

Slow semantic learning in the cerebral cortex, and its relation to the hippocampal episodic memory system

Edmund T. Rolls^{1,2,3,*}, Chenfei Zhang³, and Jianfeng Feng^{2,3}

¹Oxford Centre for Computational Neuroscience, Oxford, United Kingdom

²Department of Computer Science, University of Warwick, Coventry CV4 7AL, United Kingdom

³Institute of Science and Technology for Brain Inspired Intelligence, Fudan University, Shanghai 200403, China

*Corresponding author: Edmund T. Rolls, Department of Computer Science, University of Warwick, Coventry CV4 7AL, United Kingdom.

Email: Edmund.Rolls@oxcns.org, URL: <https://www.oxcns.org>

A key question is how new semantic representations are formed in the human brain and how this may benefit from the hippocampal episodic memory system. Here, we describe the major effective connectivity between the hippocampal memory system and the anterior temporal lobe (ATL) semantic memory system in humans. Then, we present and model a theory of how semantic representations may be formed in the human ATL using slow associative learning in semantic attractor networks that receive inputs from the hippocampal episodic memory system. The hypothesis is that if one category of semantic representations is being processed for several seconds, then a slow short-term memory trace associative biologically plausible learning rule will enable all the components during that time to be associated together in a semantic attractor network. This benefits from the binding of components provided by the hippocampal episodic memory system. The theory is modeled in a four-layer network for view-invariant visual object recognition, followed by a semantic attractor network layer that utilizes a temporal trace associative learning rule to form semantic categories based on the inputs that occur close together in time, using inputs from the hippocampal system or from the world.

Keywords: anterior temporal lobe; hippocampus; inferior temporal visual cortex; object recognition; semantic learning; VisNet.

Introduction

Key questions for understanding cortical function in humans are how semantic memories are formed and how the hippocampal episodic memory system may contribute to the formation of new semantic memories in the anterior temporal lobe (ATL). To address this, the effective connectivity from the hippocampal episodic memory system to the ATL semantic memory system in humans is measured and described. The research also provides a new biologically plausible theory and model of how the hippocampal memory system may, by binding components, contribute to the learning of new semantic representations in attractor networks in the ATL using slow learning with a biologically plausible learning rule. This biologically plausible approach is compared with previous models of semantic learning, which have typically involved deep learning of how thousands of attributes could be used to define semantic categories.

Semantic representations are based on knowledge acquired over time about, for example, semantic categories such as cars and animals. The human ATL is implicated in the storage of semantic memories (Patterson et al. 2007; Peelen and Caramazza 2012; Bonner and Price 2013; Hickok and Poeppel 2015; Kemmerer 2015; Milton et al. 2021; Rolls et al. 2022a; Rolls 2023a). This contrasts with episodic memory, which is prototypically about an event at a particular time, for example, a particular car (“What”) in a particular location (“Where”), and in which the hippocampus is implicated (Squire and Wixted 2011; Bennett and Stark 2016; Moscovitch et al. 2016; Rolls and Treves 2024; Rolls et al. 2024e). The slow learning in the semantic system contrasts with

the fast learning in the hippocampal episodic memory system (McClelland et al. 1995, 2020). Interestingly, there is evidence that the hippocampal episodic memory is implicated in learning new semantic representations, based for example, on semantic learning impairments after hippocampal damage (Duff et al. 2019).

One approach to learning semantic representations has been to predefine the categories to be learned and to use deep neural networks trained by backpropagation of error to help learn the exceptions, for example, that some birds cannot fly (Hinton 1989; Rumelhart 1990; Rumelhart and Todd 1993; McClelland et al. 1995). Another approach was to first train a deep convolutional network to categorize visual objects and then to utilize an attractor network to learn the many different semantic attributes (such as “has wheels”) that help to define each object category (eg a car) (Devereux et al. 2018). The synaptic weights between the object nodes and the semantic attribute nodes in the attractor network were trained using continuous recurrent backpropagation through time, such that correct semantic representations were recalled each time that an object was presented (Devereux et al. 2018). However, models such as these that use deep learning with backpropagation of error are unlikely to be biologically plausible, because, for example, they do not use a local synaptic learning rule (Plebe and Grasso 2019; Rolls 2023a). In the research described here, an approach to learning semantic representations is described that does use local learning rules, and that is accordingly more biologically plausible as a model that may help to account for how semantic representations are learned in the brain.

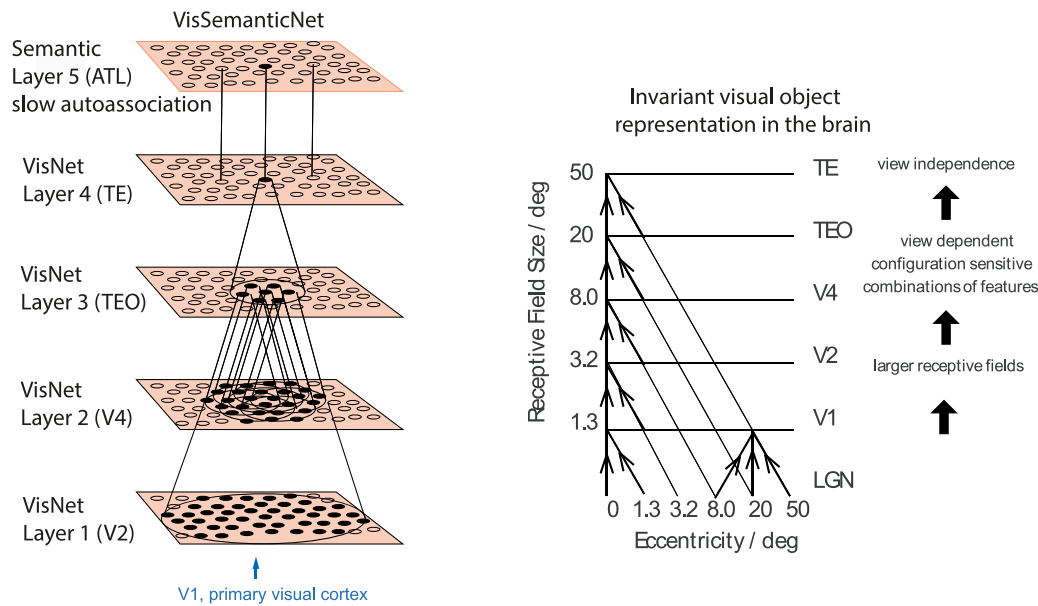


Fig. 1. The model, VisSemanticNet, for building semantic representations in the ATL, which is Layer 5 in the model. The first four layers are VisNet (Rolls 2012, 2021a), which operates by convergence from layer to layer, competitive feedforward learning that utilizes a local associative trace learning rule with a sufficiently long time constant to capture different views of an object while it transforms. Layer 5 is an autoassociation network that utilizes a local associative trace learning rule with a sufficiently long time constant to capture the different objects in one semantic category while that semantic category is being processed. Part of the input to Layer 5 could also be from other sensory systems such as auditory, somatosensory, taste, and olfactory, to contribute to building multimodal semantic representations. TEO, Posterior inferior temporal visual cortex. TE, Inferior temporal visual cortex.

Before describing the Methods and Results of the effective connectivity analyses and the results obtained with the model to be described here, the new approach and hypotheses on how semantic representations may be built using slow associative learning in attractor networks in the ATL are described in the next section. The aim here is to show how categories can be formed in semantic representations on the basis of the temporal proximity of the components. Ways in which the physical similarity of the components can contribute to the semantic representations that are formed are referred to in the Discussion.

There is a wealth of evidence the neurons in the primate hippocampus are involved in storing and later recalling episodic memories such as “what” has been shown recently and “where” has been shown (Cahusac et al. 1989; Rolls et al. 1989; Rolls et al. 2005; Rolls and Xiang 2006; Yang and Naya 2020; Yang and Naya 2023; Rolls 2023a, 2023d; Khazali et al. 2024; Rolls and Treves 2024; Rolls 2025b). The aim here is to show how this hippocampal episodic memory capability may contribute to the formation of new semantic memories in the neocortex.

Hypotheses on slow learning of semantic representations

Invariant representations of objects are formed in the inferior temporal visual cortex

The first hypothesis is that invariant representations of objects are formed in the ventrolateral stream to the inferior temporal visual cortex (Rolls et al. 2023a; Rolls 2024a). The formation of representations of objects that are invariant with respect to view, position, size, and contrast (Wallis and Rolls 1997; Rolls 2012, 2021a) is a useful prerequisite for forming semantic memories, for this greatly simplifies what needs to be learned in the semantic system. The view-invariant representations are formed here in a biologically plausible approach to object representation in which competitive networks build new invariant representations

of objects using slow learning that spans the short period of 1 to 2 s in which we might see an object transforming through different views, sizes, positions on the retina, etc. The local learning rule is associative learning, the Hebb rule, with a short-term memory trace in the post-synaptic term that spans a short time in a natural environment in which primates, including humans, may be looking at an object (Rolls 1992; Wallis and Rolls 1997; Wiskott and Sejnowski 2002; Wyss et al. 2006; Franzius et al. 2007; Rolls 2012, 2021a). VisNet is a model of these biologically plausible computations for invariant visual object representations, with the code available (Rolls 2012, 2021a, 2023a), and was used in the research described here for the first four layers of the network illustrated in Fig. 1.

Semantic representations are built in the ATL with the help of slow learning in autoassociation/attractor networks

The second hypothesis is that the invariant representations of objects are fed into an autoassociation network in the ATL (see Fig. 2) that uses slow learning that spans the longer period of several seconds in which we might be thinking about or seeing/hearing/tasting/smelling/feeling the representations in one semantic category, for example animals. Another example would be forming an episodic memory of an event and then recalling similar episodic memories, enabling the content of these related episodic (“snapshot”) memories to be associated together by slow learning into a semantic representation, which would lose the information about the individual events but form a representation of what overlapped between those events, for example, a set of people and places. Another example might be that the sensory inputs present when sitting in a room working with a colleague and a computer might be attributes that would help to define an office. Another example might be that we would keep thinking about a semantic category if what appeared to be an exceptional case within that category was encountered. The local learning

