kesner, 2006; Kesner and Rolls, 2015). The supporting empirical evidence includes the finding that the CA3 recurrent collateral system is even more widespread in primates than in rodents (Kondo et al., 2009). The theory was made directly relevant to humans by the discovery of spatial view neurons in primates that encode a location in space being looked at (Rolls et al., 1989; Feigenbaum and Rolls, 1991; Georges-François et al., 1999; Robertson et al., 1998; Rolls et al., 1997, 1998), and which combine this with information about objects (Rolls et al., 2005) and rewards (Rolls and Xiang, 2005) to enable a one-trial episodic memory based on a view of a place (Rolls and Xiang, 2006) to be implemented. The primate spatial view neurons enable a memory to be formed of a place that has been viewed but never visited, impossible with rodent place cells (Rolls, 2008; Kesner and Rolls, 2014).

A highlight of this quantitative theory of episodic memory for understanding cortical function, including how the cortex implements language, is that the number of memories that can be stored in a cortical attractor network is in the order of the number of associatively modifiable recurrent collateral connections onto any one neuron in the attractor network if sparse representations are used (Treves and Rolls, 1991; Rolls and Treves, 1998; Rolls, 2008). Thus in CA3, where there are 12,000 recurrent collaterals onto each neuron in the rat, the storage capacity is in the order of 12,000 memories (Rolls, 2008). If in the human neocortex there were 10,000 associatively modifiable recurrent collateral synapses onto a neuron in a local region 2–3 mm in diameter, then in the order of 10,000 words might be represented in such a cortical column, which is a reasonable working vocabulary. This is one of the foundations based on our understanding of quantitative aspects of memory systems in the brain for what follows on language and syntax.

3. Learning of new perceptual representations

Learning in a competitive network provides a useful way of building perceptual representations in which for example sensory inputs activate a population of neurons, which then compete with each other through inhibitory neurons, with the neurons left firing with high rates after the competition showing associative synaptic modification of the high-firing inputs to those synapses (Rolls, 1992; Rolls and Treves, 1998; Rolls, 2008). This can build useful sensory or perceptual categories, as similar inputs activate the same output neurons, and different inputs activate different output neurons (Rolls and Treves, 1998; Rolls, 2008). Competitive networks help to build the sparse distributed representations that result in high capacity in autoassociation (attractor) networks, and in pattern association networks (Rolls, 2008; Rolls and Treves, 1990). These processes may be part of what is involved in building semantic representations and even word representations, where similar inputs need to activate a small number of output neurons, though it is acknowledged that the building of semantic representations requires exceptions to be learned too (McClelland et al., 1995).

A small modification of this competitive learning process in which there is a short-term memory trace, which might be as simple as the long time constant of an NMDA receptor, or continuing firing for 100–300 ms, can enable competitive networks to learn invariant representations of objects because objects are typically viewed for short periods in different transforms before another object is viewed (Földiák, 1991; Rolls, 1992). Indeed, this type of learning with a short-term memory trace provides the basis of a

Please cite this article as: Rolls, E.T., Deco, G., Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain. Brain Research (2014), http://dx.doi.org/10.1016/j.brainres.2014.09.021

computational model of how position, view, and rotation transform invariant representations are built in the ventral stream visual pathways (Rolls, 1992; Wallis and Rolls, 1997; Rolls, 2008, 2012b).

4. Neural coding in sensory and memory systems

Understanding how information is encoded in the brain in sensory and memory systems is considered briefly, for it too may provide a foundation for starting to develop biologically plausible approaches to understanding how language might be implemented in the brain.

In most cortical systems, information is encoded by which neurons are firing, and how fast they are firing (Rolls, 2008; Rolls and Treves, 2011; Rolls, 2012b). This has been evident since even before the exciting work of Hubel and Wiesel (1968, 1977) which showed for example that in the primary visual cortex which neuron is firing conveys information about features and their location, for example about orientation of a bar or edge and its location in retinal space. The crucial point here is that it is which neurons are firing that conveys the information about the object and the spatial relations of its parts. The principle has been extended to high order visual cortical areas: in the inferior temporal visual cortex one reads off information from which neurons are firing about which face or object is being shown (Rolls et al., 1997; Rolls, 2008; Rolls and Treves, 2011). This does not at all mean that this is local or grandmother cell encoding, in which just one neuron encodes which stimulus is present. Instead, there is a sparse distributed code, in which a small proportion of neurons is firing with a particular distribution of firing rates to represent one object, and another but partially overlapping population of neurons is firing with a particular distribution of firing rates to represent another object (Rolls and Tovee, 1995; Rolls et al., 1997). In this encoding, the information increases approximately linearly with the number of neurons (up to reasonable numbers of neurons), showing that the coding by different neurons is independent (Rolls et al., 1997; Franco et al., 2007). This is a sparse distributed place code, in that it is which neurons are firing, and their relative firing rates, that encode which object is present (Rolls, 2008; Rolls and Treves, 2011). Similar encoding principles are used in the orbitofrontal cortex to encode information about taste and odour (Rolls et al., 2010), and in the hippocampus to encode information about spatial view (Rolls et al., 1998).

There are two important points to emphasise here about place coding. The first is that what is being represented is encoded by a neuronal place code in the brain, in that for example neurons in the inferior temporal visual cortex convey information about which visual object is present, in the primary taste cortex about which taste, texture, or temperature is present; in the orbitofrontal cortex about which odour is present, and in the hippocampus about which spatial view is present (Rolls and Treves, 2011). The second point is that if relational information about parts needs to be represented, as it must to define objects, then the relational information is encoded by which neurons are firing, where neurons are tuned not only to features or objects, but also to their location (Hubel and Wiesel, 1968, 1977; Aggelopoulos and Rolls, 2005; Rolls, 2008, 2012b).

Quantitative information theoretic analyses further show that relatively little information is encoded by stimulus-dependent cross-correlations between neurons, with typically 95% or more of the information being encoded by a place/firing rate code where the graded firing rates of each neuron in a sparse distributed representation are used as the encoding principle (Rolls, 2008; Rolls and Treves, 2011), as considered further in Section 6.1.1.

5. Short-term memory, attention, and decision-making

Local cortical attractor networks also provide a foundation for understanding other processes including short-term memory, attention, and decision-making (Rolls, 2008). Indeed, that was why that book was entitled *Memory, Attention, and Decision-Making: A Unifying Computational Neuroscience Approach* (Rolls, 2008). All of these processes are also relevant to language as follows, so the same neural foundation, of cortical attractor networks, is even more relevant.

Short-term memory is typically implemented in the brain by continuing neuronal firing (Fuster, 2008; Goldman-Rakic, 1996) in what appear to be attractor states (Rolls, 2008; Rolls et al., 2013). This is highly relevant to the implementation of language in the brain, for we may wish to hold the components of a sentence active, so that it can be checked and if necessary corrected during its execution (by for example a higher order syntactic thought, Rolls, 2014), and even to guide the execution of the remainder of the sentence.

Top-down attention can be implemented by biasing the operation of attractor networks to reflect the subject of our attention which is held in a short-term memory store (Rolls and Deco, 2002; Deco and Rolls, 2005a; Rolls, 2008, 2013a), and this is likely to be an important component of how our thinking and sentence production are kept focussed and on target.

Decision-making can be implemented by providing two competing inputs to an attractor network, which can then fall into an attractor basin depending on which input is stronger (Wang, 2002; Rolls, 2008; Rolls and Deco, 2010; Deco et al., 2013; Rolls, 2014). The ability to categorise a potentially ambiguous input in this way may be a very useful component for language, for example in enabling a clear interpretation to be reached if there is some ambiguity in what may be heard or meant. An advantage of this mechanism is that not only does the attractor mechanism lead to a definite result, which is better than a stalemate with no winner or decision (and which may utilise neuronal spiking related noise if the inputs are of similar strength to produce a definite outcome Rolls and Deco, 2010), but also the same mechanism allows the decision or interpretation to be held online in short-term memory to influence further (e.g. language) processing.

6. Neurodynamical hypotheses about language

When considering the computational processes underlying language, it is helpful to analyze the rules being followed (Chomsky, 1965; Jackendoff, 2002). From this, it is tempting to see what one can infer about how the computations are implemented, using for example logical operations within a rule-based system, and switches turned on during development.

In this paper, instead the approach is to take what are understood as some of the key principles of the operation of the cerebral cortex (Rolls, 2008) based on the operation of memory and sensory systems in the brain, and how information is encoded in the brain (summarised in Sections 2–5), and then to set up hypotheses about how some of these computational mechanisms might be useful and used in the implementation of language in the brain. Later in this paper we then test and elucidate some of these hypotheses by simulations of some of the neuronal network operations involved.

6.1. *Syntax and binding*

6.1.1. *Binding by synchrony*

A fundamental computational issue is how the brain implements binding of elements such as features with the correct relationship between the elements. The problem in the context of language might arise if we have neuronal populations each firing to represent a subject, a verb, and an object of a sentence. If all we had were three populations of neurons firing, how would we know which was the subject, which the verb, and which the object? How would we know that the subject was related to the verb, and that the verb operated on the object? How these relations are encoded is part of the problem of binding.

Von der Malsburg (1990) considered this computational problem, and suggested a dynamical link architecture in which neuronal populations might be bound together temporarily by increased synaptic strength which brought them into temporary synchrony. This led to a great deal of research into whether arbitrary relinking of features in different combinations is implemented in visual cortical areas (Singer et al., 1990; Engel et al., 1992; Singer and Gray, 1995; Singer, 1999; Abeles, 1991; Fries, 2009), and this has been modelled (Hummel and Biederman, 1992). However, although this approach could specify that two elements are bound, it does not specify the relation (Rolls, 2008). For example, in vision, we might know that a triangle and a square are part of the same object because of synchrony between the neurons, but we would not know the spatial relation, for example whether the circle was inside the triangle, or above, below it, etc. Similarly for language, we might know that a subject and an object were part of the same sentence, but we would not know which was the subject and which the object, that the subject operated (via a verb) on the object, etc, that is, the syntactic relations would not be encoded just by synchrony. Indeed, neurophysiological recordings show that although synchrony can occur in a dynamical system such as the brain, synchrony *per se* between neurons in high order visual cortical areas conveys little information about which objects are being represented, with 95% of the information present in the number of spikes being emitted by each of the neurons in the population (Rolls, 2008; Rolls and Treves, 2011).

Instead, in high order visual cortical areas, the spatial relations between features and objects are encoded by neurons that have spatially biased receptive fields relative to the fovea (Aggelopoulos and Rolls, 2005; Rolls, 2008, 2012b), and this feature/place coding scheme is computationally feasible (Elliffe et al., 2002). In addition, coherence and temporal synchrony do not appear to be well suited for information transmission, for in quantitative neuronal network simulations, it is found that information is transmitted between neuronal populations at much lower values of synaptic strength than those needed to achieve coherence (Rolls et al., 2012).

In this situation, I now make alternative proposals for how syntactic relations are encoded in the cortex.

