Glocal memory: a critical design principle for artificial brains and minds

June 30, 2010

Ben Goertzel[a,b][1], Joel Pitt[a], Matthew Ikle[a,c], Cassio Pennachin[a] , Liu Rui[a]

[a]*Novamente LLC, 1405 Bernerd Place, Rockville MD 20851*

[b]*Fujian Key Lab for Brain-Like Intelligent Systems, Department of Cognitive Science, Xiamen University, Xiamen, China*

[c]*Adams State College, Alamosa, CO*

Abstract

The concept of glocal memory (i.e. memory involving systematic coordination between localized memory traces and globalized dynamical-attractor-based memory traces) is reviewed, and is argued to be a critical principle for the design of artificial brains and artificial general intelligence systems. Some exploratory experiments are reviewed, involving introduction of glocal memory into Hopfield neural networks, and also into economic attention networks as utilized in the OpenCog and Novamente integrated AI architectures.

Keywords: glocal memory, artificial brains, attractor neural nets, Hopfield nets, economic attention networks, OpenCog, Novamente Cognition Engine.

[1] Corresponding author. Tel.: +1-240-505-6518; fax: +1-240-238-2630; e-mail: ben@goertzel.org.

1 Introduction

Memory lies at the heart of intelligence, and has been studied extensively in various disciplines and aspects, yet our overall understanding of the phenomenon remains extremely spotty. In this paper we present a neglected aspect of some memory systems, that we term glocality, describe some simple empirical results exploring its nature, and argue that it may be a critical property to consider when analyzing human and animal memory, and when constructing artificial brains and minds.

The notion of glocal memory has implicitly occurred in a number of prior brain theories (without use of the word "glocal"), e.g. (Calvin 1996) and Goertzel (2001), but it has not previously been explicitly developed. However the concept has risen to the fore in our recent AI work and so we feel it is apropos to currently bring it to the community's attention, and flesh it out more fully. A recent essay by the first author (Goertzel 2008*a*) reviews the glocal memory concept in a more general and less technical way than the present paper.

The basic motivation of the glocal memory idea is to overcome the dichotomy between localized memory (in which each memory item is stored in a single location within an overall memory structure) and distributed memory (in which a memory item is stored as an aspect of a multi-component memory system, in such a way that the same set of multiple components stores a large number of memories). In a glocal memory system, most memory items are stored both locally and globally, with the property that eliciting either one of the two records of an item tends to also elicit the other one.

Memory is frequently divided into multiple categories: perceptual (records of specific sensory stimuli) declarative (facts, beliefs and questions), procedural (practical how-to) and episodic (records of specific scenarios and experiences) (Anderson 1996). Glocal memory applies to multiple forms of memory; however we will focus largely on perceptual and declarative memory in our detailed analyses here, so as to conserve space and maintain simplicity of discussion.

Memory is one area where animal brain architecture differs radically from the von Neumann architecture underlying nearly all contemporary general-purpose computers. Von Neumann computers separate memory from processing, whereas in the human brain there is no such distinction. In fact, it's arguable that in most cases the brain contains no memory apart from processing: human memories are generally constructed in the course of remembering (Rosenfield 1988), which gives human memory a strong capability for "filling in gaps" of remembered experience and knowledge; and also causes problems with inaccurate remembering in many contexts (e.g Bransford & Franks 1971, Roediger & McDermott 1995). We believe the constructive aspect of memory is largely associated with its global, distributed aspect, and will enlarge on this in specific examples discussed below.

The central idea of glocal memory is that perceptual, declarative, episodic or procedural items may be stored in memory in the form of paired structures that are called (key, map) pairs. Of course the idea of a "pair" is abstract, and such pairs may manifest themselves quite differently in different sorts of memory systems (e.g. brains versus non-neuromorphic AI systems). The key is a localized version of the item, and records some significant aspects of the items in a simple and crisp way. The map is a dispersed, distributed version of the item, which represents the item as a (to some extent, dynamically shifting) combination of fragments of other items. The map includes the key as a subset; activation of the key generally (but not necessarily always) causes activation of the map; and changes in the memory item will generally involve complexly coordinated changes on the key and map level both.

In this paper, after presenting a slightly fuller formalization of the glocal memory concept, applications of the idea to various aspects of human and artificial intelligence are discussed. The hypothesis of glocal memory in the human brain, from Goertzel (1997), is updated in the light of more recent neuroscience thinking, and is fleshed out further via the design of a specific formal neural network model embodying simple glocality: the glocal Hopfield net. Results of some simple, initial computational explorations of glocal Hopfield nets are reported.

Then, the manifestation of glocal memory in more loosely brain based AI approaches is reviewed in the context of Glocal Economic Attention Networks (ECANs), rough analogues of glocal Hopfield nets that play a role in two integrated AI systems aimed at powerful artificial general intelligence: the Novamente Cognition Engine (Goertzel 2006) and OpenCog Prime (Goertzel 2008*b*) systems. Exploratory computational results regarding glocal ECANs are presented, along with discussion of the implications of these results for the coordinated global cognitive dynamics of integrated AI systems incorporating glocal ECANs.

The central theme of the paper – the importance of glocality as a property of memory systems – remains somewhat speculative. The presence of glocality in human and animal memory is strongly suggested but not firmly demonstrated by available neuroscience data; and the value of glocality in the context of artificial brains and minds is also not yet demonstrated as the whole field of artificial brain and mind building remains in its infancy. The work reported here however has been motivated by the authors' intuition that glocality is a critical property of memory in brains and in any even loosely brainlike AI system; and we have found the results of the preliminary investigations reported here to be fully consistent with this intuition.