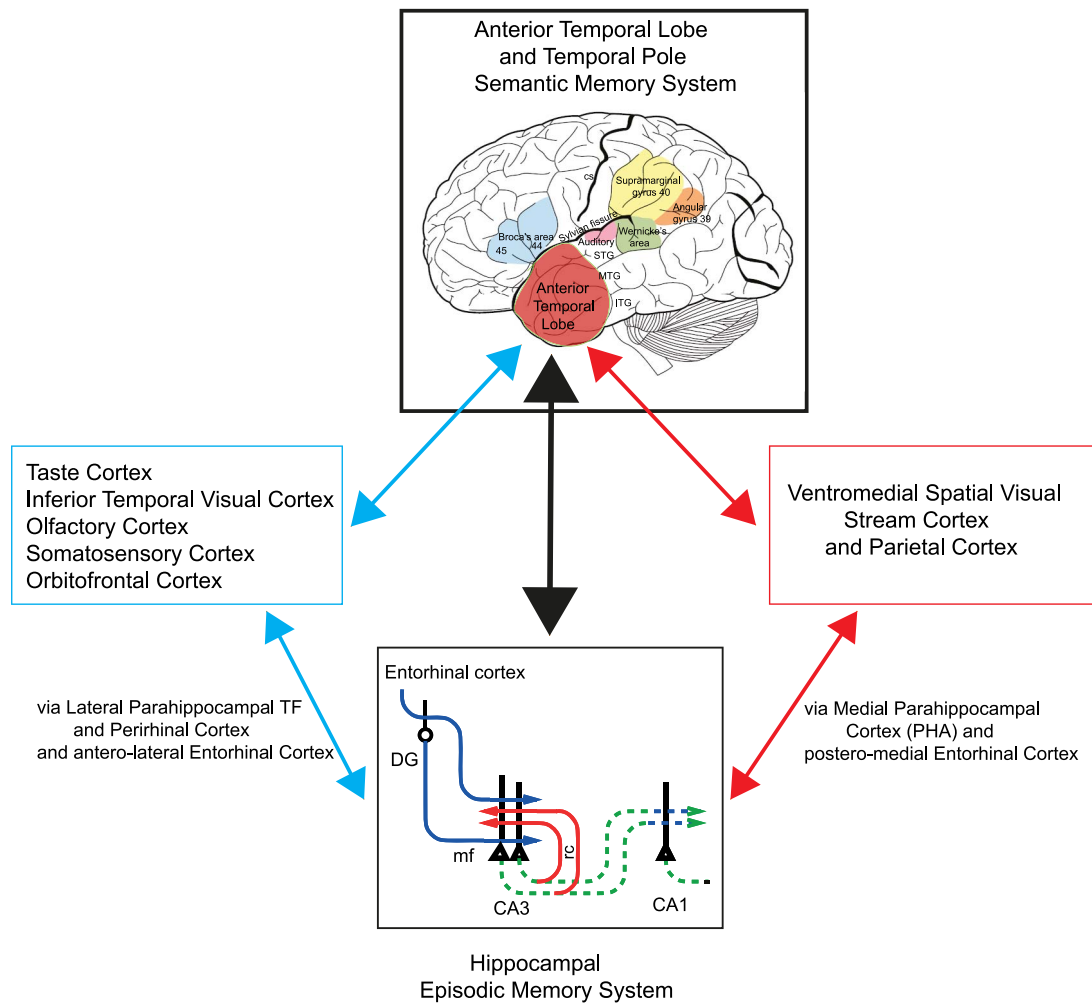


Fig. 2. Direct and hippocampal routes to the ATL semantic memory system. The ATL semantic memory system and the hippocampal episodic memory system both receive inputs from the “What” cortical systems (blue, left, taste cortex, inferior temporal visual cortex, olfactory cortex, somatosensory cortex, and reward value orbitofrontal cortex) and from the “Where” cortical systems (red, right, ventromedial spatial stream cortex for spatial view, and parietal cortex). In addition, the hippocampal episodic memory system has connectivity with the ATL semantic system. It is proposed that because the hippocampal episodic memory system can bind together elements of an episodic memory in CA3 (“What,” “Where,” and Reward), the ATL semantic memory system benefits from the hippocampal bindings of certain combinations of inputs from a number of similar episodic memories, and using slow learning can associate together the overlapping information from these episodic memories to help build a new semantic memory. Some of the other cortical areas indicated in this figure in color are involved in language, and are connected to the ATL semantic regions, which have connectivity that extends up through the temporoparietal junction (labeled here Wernicke’s area) to the inferior parietal cortex (Rolls et al. 2022a; Rolls 2023a). Further evidence for the relevant connectivity is cited in the text. On the view of the human brain: STG, Superior temporal gyrus. MTG, Middle temporal gyrus. ITG, Inferior temporal gyrus. cs, Central sulcus. On the hippocampal network architecture: DG, Dentate gyrus. mf, Mossy fibers. rc, CA3 recurrent collaterals, which provide a single autoassociation/attractor network to bind current inputs together to form an episodic memory.

rule is associative learning, the Hebb rule, with a medium-term memory trace in the post-synaptic term that spans the medium time in which we might be thinking about one semantic category or recalling related episodic memories from the hippocampus. This would enable all objects or items or episodic memories being processed in this medium-term time to become associated together to form a semantic category. Later on, presenting any one view of any one object would—via the VisNet input to the autoassociation (attractor) network—enable the semantic network to be activated, and thus for the correct semantic representation to be recalled. The medium-term trace rule learning is facilitated in principle by the fact that Layer 5 is an autoassociation network with attractor states that will tend to keep the semantic attractor neurons active while the different objects are being considered, helping a particular semantic attractor state to build up if new items are being added. The sparseness of the representation in Layer 5 is set to prevent the semantic attractors from becoming

too wide. Indeed, different attractor networks in the ATL could operate with different sparsenesses of their representations to facilitate the development of narrow and wide semantic categories, such as birds vs animals.

In line with the evidence that semantic representations are present in the ATL considered in the Introduction, the semantic attractor system just described is likely to be in the ATL. However, a single attractor network is not proposed. Instead, it appears that a large part of the ATL, extending posteriorly up into the inferior parietal cortex and including the language regions in the temporoparietal junction (“Wernicke’s area”), is involved in semantic representations, with somewhat different types of semantic representation found in different parts of this extended semantic system in the human brain, with the different parts representing somewhat different semantic categories based on the local inputs being received (Huth et al. 2016; Rolls et al. 2022a). For example, the more ventral parts of the human ATL may be

especially involved in semantic categories that can be distinguished based on visual object input; the more dorsal parts of the human ATL may be especially involved in semantic categories that can be distinguished based on auditory input; and the inferior parietal semantic regions may be especially involved in semantic categories that can be distinguished based on visuomotor actions such as tool use or bodily actions (Rolls et al. 2022a; Rolls et al. 2024b). Consistent with this somewhat distributed and somewhat specialized set of cortical regions involved in different types of semantic information, visual stimuli consisting of faces, scenes, body parts, and tools activate somewhat different collections of cortical regions that can even extend back to early cortical sensory processing regions such as the somatosensory cortex when faces are being viewed (Rolls et al. 2024b). Thus, a whole set of attractor networks in these anterior temporal cortex regions extending up into the inferior parietal cortex is proposed to implement semantic representations, each somewhat specialized for different types of semantic representation, rather than the single autoassociation network utilized in the illustrative computational model to be described here. Importantly, all of these semantic networks have connectivity with some key nodes in the language system, areas 45 and 44 (“Broca’s area”) and the closely connected nearby inferior frontal gyrus regions, and the temporoparietal junction regions (Rolls et al. 2022a).

These hypotheses are made evident in the model shown in Fig. 1 which is described in the Methods.

The hippocampal episodic memory system provides binding of components that help in the building of semantic representations in the ATL

The third hypothesis is that the hippocampal episodic memory system may be useful in forming semantic memories because of the binding of components that is performed by the hippocampal episodic memory system, with the hippocampal connectivity with the ATL semantic system shown schematically in Fig. 2.

There is evidence that hippocampal episodic memory is implicated in learning new semantic representations, based, for example, on semantic learning impairments after hippocampal damage (Duff et al. 2019). It is now proposed that the hippocampal contribution to semantic memory relates to its role in binding together using CA3 recurrent collateral connections simultaneously present information to form an episodic memory, eg the sight of a person, the sound of their voice, where they are, and who they are with (Rolls 2023a; Rolls and Treves 2024). The contribution of the hippocampal system proposed is that slow semantic learning links together a whole series of such episodes, which builds up the commonalities between the episodes, which constitute a category. Another example could be lake scenes, which might have brown water, trees, birds on the lake, boats on the lake, and buildings beside the lake. Lots of lakes might have some of these attributes in common, and they would be built together by slow learning in, for example, the ATL, with the most common attributes dominating the new semantic category, in this case lake scenes. This use of episodic “snapshots” might make the temporal slow association have to carry less of a burden of random elements becoming associated. Instead, similar episodic events would be associated together, with some of the binding done by the hippocampus. Indeed, what is associated could be some remembered past episodes with what is being perceived now, so that similar representations could become associated together by slow learning. Another example might be two movies, each filmed in a distinct location providing scene (“Where”)

information, eg London vs New York. Each movie would have different people interacting with each other, providing different “What” information for the two movies. We might then expect the episodic memories for different episodic events in each movie to become associated together into different semantic representations for the two movies. A test later would be to present one person from one of the movies and test whether participants who had seen the movies could pick another person from the same movie when the images of one person from each movie were the choices.

An implication of this proposal is that the hippocampus would need to be able to communicate its “binding” information to the semantic regions in, for example, the ATL, to contribute to generating new semantic representations that utilize different episodic memories, with the overall architecture shown in Fig. 2. New evidence for these hippocampal systems to semantic system pathways is described in the Results, which extends previous knowledge by utilizing effective connectivity with the connectivity measured between every pair of cortical regions in both directions (see Rolls et al. 2022c; Rolls et al. 2024b) and by utilizing the Human Connectome Project Multimodal Parcellation of the cortex (HCP-MMP). The HCP-MMP is chosen because it has 360 cortical regions that are well-defined using anatomy, functional connectivity, and task-related functional magnetic resonance imaging (fMRI) (Glasser et al. 2016), and because the connectivity of many of its regions has been analyzed recently using effective connectivity, functional connectivity, and diffusion tractography (Rolls et al. 2022a, 2022b, 2022c; Rolls et al. 2023a, 2023b, 2023c, 2023d, 2023e, 2023f, 2023g; Rolls et al. 2024a, 2024c).

It is noted that the first part of the recall from CA3 back to the neocortex is that each part of the episodic memory in CA1 (eg “What,” “Where,” and reward/affective value) does not need to be maintained as separate, so they are combined into a more efficient retrieval signal that recodes the separate parts of the memory into a single set of neurons firing that represent the whole episodic memory, using competitive learning (Treves and Rolls 1994; Rolls and Treves 2024). During recall of the episodic memory, the CA1 signal then recalls into firing the correct postero-medial entorhinal cortex (Syversen et al. 2024) neurons, which then recall the correct neurons in the medial parahippocampal gyrus (PHA1–3, which may be TH in macaques) for the spatial components (Rolls and Treves 2024; Rolls et al. 2024d). During recall, the CA1 signal also recalls into firing the correct anterolateral entorhinal cortex (Syversen et al. 2024) neurons, which then recall the correct neurons in the perirhinal cortex (PeEC) and lateral parahippocampal gyrus (TF) for the object components (Fig. 2).

Possible neuronal network mechanisms for hippocampal function in episodic memory to help the formation of semantic representations could be as follows: the hippocampus conveys by its output that whatever sensory inputs, etc. (eg objects, faces, places, and rewards) are currently active in semantic attractors in the ATL have occurred together in several episodic memories, while a set of similar episodic memories are recalled. Thus, whatever inputs are coming into a semantic local attractor network in the ATL is likely to be in the same semantic category (eg a set of people who work in a given building). Therefore, the hippocampal signal helps to show which items tend to co-occur in different episodes and indicates to the semantic ATL system: “link and store whatever cortical inputs are now being received.” The linking and storing might just be that the hippocampal input simultaneous with whatever sensory inputs are present exceeds a threshold in which the firing rates are high enough for long-term potentiation

in the ATL semantic memory system. Or it could be that the hippocampal afferent to the semantic attractors in the ATL have a special role in facilitating storage, perhaps because of how they terminate on cortical pyramidal cells. Of course, any cholinergic input to the neocortex simultaneously present because of reward or punishment or novelty would facilitate the learning process (Rolls 2022).

A key feature of this proposal is that the slow semantic learning makes what is stored different from any episodic memory because what is stored reflects several episodic memories recalled successively, which might, for example, have slightly different individual people present in a group in different episodes. The hippocampal process for binding described by Eichenbaum and colleagues (Sellami et al. 2017) involved associations between temporally discontinuous events to form episodic memories in the hippocampus, whereas the processes described here show how what may be temporally discontinuous hippocampal episodic memories when recalled together close in time can be used by slow semantic neocortical learning to help build new semantic neocortical representations.

Methods

Effective connectivity of the hippocampal episodic memory system with the ATL semantic system

Effective connectivity matrices were measured in 171 HCP participants imaged at 7T (Van Essen et al. 2013) with methods described in detail previously (Rolls et al. 2022a, 2022c; Rolls et al. 2023a) but summarized here. The preprocessing was performed by the HCP as described in Glasser et al. (2013), based on the updated 7T data pipeline (v3.21.0, <https://github.com/Washington-University/HCPpipelines>). To construct the effective connectivity for the regions of interest in this investigation, the lateral parahippocampal cortex TF and medial parahippocampal cortex PHA1–3, with other parts of the human brain, we utilized the 7T resting-state fMRI data from the HCP and parcellated this with the surface-based HCP-MMP atlas, which has 360 cortical regions (Glasser et al. 2016). The human lateral parahippocampal cortex region TF was chosen because TF provides the main route (but see below re PeEC) between the hippocampus and the “What” cortical systems concerned with representations of objects, faces, people, etc. (Rolls et al. 2022c; Rolls 2023a, 2024a) (Fig. 2). The human medial parahippocampal cortex regions PHA1–3 (probably corresponding to TH in macaques) were chosen because they provide the main route between the hippocampus and the “Where” cortical systems concerned with representations of spatial view and actions in space (Rolls et al. 2022c; Rolls 2023a, 2024a) (Fig. 2).