6.1.2. *Syntax using a place code*

The brief overview of encoding in Section 4 (Rolls, 2008; Rolls and Treves, 2011) leads to a hypothesis about how syntax, or the relations between the parts of a sentence, is encoded for language. The hypothesis is that a place code is used, with for example one cortical module or region used to represent subjects, another cortical module used to represent verbs, and another cortical module used to represent objects.

The size of any such neocortical module need not be large. An attractor network in the cortex need occupy no more than a local cortical area perhaps 2–3 mm in diameter within which there are anatomically dense recurrent collateral associatively modifiable connections between the neurons (Rolls, 2008). This cortical computational attractor network module would thus be about the size of a cortical column. It is an attractor network module in the sense that neurons more than a few mm away would not be sufficiently strongly activated to form part of the same attractor network (Rolls, 2008). An attractor network of this type with sparse distributed representations can store and encode approximately as many items are there are synaptic connections onto each cortical neuron from the nearby neurons (Treves and Rolls, 1991; Rolls and Treves, 1998; Rolls, 2008). The implication for language is that of order 10,000 nouns could be stored in a single cortical attractor network with 10,000 recurrent collateral connections onto each neuron. This capacity is only realised if there is only a low probability of more than one recurrent collateral connection between any pair of the neurons in a module, and this has been proposed as one of the underlying reasons for why cortical connectivity is diluted, with a probability in the order of 0.1 for connections between any pair of nearby neurons in the neocortex, and 0.02 for CA3 neurons in the rodent hippocampus (Rolls, 2012a).

The hypothesis is further that different local cortical modules encode the nouns that are the subjects of sentences, and that are the objects of sentences. A prediction is thus that there will be single neurons in a human cortical language area that respond to a noun when it is the subject but not the object of a sentence, and vice versa.

Clearly the full details of the system would be more complicated, but the general hypothesis is that adjectives and adjectival phrases that are related to the subject of a sentence will have strong connections to the subject module or modules; that adverbs and adverbial phrases that are related to the verbs of sentence will have strong connections to the verb module or modules; and that adjectives and adjectival phrases that are related to the object of a sentence will have strong connections to the object module or modules.

6.1.3. *Temporal trajectories through a state space of attractors*

To represent syntactical structure *within* the brain, what has been proposed already might be along lines that are consistent with the principles of cortical computation (Rolls, 2008). The high representational capacity would be provided for by the high capacity of a local cortical attractor network, and syntactic binding within a brain would be implemented by using a place code in which the syntactic role would be defined by which neurons are firing – for example, subjects in one cortical module or modules, and objects in another cortical module or modules.

However, a problem arises if we wish to communicate this representation to another person, for the neural implementation described so far could not be transferred to another person without transferring which neurons in the language areas were currently active, and having a well trained person as the decoder!

To transfer or communicate what is encoded in the representations to another person, and the relations or syntax, it is proposed that a number of mechanisms might be used. One might be a temporal order encoding, for example the subject–verb–object encoding that is usual in English, and which has the advantage of following the temporal order that usually underlies causality in the world. Another mechanism might be the used of inflections (usually suffixes) to words to indicate their place in the syntax, such as cases for nouns (e.g. nominative for the subject or agent, and accusative for the object or patient, dative, and genitive), and person for verbs (e.g. first, second, and third person singular and plural, to specify I, you, he/she/it, we, you, they) used to help disambiguate which noun or nouns operate on the verb. Another mechanism is the use of qualifying prepositions to indicate syntactic role of a temporally related word, with examples being 'with', 'to', and 'from'. This mechanism is used in combination with temporal order in English. In this paper, I focus on temporal order as an encoder of syntactical relations, and next set out hypotheses on how this could be implemented in the cerebral cortex based on the above computational neuroscience background.

6.1.4. *Hypotheses about the implementation of language in the cerebral cortex*

1. Subjects, verbs, and objects are encoded using sparse distributed representations (Rolls, 2008; Rolls and Treves, 2011) in localised cortical attractor networks. One cortical module with a diameter of 2–3 mm and 10,000 recurrent collateral connections per neuron could encode in the order of 10,000 items (e.g. subjects, verbs or objects) (Rolls, 2008; Treves and Rolls, 1991). One cortical module would thus be sufficient to encode all the objects, all the verbs, or all the objects (depending on the module) in most people's working vocabulary, which is of the order to several thousand nouns, or verbs.

This follows from the analysis that the capacity of an attractor network with sparse encoding a (where for binary networks a is the proportion of neurons active for any one memory pattern) is as follows, and from the fact that there are in the order of 10,000 recurrent collateral connections on each neuron (Rolls, 2008). The capacity is measured by the number of patterns p that can be stored and correctly retrieved from the attractor network

$$p \approx \frac{C}{a \ln\left(\frac{1}{a}\right)} k \qquad (1)$$

where C is the number of synapses on the dendrites of each neuron devoted to the recurrent collaterals from other neurons in the network, and k is a factor that depends weakly on the detailed structure of the rate distribution, on the connectivity pattern, etc., but is roughly in the order of 0.2–0.3 (Treves and Rolls, 1991; Rolls, 2008, 2012a).

The use of attractor networks for language-based functions is itself important. Cortical computation operates at the neuronal level by computing the similarity or dot product or correlation between an input vector of neuronal firing rates and the synaptic weights that connect the inputs to the neurons (Rolls, 2008, 2012c). The output of the neuron is a firing rate, usually between 0 and 100 spikes/s. Cortical computation at the neuronal level is thus largely analogue. This is inherently not well suited to language, in which precise, frequently apparently logical, rules are followed on symbolic representations. This gap may be bridged by attractor or autoassociation networks in the brain, which can enter discrete attractor high firing rate states that can provide for error correction in the analogue computation, and to the robustness to noise, that are often associated with the processing of discrete symbols (Treves, 2005; Rolls, 2008, 2012c). These discrete attractor states are network properties, not the properties of single neurons, and this capability is at the heart of much cortical computation including long-term memory, short-term memory, and decision-making (Rolls, 2008, 2014).

2. Place coding with sparse distributed representations is used in these attractors. The result is that the module that is active specifies the syntactic role of what is represented in it. One cortical module would be for subjects, another for verbs, another for objects, etc.

3. The presence of these weakly coupled attractors would enable linguistic operations of a certain type to be performed within the brain, but the information with the syntax could not be communicated in this form to other people. In this statement, 'weakly coupled' is clearly and quantitatively defined by attractors that have weak interconnections so that they can have different basins of attraction, yet can influence each other (Rolls, 2008) (Section B.9). The computations involved in these interactions might instantiate a 'language of thought' that would be below the level of written or spoken speech, and would involve for example constraint satisfaction within coupled attractors, and the types of brain-style computation described by Rolls (2012c) in section 2.15 on brain computation vs computation in a digital computer. The cortical computational processes could be usefully influenced and made creative by the stochastic dynamics of neuronal networks in the brain that are due to the 'noisy' Poisson-like firing of neurons (Rolls, 2008; Rolls and Deco, 2010). When someone has a hunch that they have solved a problem, this may be the computational system involved in the processing.

This might be termed a 'deep' structure or layer of linguistic processing.

4. To enable these computations that involve syntactical relations to be communicated to another person or written down to be elaborated into an extended argument, the process considered is one involving weakly forward coupled attractor networks. One such system would be to have weak forward coupling between subject–verb–object attractor networks. The exact trajectories followed (from subject to verb to object) could be set up during early language learning, by forming during such learning stronger forward than reverse connections between the attractors, by for example spike-timing dependent plasticity (Markram et al., 1997; Bi and Poo, 1998; Feldman, 2012) and experience with the order of items that is provided during language learning. Which trajectory was followed would be biased by which subject, which verb, and which object representation was currently active in the deep layer. These temporal trajectories through the word attractors would enable the syntactical relations to be encoded in the temporal order of the words.

With this relatively weak coupling between attractors implemented with integrate-and-fire neurons and low firing rates, the transition from one active attractor to the next can be relatively slow, taking 100–400 or more ms (Deco and Rolls, 2005b). This property of the system adapts it well to the production of speech, in which words are produced sequentially with a spacing in the order of 300–500 ms, a rate that is influenced by the mechanics and therefore dynamics of the speech production muscles and apparatus.

A simulation testing this system and making the details of the operation of the system and their biological plausibility clear is described in the Methods and Results sections of this paper.

5. The system for enabling the syntax to be communicated to other people or written down would have some computational advantages apart from purely the communication. In particular, once the syntax can be formally expressed in written statements, it becomes easier to perform logical operations on the statements, which become propositional, and can be tested. These logical operations, and reasoning, may not be the style of computation utilised in general by computational processes within the brain (see Rolls, 2012c Section 2.15), but may become algorithms that can be followed to achieve quantitatively precise and accurate results, as in long division, or by learning logic. Thus the importance of communication using syntax may allow other environmental tools to be applied to enable reasoning and logic that is not the natural style of neural computation.

6. To enable the system to produce words in the correct temporal order, and also to remember with a lower level of

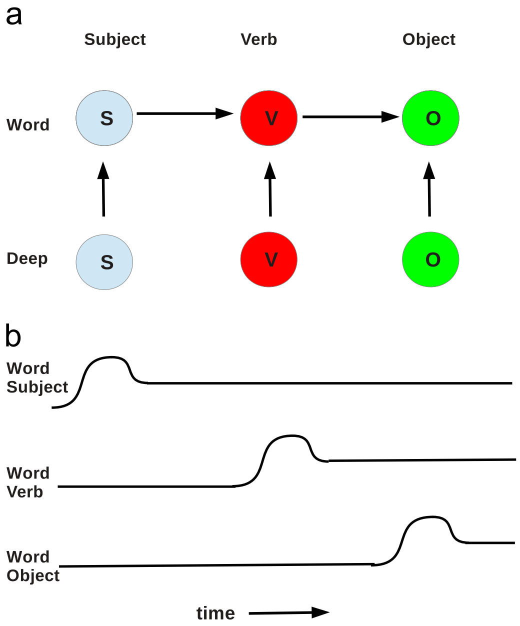

Fig. 2 – Schematic diagram of the concepts. (a) Each circle indicates a local cortical attractor network capable of storing 10,000 items. The Deep layer local cortical attractor networks use place coding to encode the syntax, that is, the syntactic role of each attractor network is encoded by where it is in the cortex. In this implementation there are separate Subject (S), Verb (V), and Object (O) networks. Syntax within the brain is implemented by place encoding. For communication, the deep representation must be converted into a sequence of words. To implement this, the Deep attractor networks provide a weak selective bias to the Word attractor networks, which have weak non-selective forward coupling S to V to O. (b) The operation in time of the system. The Deep networks fire continuously, and the syntax is implemented using place encoding. The Deep networks apply a weak selective bias to the Word networks, which is insufficient to make a word attractor fire, but is sufficiently strong to bias it later into the correct one of its 10,000 possible attractor states, each corresponding to a word. Sentence production is started by a small extra input to the Subject Word network. This with the selective bias from the Deep subject network make the Word subject network fall into an attractor, the peak firing of which is sufficient to elicit production of the subject word. Adaptation in the subject network makes its firing rate decrease, but still remain in a moderate firing rate attractor state to provide a short-term memory for the words uttered in a sentence, in case they need to be corrected or repeated. The high and moderate firing rate in the Subject Word network provides non-selective forward bias to the whole of the Object Word network, which falls into a particular attractor produced by the selective bias from the Deep Verb network, and the verb is uttered. Similar processes then lead to the correct object being uttered next. The sentence can be seen as a trajectory through a high dimensional state space of words in which the particular words in the sentence are due to the selective bias from the Deep networks, and the temporal order is determined by the weak forward non-selective connections between the networks, i.e. connections from subject-to-verb, and verb-to-object networks. The simulations show dynamical network principles by which this type of sentence encoding and also decoding could be implemented. The overall concept is that syntax within the brain can be solved by the place coding used in most other representations in the brain; and that the problems with syntax arise when this place-coded information must be transmitted to another individual, when one solution is to encode the role in syntax of a word by its temporal order in a sentence.