2 Glocal Memory

To explain the notion of glocal memory more precisely, we will introduce a simple semi-formal model of a system S that uses a memory to record information relevant to the actions it carries out. The overall concept of glocal memory should not be considered as restricted to this particular model. This model is not intended for maximal generality, but is intended to encompass a variety of current AI system designs and formal neurological models.

In this model, we will consider S's memory subsystem as a set of objects we'll call "tokens," embedded in some metric space. The metric in the space, which we will call the "basic distance" of the memory, generally will not be defined in terms of the semantics of the items stored in the memory; though it may come to shape these dynamics through the specific architecture and evolution of the memory. Note that these tokens are not intended as generally being mapped one-to-one onto meaningful items stored in the memory. The "tokens" are the raw materials that the memory arranges in various patterns in order to store items.

We assume that each token, at each point in time, may meaningfully be assigned a certain quantitative "activation level." Also, tokens may have other numerical or discrete quantities associated with them, depending on the particular memory architecture. Finally, tokens may relate other tokens, so that optionally a token may come equipped with an (ordered or unordered) list of other tokens.

To understand the meaning of the activation levels, one should think about S's memory subsystem as being coupled with an action-selection subsystem, that dynamically chooses the actions to be taken by the overall system in which the two subsystems are embedded. Each combination of actions, in each particular type of context, will generally be associated with the activation of certain tokens in memory.

Then, as analysts of the system S, we may associate each token T with an "activation vector" $v(T,t)$, whose value for each discrete time t consists of the activation of the token T at time t. So, the $50'th$ entry of the vector corresponds to the activation of the token at the $50'th$ time step.

"Items stored in memory" over a certain period of time, may then be defined as clusters in the set of activation vectors associated with memory during that period of time. Note that the system S itself may explicitly recognize and remember patterns regarding what items are stored in its memory – but, from an external analyst's perspective, the set of items in S's memory is not restricted to the ones that S has explicitly recognized as memory items.

The "localization" of a memory item may be defined as the degree to which the various tokens involved in the item are close to each other according to the metric in the memory metric-space. This degree may be formalized in various ways, but choosing a particular quantitative measure is not important here. A highly localized item may be called "local" and a not-very-localized item may be called "global."

We may define the "activation distance" of two tokens as the distance between their activation vectors. We may then say that a memory is "well aligned" to the extent that there is a correlation between the activation distance of tokens, and the basic distance of the memory metric-space.

Given the above set-up, the basic notion of glocal memory can be enounced fairly simply. A glocal memory is one:

- That is reasonably well-aligned (i.e. the correlation between activation and basic distance is significantly greater than random)
- In which most memory items come in pairs, consisting of one local item and one global item, so that activation of the local item (the "key") frequently leads in the near future to activation of the global item (the "map")

Obviously, in the scope of all possible memory structures constructible within the above formalism, glocal memories are going to be very rare and special. But, we suggest that they are important, because they are generally going to be the most effective way for intelligent systems to structure their memories.

Note also that many memories without glocal structure may be "well-aligned" in the above sense.

An example of a predominantly local memory structure, in which nearly all significant memory items are local according to the above definition, is the Cyc logical reasoning engine (Lenat & Guha 1990). To cast the Cyc knowledge base in the present formal model, the tokens are logical predicates. Cyc does not have an in-built notion of activation, but one may conceive the activation of a logical formula in Cyc as the degree to which the formula is used in reasoning or query processing during a certain interval in time. And one may define a basic metric for Cyc by associating a predicate with its extension, and defining the similarity of two predicates as the symmetric distance of their extensions. Cyc is reasonably well-aligned, but according to the dynamics of its querying and reasoning engines, it is basically a local memory structure without significant global memory structure.

On the other hand, an example of a predominantly global memory structure, in which nearly all significant memory items are global according to the above definition, is the Hopfield associative memory network (Amit 1989). Here memories are stored in the pattern of weights associated with synapses within a network of formal neurons, and each memory in general involves a large number of the neurons in the network. To cast the Hopfield net in the present formal model, the tokens are neurons and synapses; the activations are neural net activations; the basic distance between two neurons A and B may be defined as the percentage of the time that stimulating one of the neurons leads to the other one firing; and to calculate a basic distance involving a synapse, one may associate the synapse with its source and target neurons. With these definitions, a Hopfield network is a well-aligned memory, and (by intentional construction) a markedly global one. Local memory items will be very rare in a Hopfield net.

While predominantly local and predominantly global memories may have great value for particular applications, our suggestion is that they also have inherent limitations. If so, this means that the most useful memories are going to be those that involve both local and global memory items in central roles.

However, this is a more general and less risky claim than the assertion that glocal memory structure as defined above is important. Because, "glocal" as defined above doesn't just mean "neither predominantly global nor predominantly local." Rather, it refers to a specific pattern of coordination between local and global memory items – what we have called the "keys and maps" pattern.

3 Hints of Glocal Memory in the Human Brain

Our understanding of human brain dynamics is still very primitive, one manifestation of which is the fact that we really don't understand how the brain represents knowledge, except in some very simple respects. So anything anyone says about knowledge representation in the brain, at this stage, has to be considered highly speculative. Existing neuroscience knowledge does imply constraints on how knowledge representation in the brain may work, but these are relatively loose constraints. These constraints do imply that, for instance, the brain is neither a relational database (in which information is stored in a wholly localized manner) nor a collection of "grandmother neurons" that respond individually to high-level percepts or concepts; nor a simple Hopfield type neural net (in which all memories are attractors globally distributed across the whole network). But they don't tell us nearly enough to, for instance, create a formal neural net model that can confidently be said to represent knowledge in the manner of the human brain.