Effective connectivity measures the effect of one brain region on another and utilizes differences detected at different times in the signals in each connected pair of brain regions to infer the effects of one brain region on another. The method used here (see Rolls et al. 2022b, 2022c) was developed from a Hopf algorithm to enable the measurement of effective connectivity between many brain areas, as described by Deco et al. (2019). A principle is that the functional connectivity is measured at time t and time $t + \tau$, where τ is typically 2 s to take into account the time within which a change in the BOLD signal can occur, and that τ should be short to capture causality, and then the effective connectivity model is trained by error correction until it can generate the functional connectivity matrices at time t and time $t + \tau$. The local dynamics of each brain area (node) are given by Stuart–Landau oscillators, which express the normal form of

a supercritical Hopf bifurcation, describing the transition from noisy to oscillatory dynamics (Kuznetsov 2013). The equations for the model, and the procedure for measuring the effective connectivity, have been described in many previous papers (Rolls et al. 2022a, 2022b, 2022c; Rolls et al. 2023a, 2023b, 2023c, 2023d, 2023f, 2023g; Rolls et al. 2024a), and are not repeated here as the procedure was identical, though a summary is provided in [Supplementary Material 2](#). The maximum effective connectivity across the whole brain was set to 0.2. A list of the cortical regions, and illustrations of them, are included in the [Supplementary Material 2](#).

A model of view-invariant visual object representations, followed by a semantic autoassociation network in the ATL

The network for visual invariant object recognition is VisNet and consists of the first four layers, as shown in Fig. 1. This part of the whole network is described only briefly here, as full descriptions and the code are available elsewhere (Rolls 2012; Rolls and Mills 2018; Rolls 2023a). The autoassociation semantic network is Layer 5, as shown in Fig. 1, and is described in Section “The semantic autoassociation network in Layer 5 trained with a short term memory trace synaptic learning rule.”

The architecture and operation of VisNet

VisNet is a feature hierarchy network using a small radius for each neuron to receive from the previous layer (see Table 1) in order to allow any neuron in Layer 4 to be influenced by inputs in any part of Layer 1, which corresponds to V2 (Fig. 1). (Layer 2 of VisNet corresponds to V4; Layer 3 to posterior inferior temporal cortex area TEO; and Layer 4 corresponds to inferior temporal cortex area TE in macaques (Rolls 1992, 2012; Rolls and Mills 2018; Rolls 2021a, 2023a), as shown in Fig. 1, and TE2p and TE1p in humans (Rolls et al. 2023a; Rolls et al. 2023e; Rolls et al. 2024c; Rolls and Turova 2025).) The learning is competitive feedforward learning (Rolls 2023a), with no feedback of errors or supervision of the training by, for example, separate teachers for each neuron in Layer 4 for the different objects (Rolls 2023a).

In more detail, VisNet consists of a series of feedforward hierarchically connected competitive networks with convergence from layer to layer, with four layers, as illustrated in Fig. 1. Each layer in this architecture corresponds to different cortical regions, as indicated in Fig. 1. The connections to a neuron in one layer come from a small region of the preceding layer using a Gaussian distribution of connection probabilities defined by the radius, which will contain approximately 67% of the connections from the preceding layer. Table 1 shows this radius for each layer of 32×32 neurons per layer, with each neuron receiving the number of synaptic connections in Table 1 from the neurons in the preceding layer. The activation of a neuron is calculated as the synaptically weighted sum of the rate inputs it receives from the preceding layer, ie as a dot or inner product between the input rates and the synaptic weights (Rolls and Milward 2000; Rolls 2012; Rolls and Mills 2018; Rolls 2023a). The activations are converted into rates with a sigmoid activation function, with the sparseness of the representation in a layer set as described next.

In a competitive network (Rolls 2023a), mutual inhibition is required between the neurons within each layer so that for any one stimulus, only a proportion of neurons is active. The activation of the neurons in a layer is first calculated by the dot product of the synaptic weights of a neuron and the rates of the neurons in the preceding layer to which it is connected by the synaptic weights. Then, the activations are converted into rates

Table 1. VisSemanticNet architecture. Dimensions show the number of neurons in each of the five layers. # Connections show the number of synaptic connections onto each neuron. Radius shows the radius of the Gaussian profile of connectivity from the previous layer of a single neuron (see text). Sparseness shows the proportion of neurons in a layer that are above the threshold and have some firing.

	Dimensions	# Connections	Radius	Sparseness	Eta
Layer 5	32 × 32	1,024	-	0.05	0.995
Layer 4	32 × 32	200	7	0.02	0.8
Layer 3	32 × 32	200	7	0.02	0.8
Layer 2	32 × 32	200	7	0.02	0.8
Layer 1	32 × 32	272	15	0.02	0
Input layer, V1	256 × 256 × 32				

with values between 0 and 1 using a sigmoid activation function. The threshold is defined by the sparseness parameter so that if neurons have activations above the threshold, their rates will be greater than 0.5. The slope of the sigmoid was set to 10. If the neurons have binary rates, the sparseness is the proportion of neurons that are active for any one stimulus.

To help nearby neurons in a layer learn to respond to different stimuli, lateral inhibition within a layer is typically implemented. The implementation of the lateral inhibition used was convolution with a Mexican hat difference Gaussian filter, with a radius of 0.2 for the central part and a radius of 4 for the outer part, as used elsewhere with the code available (Rolls 2021a, 2023a).

The inputs to VisNet are computed to have elongated receptive fields of the type found in the primary visual cortex V1, in order to allow comparison of the neurons at different stages in VisNet to those in the brain. The Gabor filters (Daugman 1988) have four spatial frequencies, four orientations, and positive or negative. The Layer 1 neurons are connected to these with radii, as described above and in Table 1, and with the number of connections to each frequency scaled according to the spatial frequency, as described in detail elsewhere (Rolls 2012; Rolls and Mills 2018; Rolls 2021a, 2023a).

Layer 1 of VisNet is trained with a purely associative learning Equation (1) to enable feature combination neurons to be formed that represent the relative spatial locations of the features. This solves the feature binding problem, as described elsewhere (Rolls 2012, 2023a). This associative learning rule combined with the competition between neurons implements a competitive network (Rolls 2023a) to enable Layer 1 neurons to respond to the different combinations of features found locally in different spatial scenes.

$$\delta w_j = \alpha y x_j \quad (1)$$

where x_j is the j^{th} input to the neuron; y is the output from the neuron; α is the learning rate; and w_j is the synaptic weight between the j^{th} input and the neuron.

Layers 2 to 4 of VisNet are trained with a short-term memory trace rule in order to produce view (or translation or size) invariance in the representations to ensure that the same neurons coding for an object are activated by different transforms of the object, for example, by different views. The short-term memory trace learning rule has been proven to be useful in enabling invariant representations of objects and faces to be built in the ventrolateral visual cortical pathway in the VisNet and similar models (Rolls 1992; Wallis and Rolls 1997; Wiskott and Sejnowski 2002; Franzius et al. 2007; Rolls 2012, 2021a, 2023a), and this rule is used here to provide for view invariance. The short-term memory trace that enables inputs occurring close together in

time, as they would in the natural world, to become associated is implemented in the hierarchical competitive network (Rolls 2012, 2023a) model by using associative synaptic modification with a small change that allows the postsynaptic term to remain active for short periods in the order of 100 ms or more. The short-term memory trace update learning rule that we have used has the following form (Rolls 2012, 2023a):

$$\delta w_j = \alpha \bar{y}^\tau x_j \quad (2)$$

where

$$\bar{y}^\tau = (1 - \eta) y^\tau + \eta \bar{y}^{\tau-1} \quad (3)$$

where \bar{y}^τ is the trace value of the output of the neuron at time step τ and η is the trace update proportion, with 0 meaning no trace, just associative learning. The optimal value of η varies with the number of transforms of each object and is typically 0.8. Many variations of this learning rule have been explored (Rolls and Milward 2000; Rolls and Stringer 2001). The general form of the rule for computational purposes can be as shown in Equation (3), but the actual mechanism in the brain might utilize a slow synaptic eligibility trace such as provided by the NMDA receptors with their long time constant, as well as a tendency for neuronal firing to continue due to local attractor networks (Rolls 2012, 2023a). During training with the trace learning rule, in a single training epoch, all transforms of one object are presented in random sequence so that the trace rule can help learn that all of these are transforms of the same object because they occur close together in time; then all transforms of another object are shown, etc.

The network was trained with five cars (vehicles) and five animals, each with eight views, from the Amsterdam Library of Image dataset (Geusebroek et al. 2005). The stimuli are illustrated in Fig. 5, and prototypical views are shown at the base of Figs. 6–9. Eight views of each object spaced 45° apart were used, because for these objects that is sufficient to enable generalization to any horizontal view of each object, as shown in related research (Rolls 2012, 2023a). That is, after training on eight views of each object, performance is very good when cross-validation testing is performed with the eight intermediate views separated by 22.5° from the trained views. When training VisNet, the first four layers of the network shown in Fig. 1, in one training epoch, each of the ten objects was selected in random permuted sequence, and then eight views of that object were presented with the trace rule parameter η of Equation (3) set to allow the trace to last sufficiently long for the eight views of that object to be presented in random permuted sequence. As always with VisNet, and now with VisSemanticNet, the layers were trained sequentially, first Layer 1, then Layer 2, etc., with the rationale that there is no

point in training higher layers of the hierarchical network until the earlier layers have been trained; otherwise, the higher layers are performing useless learning. Consistent with this, there are biologically critical periods during development for learning, for example, binocular stereopsis in V1, after which little further learning is possible, leading to the standard clinical strategy of correcting for strabismus as soon as possible in early life (Parrey 2024). The learning rate used was 0.1 for Layers 2 to 4 and 0.0005 for Layer 1, and the number of training epochs for each layer was 175, though the operation of the system could be demonstrated with 7 training epochs.

Because the aim of the present research is to introduce and demonstrate new concepts about ways in which semantic representations may be learned, we did not attempt to train the VisNet network up to capacity with very large numbers of objects until its performance broke down. However, in the new research we are doing, we can report that when large numbers of objects are trained, then by Layer 4 of VisNet, the representation of some objects may have some correlation with physically similar objects, which is as expected. This demonstrates that physical similarities between different objects are likely to be incorporated into the semantic representations learned in Layer 5, which is another principle by which the semantic categories that are formed in Layer 5 may be influenced.

The semantic autoassociation network in Layer 5 trained with a short-term memory trace synaptic learning rule

Layer 5 was modeled as a fully connected autoassociation network (Rolls 2023a). However, instead of being trained with a purely associative (Hebbian) synaptic training rule based on the current firing rates, the network was trained on firing rates with a short-term memory trace that was set to be sufficiently long for the neurons to reflect the inputs being received while all members of one semantic category were shown, for example, all eight views of all five animals chosen in randomly permuted sequence. This enables the inputs produced by all members of one semantic category (animals or cars) to be associated together in the semantic autoassociation network. Two presentations of all members of a semantic category were typically shown in random permuted sequences to allow a stable trace-related firing to build up before purely associative synaptic modification was used to train the autoassociation network. One training epoch was sufficient. The training rule was thus as shown in Equation (4), and is purely local, as it depends on the presynaptic and postsynaptic terms only:

$$\delta w_{ij} = \alpha \bar{y}_i \bar{y}_j \quad (4)$$

where \bar{y}_i is the memory trace firing rate of neuron i ; \bar{y}_j is the memory trace firing rate of neuron j ; α is the learning rate; and w_{ij} is the synaptic weight between neurons i and j .

The trace version of the firing was set up as shown in Equation (3), with η typically 0.995. In practice, learning can start before a trace that reflects all members of a semantic category has fully built up.

The input for Layer 5 was from Layer 4. In principle, Layer 5 can include copies of what is represented at the ends of the cortical systems for different modalities, including visual, auditory, somatosensory, taste, and olfactory, so that Layer 5 can build semantic categories that reflect multimodal inputs. In practice, in the simulations described here, Layer 5 was a copy of what was present at the end of the visual cortical hierarchy, that is, in Layer 4 of the network, which corresponds to the last layer of VisNet, with that architecture indicated in Fig. 1. This helps to make it clear

that the autoassociation network that builds semantic categories, Layer 5, is conceptually and computationally different from the cortical hierarchies such as the visual hierarchy implemented in Layers 1 to 4 and that are Layers 1 to 4 of VisNet, which operate by competitive learning.

An integrate-and-fire model of VisSemanticNet

To explore the dynamics of operation of VisSemanticNet, for comparison with what is found in the brain (Rolls 2023a), an integrate-and-fire model of VisSemanticNet was implemented. The synaptic weights were imported from the rate model already described into the integrate-and-fire model, and the integrate-and-fire implementation followed very closely that which we and others have used extensively previously (Brunel and Wang 2001; Rolls and Deco 2010; Rolls 2023a; Rolls et al. 2024d), and which is therefore described in more detail in the Supplementary Material. Key aspects of the implementation were the use of the dynamics of the AMPA, NMDA, and GABA synapses in the system, as these are key to understanding the latencies of the operation of the system (Panzeri et al. 2001; Rolls 2023a). To obtain better operation of the integrate-and-fire simulation, the parameters used when training the rate model were a learning rate of 0.01 for Layers 2 to 5 with 875 training epochs; sparsenesses for Layers 1 to 5 of 0.02, 0.05, 0.05, 0.01, and 0.05; and modification of the sigmoid activation function to set firing rates in each layer that were below 0.5 to 0 (where the maximum rate was 1.0), as this produced synaptic weight matrices that reflected better the more strongly firing neurons.

