



Attractor networks

Edmund T. Rolls*

An *attractor* network is a network of neurons with excitatory interconnections that can settle into a stable pattern of firing. This article shows how attractor networks in the cerebral cortex are important for long-term memory, short-term memory, attention, and decision making. The article then shows how the random firing of neurons can influence the stability of these networks by introducing stochastic noise, and how these effects are involved in probabilistic decision making, and implicated in some disorders of cortical function such as poor short-term memory and attention, schizophrenia, and obsessive-compulsive disorder. © 2009 John Wiley & Sons, Ltd. *WIREs Cogn Sci* 2010 1 119–134

An *attractor* network is a network of neurons with excitatory interconnections that can settle into a stable pattern of firing.^{1–4} This article shows how attractor networks in the cerebral cortex are important for long-term memory, short-term memory, attention, and decision making. The article then shows how the random firing of neurons can influence the stability of these networks by introducing stochastic noise, and how these effects are involved in probabilistic decision making, and implicated in some disorders of cortical function such as poor short-term memory and attention, schizophrenia, and obsessive-compulsive disorder. Each memory pattern stored in an attractor network by associative synaptic modification consists of a subset of the neurons firing. These patterns could correspond to memories, perceptual representations or thoughts.

ATTRACTOR NETWORK ARCHITECTURE, AND THE STORAGE OF MEMORIES

The architecture of an attractor or autoassociation network is shown in Figure 1. External inputs e_i activate the neurons in the network, and produce firing y_i , where i refer to the i 'th neuron. The neurons are connected by recurrent collateral synapses w_{ij} , where j refers to the j 'th synapse on a neuron. By these synapses an input pattern on e_i is associated with itself, and thus the network is referred to as an autoassociation network. Because there is positive

feedback via the recurrent collateral connections, the network can sustain persistent firing. These synaptic connections are assumed to build up by an associative (Hebbian) learning mechanism⁵ (according to which the more two neurons are simultaneously active the stronger the neural connection becomes). The associative learning rule for the change in the synaptic weight is as shown in Eq. (1)

$$\delta w_{ij} = k \cdot y_i \cdot y_j \quad (1)$$

where k is a constant, y_i is the activation of the dendrite (the postsynaptic term), y_j is the presynaptic firing rate, and δw_{ij} is the change of synaptic weight. The inhibitory interneurons are not shown. They receive inputs from the pyramidal cells, and make negative feedback connections onto the pyramidal cells to control their activity.

In order for biologically plausible autoassociative networks to store information efficiently, heterosynaptic long-term depression (LTD) (as well as long-term potentiation) is required.^{4,6–10} This type of LTD helps to remove the correlations between the training patterns that arise because the neurons have positive-only firing rates. The effect of the LTD can be to enable the effect of the mean presynaptic firing rate to be subtracted from the patterns.^{4,6,7,9,10}

RECALL

During recall, the external input e_i is applied, and produces output firing, operating through the nonlinear activation function described below. The firing is fed back by the recurrent collateral axons shown in Figure 1 to produce activation of each output neuron through the modified synapses on each output neuron. The activation h_i produced by the recurrent collateral

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Edmund.Rolls@oxcns.org

Oxford Centre for Computational Neuroscience, Oxford, UK.

DOI: 10.1002/wcs.1

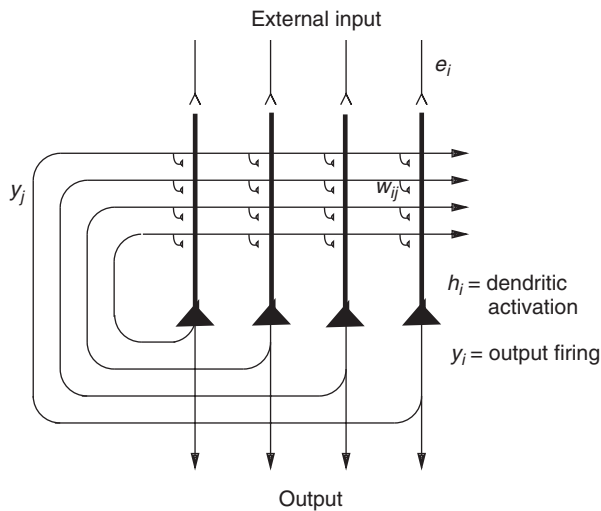


FIGURE 1 | The architecture of an autoassociative or attractor neural network (see text).

effect on the i th neuron is the sum of the activations produced in proportion to the firing rate of each axon y_j operating through each modified synapse w_{ij} , that is,

$$h_i = \sum_j y_j w_{ij} \quad (2)$$

where \sum_j indicates that the sum is over the C input axons to each neuron, indexed by j . This is a dot or inner product computation between the input firing vector y_j ($j = 1, C$) and the synaptic weight vector w_{ij} ($j = 1, C$) on neuron i , and it is because this is a vector similarity operation, closely related to a correlation, between the input vector and the synaptic weight vector that many of the properties of attractor networks arise, including completion of a memory when only a partial retrieval cue is applied.⁴ The output firing y_i is a nonlinear function of the activation produced by the recurrent collateral effect (internal recall) and by the external input e_i :

$$y_i = f(h_i; +e_i) \quad (3)$$

The activation function should be nonlinear, and may be, for example, binary threshold, linear threshold, sigmoid, etc. The threshold at which the activation function operates is set in part by the effect of the inhibitory neurons in the network (not shown in Figure 1). The threshold prevents the positive feedback inherent in the operation of attractor networks from leading to runaway neuronal firing; and allows optimal retrieval of a memory without interference from other memories stored in the synaptic weights.^{2,4}

The recall state (which could be used to implement short-term memory, or memory recall) in

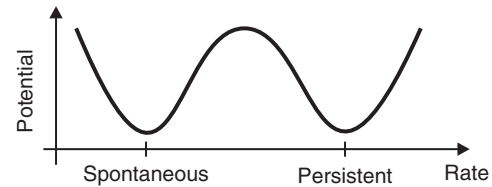


FIGURE 2 | Energy landscape of an attractor network. There are two types of stable fixed point: a spontaneous state with a low firing rate, and one or more persistent states with high firing rates in which the neurons keep firing. Each one of the high firing rate attractor states can implement a different memory.

an attractor network can be thought of as the local minimum in an energy landscape,¹ where the energy would be defined as

$$E = -\frac{1}{2} \sum_{i,j} w_{ij} (y_i - \langle y \rangle) (y_j - \langle y \rangle) \quad (4)$$

where y_i is the firing of neuron i , and $\langle y \rangle$ indicates the average firing rate. The intuition here is that if both y_i and y_j are above their average rates, and are exciting each other through a strong synapse, then the firing will tend to be stable and maintained, resulting in a low energy state that is stable. Although this energy analysis applies formally only with a fully connected network with symmetric synaptic strengths between neurons (which would be produced by an associative learning rule), it has been shown that the same general properties apply if the connectivity is diluted and becomes asymmetric.^{7,9,11–13}

Autoassociation attractor systems have two types of stable fixed point: a spontaneous state with a low firing rate, and one or more persistent states with high firing rates in which the neurons keep firing (Figure 2). Each one of the high firing rate attractor states can implement a different memory. When the system is moved to a position in the space by an external retrieval cue stimulus, it will move to the closest stable attractor state. The area in the space within which the system will move to a stable attractor state is called its basin of attraction. This is the process involved in completion of a whole memory from a partial retrieval cue.

Properties of Attractor Networks

Completion

An important and useful property of these attractor networks is that they complete an incomplete input vector, allowing recall of a whole memory from a small fraction of it. The memory recalled in response to a fragment is that stored in the memory that is closest in pattern similarity (as measured by the dot

product, or correlation). Because the recall is iterative and progressive, the recall can be perfect.

Short-Term Memory

An autoassociation or attractor memory is useful not only as a long-term memory, in for example, the memory for particular past episodes (see below), but can also be used as a short-term memory, in which iterative processing round the recurrent collateral loop keeps a representation active until another input cue is received, and this is widely used in the brain, and indeed is a prototypical property of cerebral neocortex (see below).

Graceful Degradation or Fault Tolerance

If the synaptic weight vector w_i on each neuron has synapses missing (e.g., during development), or loses synapses (e.g., with brain damage or ageing), then the activation h_i is still reasonable, because h_i is the dot product (correlation) of the input firing rate vector and the weight vector. The same argument applies if whole input axons are lost. If an output neuron is lost, then the network cannot itself compensate for this, but the next network in the brain is likely to be able to generalise or complete if its input vector has some elements missing, as would be the case if some output neurons of a preceding autoassociation network were damaged.