As a first example of the current state of knowledge, we'll discuss here a series of papers regarding the neural representation of visual stimuli (Quiroga et al. 2005, Quiroga et al. 2008), which deal with the fascinating discovery of a subset of neurons in the medial temporal lobe (MTL) that are selectively activated by strikingly different pictures of given individuals, landmarks or objects, and in some cases even by letter strings. For instance, in the 2005 paper titled "Invariant visual representation by single neurons in the human brain", it is noted that

> in one case, a unit responded only to three completely different images of the ex-president Bill Clinton. Another unit (from a different patient) responded only to images of The Beatles, another one to cartoons from The Simpson's television series and another one to pictures of the basketball player Michael Jordan.

The 2008 paper backed away from the more extreme interpretation in the title as well as the conclusion, with the title "Sparse but not 'Grandmother-cell' coding in the medial temporal lobe." As the authors emphasize there,

> Given the very sparse and abstract representation of visual information by these neurons, they could in principle be considered as 'grandmother cells'. However, we give several arguments that make such an extreme interpretation unlikely.
>
> . . .
>
> MTL neurons are situated at the juncture of transformation of percepts into constructs that can be consciously recollected. These cells respond to percepts rather than to the detailed information falling on the retina. Thus, their activity reflects the full transformation that visual information undergoes through the ventral pathway. A crucial aspect of this transformation is the complementary development of both selectivity and invariance. The evidence presented here, obtained from recordings of single-neuron activity in humans, suggests that a subset of MTL neurons possesses a striking invariant representation for consciously perceived objects, responding to abstract concepts rather than more basic metric details. This representation is sparse, in the sense that responsive neurons fire only to very few stimuli (and are mostly silent except for their preferred

> stimuli), but it is far from a Grandmother-cell representation. The fact that the MTL represents conscious abstract information in such a sparse and invariant way is consistent with its prominent role in the consolidation of long-term semantic memories.

It's interesting to note how inadequate the Quiroga et al. data really is for exploring the notion of glocal memory in the brain. Suppose it's the case that individual visual memories correspond to keys consisting of small neuronal subnetworks, and maps consisting of larger neuronal subnetworks. Then it would be not at all surprising if neurons in the "key" network corresponding to a visual concept like "Bill Clinton's face" would be found to respond differentially to the presentation of appropriate images. Yet, it would also be wrong to overinterpret such data as implying that the key network somehow comprises the "representation" of Bill Clinton's face in the individual's brain. In fact this key network would comprise only one aspect of said representation.

In the glocal memory hypothesis, a visual memory like "Bill Clinton's face" would be hypothesized to correspond to an attractor spanning a significant subnetwork of the individual's brain – but this subnetwork still might occupy only a small fraction of the neurons in the brain (say, 1/100 or less), since there are very many neurons available. This attractor would constitute the map. But then, there would be a much smaller number of neurons serving as key to unlock this map: i.e. if a few of these key neurons were stimulated, then the overall attractor pattern in the map as a whole would unfold and come to play a significant role in the overall brain activity landscape. In prior publications (e.g. Goertzel 1997) the primary author explored this hypothesis in more detail in terms of the known architecture of the cortex and the mathematics of complex dynamical attractors.

So, one possible interpretation of the Quiroga et al. data is that the MTL neurons they're measuring are part of key networks that correspond to broader map networks recording percepts. The map networks might then extend more broadly throughout the brain, beyond the MTL and into other perceptual and cognitive areas of cortex. Furthermore, in this case, if some MTL key neurons were removed, the maps might well regenerate the missing keys (as would happen e.g. in the glocal Hopfield model to be discussed in the following section).

Related and interesting evidence for glocal memory in the brain comes from a recent study of semantic memory, illustrated in Figure 1 (Patterson et al. 2007). Their research probed the architecture of semantic memory via comparing patients suffering from semantic dementia (SD) with patients suffering from three other neuropathologies, and found reasonably convincing evidence for what they call a "distributed-plus-hub" view of memory.

The SD patients they studied displayed highly distinctive symptomology; for instance, their vocabularies and knowledge of the properties of everyday objects were strongly impaired, whereas their memories of recent events and other cognitive capacities remain perfectly intact. These patients also showed highly distinctive patterns of brain damage: focal brain lesions in their anterior temporal lobes (ATL), unlike the other patients who had either less severe or more widely distributed damage in their ATLs. This led Patterson et al. to conclude that the ATL (being adjacent to the amygdala and limbic systems that process reward and emotion; and the anterior parts of the medial temporal lobe memory system, which processes episodic memory) is a "hub" for amodal semantic memory, drawing general semantic information from episodic memories based on emotional salience.

So, in this view, the memory of something like a "banana" would contain a distributed aspect, spanning multiple brain systems, and also a localized aspect, centralized in the ATL. The distributed aspect would likely contain information on various particular aspects of bananas, including their sights, smells, and touches, the emotions they evoke, and the goals and motivations they relate to. The distributed and localized aspects would influence one another dynamically, but, the data Patterson et al. gathered do not address dynamics and they don't venture hypotheses in this direction.

There is a relationship between the "distributed-plus-hub" view and Damasio's (2000) better-known notion of a "convergence zone", defined roughly as a location where the brain binds features together. A convergence zone, in Damasio's (2000) perspective, is not a "store" of information but an agent capable of decoding a signal (and of reconstructing information). He also uses the metaphor that convergence zones behave like indexes drawing information from other areas of the brain – but they are dynamic rather than static indices, containing the instructions needed to recognize and combine the features constituting the memory of something. The mechanism involved in the distributed-plus-hub model is similar to a convergence zone, but with the important difference that hubs are less local: Patterson et al.'s (2007) semantic hub may be thought of a kind of "cluster of convergence zones" consisting of a network of convergence zones for various semantic memories.