Results

Effective connectivity of the hippocampal episodic memory system with the ATL semantic system

To provide evidence on the pathways in humans via which the hippocampal episodic memory system could influence the ATL semantic memory system, Figs. 3 and 4 show how the key stages back from the hippocampus to the neocortex, the medial and lateral parahippocampal cortex, have effective connectivity to other cortical regions that include semantic regions. Figures 3 and 4 are new and are based on resting-state effective connectivity matrices measured in 171 HCP participants at 7T, with those data complemented by functional connectivity and diffusion tractography in the same participants, with methods described previously (Rolls et al. 2022a, 2022c; Rolls et al. 2023a) and summarized in the Methods section.

Figure 3 shows the effective connectivity of the human lateral parahippocampal cortex region TF with neocortical regions, for TF provides the main route (but see below re PeEC) between the hippocampus and the “What” cortical systems concerned with representations of objects, faces, people, etc. (Rolls 2024a). The studies cited in the following are those that have used the HCP-MMP of the cortex (Glasser et al. 2016; Huang et al. 2022), as explained above, which provides a well-founded framework for identifying 360 cortical regions based on neuroanatomy, functional connectivity, and task-related fMRI (Rolls 2023a). The connectivity of TF with the hippocampus is largely via the PeEC (Fig. 3) and then the entorhinal cortex.

Figure 3 shows first that there is connectivity of TF with the inferior temporal cortex, largely unimodal visual cortical regions for invariant object and face representations, TE2p and TE1p. Second, there is connectivity with the orbitofrontal cortex (47 m), which provides a route for reward value to be incorporated into

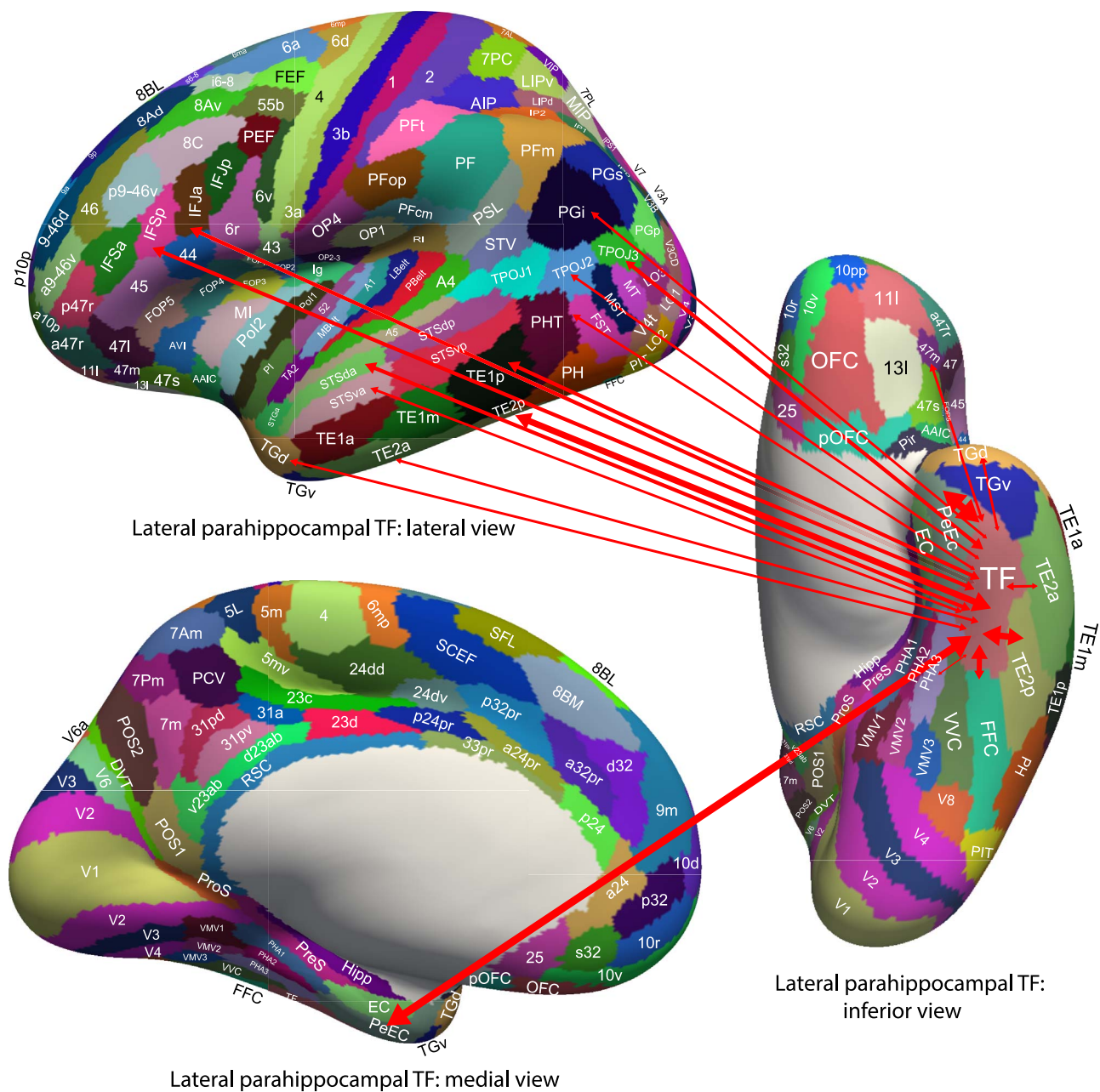


Fig. 3. Connectivity of the hippocampal episodic memory system includes connectivity with many cortical regions in and connected to ATL semantic memory systems. The connectivity of the lateral parahippocampal cortex region TF, which includes connectivity to semantic systems related to “What” representations is shown (see text). The effective connectivity between cortical regions in the HCP-MMP atlas (Glasser et al. 2016; Huang et al. 2022) is shown. A list of the cortical regions in the HCP-MMP atlas is shown in the Supplementary Material. The width of the arrows represents the magnitude of the effective connectivity, and the size of the arrowheads the magnitude in each direction. The data are from effective connectivity matrices computed previously from resting-state timeseries in 171 HCP participants imaged at 7T (Rolls et al. 2022a; Rolls et al. 2022c; Rolls et al. 2023a), but this is the first time that a cortical diagram for the connectivity of TF has been provided.

episodic memories (Rolls et al. 2022b). Third, there is a connectivity of TF with many regions in the ATL semantic network (Milton et al. 2021; Rolls et al. 2022a), including STSda, STSva, TE2a, and temporal pole TGd, all of which are multimodal (Rolls et al. 2022a). Fourth, there is connectivity of TF with temporo-parieto-occipital regions TPOJ2 and TPOJ3 that are implicated in the language (Milton et al. 2021; Rolls et al. 2022a), and that are part of what was referred to as “Wernicke’s area” (Rolls 2023a). Fifth, there is connectivity of TF with inferior parietal cortex region PGI, which is massively connected with the ATL semantic regions (Rolls et al. 2022a; Rolls et al. 2023c) and which may

encode information about actions, which is an important part of semantic representations (Rolls et al. 2022a; Rolls 2023a; Rolls et al. 2023c; Rolls et al. 2024b). Sixth, there is connectivity of TF with inferior frontal cortex regions IFSp and IFJa (Fig. 3), which are massively connected with Broca’s area regions 44 and 45, and which are suggested to add extra attractor capacity to the proposed roles of regions 44 and 45 in language-based syntactic operations on semantic information (Rolls and Deco 2015; Rolls et al. 2022a; Rolls 2023a).

Figure 4 shows the effective connectivity of the human medial parahippocampal cortex regions PHA1–3 (probably corresponding



Fig. 5. The 8 views of the 10 objects used in the training. The first five objects from the left are cars; the five objects on the right are animals. Gray-scale versions of the images shown were used in the simulations, so that object shape was relevant, and not a perhaps simpler cue such as color. The images are from the Amsterdam Library of Object Images (ALOI) dataset (Geusebroek et al. 2005).

two figures do bring out this important connectivity in humans quite clearly for the first time. This connectivity is bidirectional, as shown by the size of the arrowheads in Figs. 3 and 4, so it potentially allows existing semantic information to enter the hippocampal episodic memory system, as well as for the hippocampus to contribute to the formation of new semantic representations as described here. In a check, significantly more direct connectivity with semantic systems was not found from the PeEC and entorhinal cortex that are closer to the hippocampus, though more connectivity with the orbitofrontal cortex and anterior cingulate cortex reward-related cortical regions was found from these cortical regions (see Rolls et al. 2022b).

Learning invariant representations of objects and associating sets of objects together in an autoassociation, attractor network in Layer 5 to form semantic representations that depend on temporal proximity

VisSemanticNet, as shown in Fig. 1, was trained with five cars and five animals, each with eight views, as described in the Methods.

Layers 1 to 4 are VisNet and form view-invariant representations of each object. Figure 6 shows the correlations of the firing rates of the 1,024 neurons in Layer 4 of the network between all 8

views of all 10 stimuli. This shows that the firing rates of all the neurons in Layer 4 for all 8 views of each object were correlated with each other and not with any view of any other object. This thus demonstrates view-invariant but object-selective representations in Layer 4 and is the type of learning expected by VisNet (Wallis and Rolls 1997; Rolls 2012, 2021a, 2023a).

Figure 7 shows the correlations of the firing rates of the 1,024 neurons in Layer 5, the semantic layer, of the network between all 8 views of all 10 stimuli. Now, it is evident that the firing rates in Layer 5 to all views of the five cars are associated together, and not with any views of any animal. Correspondingly, the firing rates in Layer 5 to all views of the five animals are associated together and not with any views of any car. That is, presentation of any view of any car activates the attractor state corresponding to cars, and presentation of any view of any animal activates the attractor state corresponding to animals.

The semantic groups can deal with apparent exceptions and depend on which objects are temporally close during the training of the autoassociation, attractor, semantic net in Layer 5

Learning exceptions in semantic representations, for example, that some birds cannot fly, is a potential difficulty for the

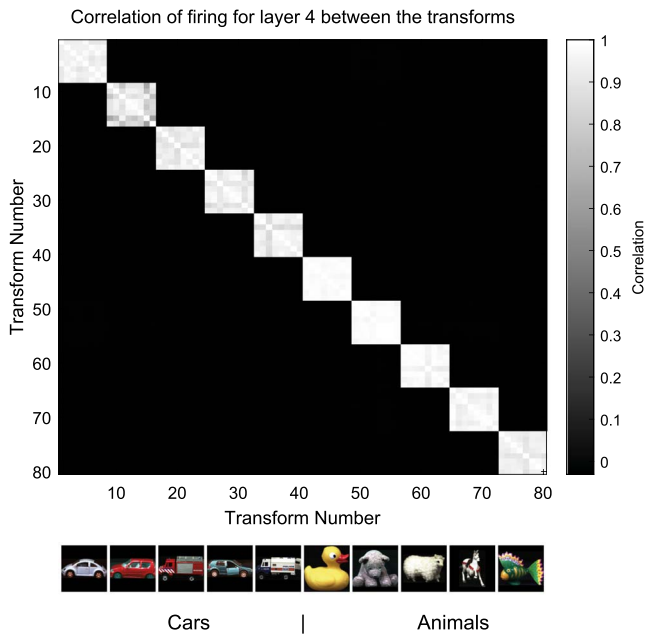


Fig. 6. The correlations between the eight views of each object, and between the views of other objects for Layer 4, the final stage of VisNet. There are eight views of each object. The first five objects are different cars (transforms 1 to 40), and the second five objects are animals (transforms 41 to 80). This shows high correlations between the views of each of the 10 objects, and that each object is represented separately from the other objects, in Layer 4. Gray-scale versions of the images shown were used in the simulations, so that object shape was relevant, and not a perhaps simpler cue such as color.