Storage Capacity, and the Sparseness of the Representation

Hopfield, using the approach of statistical mechanics, showed that in a fully connected attractor network with fully distributed binary representations (e.g., for any one pattern, half the neurons in the high firing state of 1, and the other half in the low firing state of 0 or -1), the number of stable attractor states, corresponding to the number of memories that can be successfully retrieved, is approximately $0.14C$, where C is the number of connections on each neuron from the recurrent collateral connections.^{1–3}

We (Treves and Rolls) have performed quantitative analyses of the storage and retrieval processes in attractor networks.^{7,9,11,12} We have extended previous formal models of autoassociative memory [see Ref 2] by analysing a network with graded response units, so as to represent more realistically the continuously variable rates at which neurons fire, and with incomplete connectivity.^{7,11} We have found that in general the maximum number p_{\max} of firing patterns that can be (individually) retrieved is proportional to the number C^{RC} of (associatively) modifiable recurrent collateral synapses per neuron, by a factor that increases roughly with the inverse of

the sparseness a of the neuronal representation.^a The neuronal population sparseness a of the representation can be measured by extending the binary notion of the proportion of neurons that are firing to any one stimulus or event as

$$a = \left(\sum_{i=1, n} (r_i/N)^2 / \sum_{i=1, n} (r_i^2/N) \right) \quad (5)$$

where r_i is the firing rate of the i 'th neuron in the set of N neurons. The sparseness ranges from $1/N$, when only one of the neurons responds to a particular stimulus (a local or grandmother cell representation), to a value of 1.0, attained when all the neurons are responding to a given stimulus. Approximately,

$$p_{\max} \cong \frac{C^{RC}}{a \ln(1/a)} k \quad (6)$$

where 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 0.2–0.3.⁷ For example, for $C^{RC} = 12,000$ (the number of recurrent collateral synapses on a hippocampal CA3 neuron in the rat¹⁴) and $a = 0.02$, p_{\max} is calculated to be approximately 36,000. This analysis emphasises the utility of having a sparse representation in the hippocampus, for this enables many different memories to be stored.¹⁵ The sparseness a in this equation is strictly the population sparseness.^{7,16} The population sparseness a^p would be measured by measuring the distribution of firing rates of all neurons to a single stimulus at a single time. The single neuron sparseness or selectivity a^s would be measured by the distribution of firing rates to a set of stimuli, which would take a long time. The selectivity or sparseness a^s of a single neuron measured across a set of stimuli often takes a similar value to the population sparseness a in the brain, and does so if the tuning profiles of the neurons to the set of stimuli are uncorrelated.¹⁶ These concepts are elucidated by Franco et al.¹⁶ These quantitative analyses have been confirmed numerically.¹³

The Dynamics of the Recurrent Attractor Network—Fast Recall

The analysis described above of the capacity of a recurrent network considered steady state conditions of the firing rates of the neurons. The question arises of how quickly the recurrent network would settle into its final state. If these settling processes took in the order of hundreds of ms, they would be much too slow to contribute usefully to cortical activity, whether in the hippocampus or the neocortex.^{4,10,17–19}

It has been shown that if the neurons are treated not as McCulloch–Pitts neurons which are simply

'updated' at each iteration, or cycle of time steps (and assume the active state if the threshold is exceeded), but instead are analysed and modelled as 'integrate-and-fire' neurons in real continuous time, then the network can effectively 'relax' into its recall state very rapidly, in one or two time constants of the synapses.^{9,10,20,21} This corresponds to perhaps 20 ms in the brain. One factor in this rapid dynamics of autoassociative networks with brain-like 'integrate-and-fire' membrane and synaptic properties is that with some spontaneous activity, some of the neurons in the network are close to threshold already before the recall cue is applied, and hence some of the neurons are very quickly pushed by the recall cue into firing, so that information starts to be exchanged very rapidly (within 1–2 ms of brain time) through the modified synapses by the neurons in the network. The progressive exchange of information starting early on within what would otherwise be thought of as an iteration period (of perhaps 20 ms, corresponding to a neuronal firing rate of 50 spikes/s), is the mechanism accounting for rapid recall in an autoassociative neuronal network made biologically realistic in this way. Further analysis of the fast dynamics of these networks if they are implemented in a biologically plausible way with 'integrate-and-fire' neurons, is provided in Section 7.7 of Rolls and Deco,¹⁰ in Appendix A5 of Rolls and Treves,⁹ by Treves,²⁰ by Panzeri et al.,¹⁹ and by Rolls.⁴

Continuous Attractor Networks

So far, we have considered attractor networks in which each memory pattern stored in the network is a discrete pattern. An attractor network trained with patterns that are continuous with each other can maintain the firing of its neurons to represent any location along a continuous physical dimension such as spatial position, head direction, etc. and is termed a Continuous Attractor neural network. It has the same architecture as a discrete attractor network, but uses the excitatory recurrent collateral connections between the neurons to reflect the distance between the neurons in the state space (e.g., head direction space, or the place of the animal in an environment). These networks can maintain the bubble of neural activity constant for long periods wherever it is started to represent the current state (head direction, position, etc.) of the animal, and are likely to be involved in many aspects of spatial processing and memory, including spatial vision.^{4,22–33} Global inhibition is used to keep the number of neurons in a bubble or packet of actively firing neurons relatively constant, and to help to ensure that (in typical applications) there is only one activity packet (but see³¹). Attractor networks can

operate with both continuous and discrete patterns, and this is likely to be important in episodic memory, in which typically a spatial position (e.g., a place) and discrete object-related information are components.³⁴

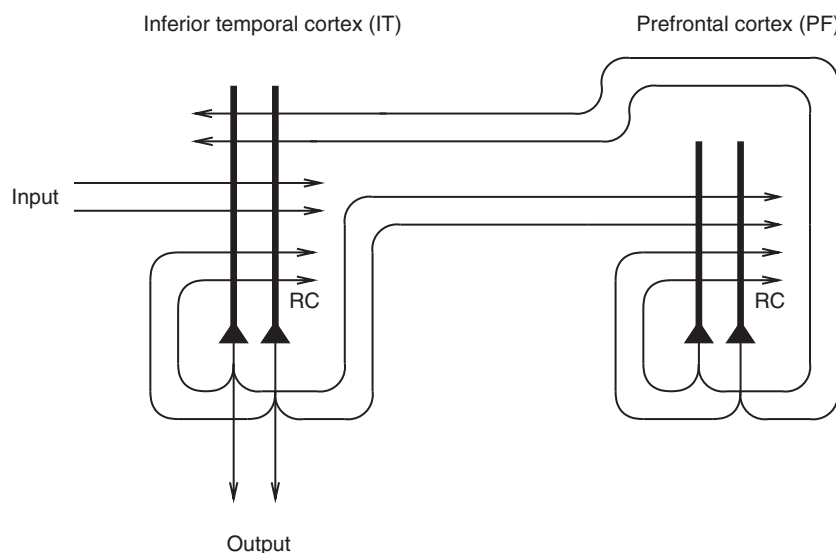
Attractor Networks for Short-Term Memory

Pyramidal neurons in the cerebral cortex have a relatively high density of excitatory connections to each other within a local area of 1–3 mm.^{4,35,36} These local recurrent collateral excitatory connections provide a positive-feedback mechanism (which is kept under control by gamma-aminobutyric acid (GABA) inhibitory interneurons) that enables a set of neurons to maintain their activity for many seconds to implement a short-term memory.³⁷ Each memory is formed by the set of the neurons in the local cortical network that were coactive when the memory was formed, resulting in strengthened excitatory connections between that set of neurons through the process of long-term potentiation, which is a property of these recurrent collateral connections.

Attractor networks appear to operate in the prefrontal cortex, an area that is important in attention and short-term memory, as shown, for example, by firing in the delay period of a short-term memory task.^{4,38–43} Short-term memory is the ability to hold information on-line during a short time period.^{39,43}

It has been proposed that whereas it is a property of all cortical areas that they have an ability to maintain neuronal activity by the attractor properties implemented by the recurrent collateral connections, the prefrontal cortex has a special role in short-term memory because it can act as an off-line store as follows (see Figure 3).⁴ First, we note that a perceptual brain area such as the inferior temporal cortex must respond to every new incoming set of objects in the world so that we can see them, and this is inconsistent with maintaining their firing in an attractor state that represents an object or objects seen seconds ago. For this reason, for a short-term memory to be maintained during periods in which new stimuli are to be perceived, there must be separate networks for the perceptual and short-term memory functions, and indeed two coupled networks, one in the inferior temporal visual cortex for perceptual functions, and another in the prefrontal cortex for maintaining the short-term memory, for example, when intervening stimuli are being shown, provide a precise model of the interaction of perceptual and short-term memory systems^{44,45} (see Figure 3). This model shows how a prefrontal cortex attractor (autoassociation) network could be triggered by a sample visual stimulus represented in the inferior temporal visual cortex in a delayed match to

FIGURE 3 | A short-term memory autoassociation network in the prefrontal cortex could hold active a working memory representation by maintaining its firing in an attractor state. The prefrontal module would be loaded with the to-be-remembered stimulus by the posterior module (in the temporal or parietal cortex) in which the incoming stimuli are represented. Backprojections from the prefrontal short-term memory module to the posterior module would enable the working memory to be unloaded, to for example, influence ongoing perception (see text). RC, recurrent collateral connections.