What is missing in Patterson et al.'s (2007) and Damasio's (2000) perspective is a vision of distributed memories as attractors. The idea of localized memories serving as indices into distributed knowledge stores is important, but is only half the picture of glocal memory: the creative, constructive, dynamical-attractor aspect of the distributed representation is the other half. The closest thing to a clear depiction of this aspect of glocal memory that seems to exist in the neuroscience literature is a portion of William Calvin's theory of the "cerebral code" (Calvin 1996). Calvin proposes a set of quite specific mechanisms by which knowledge may be represented in the brain using complexly-structured strange attractors, and by which these strange attractors may be propagated throughout the brain. Figure 2 shows one aspect of his theory: how a distributed attractor may propagate from one part of the brain to another in pieces, with one portion of the attractor getting propagated first, and then seeding the formation in the destination brain region of a close approximation of the whole attractor.

Calvin's theory may be considered a genuinely glocal theory of memory. However, it also makes a large number of other specific commitments that are not part of the notion of glocality, such as his proposal of hexagonal meta-columns in the cortex, and his commitment to evolutionary learning as the primary driver of neural knowledge creation. We find these other hypotheses interesting and highly promising, yet feel it is also important to separate out the notion of glocal memory for separate consideration.

Regarding specifics, our suggestion is that Calvin's approach may overemphasize the distributed aspect of memory, not giving sufficient due to the relatively localized aspect as accounted for in the Quiroga et al. results discussed above. In Calvin's glocal approach, global memories are attractors and local memories are parts of attractors. We suggest a possible alternative, in which global memories are attractors and local memories are particular neuronal subnetworks such as the specialized ones identified by Quiroga et al.. However, this alternative does not seem contradictory to Calvin's overall conceptual approach, even though it is different from the particular proposals made in Calvin (1996).

The above paragraphs are far from a complete survey of the relevant neuroscience literature; there are literally dozens of studies one could survey pointing toward the glocality of various sorts of human memory. Yet experimental neuroscience tools are still relatively primitive, and every one of these studies could be interpreted in various other ways. In the next couple decades, as neuroscience tools improve in accuracy, our understanding of the role of glocality in human memory will doubtless improve tremendously.

4 Glocal Hopfield Networks

The ideas in the previous section suggest that, if one wishes to construct an artificial brain, it is worth seriously considering that this artificial brain should have a glocally structured memory. One research direction that follows naturally from this notion is "glocal neural networks." In order to explore the nature of glocal neural networks in a relatively simple and tractable setting, we have formalized and implemented simple examples of "glocal Hopfield networks": palimpsest Hopfield nets with the addition of neurons representing localized memories.

Essentially, we augment the standard Hopfield net architecture by adding a set of "key neurons." These are a small percentage of the neurons in the network, and are intended to be roughly equinumerous to the number of memories the network is supposed to store. When the Hopfield net converges to an attractor A, then new links are created between the neurons that are active in A, and one of the key neurons. Which key neuron is chosen? The one that, when it is stimulated, gives rise to an attractor pattern maximally similar to A.

The ultimate result of this is that, in addition to the distributed memory of attractors in the Hopfield net, one has a set of key neurons that in effect index the attractors. Each attractor corresponds to a single key neuron. In the glocal memory model, the key neurons are the keys and the Hopfield net attractors are the maps.

This algorithm has been tested in sparse Hopfield nets, using both standard Hopfield net learning rules and Storkey's modified palimpsest learning rule (Storkey & Valabregue 1999), which provides greater memory capacity in a continuous learning context. The use of key neurons turns out to slightly increase Hopfield net memory capacity, but this isn't the main point. The main point is that one now has a local representation of each global memory, so that if one wants to create a link between the memory and something else, it's extremely easy to do so – one just needs to link to the corresponding key neuron. Or, rather, one of the corresponding key neurons: depending on how many key neurons are allocated, one might end up with a number of key neurons corresponding to each memory, not just one.

4.1 The Hopfield neural net model

Hopfield networks (Hopfield 1982) are attractor networks often used as associative memories. A Hopfield network with N neurons can be trained to store a set of bipolar patterns P, where each pattern p has N bipolar (± 1) values. A Hopfield net typically has symmetric weights with no self-connections. The weight of the connection between neurons i and j is denoted by w_{ij}.

In order to apply a Hopfield network to a given input pattern p, its activation state is set to the input pattern, and neurons are updated asynchronously, in random order, until the network converges to the closest fixed point. An often-used activation function for a neuron is:

$$y_i = sign(p_i \sum_{j \neq i} w_{ij} y_j)$$

Training a Hopfield network, therefore, involves finding a set of weights w_{ij} that stores the training patterns as attractors of its network dynamics, allowing future recall of these patterns from possibly noisy inputs. Originally, Hopfield used a Hebbian rule to determine weights:

$$w_{ij} = \sum_{p=1}^{P} p_i p_j$$

Typically, Hopfield networks are fully connected. Experimental evidence, however, suggests that the majority of the connections can be removed without significantly impacting the network's capacity or dynamics. Our experimental work in this paper uses sparse Hopfield networks.

4.2 Palimpsest Hopfield nets with a modified learning rule

In Storkey & Valabregue (1999) a new learning rule is presented, which both increases the Hopfield network capacity and turns it into a "palimpsest", i.e., a network that can continuously learn new patterns, while forgetting old ones in an orderly fashion.