Please cite this article as: Rolls, E.T., Deco, G., Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain. Brain Research (2014), http://dx.doi.org/10.1016/j.brainres.2014.09.021

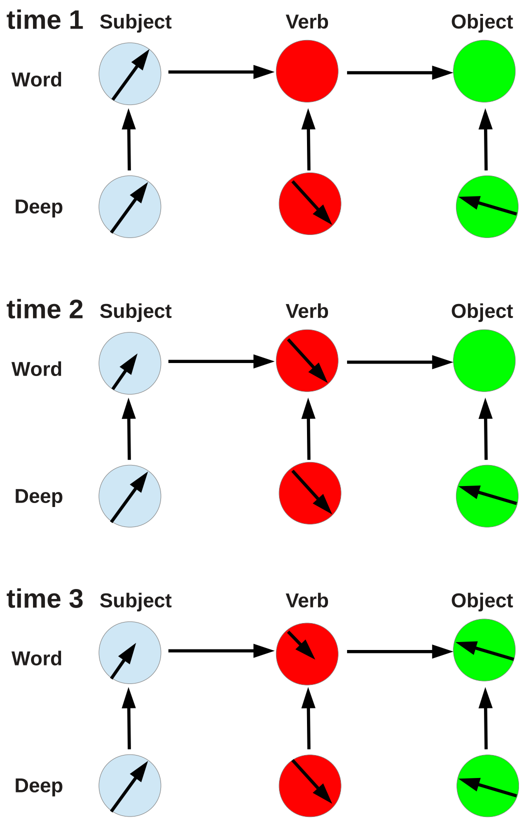

Fig. 3 – The operation of the model, illustrated at three different time steps. Time 1 is during the production of the subject work, time 2 during the verb word, and time 3 during the object word. Each circle represents an attractor network that can fall into one of 10,000 possible states indicated by the direction of the arrow vector, each state corresponding to a word (for the Word-level networks) or a semantic concept (for the Deep-level networks). The length of the arrow vector indicates the firing rate of the selected attractor. The Deep attractor networks fire continuously, with the syntactic role indicated by the particular network using place coding. The Deep networks, which represent semantics or meaning, provide a weak selective bias continuously to the Word attractor networks. The sequential operation of the Subject then Verb then Object Word networks is produced by the weak non-selective forward connections between the networks. After its initial high firing rate, a Word attractor network remains active at a lower firing rate as a result of adaptation, to provide a short-term memory for the words uttered in a sentence, in case they need to be corrected or repeated.

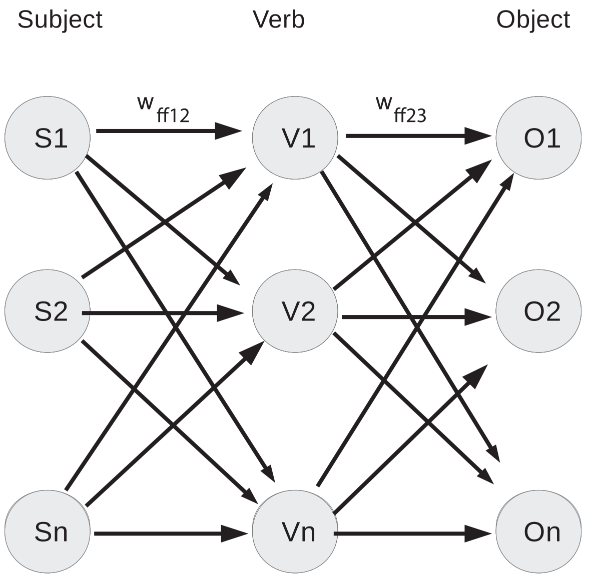

Fig. 4 – The attractor network model. There are three modules, Subject (S), Verb (V), and Object (O). Each module is a fully connected attractor network with $n=10$ pools of excitatory neurons. The ten excitatory pools each have 640 excitatory neurons, and each module has 1600 inhibitory neurons using GABA as the transmitter. Each excitatory pool has recurrent connections with strength $w_+=2.1$ to other neurons in the same pool implemented with AMPA and NMDA receptors. There are forward connections with strength w_{ff12} from all excitatory neurons in module 1 (S) to all excitatory neurons in module 2 (V). There are forward connections with strength w_{ff23} from all excitatory neurons in module 2 (V) to all excitatory neurons in module 3 (O). An external bias can be applied to any one or more of the attractor pools in each of the modules. In operation for production, a stronger bias is applied to one pool in module 1 to start the process, and then an attractor emerges sequentially in time in each of the following modules. The particular pool that emerges in each of the later modules depends on which pool in that module is receiving a weak bias from another (deeper) structure that selects the items to be included in a sentence. The syntax of the sentence, encoded in the order of the items, is determined by the connectivity and dynamics of the network. The same network can be used for decoding (see text).

neuronal firing what has just been said for monitoring in case it needs correcting, a mechanism such as spike frequency adaptation may be used, as described next.

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

When recordings are made from single neurons operating in physiological conditions in the awake behaving monkey, peristimulus time histograms of inferior temporal cortex

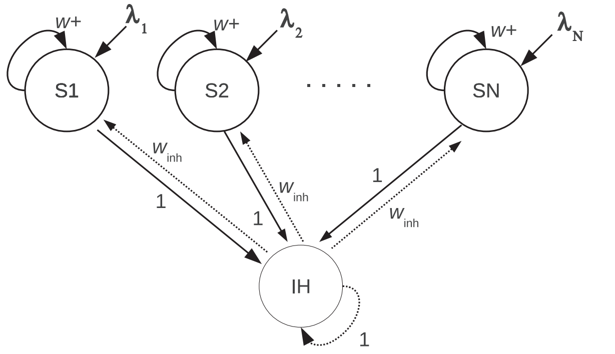

Fig. 5 – The architecture of one module containing one fully connected attractor network. The excitatory neurons are divided into $N=10$ selective pools or neuronal populations S1–SN of which three are shown, S1, S2 and SN. The synaptic connections have strengths that are consistent with associative learning. In particular, there are strong intra-pool connection strengths w_+. The excitatory neurons receive inputs from the inhibitory neurons with synaptic connection strength $w_{\text{inh}}=1$. The other connection strengths are 1. The integrate-and-fire spiking module contained 8000 neurons, with 640 in each of the 10 non-overlapping excitatory pools, and 1600 in the inhibitory pool IH. Each neuron in the network also receives external Poisson inputs λ_{ext} from 800 external neurons at a typical rate of 3 Hz/synapse to simulate the effect of inputs coming from other brain areas.

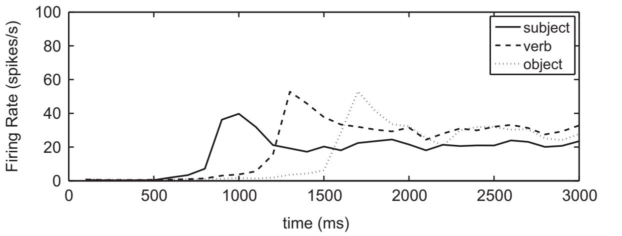

Fig. 6 – Firing rates of the biased pools in the Subject–Verb–Object modules as a function of time.

neurons to visual stimuli show only limited adaptation. There is typically an onset of the neuronal response at 80–100 ms after the stimulus, followed within 50 ms by the highest firing rate. There is after that some reduction in the firing rate, but the firing rate is still typically more than half-maximal 500 ms later (see example in Tovee et al., 1993). Thus under normal physiological conditions, firing rate adaptation can occur.

The effects of this adaptation can be studied by including a time-varying intrinsic (potassium-like) conductance in the cell membrane (Brown et al., 1990; Treves, 1993; Rolls, 2008). This can be done by specifying that this conductance, which if open tends to shunt the membrane and thus to prevent firing, opens by a fixed amount with the potential excursion associated with each spike, and then relaxes exponentially to its closed state. In this manner sustained firing driven by a constant input current occurs at lower rates after the first few spikes, in a similar way, if the relevant parameters are set appropriately, to the behaviour observed in vitro of many pyramidal cells (for example, Lanthorn et al., 1984; Mason and Larkman, 1990). The details of the implementation used are described in the Methods.

7. Results

7.1. *A production system*

The operation of the integrate-and-fire system illustrated in Fig. 4 is shown in Fig. 6 when it is producing a subject – verb – object sequence. The firing rates of attractor pools 1 in Word modules 1 (subject), 2 (verb), and 3 (object) are shown. No other pool had any increase of its firing rate above baseline. The attractor pools 1 in modules 2 and 3 received an increase above the baseline rate of 3.00 Hz per synapse (or 2400 spikes/s per neuron given that each neuron receives these inputs through 800 synapses) to 3.03 Hz per synapse throughout the trial. This itself was insufficient to move any attractor pool in modules 2 and 3 into a high firing rate state, as illustrated, until one of the attractor pools in a preceding module had entered a high firing rate attractor state. At time=500 ms, the external input into attractor pool 1 of module 1 (subject) was increased to 3.20 Hz per synapse, and maintained at this value for the rest of the trial. This produced after a little delay due to the stochastic recurrent dynamics an increase in the firing of pool 1 in module 1, which peaked at approximately 1000 ms. The initial peak firing rate of approximately 40 spikes/s was followed by a reduction due to the spike frequency adaptation to approximately 25 spikes/s, and this level was maintained for the remainder of the trial, as shown in Fig. 6. The parameters for the spike frequency adaptation for all the excitatory neurons in the network were $V_k=-80$ mV, $g_{AHP}=200$ nS, $\alpha\text{Ca}=0.002$, and $\tau\text{Ca}=300$ ms.

The increase in the firing in attractor pool 1 of module 1 influenced the neurons in module 2 via the feedforward synaptic strength of $w_{\text{ff12}}=0.55$, and, because attractor pool 1 in module 2 already had a weak external bias to help it be selected, it was attractor pool 1 in module 2 that increased its firing, with the peak rate occurring at approximately 1250 ms, as shown in Fig. 6. Again, the peak was followed by low maintained firing in this pool for the remainder of the trial.