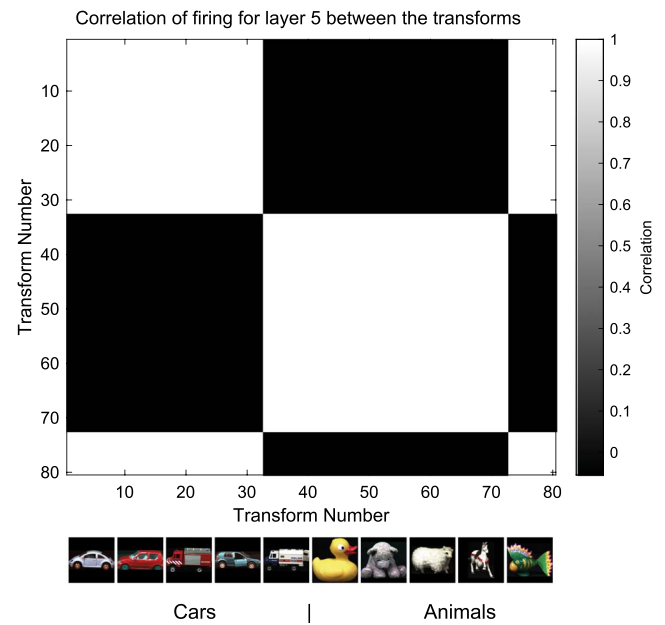


Fig. 8. The semantic groups can deal with apparent exceptions and depend on which objects are temporally close during the training of the autoassociation, attractor, and semantic net in Layer 5. The correlations between the eight views of each object, and between the views of other objects for Layer 5, the semantic level of the network. There are eight views of each object. In the ribbon of objects, Object 1 is on the left, and Object 10 is on the right. Layer 5 was trained with one semantic set of objects comprising Objects 1 to 4 (cars) and Object 10 (an animal). The second semantic set on which Layer 5 was trained was Objects 6 to 9 (animals) and Object 5 (a car). The correlation matrix for Layer 5 shows that one semantic set consists of Objects 1 to 4 (cars) and Object 10 (an animal). The correlation matrix for Layer 5 also shows that a second semantic set consists of Objects 6 to 9 (animals) and Object 5 (a car). Thus Layer 5 has formed semantic categories that depend on temporal proximity using the temporal trace associative learning rule for the Layer 5 autoassociation, attractor, and semantic network. This system can thus deal with exceptions in that, for example, the semantic category for cars can include one animal, if this set of objects is temporally associated.

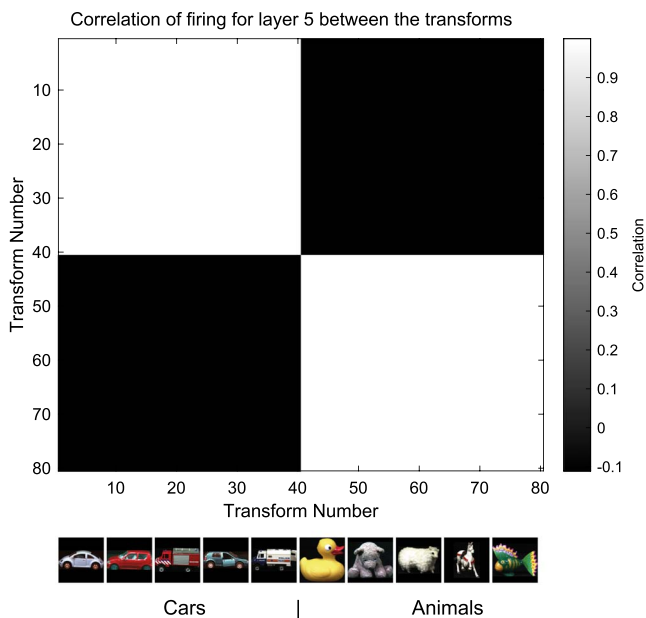


Fig. 7. The correlations between the eight views of each object, and between the views of other objects for Layer 5, the semantic level of the network. There are eight views of each object. The first five objects are different cars (transforms 1 to 40), and the second five objects are animals (transforms 41 to 80). This shows high correlations between all the views of each of the five cars, and all the views of the five animals, and no confusion between cars and animals. This shows that the two semantic categories, cars and animals, are represented separately from each other in the semantic Layer 5.

learning of semantic representations. One approach has been to predefine the categories to be learned, and to use deep neural networks trained by backpropagation of error to help learn the

exceptions (Hinton 1989; Rumelhart 1990; Rumelhart and Todd 1993; McClelland et al. 1995). To demonstrate how the current slow learning hypothesis for learning exceptions can deal with exceptions, VisSemanticNet was trained with one category that contained four cars and one animal and a second category that contained four animals and one car. The results in Fig. 8 show that the short-term memory trace learning rule can help the semantic representations formed to be the correct categories just described.

Thus, temporal proximity—implemented, for example, by a short-term memory trace learning rule in the autoassociation (attractor) network in Layer 5 of VisSemanticNet—can help with the learning of what might be considered as exceptions to the semantic rule.

The semantic groups can be arbitrary and depend on which objects are temporally close during the training of the autoassociation, attractor, semantic net in Layer 5

Another potential issue in the learning of semantic representations is whether the semantic representations can consist of what might appear to be an arbitrary group of objects. An example might be the semantic category of vehicles, which might include apparently diverse objects such as cars, trucks, bicycles, and scooters. To investigate whether the slow learning in Layer 5 of VisSemanticNet can learn what may appear to be arbitrary sets

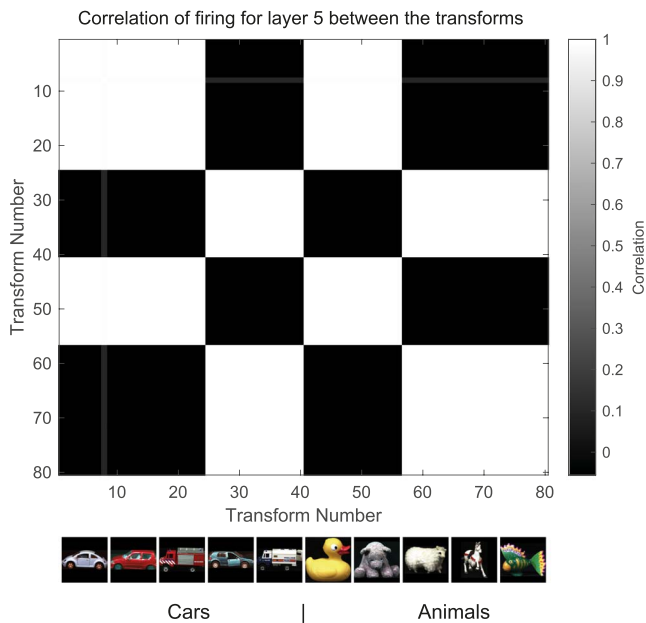


Fig. 9. The semantic groups can be arbitrary, and depend on which objects are temporally close during the training of the autoassociation, attractor, and semantic net in Layer 5. The correlations between the eight views of each object, and between the views of other objects for Layer 5, the semantic level of the network. There are eight views of each object. In the ribbon of objects, Object 1 is on the left, and Object 10 is on the right. Layer 5 was trained with one semantic set of objects comprising Objects 1 to 3 (cars) and Objects 6 and 7 (animals). The second semantic set on which Layer 5 was trained was Objects 8 to 10 (animals) and Objects 4 and 5 (cars). The correlation matrix for Layer 5 shows that one semantic set consists of Objects 1 to 3 (cars) and Objects 6 and 7 (animals). The correlation matrix for Layer 5 also shows that a second semantic set consists of Objects 8 to 10 (animals) and Objects 4 and 5 (cars). Thus, Layer 5 has formed semantic categories that depend on temporal proximity using the temporal trace associative learning rule for the Layer 5 autoassociation, attractor, and semantic network. This system can thus form arbitrary semantic groups, depending on which objects are temporally associated during the training of the semantic network in Layer 5. The weak gray line for Object 1 View 8 reflects a very weak correlation of that view with some other objects.

of objects in a category, VisSemanticNet was trained with one category that contained three cars and two animals and a second category that contained three animals and two cars. The results in Fig. 9 show that the short-term memory trace learning rule can help the semantic representations formed to be the correct categories just described.

Thus, temporal proximity—implemented, for example, by a short-term memory trace learning rule in the autoassociation (attractor) network in Layer 5 of VisSemanticNet—can help with the learning of what might be considered as an arbitrary set of members of a semantic group.

An integrate-and-fire model of VisSemanticNet to illustrate its dynamics

The operation of the integrate-and-fire model is illustrated and explained in Fig. 10, with the implementation details included in the Supplementary Material. A key point from the integrate-and-fire model is that the latencies for the neurons in each layer (from the time when input was applied to V1, which provides the input to Layer 1) increase as shown in Fig. 10a (lower right panel), with an increase in latency of 10 to 15 ms from layer to layer, which matches what is found in the macaque ventral visual pathway (Rolls 1992, 2023a), and also what is expected for

cortical computations with biophysically plausible parameters for the synaptic time constants (Panzeri et al. 2001; Rolls 2023a).

Another useful point shown in Figs. 10a and b for Layer 5 is that when any view of Object 1 or Object 2 is used as the input to the network, the neurons in Layer 5 respond at similar times, illustrating the operation of the attractor network, as Objects 1 and 2 were in the same category that was trained by slow learning for Objects 1 to 5. The Layer 5 attractor network took about 15 to 20 ms to settle toward its basin of attraction, as shown, consistent with earlier evidence for the speed of operation of integrate-and-fire attractor networks (Treves 1993; Treves et al. 1997; Rolls et al. 1997b; Battaglia and Treves 1998; Panzeri et al. 2001; Roudi and Treves 2003; Rolls and Deco 2010; Rolls 2023a; Rolls and Treves 2024).

Discussion

The hypotheses and model described here show how slow learning could contribute to the learning of semantic representations in which the correct semantic representation (in this case of car or animal) can be activated after training by any view of any of the objects in a semantic category.

The hypotheses and model are different from some previous models of semantic learning (Hinton 1989; Rumelhart and Todd 1993; McClelland et al. 1995; Devereux et al. 2018) by being more biologically plausible because a local synaptic learning rule is used rather than error backpropagation in a deep neural network.

The model is also different in that previous models required learning which of the large lists of attributes (sometimes more than 2,000) such as “can fly” are associated with a semantic category (Hinton 1989; Rumelhart and Todd 1993; McClelland et al. 1995; Devereux et al. 2018), whereas the semantic categories described here relate to processing the members of a semantic category together over time (eg while thinking about that semantic category), and in this sense use temporal statistics in the world to help define semantic categories, rather than having a formal predefined teacher for the members of each category (Devereux et al. 2018). This has advantages in that the learning that takes place is a natural extension of the trace rule learning of invariant representations of objects as implemented in, for example, VisNet, which implies economy in the use of computational principles to help understand the operation of the cerebral cortex (Rolls 2023a).

The semantic learning implemented with the approach described here does not explicitly use a list of attributes of an object, such as “can fly,” to allocate it to a semantic category (Hinton 1989; Rumelhart and Todd 1993; McClelland et al. 1995; Devereux et al. 2018; Giallanza et al. 2024). But that raises a key issue: is that how we actually learn semantic categories, by learning lists of attributes of each object, and thereby allocating the object to a semantic category? That approach may be useful later when analyzing why a particular object is a partial misfit to a semantic category, but the actual learning in the first place may instead be based on more physical descriptors that would be represented at the top of each sensory system, such as “has some straight edges” (such as cars) or “has more rounded edges” (such as animals); is “hard to touch” (such as cars), or is “soft to touch” (such as animals), tastes and smells good (ie pleasant, such as food), or tastes bad (unpleasant, eg bitter, such as many poisonous plants). I accordingly argue that it is frequently properties represented at the ends of sensory processing hierarchies that the cerebral cortex often uses in the first instance to build semantic representations in attractor networks in the ATL and connected regions. Then, associations over time as described here may be