Using this new training rule, weights are initially set to zero, and updated for each new pattern p to be learned according to:

$$h_{ij} = \sum_{k=1, k\neq i,j}^{N} w_{ik} p_k$$
$$\Delta w_{ij} = \frac{1}{n}(p_i p_j - h_{ij} p_j - h_{ji} p_i)$$

4.3 Glocal Hopfield nets

In order to transform a palimpsest Hopfield net into a glocal Hopfield net, the following steps are taken:

1. Add a fixed number of "key neurons" to the network (removing other random neurons to keep the total number of neurons constant)

2. When the network reaches an attractor, create links from the elements in the attractor to one of the key neurons

3. The key neuron chosen for the previous step is the one that most closely matches the current attractor (which may be determined in several ways, to be discussed below)

4. To avoid the increase of the number of links in the network, when new links are created in Step 2, other key-neuron links are then deleted (several approaches may be taken here, but the simplest is to remove the key-neuron links with the lowest-absolute-value weights)

Figure 3 gives pseudocode for one simple implementation of the above steps, in which step 3 is carried out simply by comparing the weights of a key neuron's links to the nodes in an attractor. A more sophisticated approach would be to select the key neuron with the highest activation during the transient interval immediately prior to convergence to the attractor.

The result of these modifications to the ordinary Hopfield net, is a Hopfield net that continually maintains a set of key neurons, each of which individually represents a certain attractor of the net.

As noted above, this turns out to slightly increase the memory capacity of the Hopfield net; but also, more importantly, it makes it far easier to link the Hopfield net into a larger system, as would occur if the Hopfield net were embedded in a genuine artificial brain. Because a neuron external to the Hopfield net may now link to a memory in the Hopfield net by linking to the corresponding key neuron.

Note that these key neurons – in spite of being "symbolic" in nature – are learned rather than preprogrammed, and are every bit as adaptive as the attractors they correspond to. Furthermore, if a key neuron is removed, the glocal Hopfield net algorithm will eventually learn it back, so the robustness properties of Hopfield nets are retained.

5 Experimental explorations

Investigation of the empirical properties of glocal Hopfield nets is a significant task and would require a substantial paper in itself. Here we restrict ourselves to giving one set of example results on a small network.

Figure 4 shows the data that was used to guide one round of experimentation with our glocal Hopfield net. We imposed a set of memories corresponding to ASCII characters on a network, and then sought to retrieve these memories using mutated versions of these characters, containing various errors.

Figure 5 shows the memory retrieval percentages obtained with and without the key neurons, for a network with neurons corresponding to the pixels in the ASCII images in Figure 4, and a sparse connection density of 10%. Note that key neurons provided a modest but significant improvement in retrieval behavior. Note also that, for the specific experiments underlying this table, we intentionally ran the algorithm in a regime where retrieval percentages were not that high. By adjusting the number of neurons, the sparseness of the network, and the number of memories imprinted, one can adjust the retrieval percentages almost arbitrarily; but generically, it seems that the key neurons do slightly improve memory retrieval regardless of parameter settings.

The above preliminary results are intriguing but merit considerable extension, with exploration of various datasets and parameter values. This work is worthwhile and will likely be pursued in a future publication, but perhaps more interesting is considering follow-up research directions beyond the strict domain of Hopfield nets.

Hopfield nets are an interesting experimental testbed for ideas about glocal memory and other topics; but ultimately Hopfield nets as such are "toy systems" rather than representing practical attempts at creating artificial brains or minds. With this in mind, the ideas of the previous section may be extended beyond Hopfield nets in a variety of ways.

For instance, one could make the neurons more biologically realistic, and seek to model specific portions of mammalian cortex using glocal Hopfield nets. Modeling rabbit olfactory cortex would be particularly interesting, given Freeman & Schneider's (1982) work on attractors and their role in rabbit olfactory memory.

On the other hand, one can also explore similar dynamics in the context of artificial cognition systems that are brain-inspired but less directly brain-based than neural net models. This is the direction we will take in the following section: we will explore a system analogous to the above glocal Hopfield net, but involving the Economic Attention Networks used in the integrative OpenCog Prime (OCP) and Novamente Cognition Engine (NCE) AI architectures.

6 Glocal Economic Attention Networks

Economic Attention Networks (ECANs) are similar in many respects to Hopfield nets, but are based on a different conceptual foundation involving the propagation of amounts of (conserved) currency rather than than neural-net activation. Further, as will be discussed in the following section, ECANs are specifically designed for integration with a diverse body of cognitive processes as embodied in integrative AI designs such as the NCE or OCP.

6.1 Economic Attention Networks and ECAN Integration into OCP

The OpenCog AGI framework, within which the current ECAN implementation exists, is a complex framework with a complex underlying theory. OpenCog is an open-source software framework designed to support the construction of multiple AI systems. The current main thrust of work within OpenCog is the implementation of a specific AGI design called OpenCogPrime (OCP), which is presented in the online wikibook (Goertzel, 2008). Much of the OpenCog software code, and many of the ideas in the OCP design, were derived from the open-sourcing of aspects of the proprietary Novamente Cognition Engine. Though they can also work as standalone systems, ECANs were initially designed to serve two main purposes within OCP. They were designed to serve as an associative memory for the network, and to facilitate effective allocation of the attention of other cognitive processes to appropriate knowledge items.

An ECAN is simply a graph, consisting of un-typed nodes and links, and also links that may be typed either HebbianLink or InverseHebbianLink. Each node and link in an ECAN is weighted with two numbers, called STI (short-term importance) and LTI (long-term importance); and each Hebbian or InverseHebbian link is weighted with a probability value. The term atom will be used to refer to either nodes or links.