The increase in the firing in attractor pool 1 of module 2 influenced the neurons in module 3 via the feedforward synaptic strength of $w_{\text{ff23}}=0.4$, and, because attractor pool 1 in module 3 already had a weak external bias to help it be selected, it was attractor pool 1 in module 3 that increased its firing, with the peak rate occurring at approximately 1650 ms, as shown in Fig. 6. Again, the peak was followed by low maintained firing in this pool for the remainder of the trial. The value of w_{ff23} was optimally a little lower than that of w_{ff12}, probably because with the parameters used the firing in module 2 was somewhat higher than that in module 1. All subsequent stages would be expected to operate with $w_{\text{ff}}=0.4$.

The network thus shows that the whole system can reliably perform a trajectory that is sequential and delayed at each step through the state space, with the item selected in each module determined by the steady bias being received from a deep structure containing the items to be included in the sentence. However, the order in which the items were produced and which specified the syntax was determined by the connectivity and dynamics of the Word network. In this

Please cite this article as: Rolls, E.T., Deco, G., Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain. Brain Research (2014), http://dx.doi.org/10.1016/j.brainres.2014.09.021

particular example, the resulting sentence might correspond to 'James chased John', which has a completely different syntax and meaning to 'John chased James', 'chased John James', etc. The peak of firing in each module could be used to produce the appropriate word in the correct order. The lower rate of ongoing firing in each module provided the basis for the items produced in the sentence to be remembered and used for monitoring, with the role of each item in the sentence made explicit by the module in which the attractor pool was still active.

7.2. *A decoding system*

In Section 7.1 the operation of the system when it is operating to produce a subject–verb–object sequence in which the temporal sequence encodes syntactic information is described. In this section, the operation of the system when it decodes a Subject–Verb–Object sentence is considered. The aim is to receive as input the temporal sequence, and to activate the correct attractor in the Subject, Verb, and Object Word attractor modules. The syntactic information in the sequence allows correct decoding of the subject and object nouns in the sentence, when the position in the sequence is the only information that enables a noun to activate a noun attractor in the subject or the object module. The deep semantic attractor modules could then be activated from the Word attractor modules, using selective, associatively modifiable, synaptic connections.

The operation of the system in decoding mode is illustrated in Fig. 7. The architecture of the network is the same as that already described. The baseline input to each of the 800 external synapses is maintained at 3 Hz per synapse in all pools in all modules throughout the sentence except where stated. Throughout the trial all pools in module 1 receive a bias of 0.24 Hz on each of the 800 external input synapses. (This corresponds to an extra 192 spikes/s received by every neuron in each of attractor pools.) The aim of this is to prepare all the attractors in module 1, the subject attractor, to respond if an input cue, a word, is received. This bias essentially sets the system into a mode where it is waiting for an input stream to arrive in module 1. All the attractors in module 1 are stable with low firing rates while only this bias is being applied.

At time 500–1000 ms module 1 pool 1 and module 3 pool 1 receive a noun recall cue as an additional input on the external synapses at an additional 0.08 Hz per synapse. (This corresponds to an extra 64 spikes/s received by every neuron in these two pools of neurons.) Module 1 pool 1 goes into an attractor, as illustrated in Fig. 7, because it is receiving a noun recall cue and the bias. Module 3 pool 1 does not enter a high firing rate attractor state, even though the word recall cue for its attractor 1 is being applied, because it is not receiving a bias. This shows how the system can decode correctly a noun due to its position in the sequence as a subject or as an object. In the simulations the bias to pool 1 can be left on, or turned off at this stage in the trial, for once an attractor state has been reached by a pool in module 1, it remains with a stable high firing rate for the remainder of the sentence. The continuing firing is implemented to ensure that the subject remains decoded in the system while the rest of the sentence is decoded, and for use even after the end of the sentence. At time 1000 ms, the noun applied to attractor pools 1 in modules 1 and 3 is removed, as it is no longer present in the environment.

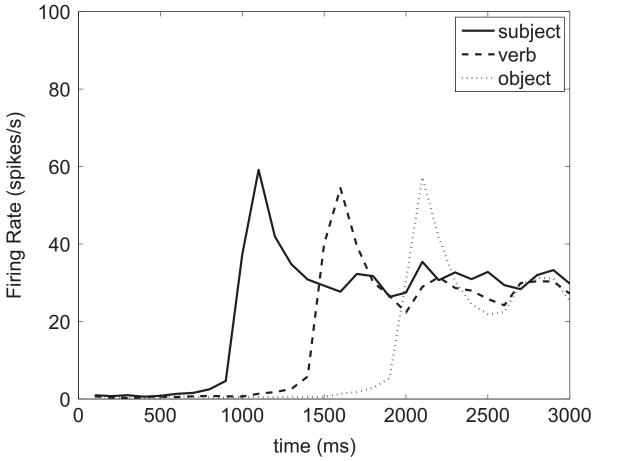

Fig. 7 – Decoding the Subject–Verb–Object sequence to produce activation in the Subject (module 1), Verb (module 2), and Object (module 3) modules. A weak bias was applied to all pools in module 1 throughout the trial (see text). Noun 1, the subject, was applied to module 1 pool 1 and module 3 pool 1 during the period 500–1000 ms. Verb 2, was applied to module 2 pool 2 during the period 1000–1500 ms. Noun 3, the object, was applied to module 1 pool 3 and module 3 pool 3 during the period 1500–2000 ms. The firing of module 1 attractor 1 neurons that reflect the decoded subject, of module 2 attractor 2 neurons that reflect the decoded verb, and of module 3 attractor 3 neurons that reflect the decoded object, are shown. None of the other 30 attractor neural populations became active.

At time 1000–1500 ms the verb recall cue is applied to module 2 pool 2, which enters an attractor. The strength of this recall cue alone (the same as before, an additional 0.08 Hz per synapse) is insufficient to cause this pool in module 2 to enter an attractor. However, all pools in module 2 are receiving now via the feed-forward connections w_{ff12} a priming input from the firing now occurring in module 1, and when the verb recall cue is applied to module 2 pool 2, the combined effects of the recall cue and the feedforward inputs cause module 2 pool 2 to enter its correct attractor state to indicate the presence of this verb in the sentence. The intention of the priming forward input from the preceding module is to provide for future expansion of the system, to allow for example correct decoding of two verbs at different positions within the sequence of words in a sentence. Module 2 pool 2 has the Verb recall cue removed at time 1500 ms as it is no longer present in the environment. Module 2 pool 2 however keeps firing in its stable high firing rate attractor state to ensure that the verb remains decoded in the system while the rest of the sentence is decoded, and for use even after the end of the sentence.

At time 1500–2000 ms the object recall cue is applied to module 3 pool 3, and, as a control and test of the syntactic operation of the system, simultaneously also to module 1 pool 3. Module 3 pool 3 enters its correct object attractor, utilising the feedforward priming inputs w_{ff23}. These priming inputs again provide for further expansion of the system, in case there is another object in the sentence in a later clause. Meanwhile module 1 remains in its pool 1 subject attractor state which is now stable in the face of interfering noun inputs because of its deep low energy basin of attraction. This shows how this noun is forced by its order in the sequence into the Object pool, demonstrating how the system is able to decode information about syntactic role that is present from the position of the item in the sequence. Module 3 pool 3 keeps firing in its stable high firing rate attractor state to ensure that the object remains decoded in the system for use even after the end of the sentence.

As noted, the architecture and overall dynamical principles of operation of the system used for decoding were the same as for encoding. The firing rate adaption was left to operate as before, though it is less useful for the decoding. The only parameters that were adjusted a little for the decoding system were $w_+ = 2.3$ (to help stability of the high firing rate attractor state in the face of for example interfering nouns); $w_{ff12} = w_{ff23} = 0.2$; bias for all pools in module $1 = 0.24$ Hz per synapse; and recall cue = 0.08 Hz per synapse.

The results just described and illustrated in Fig. 7 illustrate some of the principles of operation of the functional architecture when decoding sentences. Further results were as follows. If during the application of the noun for the subject (time 500–1000 ms in the simulations) an input effective for an attractor in module 2 was also applied, then a pool in module 2 tends to enter a high firing rate attractor state, for it is receiving both a recall cue and the forward bias from the high firing starting in module 1 and applied to module 2 via w_{ff12}. An implication is that only nouns should be applied to Subject and Object attractor modules. Having attractors for verbs that respond to different recall cues (words that are verbs) helps the system to decode the input stream into the correct modules and pools. Thus the semantics, the words being applied to the network, are important in enabling the system to respond correctly.

8. Discussion

The system described here shows how a word production system might operate using neuronal architecture of the type found in the cerebral cortex.

One possible objection to such a computational implementation is how to deal with the passive form. What I propose is that temporal order could again be used, but with different coupling between the attractors appropriate for implementing the passive voice that is again learned by early experience, and is selected instead of the active voice by top-down bias in the general way that we have described elsewhere (Deco and Rolls, 2003, 2005c). The hypothesis is thus that operation of the system for passive sentences would require a different set of connections to be used to generate the correct temporal trajectory through these or possible different modules, with the head of the sentence no longer being (in English) the subject (e.g. James in 'James chased John'), but instead the object (e.g. 'John was chased by James').

In a similar way, it is proposed that different languages are implemented by different forward connectivity between the different modules representing subjects, verbs, and objects in that language, with the connectivity for each language learned by repeated experience and forced trajectories during learning using for example spike-timing-dependent plasticity. Separate neuronal implementation of different languages is consistent with neurological evidence that after brain damage one language but not another may be impaired.

The system would require considerable elaboration to provide for adjectives and adjectival phrases qualifying the subject or the object, and for adverbs or adverbial phrases qualifying the verbs, but a possible principle is stronger synaptic connectivity between the modules in which the qualifiers are represented. To be specific, one type of implementation might have adjectives in modules that qualify subjects connected with relatively stronger synapses to subject modules than to object modules. This should be feasible given that any one attractor network capable of encoding thousands of words need occupy only 2–3 mm of neocortical area.

In such a system, the problem does arise of how the nouns in the subject attractor module can refer to the same object in the word as the nouns in the object attractor, and of the extent to which when one representation (e.g. in the subject module) is updated by modifying its properties (encoded by which neurons are active in the sparse distributed representation within a module), the representation of the same object in another module (e.g. in the object module) is updated to correspond.

The results on the operation of the system when it is decoding a sentence illustrated in Fig. 7 illustrate how the temporal sequence of the words can be used to place them into the appropriate module, for example to place a noun into a subject or an object module. Interestingly, if during the application of the noun for the subject (time 500–1000 ms in the simulations) an input effective for an attractor in module 2 (the verb module) was also applied, then a pool in module 2 tended to enter a high firing rate attractor state, for it was receiving both a recall cue and the forward bias from the high firing starting in module 1 and applied to module 2 via w_{ff12}. An implication is that only nouns should be applied to Subject and Object attractors. Having attractors for verbs that respond to different input cues (words that are verbs) helps the system to decode the input stream into the correct modules and pools. Thus the semantics, the words being applied to the network, are important in enabling the system to respond correctly.