sample task, and could keep this attractor active during a memory interval in which intervening stimuli are shown. Then when the sample stimulus reappears in the task as a match stimulus, the inferior temporal cortex module shows a large response to the match stimulus, because it is activated both by the visual incoming match stimulus, and by the consistent backprojected memory of the sample stimulus still being represented in the prefrontal cortex memory module (see Figure 3). The prefrontal attractor can be stimulated into activity by the first stimulus when it is inactive, but once in its high firing rate attractor state, it is relatively stable because of the internal positive feedback, and is not likely to be disturbed by further incoming stimuli. The internal recurrent connections must be stronger than the feedforward and feedback connections between the two cortical areas for this to work.^{4,44,45}

This computational model makes it clear that in order for ongoing perception to occur unhindered implemented by posterior cortex (parietal and temporal lobe) networks, there must be a separate set of modules that is capable of maintaining a representation over intervening stimuli. This is the fundamental understanding offered for the evolution and functions of the dorsolateral prefrontal cortex, and it is this ability to provide multiple separate short-term attractor memories that provides I suggest the basis for its functions in planning.⁴

The impairments of attention induced by prefrontal cortex damage may be accounted for in large part by an impairment in the ability to hold the object of attention stably and without distraction in the short-term memory systems in the prefrontal cortex.^{4,38,46}

Attractor Networks Involved in Attention

Short-term memory, and thus attractor networks, are fundamental to top-down attention in the sense that whatever requires attention (e.g., a spatial location) has to be maintained in a short-term memory. The short-term memory then biases competition between the multiple bottom-up items in the stimulus input. The result is an advantage in the neuronal competition between the multiple inputs for the item that receives top-down bias from the short-term memory.^{4,10,47,48} The overall network architecture within the brain by which this is realised is illustrated in Figure 4, in which the prefrontal cortex acts as the short-term memory, which via the top-down backprojections can bias competition in the perceptual areas such as the inferior temporal visual cortex and parietal cortex to implement object and spatial attention.^{4,10,48,49}

Attractor Networks Formed by Forward and Backward Connections Between Cortical Areas

Although one usually thinks of attractors as being formed in the cerebral neocortex by the recurrent collateral connections within a local area of cerebral cortex, the forward and backward connections between two cortical areas can also potentially form an attractor network as can be seen from Figure 3, provided that the forward and backward synapses are associatively modifiable, as seems likely. The forward and backward connections between cortical areas in a hierarchy can thus potentially contribute to the attractor properties of connected cortical areas. An interesting implication is that when a decision is taken, by mechanisms described later, a number

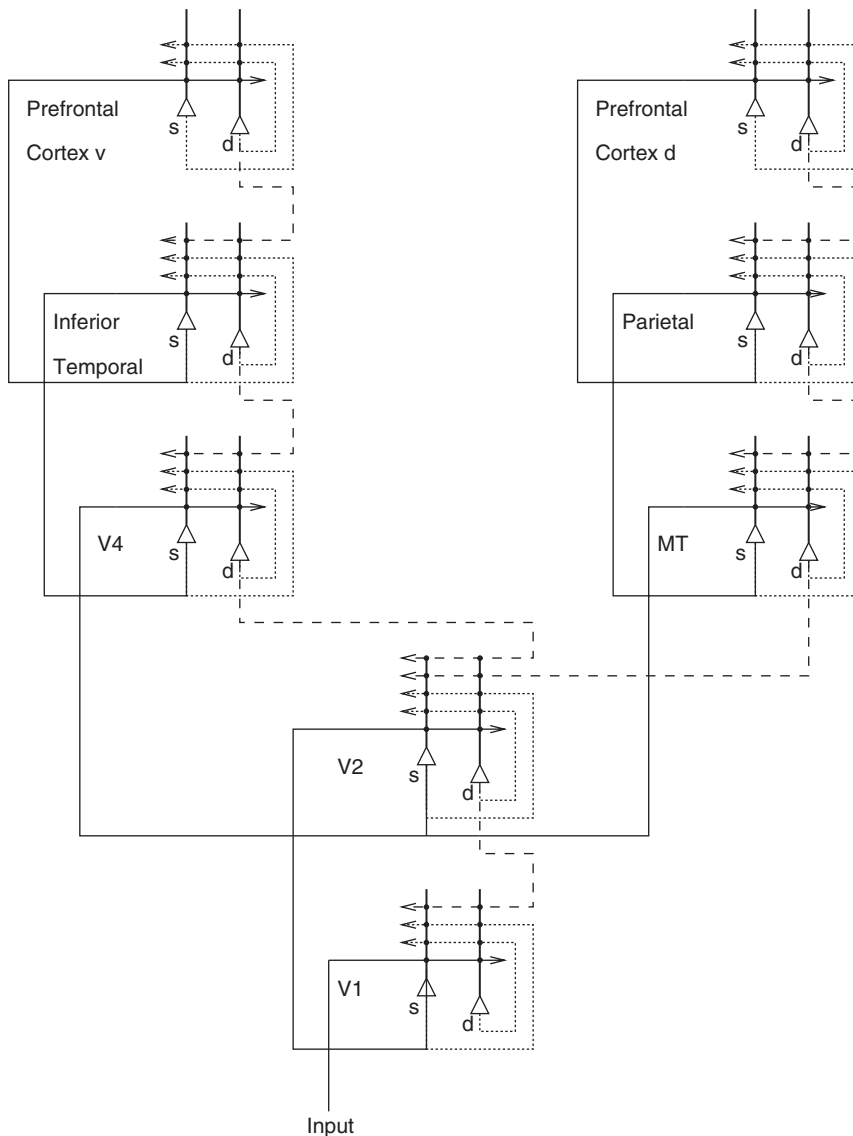


FIGURE 4 | The overall architecture of a model of object and spatial processing and attention, including the prefrontal cortical areas that provide the short-term memory required to hold the object or spatial target of attention active. Forward connections are indicated by solid lines; backprojections, which could implement top-down processing, by dashed lines; and recurrent connections within an area by dotted lines. The triangles represent pyramidal cell bodies, with the thick vertical line above them the dendritic trees. The cortical layers in which the cells are concentrated are indicated by s (superficial, layers 2 and 3) and d (deep, layers 5 and 6). The prefrontal cortical areas most strongly reciprocally connected to the inferior temporal cortex 'what' processing stream are labelled v to indicate that they are in the more ventral part of the lateral prefrontal cortex, area 46, close to the inferior convexity in macaques. The prefrontal cortical areas most strongly reciprocally connected to the parietal visual cortical 'where' processing stream are labelled d to indicate that they are in the more dorsal part of the lateral prefrontal cortex, area 46, in and close to the banks of the principal sulcus in macaques (after Rolls and Deco¹⁰).

of connected cortical areas may contribute to the settling process into an attractor state, and thus to the decision.⁴

HIERARCHICALLY CONNECTED ATTRACTOR NETWORKS: ACTION SELECTION IN THE PREFRONTAL CORTEX

A series of attractor networks can be connected together by forward and backward projections, and interesting properties can arise if the forward connections are stronger than the backward connections. One such scenario is illustrated in Figure 5, a model of the prefrontal cortex in which a set of neurons closer to the sensory input can be activated by inputs from other cortical areas. These 'sensory pools' or populations of

neurons can implement continuing firing and thus a short-term memory of the sensory stimuli, as described above. However, these sensory populations project forward to further populations which represent different combinations of sensory inputs, the associative pools in Figure 5. These intermediate pools have short-term memory properties in their own right, but also connect forward to a set of neurons with more motor-related properties, labelled premotor pools in Figure 5. This hierarchical attractor system can if triggered by a sensory input select an action or motor output as activity cascades through the system, and can also account for the maintenance of neuronal activity during delay periods. Even more, top-down inputs shown as coming from rule attractor pools in Figure 5 can bias the intermediate combination-responding pools of neurons to determine which action is selected by a sensory