The equations of an ECAN explain how the STI, LTI and Hebbian probability values get updated over time. As alluded to above, the metaphor underlying these equations is the interpretation of STI and LTI values as (separate) artificial currencies. The fact that STI and LTI are currencies means that, except in unusual instances where the ECAN controller decides to introduce inflation or deflation and explicitly manipulate the amount of currency in circulation, the total amounts of STI and LTI in the system are conserved. This fact makes the dynamics of an ECAN dramatically different than that of an attractor neural network.

Conceptually, the STI value of an Atom is interpreted to indicate the immediate urgency of the Atom to the ECAN at a certain point in time; whereas the LTI value of an Atom indicates the amount of value the ECAN perceives in the retention of the Atom in memory (RAM). The ECAN equations also contain the notion of an Attentional Focus (AF), consisting of those Atoms in the ECAN with the highest STI values. These atoms play a privileged role in the system and, as such, are treated using an alternate set of equations.

Conceptually, the probability value of a HebbianLink from A to B is the odds that if A is in the AF, so is B; and correspondingly, the InverseHebbianLink from A to B is weighted with the odds that if A is in the AF, then B is not. An ECAN will often be coupled with a "Forgetting" process that removes low-LTI Atoms from memory according to certain heuristics. A critical aspect of the ECAN equations is that Atoms periodically spread their STI and LTI to other Atoms that connect to them via Hebbian and InverseHebbianLinks; this is the ECAN analogue of activation spreading in neural networks.

6.2 Glocal Economic Attention Networks

In order to transform ordinary ECANs into glocal ECANs, one may proceed in essentially the same manner as with glocal Hopfield nets above. In the langauge normally used to describe NCE and OCP, this would be termed a "map encapsulation" heuristic. As with glocal Hopfield nets, one may proceed most simply via creating a fixed pool of nodes intended to provide locally-representative keys for the maps formed as attractors of the network. Links may then be formed to these key nodes, with weights and STI and LTI values adapted by the usual ECAN algorithms.

6.3 Experimental explorations

To compare the performance of ECANs with Hopfield networks, we set up the ECAN to have the same topology as a Hopfield network. That is, a number of nodes take on the equivalent role of the neurons that are presented patterns to be stored. These patterns are imprinted by setting the corresponding nodes of active bits to have their STI within the AF, whereas nodes corresponding to inactive bits of the pattern are below the AF threshold. Link weight updating then occurs, using one of several update rules, but in this case the update rule of Storkey & Valabregue (1999) was used. Attention spread used a diffusion approach by representing the weights of Hebbian links between pattern nodes within a left stochastic Markov matrix, and multiplying it by the vector of normalised STI values to give a vector representing the new distribution of STI.

To explore the effects of key nodes on ECAN Hopfield networks, we used the palimpsest testing scenario of Storkey & Valabregue (1999), where all the local neighbours of the imprinted pattern, within a single bit change, are tested. Each neighbouring pattern is used as input to try and retrieve the original pattern. If all the retrieved pattern are the same as the original (within a given tolerance) then the pattern is deemed successfully retrieved and recall of the previous pattern is attempted via its neighbours. The number of patterns this can repeat for successfully is called the palimpsest storage of the network.

Figure 6 shows the results of one particular set up with 10 by 10 pattern nodes, a Hebbian link density of 30%, and with 1% of links being forgotten before each pattern is imprinted. The results demonstrate that, regardless of the presence or absence of key nodes, the recall is successful for all patterns up till about the 30th. Thereafter the storage is erratic, and it is difficult to tell whether the presence of key nodes is beneficial to palimpsest storage. However, when the mean palimpsest storage is calculated for each of 0, 1, 5 and 10 key nodes we find that the storage is 22.6, 22.4, 24.9, and 26.0 patterns respectively, indicating that key nodes do improve memory recall on average.

As in the case of our results on glocal Hopfield networks, these results are preliminary but promising, and will likely be pursued in a future paper that also explores the effect of link density, among other factors.

7 Glocality in Integrative AGI Systems

One of the main motivations for the development of the glocal memory concept has been the design of artificial memories, which is a task different in many ways from the analysis of modeling of naturally occurring memories. In our work on the Novamente Cognition Engine (Goertzel 2006) and OpenCog (Goertzel 2008*b*) AI systems, we've been motivated by the glocal memory concept to design memory approaches that are explicitly glocal in nature.

The glocality concept hits straight at the center of one of the biggest debates of theoretical AI: symbolic versus subsymbolic knowledge representation (Hofstadter 1979, Hammer & Hitzler 2007, d'Avila Garcez et al. 2008). This dichotomy is often discussed but rarely drawn in a formal and rigorous way, and we have argued elsewhere that it is actually a largely bogus dichotomy (Goertzel et al. 2009). Traditionally, logic-based AI systems are viewed as "symbolic", and neural net systems are viewed as "subsymbolic." But this distinction has gotten fuzzier and fuzzier in recent years, with developments such as

- Logic-based systems being used to control embodied agents (hence using logical terms to deal with data that is apparently perception or actuation-oriented in nature, rather than being symbolic in the semiotic sense), see Santore & Shapiro (2003) and Goertzel (2008*b*).

- Hybrid systems combining neural net and logical parts, or using logical or neural net components interchangeably in the same role (Lebiere & Anderson (in preparation)).

- Neural net systems being used for strongly symbolic tasks such as automated grammar learning (Elman, (1991), plus more recent work.)

In our own AI systems referenced above, we have explicitly sought to span the symbolic/subsymbolic pseudo-dichotomy, via creating integrative systems that combine logic-based aspects with neural-net-like aspects. These work, not in the manner of multimodular systems, but via what we call a probabilistic hypergraph. The nodes and links in this hypergraph are typed, like a standard semantic network approach for knowledge representation, so they're able to handle all sorts of knowledge, from the most concrete perception and actuation related knowledge to the most abstract relationships. We also attach uncertain-logical truth values and neural-net-like weight and activation values to the same nodes and links. The concept of glocality lies at the heart of this combination, in a way that spans the pseudo-dichotomy:

- Local knowledge is represented in abstract logical relationships stored in explicit logical form, and also in Hebbian-type associations between nodes and links.