Indeed, overall the system might be thought of as having different modules and pools for different types of word (subject noun, object noun, verb, adverb, adjective, etc) and using the match of the incoming word to the word defined in a module to provide an important cue to which module should have an attractor activated, and then adding to this the temporal sequence sensitivity also considered here to help disambiguate the syntax, for example whether a noun is a subject or an object. Thus the semantics, the word as a noun, verb, or potentially adjective or adverb, help the dynamics because the cue details (adverb, noun, verb, adjective etc) provides constraints on the trajectories and dynamics of the system, and thus on how it decodes input sequences.

Language would thus be brittle if there were not subject-noun, object-noun, verb, adjective, adverb etc pools. An inflected language helps words to activate the correct pools. If inflections are lost or not present in a language, then the order in a sequence can compensate to some extent, but the system still relies on the words activating selective pools, with the temporal dynamics used for example to disambiguate matters if a noun might otherwise be a subject or an object, or an adjective might qualify a subject vs an object, etc. Moreover, because language is often irregular, particular words must also favour particular dynamics/relations.

In such a stochastic dynamical system (Rolls, 2008; Rolls and Deco, 2010), speech or writing errors such as words appearing in the incorrect order, word substitution, and repetition of the same word, could be easily accounted for by failures in the stochastically influenced (Rolls, 2008; Rolls and Deco, 2010) transitions from one attractor to the next, and in the word selected in each attractor. This supports the proposed account of the cortical implementation of language.

Overall, in this paper the coding and dynamical principles of the operation of the cerebral cortex have been considered in the context of how they may be relevant to the implementation of language in the cerebral cortex. It has been proposed that the high capacity of local attractor networks in the neocortex would provide a useful substrate for representations of words with different syntactical roles, for example subject noun, object noun, adjective modifying a subject, adjective modifying an object, verb, and adverb. With this as a principle of operation, in an inflected language the words produced can have the appropriate suffix (typically) added to specify the module from which it originated and therefore its syntactic role in the sentence. In an inflected language, the inflections added to words can indicate case, person etc, and during decoding of the sentence (when listening or reading) these inflections can be used to help the word to activate attractors in the correct module. In a language without inflections, or that is losing inflections, the order in the sequence can be used to supplement the information present in the word to help activate a representation in the correct attractor. Examples of how cortical dynamics might help in this process both during production and during decoding are provided in the simulations in this paper.

Interestingly, at least during decoding, temporal dynamics alone was found to be brittle in enabling words to be decoded by the correct module, and the system was found to work much more robustly if words find matches only in different specialised modules, for example with nouns being decodable by only subject noun or object noun modules, verbs only be verb modules, etc. The actual decoding is of course a great strength of the type of attractor neuronal network approach described here, for attractor networks are beautifully suited to performing such decoding based on the vector dot product similarity of the recall cue to what is stored in the network (Rolls, 2008; Hopfield, 1982; Amit, 1989). The implication is that content addressable specialised word attractor cortical modules are important in the implementation of language, and that temporal dynamics utilising the order of a word in the sequence can be used to help disambiguate the syntactic role of a word being received by enabling it to activate a representation in the correct module, using mechanisms of the general type described.

This raises the interesting point that in the present proposal, the syntactic role of a representation is encoded in the brain by the particular cortical module that is active, with different cortical modules for different parts of speech. An implication is that for the internal operations of this syntactical system, the syntax is encoded by the module within which the representation is active, and this is a form of place coding. Much may be computed internally by such a system based on this specification of the syntactic role of each module in the thought process. The problem arises when these thoughts must be communicated to others. Then a production system is needed, and in this system the syntactic role of which module the representation arises from can be specified partly by the word itself (with noun words indicating that they arise from a subject or object noun representations, verb words indicating that they arise from a verb module); and this specification is supported by inflection and/or by temporal order information to help disambiguate the module from which the word originates. Then during decoding of a sentence, the word again allows it to match only certain modules, with inflection and/or order in the sequence being used to disambiguate the module that should be activated by the word.

Thus an internal language of thought may be implemented by allocating different cortical modules to different syntactic roles, and using place encoding. However, when language becomes externalised in the process of communication, the way in which language can be used as a computational mechanism may be enhanced. Once a language has the rules that allow syntactic role to be expressed in for example written form, this enables formal syntactic operations including logic to be checked in an extended argument or algorithm or proof, and this then provides language with much greater power, providing a basis for formal reasoned extended argument, which may not be a general property of neuronal network operations (Rolls, 2008), and which facilitates the use of the reasoned route to action (Rolls, 2014).

One of the hypotheses considered here is that place coding in quite small cortical modules approximately the size of a cortical column (i.e. a region within which there is a high density of local recurrent collaterals to support attractor functionality, and within which local inhibitory neurons operate) may be used to encode the syntactic role of a word might not easily reveal itself at the brain lesion or functional neuroimaging levels, which generally operate with less resolution than this. For example, such studies of the effects of brain damage and of activations do not provide clear evidence for segregation by syntactic role, such as noun-selective vs verb-selective areas (Vigliocco et al., 2011). Although effects produced by nouns vs verbs do segregate to some extent into the temporal and frontal lobes, this may be because the semantic associations of nouns with objects, and verbs with actions, will tend to activate different cortical areas because of the semantic not purely because of the syntactic difference (Vigliocco et al., 2011). One of the hypotheses developed here is that nouns as subjects and nouns as objects may use different place coding, and one way that this might become evident in future is if single neuron recordings from language areas support this. Indeed, it is a specific and testable prediction of the approach described here that some neurons in language-related cortical areas will have responses to words that depend on the syntactic role of the word. For example, such a neuron might respond preferentially to a noun word when it is a subject compared to when it is an object.

It will be interesting in future to investigate whether this approach based on known principles of cortical computation can be extended to show whether it could provide at least a part of the biological foundation for the implementation of language in the brain. Because of its use of place coding, the system would not be recursive. But language may not be recursive, and indeed some of the interesting properties but also limitations of language may be understood in future as arising from the limitations of its biological implementation.

We note that there are a number of biological mechanisms that might implement the slow transitions from one attractor state to another, some investigated by Deco and Rolls (2005b), and that we are not wedded to any particular mechanism. The speed with which the trajectory through the state space of attractors is executed will depend on factors such as the magnitude of the inputs including the biassing inputs, the strengths of the synapses between the modules, the NMDA dynamics, the effects of finite size noise which will be influenced by the number of spiking neurons, the dilution of the connectivity, and the graded nature of the firing (Rolls and Deco, 2010; Webb et al., 2011; Rolls and Webb, 2012). An implication is that the speed of speech production might be influenced by factors that influence for example NMDA receptors, such as dopamine via D1 receptors (Rolls and Deco, 2010).

One such extension in the future of the approach taken in this paper would be to extend the implementation from a single cortical attractor network for each linguistic type (subject nouns, object nouns, verbs, etc) to a set of attractor networks for each linguistic type. If each single local cortical attractor network could store say S patterns (the p referred to above), then how would the system operate with M such attractor nets or modules? There would be two types of connection in such a system. One would be the synaptic connections between the neurons in each attractor network or module. The other connections would be the typically weaker connections between cortical modules (see Rolls, 2008). The whole system of coupled interacting attractor nets is known as a Potts attractor (Treves, 2005; Kropff and Treves, 2005). In a language system representing for example nouns, one attractor net or 'unit' in Potts terminology might contain properties such as shape, another colour, another texture, etc, and the full semantic description of the object might be represented by which attractors in which of the M modules are active, that is in a high firing rate state. One advantage of such a system is that all of the properties of an object could be encoded in this way, so we could specify whether the hat is round, red, smooth in texture, etc. Associated with this advantage, the total capacity of the system, that is the number of possible objects that could be represented, is now proportional to S^2. Thus if a single attractor network could store $S=10{,}000$ items (10^4), a Potts system with 5 such modules might be able to represent of order 5.10^8 such objects. In more detail, the number of patterns P_c that can be represented over the Potts attractor is

$$P_c \approx c_M S^2 / a_M \tag{2}$$

where c_M is the number of other modules on average with which a module is connected, S is the number of different attractor states within any one module, and a_M is the proportion of the attractor modules in which there is an active attractor, i.e. a high firing rate attractor state (Treves, 2005; Kropff and Treves, 2005). Such a Potts system only works well if (1) long-range connections between the different modules are non-uniformly distributed and (2) only a sparse set of the modules (measured by a_M) is in a high firing rate attractor state (Treves, 2005). In principle, the dynamical system described here could be replaced by substituting each cortical word type module (e.g. that for subject nouns) with a Potts attractor system each with several attractor network modules coding for different properties of features such as shape, colour, texture, etc.

An overview at present is that such a Potts system might be useful for a semantic network (Treves, 2005), which might correspond to the deep network that biases the word modules in the architecture described in this paper. That semantic system would then bias the word modules (with one cortical module for each type of word, subject noun, etc), and this architecture might have the advantage that word representations may be more uncorrelated than are semantic representations, which would keep the word representation capacity high. However, in the Potts system simulated so far to model language, the units corresponded to semantic features not to words; and place coding was not used, with instead the syntactic roles of the semantic representations requiring further Potts units to specify the syntactic roles of the semantic units (Pirmoradian and Treves, 2013).

Treves (2005) also considered how such a Potts system might have dynamics that might display some of the properties of language. Adaptation was introduced into each of the attractor networks. After a population of neurons had been active in a high firing rate state in one attractor module for a short time, due to adaptation in that neuronal population, the system then jumped to another attractor state in the same module, or in another connected module a jump might occur to another attractor state, because its inputs had changed due to the adaptation in the other module. The result was a latching process that resulted in complex dynamics of the overall system (Treves, 2005; Song et al., 2014). Whether that complex dynamics is how language is produced has not been proven yet. My view here is that there are sensory inputs from the world, or remembered states that enter short-term memory, and that these states in the deep, semantic, networks then bias the word networks to produce the sequential stream of language. That keeps the processing not a random trajectory through a state space perhaps biased by the statistics of the correlations within a language, but instead a trajectory useful for communication that reflects the states produced by the world or recalled into memory and that can then be communicated to others as a stream of words.

In conclusion, we have shown in this paper how some of the principles of cortical computation used to implement episodic memory, short-term memory, perception, attention, and decision-making (Rolls, 2008) might contribute to the implementation of language including syntax in the cortex.