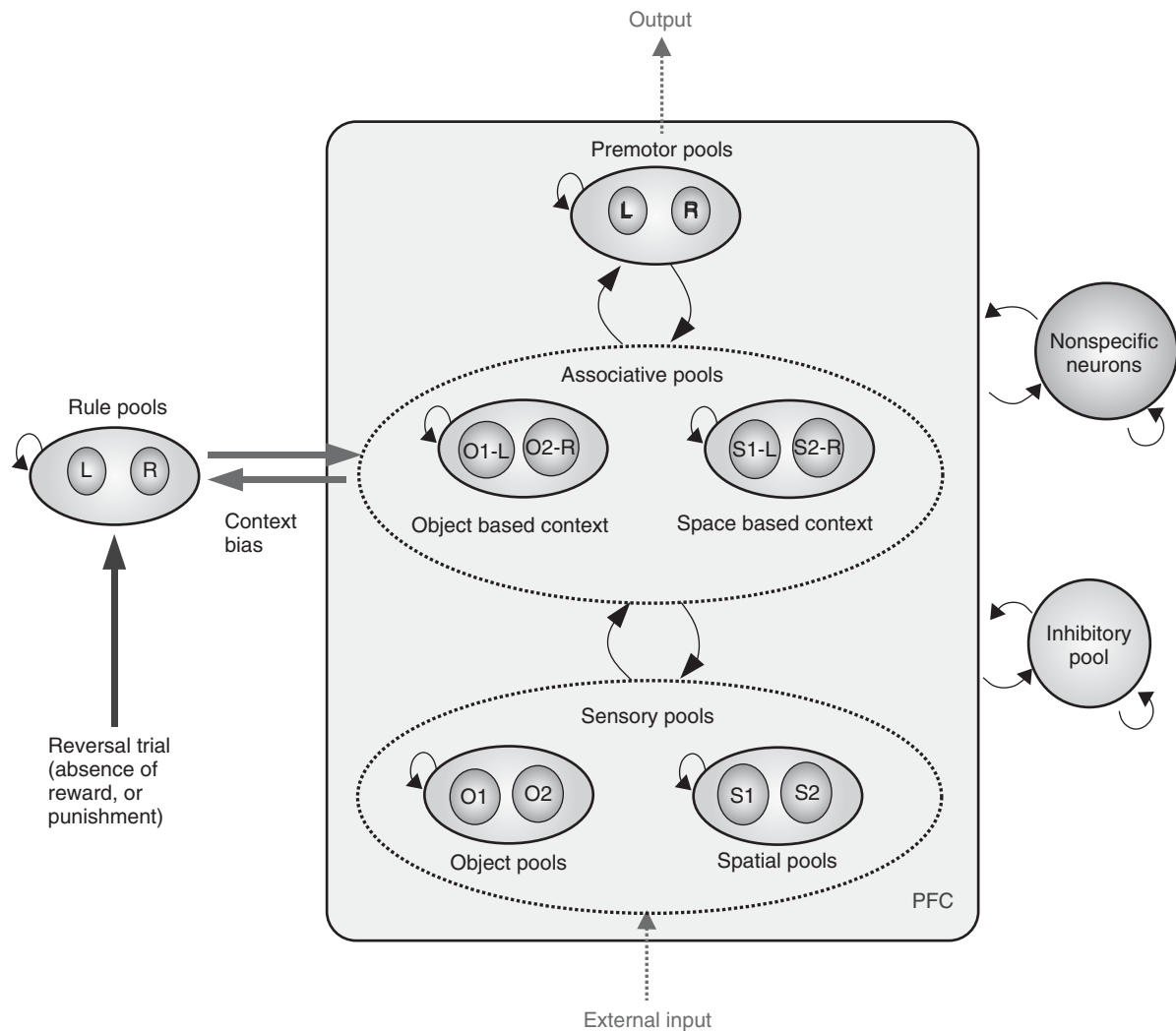


FIGURE 5 | Network architecture of the prefrontal cortex unified model of attention, working memory, action selection, and decision making. There are sensory neuronal populations or pools for object type (O1 or O2) and spatial position (S1 or S2). These connect hierarchically (with stronger forward than backward connections) to the intermediate or ‘associative’ pools in which neurons may respond to combinations of the inputs received from the sensory pools for some types of mapping such as reversal, as described by Deco and Rolls.⁵⁰ For the simulation of the data of Asaad et al.,⁵² these intermediate pools respond to O1-L, O2-R, S1-L, or S2-R. These intermediate pools receive an attentional bias, which in the case of this particular simulation biases either the O pools or the S pools. The intermediate pools are connected hierarchically to the premotor pools, which in this case code for a Left or Right response. Each of the pools is an attractor network in which there are stronger associatively modified synaptic weights between the neurons that represent the same state (e.g., object type for a sensory pool, or response for a premotor pool) than between neurons in the other pools or populations. However, all the neurons in the network are associatively connected by at least weak synaptic weights. The attractor properties, the competition implemented by the inhibitory interneurons, and the biasing inputs result in the same network implementing both short-term memory and biased competition, and the stronger feed forward than feedback connections between the sensory, intermediate, and premotor pools results in the hierarchical property by which sensory inputs can be mapped to motor outputs in a way that depends on the biasing contextual or rule input (after Deco and Rolls⁵⁰).

input, depending on the rule currently held in the rule attractor. This provides a computational model for action selection in the prefrontal cortex depending on the current rule or context⁵⁰, and is consistent, and indeed was based on, the properties of neurons recorded in the prefrontal cortex during action selection tasks with short-term memory requirements.^{51,52}

STABILITY OF ATTRACTOR STATES

Using an integrate-and-fire approach, the individual neurons, synapses and ion channels that comprise an attractor network, can be simulated, and when a threshold is reached the cell fires (see Figure 6(a) and Supporting Information). The firing times of the

neurons can be approximately like those of neurons in the brain, approximately Poisson distributed, that is the firing time is approximately random for a given mean rate. The random firing times of neurons are one source of noise in the attractor network, and can influence the stability of the network.^{53–56} The attractor dynamics can be pictured by effective energy landscapes, which indicate the basin of attraction by valleys, and the attractor states or fixed points by the bottom of the valleys. The stability of an attractor is characterised by the average time in which the system stays in the basin of attraction under the influence of noise, which provokes transitions to other attractor states. Noise results from the interplay between the Poissonian character of the spikes, and the finite-size effect because of the limited numbers of neurons in the network. Two factors determine the stability. First, if the depths of the attractors are shallow (as in the left compared to the right valley in Figure 6(b)), less force is needed to move a ball from the shallow valley to the next. Second, a high level of noise increases the likelihood that the system will jump over an energy boundary from one state to another. We envision that the brain, as a dynamical system, has characteristics of such an attractor system, including statistical fluctuations.

This type of model can then be applied to the prefrontal cortex and used to link these low-level neuronal properties to the cognitive functions such as short-term memories that result from the interactions between thousands of neurons in the whole network. In order to maintain a short-term memory, these interactions have to remain stable, and several factors influence the stability of such a short-term memory attractor state with noise inherent in its operation.

First, the stable states of the network are the ‘low energy’ states in which one set of the neurons, connected by strengthened recurrent collateral synapses, and representing one memory, is activated (see Figures 1 and 2). The higher the firing rates of this set of neurons, the stronger will be the negative feedback inhibition by the GABA inhibitory interneurons to the other excitatory (pyramidal) neurons in the network. This will keep the short-term memory state stable, and will prevent distracting inputs to the other, inhibited, neurons in the network from taking over.⁵⁷ Any factor that reduces the currents through the N-Methyl-D-Aspartate (NMDA) receptors (NMDARs) on the pyramidal cells, as appears to be the case in patients with schizophrenia,⁵⁸ would decrease the firing rates of the set of activated neurons and tend to make the network more distractible.^{59–61}

Second, the strong synaptic connections implemented by the recurrent collateral synapses between

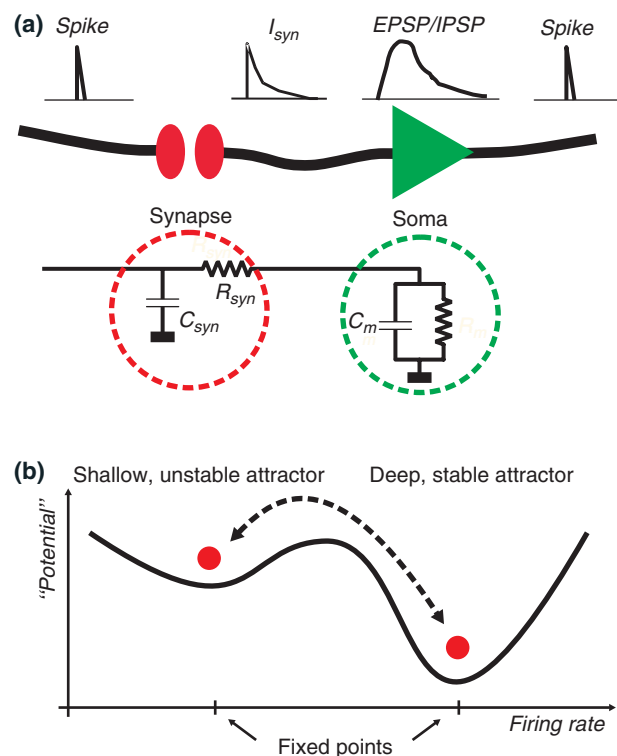


FIGURE 6 | (a) Using an integrate-and-fire approach, the individual neurons, synapses, and ion channels that comprise an attractor network, can be simulated, and when a threshold is reached the cell fires. (b) The attractor dynamics can be pictured by effective energy landscapes, which indicate the basin of attraction by valleys, and the attractor states or fixed points by the bottom of the valleys. The stability of an attractor is characterised by the average time in which the system stays in the basin of attraction under the influence of noise, which provokes transitions to other attractor states. Two factors determine the stability. First, if the depths of the attractors are shallow (as in the left compared to the right valley), less force is needed to move a ball from the shallow valley to the next. Second, a high level of noise increases the likelihood that the system will jump over an energy boundary from one state to another.

the excitatory neurons in the network (e.g., the pyramidal cells in the prefrontal cortex) also tend to promote stability, by enhancing the firing of the neurons that are active for a short-term memory.⁶² This helps to keep the energy low in the Hopfield equation [see Eq. (4)], and thus to make it difficult to jump from one energy minimum over a barrier to a different energy minimum that represents a different memory.