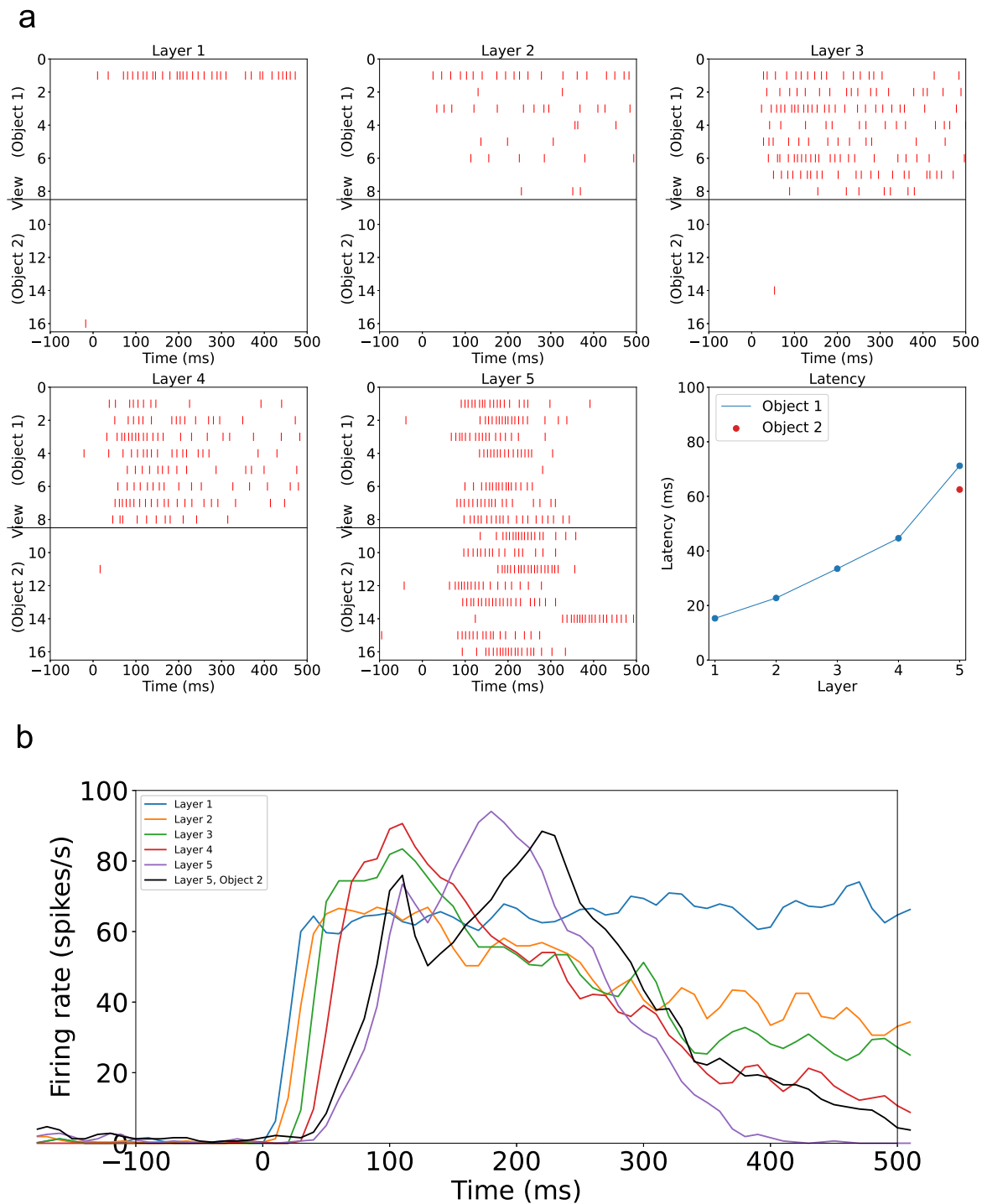


Fig. 10. Integrate-and-fire simulation of VisSemanticNet. **a)** Rastergrams for Layers 1 to 5. For each layer, the activity of a single neuron responsive to Object 1 is illustrated. Each row of the rastergram shows the responses of the neuron to one of the eight views of Object 1 (Views 1 to 8), or of Object 2 (Views 9 to 16). For Layer 1, the neuron responds to only View 1 of Object 1, as the learning in Layer 1 is purely associative, whereas in Layers 2 to 4, the short-term memory trace learning rule builds from layer-to-layer view-invariant representations, so the neurons learn to represent an object with view invariance by Layer 4. For Layer 5, the neuron responds to all eight views of both Object 1 and Object 2, as this is the semantic autoassociative layer in which Objects 1 to 5 are associated together to form a semantic category. When Object 1 was used as the input stimulus, the time for recall of the neurons in Layer 5 connected to Layer 4 Object 2 neurons was similar (red point) to that for neurons in Layer 5 receiving Object 1 input, demonstrating the operation of the attractor network in Layer 5 given that Object 1 and Object 2 were in the same category. The Layer 5 neuron did not respond to any views of Objects 6 to 10 (not illustrated), as Objects 6 to 10 are in a different semantic category. Object 1 was presented to V1 at time zero, and V1 connected to Layer 1. The latencies of the responses in each layer shown in the lower right panel are calculated as the time from the presentation of the stimulus in V1 to the response at half the maximum of the firing rates of the neurons. **b)** The mean firing rates across neurons of the neurons in each layer that respond to Object 1. A three-point smooth was applied. (For Layer 1, the firing rates are averaged across views for the neurons that responded best to each view.) The mean firing rates were used to calculate the latencies shown in (a) lower right panel.

useful in helping to build semantic representations. These may be later refined by thinking about exceptions, but again, the type of autoassociation process described here could still be involved. Indeed, it is demonstrated in the Results that VisSemanticNet can deal with exceptions and that VisSemanticNet can deal with arbitrary semantic categories that contain what at least physically may be unrelated objects.

In fact, the categories formed in the ways described here could be made more distinct by being, in turn, associated with word descriptors, such as “is a car,” “is a truck,” and as words help to define distinct categories, this could, by associative learning with word labels, help to refine and orthogonalize the semantic categories so far established based on sensory descriptors.

For the approach described here, the inputs received by the semantic layer (5 in Fig. 1) might reflect some of the different sensory attributes that distinguish cars from animals. At present, the first four layers, VisNet, are designed to produce view-invariant representations of objects with relatively uncorrelated, view-invariant representations of different objects. It will be interesting in the future to relax the object discriminative ability of VisNet to enable it to represent in its final layer (Layer 4) more about the different properties of the objects that it represents, so that they can be used to help the semantic networks that follow it to build useful semantic categories based on the type of sensory information being received. It will further be of interest to investigate how far the short-term memory trace learning rule in an autoassociative semantic net as used in Layer 5 of VisSemanticNet can re-organize categories based on the temporal associations despite a tendency to form semantic categories based on physical properties of stimuli, such as that they move.

The integrate-and-fire model of VisSemanticNet illustrated and explained in Fig. 10 shows an increase in latency of 10 to 15 ms from layer to layer, which matches what is found in the macaque ventral visual pathway (Rolls 1992, 2023a), which is useful validation. Further, the dynamics of these systems are dominated by the synaptic time constants and are as expected given that $\tau_{\text{AMPA}}=6$ ms, $\tau_{\text{NMDAdecay}}=100$ ms, and $\tau_{\text{GABA}}=10$ ms (Rolls and Deco 2010). Another useful point shown in Figs. 10a and b for Layer 5 is that the associative semantic recall implemented in Layer 5 by its attractor network (eg neurons connected to Object 2 when Object 1 is presented) takes in the order of 15 to 20 ms, which is consistent with attractor dynamics when the firing rates are not very high (Panzeri et al. 2001; Rolls and Deco 2010; Rolls 2023a; Rolls et al. 2024d).

A key issue in the conceptual framework developed here is that there is not a single attractor network for the semantics (only one is shown in Fig. 1 to help elucidate the concepts), but probably hundreds, in the cortical regions extending from the temporal pole through the anterior temporal cortex through the temporoparietal junction to the inferior parietal cortex (Rolls et al. 2022a; Rolls 2023a). Each local cortical attractor network operates within a range of approximately 2 mm, corresponding to the high density of local recurrent excitatory collaterals between pyramidal cells (Rolls 2023a), so there is plenty of cortical area for very many separate local attractor networks in the ATL to inferior parietal cortex semantic regions (Rolls et al. 2022a; Rolls 2023a). Moreover, each local attractor network can potentially store up to in the order of 10,000 representations (though fewer if they are correlated, as will be the case for semantic representations), given that there are in the order of 10,000 excitatory recurrent collateral synapses on each neuron and a sparse representation (Treves 1991; Treves and Rolls 1991; Rolls 2023a; Rolls and Treves 2024). The concept is advanced that each of these semantic

attractor networks will have different inputs from different parts of the cortex (visual object, visual face, visual scene, auditory, somatosensory, taste, olfactory, and action ...), because of the different connectivity of each region in the ATL, as we have shown (Rolls et al. 2022a). This results, together with the probabilistic local connectivity typical of the cerebral cortex (Rolls 2023a), in somewhat different semantic representations in each differently connected local attractor network. This is computationally useful, for though each single attractor network may be able to perform moderately well at storing hierarchically organized representations (Parga and Virasoro 1986), having the semantics set up in different attractor networks may help. For example, more general semantic concepts may be represented more anteriorly in the temporal lobe toward or in the temporal pole, being fed from more posterior semantic regions in the ATL that receive inputs from more unimodal sensory cortical processing streams (Rolls et al. 2022a). All of these semantic cortical areas in the ATL and inferior parietal cortex would need connectivity with cortical regions that may be involved in syntax, such as regions in and near regions 45 and 44, and that is the case (Rolls et al. 2022a; Rolls 2023a). Further support for this approach is that it is possible to store large numbers of memories with a semantic structure in a set of coupled attractor networks analyzed in the framework of a Potts network (Boboeva et al. 2018).

We now consider some of the concepts raised by the research described here.

First, although the example was the formation of a semantic category using visual stimuli, the theory and model apply fully to semantic categories formed by multimodal convergence, for example of visual, auditory, olfactory, somatosensory, and taste inputs. Indeed, one key cortical region for this multimodal convergence is the ATL, strongly involved in semantic representations (Peelen and Caramazza 2012; Milton et al. 2021; Rolls 2023a), which has connectivity from each sensory system to its cortical regions implicated in semantic representations (Rolls et al. 2022a). The semantic systems even extend up into the inferior parietal cortex, as shown by cortical connectivity (Rolls et al. 2022a), with the inferior parietal cortex probably encoding movement and action-related aspects of semantic representations (Wurm and Caramazza 2019). Indeed, different sets of cortical regions are recruited into activity depending on what attributes are relevant to a semantic representation (Peelen and Caramazza 2012; Fairhall and Caramazza 2013a, 2013b; Huth et al. 2016; Rolls et al. 2024b). For example, when viewing stationary images of tools, visual cortical movement-related and action-related regions are activated (Rolls et al. 2024b). Multimodal semantic representations of these types, perhaps formed in the way described here, are, it is suggested, what “concept cells” recorded in humans (Quian 2012) and perhaps in non-human primates (Rolls 2023c) reflect. Further, category-selective single neurons, responding not only to faces, objects, and scenes as we discovered in macaques (Perrett et al. 1979; Perrett et al. 1982; Rolls et al. 1997a; Booth and Rolls 1998; Rolls 2025c), but also, for example, to cars, have been described in the human medial temporal lobe (Kreiman et al. 2000).

Second, the principle described here is that temporal association of sets of attributes captured by cortical attractor networks that form associations across time in the order of a few seconds is useful in building semantic associations. An associative learning rule with a short-term memory is described here for building semantic categories in Equation (4). Other factors may also, of course, contribute to building semantic representations, such as physical visual similarities between all cars and separately

between all animals, though this was not utilized here, in that, as shown in Fig. 6, the representations in Layer 4 of VisNet of each object were orthogonal to each other. Another example is that semantic representations could reflect the similarity of certain visual scenes with each other. This is generally consistent with the idea that conceptual structure can arise from learning about the statistical structure of the environment across various modalities of perception and action (Giallanza et al. 2024), with, in our case, an emphasis on the temporal statistics of the environment. The type of learning described here that involves temporal associativity may be especially useful in the initial forming of semantic representations that might be described as “semantic categories,” and with some subsequent updating, these semantic representations may be further refined by other linguistic processes later on. To elucidate, there may be an asymmetry involved in some semantic knowledge; for example, we know that vehicles can be used to transport animals, but not (usually, at least nowadays) the other way around. Hence, temporal associativity may be useful in forming some semantic categories as described here, but other processes, which often involve language, may be involved in some more complex semantic knowledge.