- Global knowledge is represented in large-scale patterns of node and link weights, which lead to large-scale patterns of network activity, which often take the form of attractors qualitatively similar to Hopfield net attractors. These attractors are called maps.

This hypergraph is called an AtomSpace in the Novamente Cognition Engine (NCE) and OpenCog designs. Figure 7 illustrates a small hypergraph with a map.

Both the NCE and OpenCog are multi-agent systems. They possess a number of cognitive processes, called MindAgents, which continuously process the AtomSpace, modifying existing nodes and links, creating new ones, and removing old ones that haven't been useful to the system. Figure 8 depicts one such system.

The AtomSpace is continuously acted on by a variety of cognitive processes, encapsulated in software objects called MindAgents. Some MindAgents work together to carry out probabilistic logical reasoning according to the mathematics given in Goertzel (2008*b*); others spread neural-net-like weights and activation values (STI and LTI) according to the ECAN equations. The attractors of this nonlinear activation spreading process constitute global memories; and there are then explicit MapEncapsulation MindAgents that identify these attractors and build "key nodes" corresponding to them, according to the glocal ECAN process described above. The logical inference and activation spreading processes feed off each other in particular ways, so that the formation and maintenance of the glocal memory is a result of the integrated behavior of the system's multiple cognitive dynamics.

The result of all this is that a concept like "cat" might be represented as a combination of:

- A small number of logical relationships and strong associations, that constitute the "key" subnetwork for the "cat" concept.
- A large network of weak associations, binding together various nodes and links of various types and various levels of abstraction, representing the "cat map".

The activation of the key will generally cause the activation of the map, and the activation of a significant percentage of the map will cause the activation of the rest of the map, including the key. Furthermore, if the key were for some reason forgotten, then after a significant amount of effort, the system would likely to be able to reconstitute it (perhaps with various small changes) from the information in the map. We conjecture that this particular kind of glocal memory will turn out to be very powerful for AI, due to its ability to combine the strengths of formal logical inference with those of self-organizing attractor neural networks.

8 Promising Directions for Future Work

This paper has woven together a number of major themes – glocal memory in the brain, glocal Hopfield nets, glocal ECANs and their role in integrative AI systems – and every one of these is at an early stage, requiring significant further investigation. Among the many avenues for future work, the ones that seem most fascinating to us are:

- combining analysis of multiple forms of brain imaging data so as to go beyond the insights of Quiroga et al. (2008) and explore the hypothesis of glocal concept memory in the human brain
- extending the mathematical analyses of ANNs and ECANs to encompass the process of node/link formation according to glocal memory related algorithms
- empirically exploring the contribution of glocal ECANs to the NCE and OCP integrative AGI systems
- exploring hierarchical keys, leading to structured memories that may act similarly to the stacking of restricted Boltzmann machines as described in (Hinton et al. 2006). Boltzmann machines can be seen as stochastically updating counterparts to Hopfield network.

In the process of making relevant technical discoveries along these and other avenues of exploration, the core hypothesis presented here – that glocal memory is a critical concept for artificial brains and minds – will be crispened, and its limits of validity will be explored and more fully comprehended.

9 Acknowledgments

The authors would like to acknowledge Linas Vepstas, Gustavo Gama, and Jared Wigmore among others for their contributions to the open-source OpenCog framework. Ben Goertzel would like to thank Jeffrey Epstein for his research support. Joel Pitt would like to thank the Singularity Institute for Artificial Intelligence for their research funding.

References

Amit, D. J. (1989). *Modeling brain function – the world of attractor neural networks*. Cambridge University Press, New York, USA.

Anderson, J. R. (1996). *The Architecture of Cognition*. Lawrence Erlbaum Associates.

Bransford, J. D. & Franks, J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology* **2**, 331–350.

Calvin, W. (1996). *The Cerebral Code*. MIT Press.

Damasio, A. (2000). *The Feeling of What Happens*. Harvest Books.

d'Avila Garcez, A. S., Lamb, L. C. & Gabbay, D. M. (2008). *Neural-Symbolic Cognitive Reasoning. Cognitive Technologies.*. Springer.

Elman, J. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning* **7**, 195–226.

Freeman, W. J. & Schneider, W. (1982). Changes in spatial patterns of rabbit olfactory eeg with conditioning to odors. *Psychophysiology* **19**, 44–56.

Goertzel, B. (1997). *From Complexity to Creativity*. Plenum Press.

Goertzel, B. (2001). *Creating Internet Intelligence*. Plenum Press.

Goertzel, B. (2006). Virtual easter egg hunting: A thought-experiment in embodied social learning, cognitive process integration, and the dynamic emergence of the self. *AGI-06*.

Goertzel, B. (2008*a*). Glocal memory: A new perspective on knowledge representation, neurodynamics, distributed cognition, and the nature of mind. *Dynamical Psychology*.

Goertzel, B. (2008*b*). *OpenCog Prime: Design for a Thinking Machine*. Online wikibook at http://opencog.org.

Goertzel, B., Iklé, M., Goertzel, I. & Heljakka, A. (2009). *Probabilistic Logic Networks: A Comprehensive Framework for Uncertain Inference*. Springer.

Hammer, B. & Hitzler, P., eds (2007). *Perspectives of Neural-Symbolic Integration. Studies in Computational Intelligence, Vol. 77.*. Springer.