9. Experimental Procedures

9.1. *An integrate-and-fire network with three attractor network modules connected by stronger forward than backward connections*

The computational neuroscience aspects of the hypotheses described above were investigated with an integrate-and-fire network with three attractor network modules connected by stronger forward than backward connections, the operation of which is illustrated conceptually in Figs. 2 and 3, and the architecture of which is illustrated in Fig. 4. Each module is an integrate-and-fire attractor network with n possible attractor states. For the simulations described there were $n=10$ orthogonal attractor states in each module, each implemented by a population of neurons with strong excitatory connections between the neurons with value $w_+=2.1$ for the NMDA and AMPA synapses. There were $N=8000$ neurons in each module, of which 0.8 (i.e. 6400) were excitatory, and 1600 were inhibitory. The first module could represent one of 10 Subjects (S), the second module one of 10 verbs (V), and the third module one of 10 objects (O). In the cerebral cortex, each module might be expected to be able to encode up to 10,000 items using sparse distributed representations, and assuming of order 10,000 excitatory recurrent collateral connection onto each neuron (Treves and Rolls, 1991, 1994; Rolls, 2008). The parameters in each module were set so that the spontaneous firing state with no input applied was stable, and so that only one of its n possible attractor states was active at any time when inputs were applied (Rolls, 2008; Rolls and Deco, 2010; Deco et al., 2013) (see Appendix).

In the model, there are forward connections from all excitatory neurons in one module to all excitatory neurons in the next module, with uniform strength w_{ff}. The role of these forward connections is to produce some non-specific input to the next module of the network when the previous module enters an attractor state with one of its neuronal pools. The role of this forward input is to encourage the next module in the series to start firing with some small delay after the previous module, to enable a temporal trajectory through a state space of an active attractor pool in one module to an active attractor pool in the next module. In the cerebral cortex, there are typically both stronger forward than backward connections between modules in different cortical areas, and connections in both directions between modules in the same cortical area (Rolls, 2008). In the latter case, the hypothesis is that the stronger connections in one direction than the reverse between nearby modules might be set up by spike-timing dependent plasticity (Markram et al., 1997; Bi and Poo, 1998; Feldman, 2012) based on the temporal order in which the relevant modules were normally activated during a stage when a language was being learned. For the simulations, only weak forward connections were implemented for simplicity.

During operation, one attractor pool in each module receives a continuous bias throughout a trial from a Deep network (see Figs. 2 and 3). For example the Subject Word module might receive bias from a Deep attractor pool representing 'James', the Verb Word module might receive bias from an attractor pool representing 'chased', and the Object Word module might receive bias from an attractor pool representing 'John'. These biases would represent the deep structure of the sentence, what it is intended to say, but not the generation of the sentence, which is the function of the network shown in Fig. 4. The biases in all the modules apart from the first are insufficient to push any attractor pool in a Word module into an attractor state. In the first (or head) Word module, the bias is sufficiently strong to make the attractor pool being biased enter a high firing rate attractor state. The concept is that because of the forward connections to all neurons in the second (Verb Word) module, the attractor pool in the second module receiving a steady bias (in our example, the 'chased' pool) then has sufficient input for it to gradually enter an attractor. The same process is then repeated for the biased attractor in Word module 3, which then enters a high firing rate state. Due to the slow stochastic dynamics of the network, there are delays between the firing in each of the Word modules. It is this that provides the sequentiality to the process that generates the words in the sentence in the correct order.

In addition, a concept of the cortical dynamics of this system is that each module should maintain a level of continuing firing in its winning attractor for the remainder of the sentence, and even for a few seconds afterwards. The purpose of this is to enable correction of the process (by for example a higher order thought or monitoring process, see Rolls, 2014) if the process needs to be corrected. The maintenance of the attractors in a continuing state of firing enables monitoring of exactly which attractors did occur in the trajectory through the state space, in case there was a slip of the tongue. (Indeed, 'slips of the tongue' or speech production errors are accounted for in this framework by the somewhat noisy trajectory through the state space that is likely to occur because of the close to Poisson spiking times of the neurons for a given mean rate, which introduces noise into the system Rolls, 2008; Rolls and Deco, 2010; Rolls, 2014). The main parameters of each module that enable this to be achieved are w_+, and the external bias entering each attractor pool.

However, although it is desired to have a short-term memory trace of previous activity during and for a short time after a sentence, it is also important that each word is uttered at its correct time in the sentence, for this carries the syntactic relations in this system. To achieve the production of the word at the correct time, the firing of each attractor has a mechanism to produce high firing initially for perhaps 200–300 ms, and then lower firing later to main an active memory trace of previous neuronal activity. The mechanism used to achieve this initial high firing when a neuronal pool enters a high firing rate state, was spike frequency adaptation, a common neuronal process which is described later, and which has been implemented and utilised previously (Liu and Wang, 2001; Deco and Rolls, 2005b; Rolls and Deco, 2014).

9.2. The operation of a single attractor network module

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

Each Word attractor network contains 6400 excitatory, and 1600 inhibitory neurons, which is consistent with the observed proportions of pyramidal cells and interneurons in the cerebral cortex (Abeles, 1991; Braitenberg and Schütz, 1991). The connection strengths are adjusted using mean-field analysis (Brunel and Wang, 2001; Deco and Rolls, 2006; Rolls and Deco, 2010), so that the excitatory and inhibitory neurons exhibit a spontaneous activity of 3 Hz and 9 Hz, respectively (Wilson et al., 1994; Koch and Fuster, 1989). The recurrent excitation mediated by the AMPA and NMDA receptors is dominated by the NMDA current to avoid instabilities during delay periods (Wang, 2002).

The architecture of the cortical network module illustrated in Fig. 5 has 10 selective pools each with 640 neurons. The connection weights between the neurons within each pool or population are called the intra-pool connection strengths w_+, which were set to 2.1 for the simulations described. All other weights including w_{inh} were set to 1.

All the excitatory neurons in each attractor pool S1, S2 … SN receive an external bias input λ_1, λ_2 … λ_N. This external input consists of Poisson external input spikes via AMPA receptors which are envisioned to originate from 800 external neurons. One component of this bias which is present by default arrives at an average spontaneous firing rate of 3 Hz from each external neuron onto each of the 800 synapses for external inputs, consistent with the spontaneous activity observed in the cerebral cortex (Wilson et al., 1994; Rolls and Treves, 1998; Rolls, 2008). The second component is a selective bias from a deep structure system which provides a bias present throughout a trial to one of the attractor pools in each module, corresponding to the subject, verb, or object (depending on the module) to be used in the sentence being generated. This bias makes it more likely that the attractor pool will become active if there are other inputs, but is not sufficiently strong (except in the first module) to initiate a high firing rate attractor state. (This selective bias might be set up by associative synaptic modification between the Deep and the Word modules.) In addition, all excitatory neurons in a module receive inputs with a uniform synaptic strength of w_{ff} from all the excitatory neurons in the preceding module, as illustrated in Fig. 4.

Both excitatory and inhibitory neurons are represented by a leaky integrate-and-fire model (Tuckwell, 1988). The basic state variable of a single model neuron is the membrane potential. It decays in time when the neurons receive no synaptic input down to a resting potential. When synaptic input causes the membrane potential to reach a threshold, a spike is emitted and the neuron is set to the reset potential at which it is kept for the refractory period. The emitted action potential is propagated to the other neurons in the network. The excitatory neurons transmit their action potentials via the glutamatergic receptors AMPA and NMDA which are both modeled by their effect in producing exponentially decaying currents in the postsynaptic neuron. The rise time of the AMPA current is neglected, because it is typically very short. The NMDA channel is modeled with an alpha function including both a rise and a decay term. In addition, the synaptic function of the NMDA current includes a voltage dependence controlled by the extracellular magnesium concentration (Jahr and Stevens, 1990). The inhibitory postsynaptic potential is mediated by a GABA_A receptor model and is described by a decay term. A detailed mathematical description is provided in the Appendix.

9.3. Spike frequency adaptation mechanism

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

Acknowledgments

Discussions on early aspects of these ideas with Alessandro Treves (SISSA, Trieste, Italy) were very stimulating, as always.

Appendix A

A.1. Implementation of neural and synaptic dynamics

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

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

$$C_\text{m} \frac{dV(t)}{dt} = -g_\text{m}(V(t) - V_\text{L}) - I_{\text{syn}}(t), \qquad (3)$$

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

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

Each attractor network is fully connected with $N_E = 6400$ excitatory neurons and $N_I = 1600$ inhibitory neurons, which is consistent with the observed proportions of the pyramidal neurons and interneurons in the cerebral cortex (Braitenberg and Schütz, 1991; Abeles, 1991). The synaptic current impinging on each neuron is given by the sum of recurrent excitatory currents ($I_{AMPA,rec}$ and $I_{NMDA,rec}$), the external excitatory current ($I_{AMPA,ext}$) and the inhibitory current (I_{GABA}):

$$I_{syn}(t) = I_{AMPA,ext}(t) + I_{AMPA,rec}(t) + I_{NMDA,rec}(t) + I_{GABA}(t). \tag{4}$$

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

$$I_{AMPA,ext}(t) = g_{AMPA,ext}(V(t) - V_E) \sum_{j=1}^{N_{ext}} s_j^{AMPA,ext}(t) \tag{5}$$

$$I_{AMPA,rec}(t) = g_{AMPA,rec}(V(t) - V_E) \sum_{j=1}^{N_E} w_{ji}^{AMPA} s_j^{AMPA,rec}(t) \tag{6}$$

$$I_{NMDA,rec}(t) = \frac{g_{NMDA}(V(t) - V_E)}{1 + [Mg^{++}]\exp(-0.062\,V(t))/3.57} \times \sum_{j=1}^{N_E} w_{ji}^{NMDA} s_j^{NMDA}(t) \tag{7}$$

$$I_{GABA}(t) = g_{GABA}(V(t) - V_I) \sum_{j=1}^{N_I} w_{ji}^{GABA} s_j^{GABA}(t) \tag{8}$$

where $V_E = 0$ mV, $V_I = -70$ mV, w_j are the synaptic weights, s_j's the fractions of open channels for the different receptors and g's the synaptic conductances for the different channels. The NMDA synaptic current depends on the membrane potential and the extracellular concentration of Magnesium ($[Mg^{++}] = 1$ mM Jahr and Stevens, 1990). The values for the synaptic conductances for excitatory neurons are $g_{AMPA,ext} = 2.08$ nS, $g_{AMPA,rec} = 0.013$ nS, $g_{NMDA} = 0.041$ nS and $g_{GABA} = 0.16$ nS; and for inhibitory neurons $g_{AMPA,ext} = 1.62$ nS, $g_{AMPA,rec} = 0.01$ nS, $g_{NMDA} = 0.03$ nS and $g_{GABA} = 0.12$ nS. These values are obtained from the ones used by Brunel and Wang (2001) by correcting for the different numbers of neurons. The conductances were calculated so that in an unstructured network the excitatory neurons have a spontaneous spiking rate of 3 Hz and the inhibitory neurons a spontaneous rate of 9 Hz. The fractions of open channels are described by

$$\frac{ds_j^{AMPA,ext}(t)}{dt} = -\frac{s_j^{AMPA,ext}(t)}{\tau_{AMPA}} + \sum_k \delta\left(t - t_j^k\right) \tag{9}$$

$$\frac{ds_j^{AMPA,rec}(t)}{dt} = -\frac{s_j^{AMPA,rec}(t)}{\tau_{AMPA}} + \sum_k \delta\left(t - t_j^k\right) \tag{10}$$

$$\frac{ds_j^{NMDA}(t)}{dt} = -\frac{s_j^{NMDA}(t)}{\tau_{NMDA,decay}} + \alpha x_j(t)\left(1 - s_j^{NMDA}(t)\right) \tag{11}$$

$$\frac{dx_j(t)}{dt} = -\frac{x_j(t)}{\tau_{NMDA,rise}} + \sum_k \delta\left(t - t_j^k\right) \tag{12}$$

$$\frac{ds_j^{GABA}(t)}{dt} = -\frac{s_j^{GABA}(t)}{\tau_{GABA}} + \sum_k \delta\left(t - t_j^k\right), \tag{13}$$