Third, the operation of the network is inherently noisy and probabilistic owing to the random spiking of the individual neurons in the network and the finite size of the network.^{63–67} The random spiking will sometimes (i.e., probabilistically) be large in neurons that are not among those in the currently active set that represents the short-term memory in

mind; this chance effect, perhaps in the presence of a distracting stimulus, might make the network jump over an energy barrier between the memory states into what becomes a different short-term memory, resulting in distraction. In a different scenario, the same type of stochastic noise could make the network jump from a spontaneous state of firing in which there is no item in short-term memory, to an active state in which one of the short-term memories becomes active. In the context of schizophrenia, this might represent an intrusive thought or hallucination.⁶¹ The effects of noise operating in this way would be more evident if the firing rates are low (resulting in a low energy barrier over which to jump); or if the GABA inhibition is reduced, as suggested by *post-mortem* studies of patients with schizophrenia,^{68,69} which would make the spontaneous firing state less stable. GABA interneurons normally inhibit the neurons that are not in the active set that represent a memory, but hypofunction of the NMDARs on GABA interneurons could diminish this inhibition.⁵⁸

Fourth, the stability of the attractor state is enhanced by the long time constants (around 100 ms) of the NMDARs in the network.^{70–73} The contribution of these long time constants (long in relation to those of the alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionate (AMPA) excitatory receptors, which are in the order of 5–10 ms) is to smooth out in time the statistical fluctuations that are caused by the random spiking of populations of neurons in the network, and thus to make the network more stable and less likely to jump to a different state. The different state might represent a different short-term memory; or the noise might return the active state back to the spontaneous level of firing, producing failure of the short-term memory and failure to maintain attention. Further, once a neuron is strongly depolarised, the voltage dependence of the NMDAR may tend to promote further firing.⁷⁰ If the NMDARs were less efficacious, as has been observed in patients with schizophrenia,⁵⁸ the short-term memory network would be less stable because the effective time constant of the whole network would be reduced, owing to the greater relative contribution of the short time constant AMPA receptors to the effects implemented through the recurrent collateral excitatory connections between the pyramidal cells.^{71,72,74}

ATTRACTOR NETWORK STABILITY AND PSYCHIATRIC DISORDERS

It is hypothesised that some of the cognitive symptoms of schizophrenia, including poor short-term memory and attention, can be related to a reduced depth in

the basins of attraction of the attractor networks in the prefrontal cortex that implement these functions.^{60,61,75} The reduced depth of the basins of attraction may be related to hypoglutamatergia,^{58,76} and/or changes in dopaminergic function which act partly by influencing glutamatergic function.^{61,77–80} The negative and positive symptoms of schizophrenia may be related to similar underlying changes, but expressed in different parts of the brain such as the orbitofrontal and anterior cingulate cortex, and the temporal lobes.^{4,60,61,75}

Obsessive–compulsive disorder has been linked to overstability in cortical attractor networks involved in short-term memory, attention, and action selection, related it is hypothesised at least in part to hyperglutamatergia.⁶²

Attractor Networks, Noise, and Decision-Making

Recently, a series of biologically plausible models, motivated and constrained by neurophysiological data, have been formulated to establish an explicit link between probabilistic decision-making and the way in which the noisy (i.e., stochastic) firing of neurons influences which attractor state, representing a decision, is reached when there are two or more competing inputs or sources of evidence to the attractor network.^{55,56,66,81–84} The way in which these decision-making attractor network models operate is as follows.

An attractor network of the type illustrated in Figure 7(a) is set up to have two possible high firing rate attractor states, one for each of the two decisions. The evidence for each decision (1 vs. 2) biases each of the two attractors via the external inputs λ_1 and λ_2 . The attractors are supported by strengthened synaptic connections in the recurrent collateral synapses between the (e.g., cortical pyramidal) neurons activated when λ_1 is applied, or when λ_2 is applied. (This is an associative or Hebbian process set up during a learning stage by a process like long-term potentiation.) Inhibitory interneurons (not shown in Figure 7(a)) receive inputs from the pyramidal neurons and make negative feedback connections onto the pyramidal cells to control their activity. When inputs λ_1 and λ_2 are applied, there is positive feedback via the recurrent collateral connections, and competition implemented through the inhibitory interneurons so that there can be only one winner. The network starts in a low spontaneous state of firing. When λ_1 and λ_2 are applied, there is competition between the two attractors, each of which is pushed towards a high firing rate state,

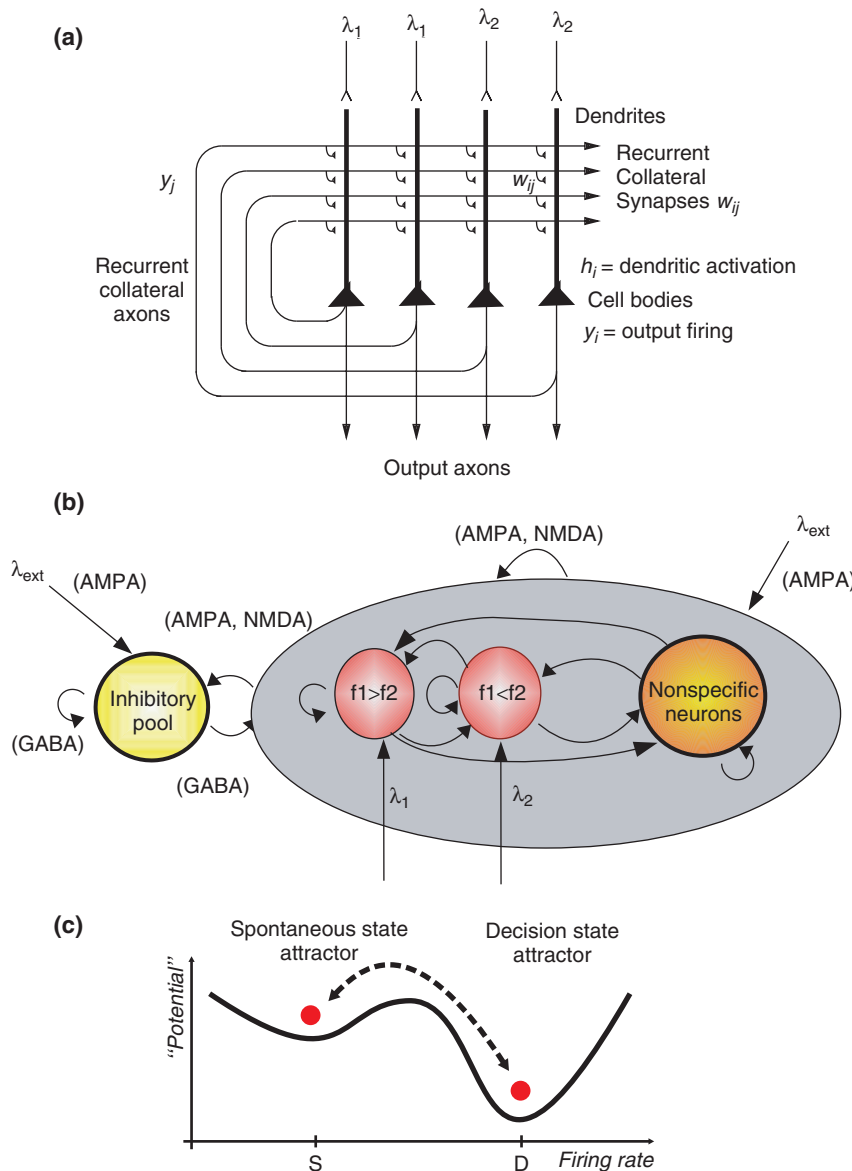


FIGURE 7 | (a) Attractor or autoassociation network architecture for decision making. The evidence for decision 1 is applied via the λ_1 inputs, and for decision 2 via the λ_2 inputs. The synaptic weights w_{ij} have been associatively modified during training in the presence of λ_1 and at a different time of λ_1 . When λ_1 and λ_2 are applied, each attractor competes through the inhibitory interneurons (not shown), until one wins the competition, and the network falls into one of the high firing rate attractors that represents the decision. The noise in the network caused by the random spiking of the neurons means that on some trials, for given inputs, the neurons in the decision 1 attractor are more likely to win, and on other trials the neurons in the decision 2 attractor are more likely to win. This makes the decision-making probabilistic, for, as shown in (c), the noise influences when the system will jump out of the spontaneous firing stable (low energy) state S, and whether it jumps into the high firing state for decision 1 or decision 2 (D). (b) The architecture of the integrate-and-fire network used to model vibrotactile decision making (see text).