Third, the principle described here is biologically plausible in that it involves multimodal convergence in connected cortical regions in the ATL and inferior parietal cortex, as already demonstrated (Rolls et al. 2022a), together with a local learning rule in which the information needed to alter the synaptic weights is present in the presynaptic and postsynaptic firing Equation (4). It is confirmed that the learning in the system described here is all local, and in that sense biologically plausible, with the synaptic weights changed as shown in Equation (2) (for VisNet) and Equation (4) (for Layer 5), which applies even if Layer 5 consists of many separate networks in the ATL. Another factor that is likely to be important in at least humans is linguistic analysis of all the thousands of attributes that can be used to describe objects and semantic categories, and the difficult learning task that this implies, which is typically supported by error backpropagation learning in deep neural networks (Hinton 1989; Rumelhart 1990; Rumelhart and Todd 1993; McClelland et al. 1995; Devereux et al. 2018; Saxe et al. 2019). However, error backpropagation learning in deep neural networks may not be biologically plausible in that it does not involve only the firing rates of the pre- and postsynaptic neurons, but instead, errors calculated based on the error measured at the output backpropagated through all the layers, neurons, and synapses in the preceding cortical regions (LeCun et al. 2015; Lillicrap et al. 2020; Rolls 2023a). Another problem with backpropagation of error learning in a biological context is that if something new needs to be learned, this can disrupt what has already been learned, with much retraining needed (Song et al. 2024), which is not very biologically adaptive. Indeed, an advantage of the principle of learning semantic representations described here is that it does not depend on lists of thousands of linguistic attributes such as “can fly” that are used to build semantic categories (Hinton 1989; Rumelhart 1990; Rumelhart and Todd 1993; McClelland et al. 1995; Devereux et al. 2018), but instead on what input representations are co-active in short periods of time of a few seconds, and so can contribute to semantic representations in most animals without explicit linguistic representations. The point being made here is that it is not clear that the way in which semantic representations are learned, referred to above (Hinton 1989; Rumelhart 1990; Rumelhart and Todd 1993; McClelland et al. 1995; Devereux et al. 2018; Saxe et al. 2019), is by using lists of thousands of attributes to find a particular weighted combination that is able to represent and

define a semantic category. Further, one alternative type of more biologically plausible approach is described here, which has the advantage that it can be applied to animals with semantic representations but without human language. Another type of artificial neural network that has been shown to produce some results similar to those in classical biologically plausible computational models of the hippocampal system (Rolls and Treves 2024) is the Transformer model (Whittington et al. 2022), but that has not been shown to be biologically plausible, and what was learned was about place cells, which is not a good model of primate including human hippocampal function in which spatial view cells provide representations of locations in viewed scenes (Rolls 2023d, 2025c, 2025a). The present authors do not exclude the possibility that some form of deep learning might be implemented in the brain, but note that more work is needed on this (LeCun et al. 2015; Lillicrap et al. 2020; Rolls 2023a, 2024b; Song et al. 2024).

In the training of the autoassociation network in Layer 5 in the training examples used here in Figs. 7–9, the five objects in one category were presented close together in one time period, and the five objects in the other category were presented close together in another time period, and separate categories were correctly learned by the Layer 5 network. If the training was not perfect, and for example in Fig. 7 where only cars were trained, instead one animal, such as a dog, sometimes occurred in the car group, then some association between that car category and a dog would be learned by the semantic network in Layer 5. That is, in fact, a natural component of semantic representations that they might reflect, for example, that dogs may jump into cars and fish, for example, do not. So some mixed training of that type might be advantageous to learning the associations between semantic items (in this case, cars and dogs) that occur naturally in the world.

In summary, this research describes the effective connectivity from the hippocampal episodic memory system to the ATL semantic memory system in humans (Figs. 3 and 4). The research also provides a new biologically plausible theory and model of how the hippocampal memory system may, by binding components, contribute to the learning of new semantic representations in attractor networks in the ATL using slow learning with a biologically plausible learning rule. It is emphasized that the temporal proximity described here that is used to build semantic representations could include the successive recall of a set of related episodic memories, with each episodic memory perhaps being about temporally quite distant events. For example, each episode might be about giving a different particular lecture each 1 week apart, but when recalled successively when thinking about each lecture, a semantic representation of what it is like giving lectures could be formed in the way described here. Another example might be many different bicycle rides, which, when recalled successively later, would help to build a semantic representation of what going for a bicycle ride is like. These examples show how the hippocampal episodic memory system could in the way described here help to build new semantic memories in neocortical semantic regions. This biologically plausible approach is compared with previous models of semantic learning, which have typically involved deep learning of how complex mappings of thousands of attributes could be used to define semantic categories.

Acknowledgments

Dr Alex Clarke of the Department of Psychology, University of Warwick, UK, is warmly thanked for a discussion of the paper by

Devereux, Clarke, and Tyler (Devereux et al. 2018), and Dr Sylvia Wirth (CNRS Institut des Sciences Cognitives Marc Jeannerod, Lyon, France) for discussions of the relations between episodic and semantic memory. Use of the Amsterdam Library of Object Images (ALOI) (Geusebroek et al. 2005) (<http://staff.science.uva.nl/aloi/>) is also gratefully acknowledged.

Author contributions

The first author Edmund T. Rolls designed the theory and model, programmed it, and wrote the paper. Dr Chenfei Zhang discussed the concepts, wrote the code for the integrate-and-fire modelling, and performed the Gabor filtering of the Amsterdam Library of Object Images (ALOI) objects using the Matlab code available with VisNet at <https://www.oxcns.org>. Jianfeng Feng provided the funding for the research. All authors approved the paper.

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

Funding

The work was supported by the following grants. Professor J. Feng: National Key R&D Program of China (No. 2019YFA0709502); 111 Project (No. B18015); Shanghai Municipal Science and Technology Major Project (No. 2018SHZDZX01), ZJLab, and Shanghai Center for Brain Science and Brain-Inspired Technology; and National Key R&D Program of China (No. 2018YFC1312904). The funding sources had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

Conflict of interest statement: None declared.

Data availability

The Matlab code for the first four layers of the model, VisNet, is available at <https://www.oxcns.org>. The implementation of the autoassociation (attractor) networks is described by Rolls (2016, 2021b, 2023a), with sample Matlab code made available at <https://www.oxcns.org/NeuronalNetworkSimulationSoftware.html>.

Ethical permissions

No data were collected as part of the research described here.

References

- Battaglia FP, Treves A. 1998. Stable and rapid recurrent processing in realistic auto-associative memories. *Neural Comput.* 10:431–450. <https://doi.org/10.1162/089976698300017827>.
- Bennett IJ, Stark CE. 2016. Mnemonic discrimination relates to perforant path integrity: an ultra-high resolution diffusion tensor imaging study. *Neurobiol Learn Mem.* 129:107–112. <https://doi.org/10.1016/j.nlm.2015.06.014>.
- Boboeva V, Brasselet R, Treves A. 2018. The capacity for correlated semantic memories in the cortex. *Entropy (Basel)*. 20:824. <https://doi.org/10.3390/e20110824>.
- Bonner MF, Price AR. 2013. Where is the anterior temporal lobe and what does it do? *J Neurosci.* 33:4213–4215. <https://doi.org/10.1523/JNEUROSCI.0041-13.2013>.
- Booth MCA, Rolls ET. 1998. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb Cortex.* 8:510–523. <https://doi.org/10.1093/cercor/8.6.510>.
- Brunel N, Wang XJ. 2001. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J Comput Neurosci.* 11:63–85. <https://doi.org/10.1023/A:1011204814320>.
- Cahusac PMB, Miyashita Y, Rolls ET. 1989. Responses of hippocampal formation neurons in the monkey related to delayed spatial response and object-place memory tasks. *Behav Brain Res.* 33:229–240. [https://doi.org/10.1016/S0166-4328\(89\)80118-4](https://doi.org/10.1016/S0166-4328(89)80118-4).
- Daugman JG. 1988. Complete discrete 2-D Gabor transforms by neural networks for image-analysis and compression. *IEEE Trans Acoust Speech Signal Process.* 36:1169–1179. <https://doi.org/10.1109/29.1644>.
- Deco G et al. 2019. Awakening: predicting external stimulation to force transitions between different brain states. *Proc Natl Acad Sci.* 116:18088–18097. <https://doi.org/10.1073/pnas.1905534116>.
- Devereux BJ, Clarke A, Tyler LK. 2018. Integrated deep visual and semantic attractor neural networks predict fMRI pattern-information along the ventral object processing pathway. *Sci Rep.* 8:10636.
- Duff MC, Covington NV, Hilverman C, Cohen NJ. 2019. Semantic memory and the hippocampus: revisiting, reaffirming, and extending the reach of their critical relationship. *Front Hum Neurosci.* 13:471. <https://doi.org/10.3389/fnhum.2019.00471>.
- Fairhall SL, Caramazza A. 2013a. Brain regions that represent amodal conceptual knowledge. *J Neurosci.* 33:10552–10558.
- Fairhall SL, Caramazza A. 2013b. Category-selective neural substrates for person- and place-related concepts. *Cortex.* 49:2748–2757.
- Franzius M, Sprekeler H, Wiskott L. 2007. Slowness and sparseness lead to place, head-direction, and spatial-view cells. *PLoS Comput Biol.* 3:e166. <https://doi.org/10.1371/journal.pcbi.0030166>.
- Geusebroek JM, Burghouts GJ, Smeulders AWM. 2005. The Amsterdam library of object images. *Int J Comput Vis.* 61:103–112. <https://doi.org/10.1023/B:VISI.0000042993.50813.60>.
- Giallanza T, Campbell D, Cohen JD, Rogers TT. 2024. An integrated model of semantics and control. *Psychol Rev.* <https://doi.org/10.1037/rev0000485>.
- Glasser MF et al. 2013. The minimal preprocessing pipelines for the human connectome project. *NeuroImage.* 80:105–124. <https://doi.org/10.1016/j.neuroimage.2013.04.127>.
- Glasser MF et al. 2016. A multi-modal parcellation of human cerebral cortex. *Nature.* 536:171–178. <https://doi.org/10.1038/nature18933>.
- Hickok G, Poeppel D. 2015. Neural basis of speech perception. *Handb Clin Neurol.* 129:149–160.
- Hinton GE. 1989. Learning distributed representations of concepts. In: *Parallel distributed processing. Implications for psychology and neurobiology* Morris RGM (ed). Clarendon Press, pp 46–61.
- Huang C-C, Rolls ET, Hsu C-CH, Feng J, Lin C-P. 2021. Extensive cortical connectivity of the human hippocampal memory system: beyond the “what” and “where” dual-stream model. *Cereb Cortex.* 31:4652–4669. <https://doi.org/10.1093/cercor/bhab113>.
- Huang CC, Rolls ET, Feng J, Lin CP. 2022. An extended human connectome project multimodal parcellation atlas of the human cortex and subcortical areas. *Brain Struct Funct.* 227:763–778. <https://doi.org/10.1007/s00429-021-02421-6>.
- Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL. 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature.* 532:453–458. <https://doi.org/10.1038/nature17637>.
- Kemmerer D. 2015. *Cognitive neuroscience of language*. New York: Psychology Press, DOI: <https://doi.org/10.3758/s13423-014-0784-1>.