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

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

A.2. *Calcium-dependent spike frequency adaptation mechanism*

A specific implementation of the spike-frequency adaptation mechanism using Ca^{++}-activated K^+ hyper-polarising currents (Liu and Wang, 2001) is described next, and was used by Deco and Rolls (2005b). We assume that the intrinsic gating of K^+ After-Hyper-Polarising current (I_{AHP}) is fast, and therefore its slow activation is due to the kinetics of the cytoplasmic Ca^{2+} concentration. This can be introduced in the model by adding an extra current term in the integrate-and-fire model, i.e. by adding I_{AHP} on the right hand side of equation (14, which describes the evolution of the subthreshold membrane potential $V(t)$ of each neuron:

$$C_m \frac{dV(t)}{dt} = -g_m(V(t) - V_L) - I_{syn}(t) \tag{14}$$

where $I_{syn}(t)$ is the total synaptic current flow into the cell, V_L is the resting potential, C_m is the membrane capacitance, and g_m is the membrane conductance. The extra current term that is introduced into this equation is as follows:

$$I_{AHP} = -g_{AHP}[Ca^{2+}](V(t) - V_K) \tag{15}$$

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

$$\frac{d[Ca^{2+}]}{dt} = -\frac{[Ca^{2+}]}{\tau_{Ca}} \tag{16}$$

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

Please cite this article as: Rolls, E.T., Deco, G., Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain. Brain Research (2014), http://dx.doi.org/10.1016/j.brainres.2014.09.021

Table 1 – Parameters used for each module in the integrate-and-fire simulations.

N_E	6400
N_I	1600
r	0.1
w_+	2.1
w_I	1.0
N_{ext}	800
ν_{ext}	2.4 kHz
C_m (excitatory)	0.5 nF
C_m (inhibitory)	0.2 nF
g_m (excitatory)	25 nS
g_m (inhibitory)	20 nS
V_L	−70 mV
V_{thr}	−50 mV
V_{reset}	−55 mV
V_E	0 mV
V_I	−70 mV
$g_{AMPA,ext}$ (excitatory)	2.08 nS
$g_{AMPA,rec}$ (excitatory)	0.01 nS
g_{NMDA} (excitatory)	0.041 nS
g_{GABA} (excitatory)	0.156 nS
$g_{AMPA,ext}$ (inhibitory)	1.62 nS
$g_{AMPA,rec}$ (inhibitory)	0.01 nS
g_{NMDA} (inhibitory)	0.032 nS
g_{GABA} (inhibitory)	0.122 nS
$\tau_{NMDA,decay}$	100 ms
$\tau_{NMDA,rise}$	2 ms
τ_{AMPA}	2 ms
τ_{GABA}	10 ms
α	0.5 ms^{-1}

$g_{AHP}=0$ nS simulates the effect of low levels of acetylcholine in normal aging.

A.3. *The model parameters used in the simulations of memory*

The fixed parameters of the model are shown in Table 1, and not only provide information about the values of the parameters used in the simulations, but also enable them to be compared to experimentally measured values.

REFERENCES

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

Abeles, A., 1991. Corticonics. Cambridge University Press, New York.

Aggelopoulos, N.C., Rolls, E.T., 2005. Natural scene perception: inferior temporal cortex neurons encode the positions of different objects in the scene. Eur. J. Neurosci. 22, 2903–2916.

Amit, D.J., 1989. Modeling Brain Function. The World of Attractor Neural Networks. Cambridge University Press, Cambridge.

Bi, G.-Q., Poo, M.-M., 1998. Activity-induced synaptic modifications in hippocampal culture, dependence on spike timing, synaptic strength and cell type. J. Neurosci. 18, 10464–10472.

Braitenberg, V., Schütz, A., 1991. Anatomy of the Cortex. Springer Verlag, Berlin.

Brown, D.A., Gähwiler, B.H., Griffith, W.H., Halliwell, J.V., 1990. Membrane currents in hippocampal neurons. Prog. Brain Res. 83, 141–160.

Brunel, N., Wang, X.J., 2001. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. J. Comput. Neurosci. 11, 63–85.

Cerasti, E., Treves, A., 2010. How informative are spatial CA3 representations established by the dentate gyrus? PLoS Comput. Biol. 6, e1000759.

Chomsky, N., 1965. Aspects of the Theory of Syntax. MIT Press, Cambridge, Massachusetts.

Deco, G., Rolls, E.T., 2003. Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. Eur. J. Neurosci. 18, 2374–2390.

Deco, G., Rolls, E.T., 2005a. Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons. J. Neurophysiol. 94, 295–313.

Deco, G., Rolls, E.T., 2005b. Sequential memory: a putative neural and synaptic dynamical mechanism. J. Cogn. Neurosci. 17, 294–307.

Deco, G., Rolls, E.T., 2005c. Synaptic and spiking dynamics underlying reward reversal in the orbitofrontal cortex. Cereb. Cortex 15, 15–30.

Deco, G., Rolls, E.T., 2006. A neurophysiological model of decision-making and Weber's law. Eur. J. Neurosci. 24, 901–916.

Deco, G., Rolls, E.T., Albantakis, L., Romo, R., 2013. Brain mechanisms for perceptual and reward-related decision-making. Prog. Neurobiol. 103, 194–213.

Elliffe, M.C.M., Rolls, E.T., Stringer, S.M., 2002. Invariant recognition of feature combinations in the visual system. Biolog. Cybern. 86, 59–71.

Engel, A.K., Konig, P., Kreiter, A.K., Schillen, T.B., Singer, W., 1992. Temporal coding in the visual system: new vistas on integration in the nervous system. Trends Neurosci. 15, 218–226.

Feigenbaum, J.D., Rolls, E.T., 1991. Allocentric and egocentric spatial information processing in the hippocampal formation of the behaving primate. Psychobiology 19, 21–40.

Feldman, D.E., 2012. The spike-timing dependence of plasticity. Neuron 75, 556–571.

Földiák, P., 1991. Learning invariance from transformation sequences. Neural Comput. 3, 194–200.

Franco, L., Rolls, E.T., Aggelopoulos, N.C., Jerez, J.M., 2007. Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex. Biolog. Cybern. 96, 547–560.

Fries, P., 2009. Neuronal gamma-band synchronization as a fundamental process in cortical computation. Annu. Rev. Neurosci. 32, 209–224.

Fuhrmann, G., Markram, H., Tsodyks, M., 2002. Spike frequency adaptation and neocortical rhythms. J. Neurophysiol. 88, 761–770.

Fuster, J.M., 2008. The Prefrontal Cortex, fourth ed. Academic Press, London.

Gardner-Medwin, A.R., 1976. The recall of events through the learning of associations between their parts. Proc. R. Soc. Lond. Ser. B 194, 375–402.

Georges-François, P., Rolls, E.T., Robertson, R.G., 1999. Spatial view cells in the primate hippocampus: allocentric view not head direction or eye position or place. Cereb. Cortex 9, 197–212.

Goldman-Rakic, P., 1996. The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. Philos. Trans. R. Soc. Lond. B 351, 1445–1453.

Hertz, J., Krogh, A., Palmer, R.G., 1991. Introduction to the Theory of Neural Computation. Addison Wesley, Wokingham, U.K.

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

Please cite this article as: Rolls, E.T., Deco, G., Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain. Brain Research (2014), http://dx.doi.org/10.1016/j.brainres.2014.09.021

Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. Proc. Nat. Acad. Sci. U. S. A. 79, 2554–2558.

Hubel, D.H., Wiesel, T.N., 1968. Receptive fields and functional architecture of monkey striate cortex. J. Physiol. Lond. 195, 215–243.

Hubel, D.H., Wiesel, T.N., 1977. Functional architecture of the macaque monkey visual cortex. Proc. R. Soc. Lond. [B] 198, 1–59.

Hummel, J.E., Biederman, I., 1992. Dynamic binding in a neural network for shape recognition. Psychol. Rev. 99, 480–517.

Jackendoff, R., 2002. Foundations of Language. Oxford University Press, Oxford.

Jahr, C., Stevens, C., 1990. Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. J. Neurosci. 10, 3178–3182.

Kesner, R.P., Rolls, E.T., 2015. A theory of hippocampal function, and tests of the theory: new developments. In review.

Koch, K.W., Fuster, J.M., 1989. Unit activity in monkey parietal cortex related to haptic perception and temporary memory. Exp. Brain Res. 76, 292–306.

Kohonen, T., 1977. Associative Memory: A System Theoretical Approach. Springer, New York.

Kohonen, T., Oja, E., Lehtio, P., 1981. Storage and processing of information in distributed memory systems. In: Hinton, G.E., Anderson, J.A. (Eds.), Parallel Models of Associative Memory. Erlbaule, Hillsdale, NJ, pp. 105–143 chapter 4.

Kondo, H., Lavenex, P., Amaral, D.G., 2009. Intrinsic connections of the macaque monkey hippocampal formation: II. CA3 connections. J. Comp. Neurol. 515, 349–377.

Kropff, E., Treves, A., 2005. The storage capacity of Potts models for semantic memory retrieval. J. Stat. Mech. Theory Exp. 2005, P08010.

Lanthorn, T., Storm, J., Andersen, P., 1984. Current-to-frequency transduction in CA1 hippocampal pyramidal cells: slow prepotentials dominate the primary range firing. Exp. Brain Res. 53, 431–443.

Liu, Y., Wang, X.-J., 2001. Spike-frequency adaptation of a generalized leaky integrate-and-fire model neuron. J. Comput. Neurosci. 10, 25–45.

MacDonald, C.J., Lepage, K.Q., Eden, U.T., Eichenbaum, H., 2011. Hippocampal “time cells” bridge the gap in memory for discontiguous events. Neuron 71, 737–749.

Malsburg, C.v.d., 1990. A neural architecture for the representation of scenes. In: McGaugh, J.L., Weinberger, N.M., Lynch, G. (Eds.), Brain Organization and Memory: Cells, Systems and Circuits. Oxford University Press, New York, pp. 356–372 chapter 18.

Markram, H., Lübke, J., Frotscher, M., Sakmann, B., 1997. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science 275, 213–215.

Marr, D., 1971. Simple memory: a theory for archicortex. Philos. Trans. R. Soc. Lond. [B] 262, 23–81.