and eventually, depending on the relative strength of the two inputs, and the noise in the network caused by the random firing times of the neurons, one of the attractors will win the competition, and it will reach a high firing rate state, with the firing of the neurons in the other attractor inhibited to a low firing rate. The process is illustrated in Figure 8. The result is a binary decision, with one group of neurons because of the positive feedback firing at a high firing rate, and the neurons corresponding to the other decision firing with very low rates. Because it is a nonlinear positive-feedback system, the final firing rates are in what is effectively a binary decision state, of high firing rate or low firing rate, and do not reflect the exact relative values of the two inputs

λ_1 and λ_2 once the decision is reached. The noise in the network because of the random spiking of the neurons is important to the operation of the network, because it enables the network to jump out of a stable spontaneous rate of firing to a high firing rate, and to do so probabilistically, depending on whether on a particular trial there is relatively more random firing in the neurons of one attractor than the other attractor. This can be understood in terms of energy landscapes, where each attractor (the spontaneous state, and the two high firing rate attractors) is a low energy basin, and the spiking noise helps the system to jump over an energy barrier into another energy minimum, as illustrated in Figure 7(c). If λ_1 and λ_2 are equal, then the decision that is taken is random and probabilistic,

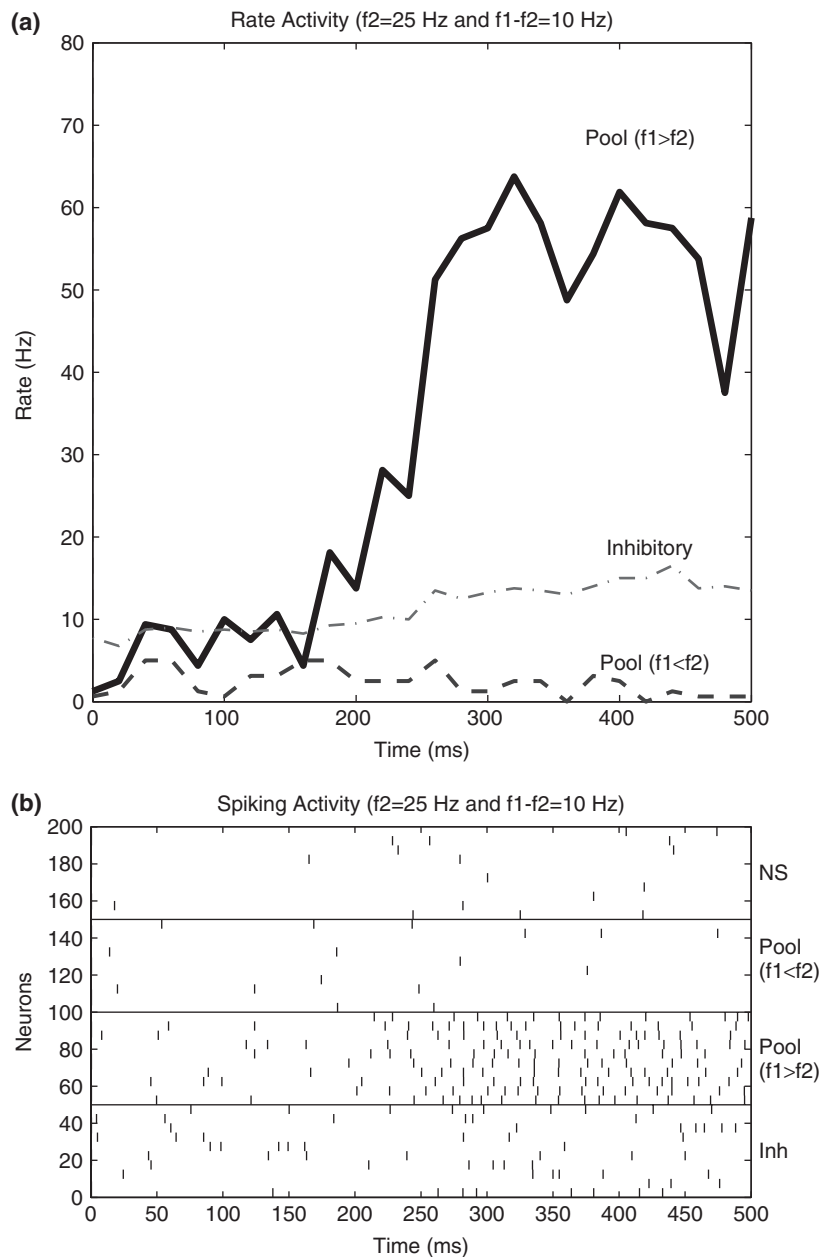


FIGURE 8 | Decision making in a model of vibrotactile decision making. Dynamical evolution of the network activity of Ventral Premotor Cortex neurons during the comparison period between vibrotactile frequency $f_1 = 30$ Hz and frequency $f_2 = 20$ Hz. (a) The evolution as a function of time of the spiking rate of the populations ($f_1 > f_2$) (corresponding to a decision that f_1 is greater than f_2), ($f_1 < f_2$), and the inhibitory population. (b) The corresponding rastergrams of 10 randomly selected neurons for each pool (population of neurons) in the network. Each vertical line corresponds to the generation of a spike. The spatio-temporal spiking activity shows the transition to the correct final single-state attractor, i.e., a transition to the correct final attractor encoding the result of the discrimination ($f_1 > f_2$) (after Deco and Rolls⁶⁶).

with the noise in each attractor determining which decision is taken on a particular trial. If one of the inputs is larger than the other, then the decision is biased towards it, but is still probabilistic. Because this is an attractor network, it has short-term memory properties implemented by the recurrent collaterals, which tend to promote a state once it is started, and these help it to accumulate evidence over time, an important part of a decision-making mechanism, and also to maintain the firing once it has reached the decision state, enabling a suitable action to be implemented even if this takes some time.

This approach to decision making shows how noise in the brain can be useful, how we can account for probabilistic choice and the way it is influenced by the odds as in the Matching Law, how we can account for reaction times for easy versus difficult decisions, and even how Weber's Law may be implemented in the brain.^{4,55,56,66}

Attractor Networks, Noise, and Signal Detection

A similar approach has been taken to the detection of signals, where noise caused by the stochastic firing of

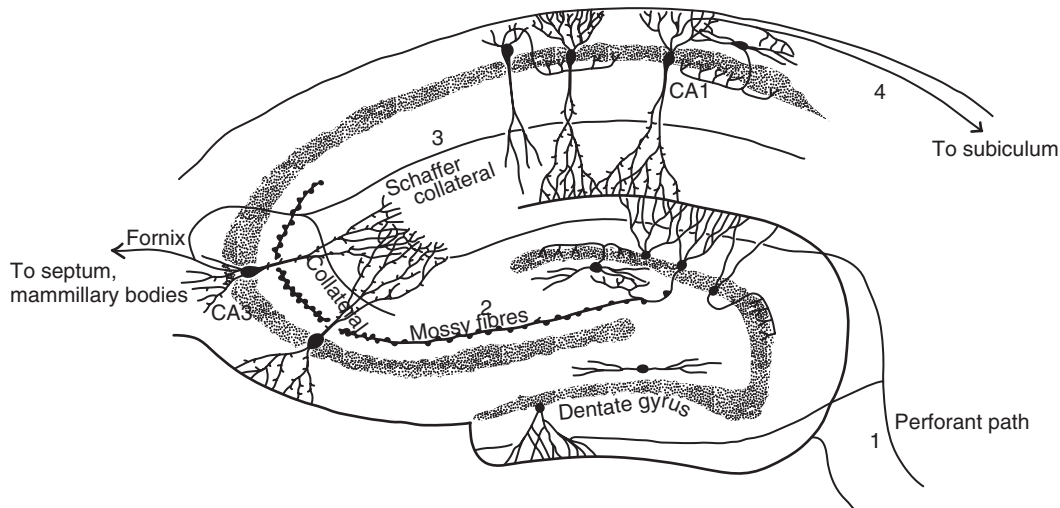


FIGURE 9 | Representation of connections within the hippocampus. Inputs reach the hippocampus through the perforant path (1) which makes synapses with the dendrites of the dentate granule cells and also with the apical dendrites of the CA3 pyramidal cells. The dentate granule cells project via the mossy fibres (2) to the CA3 pyramidal cells. The well-developed recurrent collateral system of the CA3 cells is indicated. The CA3 pyramidal cells project via the Schaffer collaterals (3) to the CA1 pyramidal cells, which in turn have connections (4) to the subiculum.

neurons influences how an attractor network may or may not fall into a basin of attraction that represents the detection of a signal.^{55,56}

Hippocampal Versus Neocortical Attractor Networks

The neocortex has local recurrent collateral connections between the pyramidal that achieve a high density only for a few millimetres across the cortex. It is hypothesised that this enables the neocortex to have many local attractor networks, each concerned with a different type of processing, short-term memory, long-term memory, decision making, etc.⁴ This is important, for recall that the capacity of an attractor network is set to first order by the number of connections onto a neuron from other neurons in the network. If there were widespread recurrent collateral connections in the neocortex so that the whole neocortex operated as a single attractor, the total memory capacity of the neocortex would be only that of a single attractor network (of order thousands of memories), and this possibility is thus ruled out.^{4,85} There are great advantages in having large numbers of local but weakly couple neocortical attractor networks, and some have been described above, and many more are described by Rolls.⁴