- Khazali MF et al. 2024. A preserved neural code for temporal order between memory formation and recall in the human medial temporal lobe. *bioRxiv*. <https://doi.org/10.52054/FVVO.16.4.051>.
- Kreiman G, Koch C, Fried I. 2000. Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat Neurosci*. 3:946–953. <https://doi.org/10.1038/78868>.
- Kuznetsov YA. 2013. *Elements of applied bifurcation theory*. Springer Science and Business Media.
- LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature*. 521:436–444. <https://doi.org/10.1038/nature14539>.
- Lillicrap TP, Santoro A, Marris L, Akerman CJ, Hinton G. 2020. Back-propagation and the brain. *Nat Rev Neurosci*. 21:335–346. <https://doi.org/10.1038/s41583-020-0277-3>.
- McClelland JL, McNaughton BL, O'Reilly RC. 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. <https://doi.org/10.1037/0033-295X.102.3.419>.
- McClelland JL, McNaughton BL, Lampinen AK. 2020. Integration of new information in memory: new insights from a complementary learning systems perspective. *Philos Trans R Soc Lond Ser B Biol Sci*. 375:20190637.
- Milton CK et al. 2021. Parcellation-based anatomic model of the semantic network. *Brain Behav*. 11:e02065.
- Moscovitch M, Cabeza R, Winocur G, Nadel L. 2016. Episodic memory and beyond: the hippocampus and neocortex in transformation. *Annu Rev Psychol*. 67:105–134. <https://doi.org/10.1146/annurev-psych-113011-143733>.
- Panzeri S, Rolls ET, Battaglia F, Lavis R. 2001. Speed of feedforward and recurrent processing in multilayer networks of integrate-and-fire neurons. *Network*. 12:423–440.
- Parga N, Virasoro MA. 1986. The ultrametric organization of memories in a neural network. *J Phys*. 47:1857–1864. <https://doi.org/10.1051/jphys:0198600470110185700>.
- Parrey MUR. 2024. From cats to the cortex: unravelling the hierarchical processing system of vision and brain lasticity. *Cureus*. 16:e68424. <https://doi.org/10.7759/cureus.68424>.
- Patterson K, Nestor PJ, Rogers TT. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci*. 8:976–987. <https://doi.org/10.1038/nrn2277>.
- Peelen MV, Caramazza A. 2012. Conceptual object representations in human anterior temporal cortex. *J Neurosci*. 32:15728–15736. <https://doi.org/10.1523/JNEUROSCI.1953-12.2012>.
- Perrett DI, Rolls ET, Caan W. 1979. Temporal lobe cells of the monkey with visual responses selective for faces. *Neurosci Lett*. S3:S358.
- Perrett DI, Rolls ET, Caan W. 1982. Visual neurons responsive to faces in the monkey temporal cortex. *Exp Brain Res*. 47:329–342. <https://doi.org/10.1007/BF00239352>.
- Plebe A, Grasso G. 2019. The unbearable shallow understanding of deep learning. *Mind Mach*. 29:515–553. <https://doi.org/10.1007/s11023-019-09512-8>.
- Quian QR. 2012. Concept cells: the building blocks of declarative memory functions. *Nat Rev Neurosci*. 13:587–597.
- Rolls ET. 1992. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philos Trans R Soc Lond B*. 335:11–21.
- Rolls ET. 2012. Invariant visual object and face recognition: neural and computational bases, and a model, VisNet. *Front Comput Neurosci*. 6:35. <https://doi.org/10.3389/fncom.2012.00035>.
- Rolls ET. 2016. *Cerebral cortex: principles of operation*. Oxford University Press, Oxford.
- Rolls ET. 2021a. Learning invariant object and spatial view representations in the brain using slow unsupervised learning. *Front Comput Neurosci*. 15:686239.
- Rolls ET. 2021b. *Brain computations: what and how*. Oxford University Press, Oxford.
- Rolls ET. 2022. The hippocampus, ventromedial prefrontal cortex, and episodic and semantic memory. *Prog Neurobiol*. 217:102334. <https://doi.org/10.1016/j.pneurobio.2022.102334>.
- Rolls ET. 2023a. *Brain computations and connectivity*. Oxford University Press, Open Access, Oxford.
- Rolls ET. 2023b. Emotion, motivation, decision-making, the orbitofrontal cortex, anterior cingulate cortex, and the amygdala. *Brain Struct Funct*. 228:1201–1257.
- Rolls ET. 2023c. Hippocampal spatial view cells, place cells, and concept cells: view representations. *Hippocampus*. 33:667–687.
- Rolls ET. 2023d. Hippocampal spatial view cells for memory and navigation, and their underlying connectivity in humans. *Hippocampus*. 33:533–572.
- Rolls ET. 2024a. Two what, two where, visual cortical streams in humans. *Neurosci Biobehav Rev*. 160:105650.
- Rolls ET. 2024b. The memory systems of the human brain and generative artificial intelligence. *Heliyon*. 10:e31965.
- Rolls ET. 2025a. A theory and model of scene representations with hippocampal spatial view cells. *Hippocampus*. 35:e70013. <https://doi.org/10.1002/hipo.70013>.
- Rolls ET. 2025b. *Neuroscience discoveries*. Oxford University Press, Oxford.
- Rolls ET. 2025c. Hippocampal discoveries: spatial view cells, connectivity, and computations for memory and navigation, in primates including humans. *Hippocampus*. 35:e23666.
- Rolls ET, Deco G. 2010. *The noisy brain: stochastic dynamics as a principle of brain function*. Oxford University Press.
- Rolls ET, Deco G. 2015. Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain. *Brain Res*. 1621:316–334. <https://doi.org/10.1016/j.brainres.2014.09.021>.
- Rolls ET, Mills WPC. 2018. Non-accidental properties, metric invariance, and encoding by neurons in a model of ventral stream visual object recognition, VisNet. *Neurobiol Learn Mem*. 152:20–31. <https://doi.org/10.1016/j.nlm.2018.04.017>.
- Rolls ET, Milward T. 2000. A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Comput*. 12:2547–2572. <https://doi.org/10.1162/089976600300014845>.
- Rolls ET, Stringer SM. 2001. Invariant object recognition in the visual system with error correction and temporal difference learning. *Network*. 12:111–129.
- Rolls ET, Treves A. 2024. A theory of hippocampal function: new developments. *Prog Neurobiol*. 238:102636. <https://doi.org/10.1016/j.pneurobio.2024.102636>.
- Rolls ET, Turova TS. 2025. Visual cortical networks for 'What' and 'Where' to the human hippocampus revealed with dynamical graphs. *Cereb Cortex*. 35:bhaf106. <https://doi.org/10.1093/cercor/bhaf106>.
- Rolls ET, Xiang J-Z. 2006. Spatial view cells in the primate hippocampus, and memory recall. *Rev Neurosci*. 17:175–200. <https://doi.org/10.1515/revneuro.2006.17.1-2.175>.
- Rolls ET et al. 1989. Hippocampal neurons in the monkey with activity related to the place in which a stimulus is shown. *J Neurosci*. 9:1835–1845. <https://doi.org/10.1523/JNEUROSCI.09-06-01835.1989>.

- Rolls ET, Robertson RG, Georges-François P. 1997a. Spatial view cells in the primate hippocampus. *Eur J Neurosci*. 9:1789–1794.
- Rolls ET, Treves A, Foster D, Perez-Vicente C. 1997b. Simulation studies of the CA3 hippocampal subfield modelled as an attractor neural network. *Neural Netw*. 10:1559–1569.
- Rolls ET, Xiang J-Z, Franco L. 2005. Object, space and object-space representations in the primate hippocampus. *J Neurophysiol*. 94:833–844. <https://doi.org/10.1152/jn.01063.2004>.
- Rolls ET, Deco G, Huang C-C, Feng J. 2022a. The human language effective connectome. *NeuroImage*. 258:119352.
- Rolls ET, Deco G, Huang CC, Feng J. 2022b. The human orbitofrontal cortex, vmPFC, and anterior cingulate cortex effective connectome: emotion, memory, and action. *Cereb Cortex*. 33:330–356.
- Rolls ET, Deco G, Huang CC, Feng J. 2022c. The effective connectivity of the human hippocampal memory system. *Cereb Cortex*. 32:3706–3725.
- Rolls ET, Deco G, Huang C-C, Feng J. 2023a. Multiple cortical visual streams in humans. *Cereb Cortex*. 33:3319–3349.
- Rolls ET, Deco G, Huang C-C, Feng J. 2023b. Human amygdala compared to orbitofrontal cortex connectivity, and emotion. *Prog Neurobiol*. 220:102385.
- Rolls ET, Deco G, Huang CC, Feng J. 2023c. The human posterior parietal cortex: effective connectome, and its relation to function. *Cereb Cortex*. 33:3142–3170.
- Rolls ET, Deco G, Huang CC, Feng J. 2023d. Prefrontal and somatosensory-motor cortex effective connectivity in humans. *Cereb Cortex*. 33:4939–4963.
- Rolls ET, Deco G, Zhang Y, Feng J. 2023e. Hierarchical organization of the human ventral visual streams revealed with magnetoencephalography. *Cereb Cortex*. 33:10686–10701.
- Rolls ET, Rauschecker JP, Deco G, Huang CC, Feng J. 2023f. Auditory cortical connectivity in humans. *Cereb Cortex*. 33:6207–6227.
- Rolls ET, Wirth S, Deco G, Huang C-C, Feng J. 2023g. The human posterior cingulate, retrosplenial and medial parietal cortex effective connectome, and implications for memory and navigation. *Hum Brain Mapp*. 44:629–655.
- Rolls ET, Deco G, Huang CC, Feng J. 2024a. The connectivity of the human frontal pole cortex, and a theory of its involvement in exploit versus explore. *Cereb Cortex*. 34:1–19.
- Rolls ET, Feng J, Zhang R. 2024b. Selective activations and functional connectivities to the sight of faces, scenes, body parts and tools in visual and non-visual cortical regions leading to the human hippocampus. *Brain Struct Funct*. 229:1471–1493.
- Rolls ET et al. 2024c. A ventromedial visual cortical 'Where' stream to the human hippocampus for spatial scenes revealed with magnetoencephalography. *Commun Biol*. 7:1047.
- Rolls ET, Zhang C, Feng J. 2024d. Hippocampal storage and recall of neocortical 'What' - 'Where' representations. *Hippocampus*. 34:608–624.
- Rolls ET, Zhang R, Deco G, Vatansever D, Feng J. 2024e. Selective brain activations and connectivities related to the storage and recall of human object-location, reward-location, and word-pair episodic memories. *Hum Brain Mapp*. 45:e70056.
- Roudi Y, Treves A. 2003. Disappearance of spurious states in analog associative memories. *Phys Rev E Stat Nonlinear Soft Matter Phys*. 67:041906.
- Rumelhart DE (ed). 1990. *Brain style computation: learning and generalization*. Academic Press.
- Rumelhart DE, Todd PM. 1993. Learning and connectionist representations. *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience*; 3–30.
- Saxe AM, McClelland JL, Ganguli S. 2019. A mathematical theory of semantic development in deep neural networks. *Proc Natl Acad Sci USA*. 116:11537–11546. <https://doi.org/10.1073/pnas.1820261116>.
- Sellami A et al. 2017. Temporal binding function of dorsal CA1 is critical for declarative memory formation. *Proc Natl Acad Sci USA*. 114:10262–10267. <https://doi.org/10.1073/pnas.1619657114>.
- Song Y et al. 2024. Inferring neural activity before plasticity as a foundation for learning beyond backpropagation. *Nat Neurosci*. 27:348–358. <https://doi.org/10.1038/s41593-023-01514-1>.
- Squire LR, Zola-Morgan JT. 2011. The cognitive neuroscience of human memory since H.M. *Annu Rev Neurosci*. 34:259–288. <https://doi.org/10.1146/annurev-neuro-061010-113720>.
- Syversen IF et al. 2024. A combined DTI-fMRI approach for optimizing the delineation of posteromedial versus anterolateral entorhinal cortex. *Hippocampus*. 34:659–672. <https://doi.org/10.1002/hipo.23639>.
- Treves A. 1991. Dilution and sparse coding in threshold-linear nets. *J Phys A*. 24:327–335. <https://doi.org/10.1088/0305-4470/24/1/038>.
- Treves A. 1993. Mean-field analysis of neuronal spike dynamics. *Network*. 4:259–284.
- Treves A, Rolls ET. 1991. What determines the capacity of autoassociative memories in the brain? *Network*. 2:371–397.
- Treves A, Rolls ET. 1994. A computational analysis of the role of the hippocampus in memory. *Hippocampus*. 4:374–391. <https://doi.org/10.1002/hipo.450040319>.
- Treves A, Rolls ET, Simmen M. 1997. Time for retrieval in recurrent associative memories. *Physica D*. 107:392–400. [https://doi.org/10.1016/S0167-2789\(97\)00107-3](https://doi.org/10.1016/S0167-2789(97)00107-3).
- Van Essen DC et al. 2013. The WU-Minn human connectome project: an overview. *NeuroImage*. 80:62–79. <https://doi.org/10.1016/j.neuroimage.2013.05.041>.
- Wallis G, Rolls ET. 1997. Invariant face and object recognition in the visual system. *Prog Neurobiol*. 51:167–194. [https://doi.org/10.1016/S0301-0082\(96\)00054-8](https://doi.org/10.1016/S0301-0082(96)00054-8).
- Whittington JCR, Warren JD, Behrens TEJ. 2022. Relating transformers to models and neural representations of the hippocampal formation. <https://doi.org/10.48550/arXiv.2112.04035>.
- Wiskott L, Sejnowski TJ. 2002. Slow feature analysis: unsupervised learning of invariances. *Neural Comput*. 14:715–770. <https://doi.org/10.1162/089976602317318938>.
- Wurm MF, Caramazza A. 2019. Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. *Nat Commun*. 10:289.
- Wyss R, König P, Verschure PF. 2006. A model of the ventral visual system based on temporal stability and local memory. *PLoS Biol*. 4:e120. <https://doi.org/10.1371/journal.pbio.0040120>.
- Yang C, Naya Y. 2020. Hippocampal cells integrate past memory and present perception for the future. *PLoS Biol*. 18:e3000876. <https://doi.org/10.1371/journal.pbio.3000876>.
- Yang C, Naya Y. 2023. Sequential involvements of the perirhinal cortex and hippocampus in the recall of item-location associative memory in macaques. *PLoS Biol*. 21:e3002145. <https://doi.org/10.1371/journal.pbio.3002145>.