Mason, A., Larkman, A., 1990. Correlations between morphology and electrophysiology of pyramidal neurones in slices of rat visual cortex. I. Electrophysiology. J. Neurosci. 10, 1415–1428.

McClelland, J.L., McNaughton, B.L., O'Reilly, R.C., 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol. Rev. 102, 419–457.

McCormick, D., Connors, B., Lighthall, J., Prince, D., 1985. Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons in the neocortex. J. Neurophysiol. 54, 782–806.

McNaughton, B.L., Morris, R.G.M., 1987. Hippocampal synaptic enhancement and information storage within a distributed memory system. Trends Neurosci. 10, 408–415.

Nicoll, R.A., 1988. The coupling of neurotransmitter receptors to ion channels in the brain. Science 241, 545–551.

Pirmoradian, S., Treves, A., 2013. Encoding words into a Potts attractor network. In: Mayor, J., Gomez, P. (Eds.), Proceedings of the Thirteenth Neural Computation and Psychology Workshop (NCPW13) on Computational Models of Cognitive Processes, World Scientific Press, Singapore, pp. 29–42.

Robertson, R.G., Rolls, E.T., Georges-François, P., 1998. Spatial view cells in the primate hippocampus: effects of removal of view details. J. Neurophysiol. 79, 1145–1156.

Rolls, E.T., 1987. Information representation, processing and storage in the brain: analysis at the single neuron level. In: Changeux, J.-P., Konishi, M. (Eds.), The Neural and Molecular Bases of Learning. Wiley, Chichester, pp. 503–540.

Rolls, E.T., 1989a. Functions of neuronal networks in the hippocampus and cerebral cortex in memory. In: Cotterill, R. (Ed.), Models of Brain Function. Cambridge University Press, Cambridge, pp. 15–33.

Rolls, E.T., 1989b. Functions of neuronal networks in the hippocampus and neocortex in memory. In: Byrne, J.H., Berry, W.O. (Eds.), Neural Models of Plasticity: Experimental and Theoretical Approaches. Academic Press, San Diego, CA, pp. 240–265 chapter 13.

Rolls, E.T., 1989c. Parallel distributed processing in the brain: implications of the functional architecture of neuronal networks in the hippocampus. In: Morris, R.G.M. (Ed.), Parallel Distributed Processing: Implications for Psychology and Neurobiology. Oxford University Press, Oxford, pp. 286–308 chapter 12.

Rolls, E.T., 1989d. The representation and storage of information in neuronal networks in the primate cerebral cortex and hippocampus. In: Durbin, R., Miall, C., Mitchison, G. (Eds.), The Computing Neuron. Addison-Wesley, Wokingham, England, pp. 125–159 chapter 8.

Rolls, E.T., 1990a. Functions of the primate hippocampus in spatial processing and memory. In: Olton, D.S., Kesner, R.P. (Eds.), Neurobiology of Comparative Cognition. L. Erlbaum, Hillsdale, NJ, pp. 339–362 chapter 12.

Rolls, E.T., 1990b. Theoretical and neurophysiological analysis of the functions of the primate hippocampus in memory. Cold Spring Harbor Symposia in Quantitative Biology 55, 995–1006.

Rolls, E.T., 1992. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. Philos. Trans. R. Soc. 335, 11–21.

Rolls, E.T., 1995. A model of the operation of the hippocampus and entorhinal cortex in memory. Int. J. Neural Syst. 6 (Suppl.), 51–70.

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

Rolls, E.T., 2010. A computational theory of episodic memory formation in the hippocampus. Behav. Brain Res. 215, 180–196.

Rolls, E.T., 2012a. Advantages of dilution in the connectivity of attractor networks in the brain. Biol. Inspir. Cognit. Archit. 1, 44–54.

Rolls, E.T., 2012b. Invariant visual object and face recognition: neural and computational bases, and a model, VisNet. Front. Comput. Neurosci. 6 (35), 1–70.

Rolls, E.T., 2012c. Neuroculture: On the Implications of Brain Science. Oxford University Press, Oxford.

Rolls, E.T., 2013a. A biased activation theory of the cognitive and attentional modulation of emotion. Front. Hum. Neurosci. 7, 74.

Rolls, E.T., 2013b. The mechanisms for pattern completion and pattern separation in the hippocampus. Front. Syst. Neurosci. 7, 74.

Rolls, E.T., 2014. Emotion and Decision-Making Explained. Oxford University Press, Oxford.

Please cite this article as: Rolls, E.T., Deco, G., Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain. Brain Research (2014), http://dx.doi.org/10.1016/j.brainres.2014.09.021

Rolls, E.T., Critchley, H., Verhagen, J.V., Kadohisa, M., 2010. The representation of information about taste and odor in the primate orbitofrontal cortex. Chemosens. Percept. 3, 16–33.

Rolls, E.T., Deco, G., 2002. Computational Neuroscience of Vision. Oxford University Press, Oxford.

Rolls, E.T., Deco, G., 2010. The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function. Oxford University Press, Oxford.

Rolls, E.T., Deco, G., 2014. Stochastic cortical neurodynamics underlying the memory and cognitive changes in aging.

Rolls, E.T., Dempere-Marco, L., Deco, G., 2013. Holding multiple items in short term memory: a neural mechanism. PLoS One 8, e61078.

Rolls, E.T., Kesner, R.P., 2006. A theory of hippocampal function, and tests of the theory. Prog. Neurobiol. 79, 1–48.

Rolls, E.T., Miyashita, Y., Cahusac, P.M.B., Kesner, R.P., Niki, H., Feigenbaum, J., Bach, L., 1989. Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. J. Neurosci. 9, 1835–1845.

Rolls, E.T., Robertson, R.G., Georges-François, P., 1997. Spatial view cells in the primate hippocampus. Eur. J. Neurosci. 9, 1789–1794.

Rolls, E.T., Tovee, M.J., 1995. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. J. Neurophysiol. 73, 713–726.

Rolls, E.T., Treves, A., 1990. The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. Network 1, 407–421.

Rolls, E.T., Treves, A., 1998. Neural Networks and Brain Function. Oxford University Press, Oxford.

Rolls, E.T., Treves, A., 2011. The neuronal encoding of information in the brain. Prog. Neurobiol. 95, 448–490.

Rolls, E.T., Treves, A., Robertson, R.G., Georges-François, P., Panzeri, S., 1998. Information about spatial view in an ensemble of primate hippocampal cells. J. Neurophysiol. 79, 1797–1813.

Rolls, E.T., Treves, A., Tovee, M.J., 1997. The representational capacity of the distributed encoding of information provided by populations of neurons in the primate temporal visual cortex. Exp. Brain Res. 114, 149–162.

Rolls, E.T., Treves, A., Tovee, M., Panzeri, S., 1997. Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. J. Comput. Neurosci. 4, 309–333.

Rolls, E.T., Webb, T.J., 2012. Cortical attractor network dynamics with diluted connectivity. Brain Res. 1434, 212–225.

Rolls, E.T., Webb, T.J., Deco, G., 2012. Communication before coherence. Eur. J. Neurosci. 36, 2689–2709.

Rolls, E.T., Xiang, J.-Z., 2005. Reward-spatial view representations and learning in the primate hippocampus. J. Neurosci. 25, 6167–6174.

Rolls, E.T., Xiang, J.-Z., 2006. Spatial view cells in the primate hippocampus, and memory recall. Rev. Neurosci. 17, 175–200.

Rolls, E.T., Xiang, J.-Z., Franco, L., 2005. Object, space and object-space representations in the primate hippocampus. J. Neurophysiol. 94, 833–844.

Sah, P., 1996. Ca^{2+}-activated K^{+} currents in neurones: types, physiological roles and modulation. Trends Neurosci. 19, 150–154.

Sah, P., Faber, E.S., 2002. Channels underlying neuronal calcium-activated potassium currents. Prog. Neurobiol. 66, 345–353.

Salin, P., Prince, D., 1996. Spontaneous GABA-A receptor mediated inhibitory currents in adult rat somatosensory cortex. J. Neurophysiol. 75, 1573–1588.

Singer, W., 1999. Neuronal synchrony: a versatile code for the definition of relations? Neuron 24, 49–65.

Singer, W., Gray, C., Engel, A., Konig, P., Artola, A., Brocher, S., 1990. Formation of cortical cell assemblies. Cold Spring Harbor Symposium on Quantitative Biology 55, 939–952.

Singer, W., Gray, C.M., 1995. Visual feature integration and the temporal correlation hypothesis. Annu. Rev. Neurosci. 18, 555–586.

Song, S., Yao, H., Treves, A., 2014. A modular latching chain. Cognit. Neurodyn. 8, 37–46.

Spruston, N., Jonas, P., Sakmann, B., 1995. Dendritic glutamate receptor channel in rat hippocampal CA3 and CA1 pyramidal neurons. J. Physiol. 482, 325–352.

Stella, F., Cerasti, E., Treves, A., 2013. Unveiling the metric structure of internal representations of space. Front. Neural Circuits 7, 81.

Tovee, M.J., Rolls, E.T., Treves, A., Bellis, R.P., 1993. Information encoding and the responses of single neurons in the primate temporal visual cortex. J. Neurophysiol. 70, 640–654.

Treves, A., 1990. Graded-response neurons and information encodings in autoassociative memories. Phys. Rev. A 42, 2418–2430.

Treves, A., 1993. Mean-field analysis of neuronal spike dynamics. Network 4, 259–284.

Treves, A., 2005. Frontal latching networks: a possible neural basis for infinite recursion. Cognit. Neuropsychol. 22, 276–291.

Treves, A., Rolls, E.T., 1991. What determines the capacity of autoassociative memories in the brain? Network 2, 371–397.

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

Treves, A., Rolls, E.T., 1994. A computational analysis of the role of the hippocampus in memory. Hippocampus 4, 374–391.

Tuckwell, H., 1988. Introduction to Theoretical Neurobiology. Cambridge University Press, Cambridge.

Vigliocco, G., Vinson, D.P., Druks, J., Barber, H., Cappa, S.F., 2011. Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies. Neurosci. Biobehav. Rev. 35, 407–426.

Wallis, G., Rolls, E.T., 1997. Invariant face and object recognition in the visual system. Prog. Neurobiol. 51, 167–194.

Wang, X.J., 2002. Probabilistic decision making by slow reverberation in cortical circuits. Neuron 36, 955–968.

Webb, T.J., Rolls, E.T., Deco, G., Feng, J., 2011. Noise in attractor networks in the brain produced by graded firing rate representations. PLoS One 6, e23620.

Wilson, F.A.W., O'Scalaidhe, S.P., Goldman-Rakic, P., 1994. Functional synergism between putative gamma-aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex. Proc. Natl. Acad. Sci. 91, 4009–4013.

Xiang, Z., Huguenard, J., Prince, D., 1998. GABA-A receptor mediated currents in interneurons and pyramidal cells of rat visual cortex. J. Physiol. 506, 715–730.

Please cite this article as: Rolls, E.T., Deco, G., Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain. Brain Research (2014), http://dx.doi.org/10.1016/j.brainres.2014.09.021