However, it has been suggested that one network in the brain, the hippocampal CA3 network, does operate as a *single* attractor network^{4,9,15,86–94} (with related approaches, although not emphasising the relative importance of a single attractor network (or

CA3⁹⁵) including.^{96–99}) Part of the anatomical basis for this is that the recurrent collateral connections between the CA3 neurons are very widespread, and have a chance of contacting any other CA3 neuron in the network (see Figure 9).^{14,100} The underlying theory is that the associativity in the network allows any one set of active neurons, perhaps representing one part of an episodic memory, to have a fair chance of making modifiable synaptic contacts with any other set of CA3 neurons perhaps representing another part of an episodic memory. (An episodic memory is a memory of a single event or episode, such as where one was at dinner, with whom, what was eaten, and what was discussed.) This widespread connectivity providing for a single attractor network means that any one part of an episodic memory can be associated with any other part of an episodic or event memory. (This is what I mean by calling this an arbitrary memory, in that any arbitrary set of events can be associated with any other.) Now this functionality would be impossible in the neocortex, as the connections are local. This is thus a special contribution that the hippocampus can make to event or episodic memory.^{4,93,101,102}

CONCLUSION

I propose that attractor networks are fundamental design features of the neocortex and hippocampal cortex (and not of, e.g., the cerebellar cortex or basal ganglia). In the neocortex the attractor networks are local and therefore there can be many of

them. They allow many items of information to be held on-line, and thus provide the basis and/or underpinning for powerful computations that require short-term memory, working memory (which involves the manipulation of items in short-term memory), planning, attention, and even language (which requires multiple items to be held on-line during the parsing of a sentence). In the hippocampal cortex, an attractor network in the CA3 region allows associations between any events that co-occur, and thus provides a basis for the memory of particular episodes,

and the recall of an episodic memory from any part.

NOTES

^aEach memory is precisely defined in the theory: it is a set of firing rates of the population of neurons (which represent a memory) that can be stored and later retrieved, with retrieval being possible from a fraction of the originally stored set of neuronal firing rates.

REFERENCES

- Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A* 1982, 79:2554–2558.
- Amit DJ. *Modeling Brain Function*. Cambridge: Cambridge University Press; 1989.
- Hertz J, Krogh A, Palmer RG. *An Introduction to the Theory of Neural Computation*. Wokingham: Addison-Wesley; 1991.
- Rolls ET. *Memory, Attention, and Decision-making: A Unifying Computational Neuroscience Approach*. Oxford: Oxford University Press; 2008.
- Hebb DO. *The Organization of Behavior: A Neuropsychological Theory*. New York: John Wiley & Sons; 1949.
- Rolls ET, Treves A. The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network* 1990, 1:407–421.
- Treves A, Rolls ET. What determines the capacity of autoassociative memories in the brain? *Network* 1991, 2:371–397.
- Fazeli MS, Collingridge GL, eds. *Cortical Plasticity: LTP and LTD*. Oxford: Bios; 1996.
- Rolls ET, Treves A. *Neural Networks and Brain Function*. Oxford: Oxford University Press; 1998.
- Rolls ET, Deco G. *Computational Neuroscience of Vision*. Oxford: Oxford University Press; 2002.
- Treves A. Graded-response neurons and information encodings in autoassociative memories. *Phys Rev A* 1990, 42:2418–2430.
- Treves A, Rolls ET. Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* 1992, 2:189–199.
- Rolls ET, Treves A, Foster D, Perez-Vicente C. Simulation studies of the CA3 hippocampal subfield modelled as an attractor neural network. *Neural Netw* 1997, 10:1559–1569.
- Amaral DG, Ishizuka N, Claiborne B. Neurons, numbers and the hippocampal network. *Prog Brain Res* 1990, 83:1–11.
- Treves A, Rolls ET. A computational analysis of the role of the hippocampus in memory. *Hippocampus* 1994, 4:374–391.
- Franco L, Rolls ET, Aggelopoulos NC, Jerez JM. Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex. *Biol Cybern* 2007, 96:547–560.
- Rolls ET. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philos Trans R Soc Lond B Biol Sci* 1992, 335:11–21.
- Rolls ET. Consciousness absent and present: a neurophysiological exploration. *Prog Brain Res* 2003, 144:95–106.
- Panzeri S, Rolls ET, Battaglia F, Lavis R. Speed of information retrieval in multilayer networks of integrate-and-fire neurons. *Netw Comput Neural Syst* 2001, 12:423–440.
- Treves A. Mean-field analysis of neuronal spike dynamics. *Network* 1993, 4:259–284.
- Battaglia FP, Treves A. Stable and rapid recurrent processing in realistic auto-associative memories. *Neural Comput* 1998, 10:431–450.
- Amari S. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol Cybern* 1977, 27:77–87.
- Zhang K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci* 1996, 16:2112–2126.
- Taylor JG. Neural “bubble” dynamics in two dimensions: foundations. *Biol Cybern* 1999, 80:393–409.
- Samsonovich A, McNaughton BL. Path integration and cognitive mapping in a continuous attractor neural network model. *J Neurosci* 1997, 17:5900–5920.

26. Battaglia FP, Treves A. Attractor neural networks storing multiple space representations: a model for hippocampal place fields. *Phys Rev E* 1998, 58:7738–7753.
27. Stringer SM, Trappenberg TP, Rolls ET, Araujo IET. Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells. *Netw Comput Neural Syst* 2002, 13:217–242.
28. Stringer SM, Rolls ET, Trappenberg TP, Araujo IET. Self-organizing continuous attractor networks and path integration. Two-dimensional models of place cells. *Netw Comput Neural Syst* 2002, 13:429–446.
29. Stringer SM, Rolls ET, Trappenberg TP. Self-organizing continuous attractor network models of hippocampal spatial view cells. *Neurobiol Learn Mem* 2005, 83:79–92.
30. Stringer SM, Rolls ET, Trappenberg TP, de Araujo IET. Self-organising continuous attractor networks and motor function. *Neural Netw* 2003, 16:161–182.
31. Stringer SM, Rolls ET, Trappenberg TP. Self-organising continuous attractor networks with multiple activity packets, and the representation of space. *Neural Netw* 2004, 17:5–27.
32. Rolls ET, Stringer SM. Spatial view cells in the hippocampus, and their idiothetic update based on place and head direction. *Neural Netw* 2005, 18:1229–1241.
33. Stringer SM, Rolls ET. Self-organizing path integration using a linked continuous attractor and competitive network: path integration of head direction. *Netw Comput Neural Syst* 2006, 17:419–445.
34. Rolls ET, Stringer SM, Trappenberg TP. A unified model of spatial and episodic memory. *Proc R Soc Lond B Biol Sci* 2002, 269:1087–1093.
35. Braitenberg V, Schütz A. *Anatomy of the Cortex*. Berlin: Springer-Verlag; 1991.
36. Abeles M. *Corticonics - Neural Circuits of the Cerebral Cortex*. New York: Cambridge University Press; 1991.
37. Goldman-Rakic PS. Cellular basis of working memory. *Neuron* 1995, 14:477–485.
38. Goldman-Rakic PS. The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Philos Trans R Soc Lond B Biol Sci* 1996, 351:1445–1453.
39. Fuster JM. Executive frontal functions. *Exp Brain Res* 2000, 133(1):66–70.
40. Fuster JM, Alexander GE. Neuron activity related to short-term memory. *Science* 1971, 173:652–654.
41. Kubota K, Niki H. Prefrontal cortical unit activity and delayed alternation performance in monkeys. *J Neurophysiol* 1971, 34(3):337–347.
42. Funahashi S, Bruce CJ, Goldman-Rakic PS. Mnemonic coding of visual space in monkey dorsolateral prefrontal cortex. *J Neurophysiol* 1989, 61:331–349.
43. Fuster JM. *Memory in the Cerebral Cortex*. Cambridge, MA: MIT Press; 1995.
44. Renart A, Parga N, Rolls ET. A recurrent model of the interaction between the prefrontal cortex and inferior temporal cortex in delay memory tasks. In: Solla SA, Leen TK, Mueller K-R, eds. *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press; 2000, 171–177.
45. Renart A, Moreno R, de la Rocha J, Parga N, Rolls ET. A model of the IT-PF network in object working memory which includes balanced persistent activity and tuned inhibition. *Neurocomputing* 2001, 38–40:1525–1531.
46. Goldman-Rakic PS, Leung H-C. Functional architecture of the dorsolateral prefrontal cortex in monkeys and humans. In: Stuss DT, Knight RT, eds. *Principles of Frontal Lobe Function*. New York: Oxford University Press; 2002, 85–95.
47. Desimone R, Duncan J. Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 1995, 18:193–222.
48. Deco G, Rolls ET. Attention, short-term memory, and action selection: a unifying theory. *Prog Neurobiol* 2005, 76:236–256.
49. Deco G, Rolls ET. Neurodynamics of biased competition and co-operation for attention: a model with spiking neurons. *J Neurophysiol* 2005, 94:295–313.
50. Deco G, Rolls ET. Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. *Eur J Neurosci* 2003, 18:2374–2390.
51. Asaad WF, Rainer G, Miller EK. Neural activity in the primate prefrontal cortex during associative learning. *Neuron* 1998, 21:1399–1407.
52. Asaad WF, Rainer G, Miller EK. Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol* 2000, 84:451–459.
53. Tuckwell H. *Introduction to Theoretical Neurobiology*. Cambridge: Cambridge University Press; 1988.
54. Jackson BS. Including long-range dependence in integrate-and-fire models of the high interspike-interval variability of cortical neurons. *Neural Comput* 2004, 16(10):2125–2195.
55. Deco G, Rolls ET, Romo R. Stochastic dynamics as a principle of brain function. *Prog Neurobiol* 2009, 88:1–16.
56. Rolls ET, Deco G. *Stochastic dynamics as a principle of brain function: noise in the Brain*, 2009.
57. Brunel N, Wang XJ. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J Comput Neurosci* 2001, 11:63–85.
58. Coyle JT. Glutamate and schizophrenia: beyond the dopamine hypothesis. *Cell Mol Neurobiol* 2006, 26(4–6):365–384.

59. Durstewitz D, Seamans JK, Sejnowski TJ. Neurocomputational models of working memory. *Nat Neurosci* 2000, 3:1184–1191.
60. Loh M, Rolls ET, Deco G. A dynamical systems hypothesis of schizophrenia. *PLoS Comput Biol* 2007, 3(11):e228. doi:10.1371/journal.pcbi.0030228.
61. Rolls ET, Loh M, Deco G, Winterer G. Computational models of schizophrenia and dopamine modulation in the prefrontal cortex. *Nat Rev Neurosci* 2008, 9:696–709.
62. Rolls ET, Loh M, Deco G. An attractor hypothesis of obsessive-compulsive disorder. *Eur J Neurosci* 2008, 28:782–793.
63. Brunel N, Hakim V. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput*. 1999, 11(7):1621–1671.
64. Mattia M, Del Giudice P. Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. *Phys Rev E* 2002, 66:51917–51919.
65. Mattia M, Del Giudice P. Finite-size dynamics of inhibitory and excitatory interacting spiking neurons. *Phys Rev E Stat Nonlin Soft Matter Phys* 2004, 70(5 pt 1): 052903.
66. Deco G, Rolls ET. Decision-making and Weber's Law: a neurophysiological model. *Eur J Neurosci* 2006, 24:901–916.
67. Faisal AA, Selen LP, Wolpert DM. Noise in the nervous system. *Nat Rev Neurosci* 2008, 9(4):292–303.
68. Benes FM. Emerging principles of altered neural circuitry in schizophrenia. *Brain Res Brain Res Rev* 2000, 31(2–3):251–269.
69. Hashimoto T, Arion D, Unger T, Maldonado-Avilés JG, Morris HM, et al. Alterations in GABA-related transcriptome in the dorsolateral prefrontal cortex of subjects with schizophrenia. *Mol Psychiatry* 2008, 13(2): 147–161.
70. Lisman JE, Fellous JM, Wang XJ. A role for NMDA-receptor channels in working memory. *Nat Neurosci* 1998, 1(4):273–275.
71. Wang X-J. Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *J Neurosci* 1999, 19(21):9587–9603.
72. Compte A, Brunel N, Goldman-Rakic PS, Wang XJ. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb Cortex* 2000, 10(9): 910–923.
73. Wang XJ. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci* 2001, 24(8): 455–463.
74. Tegner J, Compte A, Wang XJ. The dynamical stability of reverberatory neural circuits. *Biol Cybern* 2002, 87(5–6):471–481.
75. Rolls ET. *Emotion Explained*. Oxford: Oxford University Press; 2005.
76. Coyle JT, Tsai G, Goff D. Converging evidence of NMDA receptor hypofunction in the pathophysiology of schizophrenia. *Ann N Y Acad Sci* 2003, 1003:318–327.
77. Seamans JK, Yang CR. The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Prog Neurobiol* 2004, 74(1):1–58.
78. Durstewitz D. A few important points about dopamine's role in neural network dynamics. *Pharmacopsychiatry* 2006, 39(Suppl 1):S72–S75.
79. Durstewitz D. Dopaminergic modulation of prefrontal cortex network dynamics. In: Tseng K-Y, Atzori M, eds. *Monoaminergic Modulation of Cortical Excitability*. New York: Springer; 2007, 217–234.
80. Winterer G, Weinberger DR. Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends Neurosci* 2004, 27(11): 683–690.
81. Wang XJ. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 2002, 36:955–968.
82. Brody CD, Romo R, Kepecs A. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Curr Opin Neurobiol* 2003, 13:204–211.
83. Machens CK, Romo R, Brody CD. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* 2005, 307:1121–1124.
84. Wong KF, Wang XJ. A recurrent network mechanism of time integration in perceptual decisions. *J Neurosci* 2006, 26(4):1314–1328.
85. O'Kane D, Treves A. Why the simplest notion of neocortex as an autoassociative memory would not work. *Network* 1992, 3:379–384.
86. Rolls ET. 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*. Chichester: John Wiley & Sons; 1987, 503–540.
87. Rolls ET. Functions of neuronal networks in the hippocampus and neocortex in memory. In: Byrne JH, Berry WO, eds. *Neural Models of Plasticity: Experimental and Theoretical Approaches*. San Diego, CA: Academic Press; 1989, 240–265.
88. Rolls ET. 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*. Wokingham: Addison-Wesley; 1989, 125–159.
89. Rolls ET. Functions of neuronal networks in the hippocampus and cerebral cortex in memory. In: Cotterill RMJ, ed. *Models of Brain Function*. Cambridge: Cambridge University Press; 1989, 15–33.
90. Rolls ET. Theoretical and neurophysiological analysis of the functions of the primate hippocampus in

- memory. *Cold Spring Harb Symp Quant Biol* 1990, 55:995–1006.
91. Rolls ET. Functions of the primate hippocampus in spatial processing and memory. In: Olton DS, Kesner RP, eds. *Neurobiology of Comparative Cognition*. Hillsdale, NJ: L. Erlbaum; 1990, 339–362.
 92. Rolls ET. Functions of the primate hippocampus in spatial and non-spatial memory. *Hippocampus* 1991, 1:258–261.
 93. Rolls ET, Kesner RP. A computational theory of hippocampal function, and empirical tests of the theory. *Prog Neurobiol* 2006, 79:1–48.
 94. Rolls ET. An attractor network in the hippocampus: theory and neurophysiology. *Learn Mem* 2007, 14:714–731.
 95. Marr D. Simple memory: a theory for archicortex. *Philos Trans R Soc Lond B Biol Sci* 1971, 262: 23–81.
 96. McNaughton BL, Morris RGM. Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci* 1987, 10(10):408–415.
 97. Levy WB. A computational approach to hippocampal function. In: Hawkins RD, Bower GH, eds. *Computational Models of Learning in Simple Neural Systems*. San Diego, CA: Academic Press; 1989, 243–305.
 98. McNaughton BL. Associative pattern completion in hippocampal circuits: new evidence and new questions. *Brain Res Rev* 1991, 16:193–220.
 99. McClelland JL, McNaughton BL, O'Reilly RC. 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* 1995, 102:419–457.
 100. Ishizuka N, Weber J, Amaral DG. Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. *J Comp Neurol* 1990, 295:580–623.
 101. Rolls ET. The primate hippocampus and episodic memory. In: Dere E, Easton A, Nadel L, Huston JP, eds. *Handbook of Episodic Memory*. Amsterdam: Elsevier; 2008, 417–438.
 102. Rolls ET, Xiang J-Z. Spatial view cells in the primate hippocampus, and memory recall. *Rev Neurosci* 2006, 17:175–200.

FURTHER READING

- Amit DJ. *Modeling Brain Function*. Cambridge: Cambridge University Press; 1989.
- Deco G, Rolls ET, Romo R. *Stochastic dynamics as a principle of brain function*. *Prog Neurobiol* 2009, 88: 1–16.
- Hertz J, Krogh A, Palmer RG. *An Introduction to the Theory of Neural Computation*. Wokingham: Addison-Wesley; 1991.
- Rolls ET. *Memory, Attention, and Decision-Making: A Unifying Computational Neuroscience Approach*. Oxford: Oxford University Press; 2008.
- Rolls ET, Deco G. *Computational Neuroscience of Vision*. Oxford: Oxford University Press; 2002.
- Rolls ET, Loh M, Deco G, Winterer G. Computational models of schizophrenia and dopamine modulation in the prefrontal cortex. *Nat Rev Neurosci* 2008, 9:696–709.
- Rolls ET, Deco G. *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*. Oxford: Oxford University Press; 2010.