Hinton, G. E., Osindero, S. & Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation* **18**, 1527–1554.

Hofstadter, D. (1979). *GÃűdel, Escher, Bach: an Eternal Golden Braid*. Basic Books.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* **79**, 2554–2558.

Lebiere, C. & Anderson, J. R. ((in preparation)). The case for a hybrid architecture of cognition.

Lenat, D. & Guha, R. V. (1990). *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Addison-Wesley.

Patterson, K., Nestor, P. J. & Rogers, T. T. (2007). Where do you know what you know? the representation of semantic knowledge in the human brain.. *Nature Reviews Neuroscience* **8**, 976–987.

Quiroga, R. Q., amd G. Kreiman, L. R., Koch, C. & Fried, I. (2005). Invariant visual representation by single-neurons in the human brain.. *Nature* **435**, 1102–1107.

Quiroga, R. Q., Kreiman, G., Koch, C. & Fried, I. (2008). Sparse but not "grandmother-cell" coding in the medial temporal lobe.. *Trends in Cognitive Sciences* **12**, 87–91.

Roediger, H. L. & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **21**, 803–814.

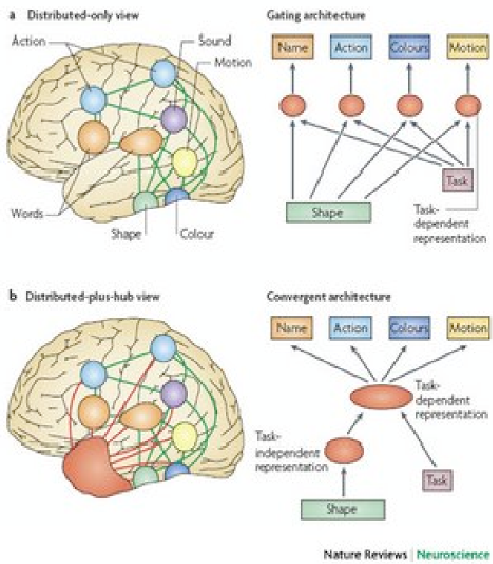
Rosenfield, I. (1988). *The Invention of Memory: A New View of the Brain.* Basic Books.

Santore, J. F. & Shapiro, S. C. (2003). Crystal cassie: Use of a 3-d gaming environment for a cognitive agent. *in* 'Papers of the IJCAI 2003 Workshop on Cognitive Modeling of Agents and Multi-Agent Interactions'.

Storkey, A. J. & Valabregue, R. (1999). The basins of attraction of a new hopfield learning rule. *Neural Networks* **12**, 869–876.

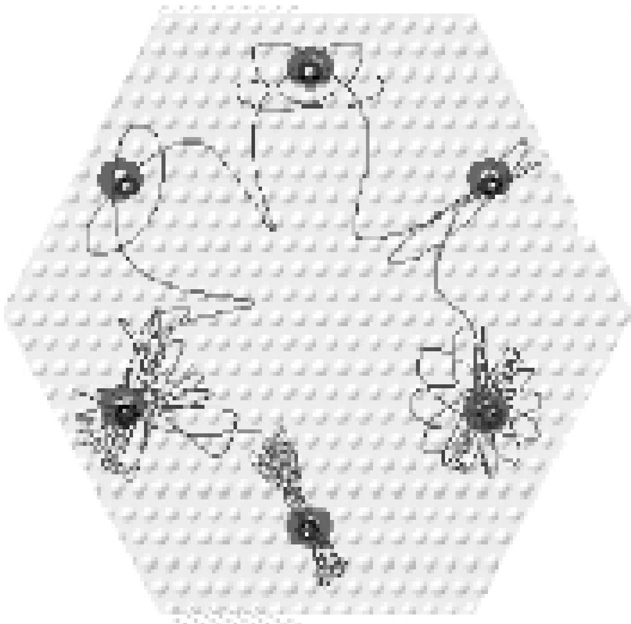
A Figures

A.1 Figure 1



The top image shows a more conventional "distributed-only" view in which the neural patterns corresponding to the concept are spread across multiple sensory and motor areas, and binding occurs in a purely distributed and self-organizing manner, roughly similar to a Hopfield net. The bottom panel shows an alternative, glocally-flavored "distributed plus hub" view, which involves the same distributed regions coordinated by a single "hub" within the ATL. Figure drawn from Patterson et al. (2007).

A.2 Figure 2



From William Calvin's "Cerebral Code", illustrating a posited mechanism via which attractors in networks of hexagonal cortical metacolumns may be communicated from one part of the brain to another. Glocality exists here because memories are posited to be represented by multi-lobed strange attractors, and it is suggested that communication of a small portion of an attractor suffices to "seed" the approximate replication of the full attractor in the target brain region.

A.3 Figure 3

Create filter matrix F to determine which links should be established initially;

Create WeightMatrix w (use adjacency list to present the partial connected network);

for *each Pattern p* **do**

  Calculate h(i,j)= $\sum_{k=1,k\neq i,j}^{n} w_{\mathrm{ik}} p_k$;

  for *i=0 to n-1, j=0 to n-1* **do**

    $\Delta w_{i,j} = (p_i \times p_j - h_{i,j} \times p_j - h_{j,i} \times p_i) F_{i,j}$;

  end

  Double min=MAX_VALUE;

  Double inaccuracy;

  for *i=0 to n-1* **do**

    Inaccuracy=0;

    inaccuracy= $\sum_{j=1}^{n} |h_{i,j} \times p_j| + |h_{j,i} \times p_i|$;

    if *inaccuracy<min* **then**

      key=i;

      min=inaccuracy;

    end

  end

  for *i=0 to n-1* **do**

    Set $\Delta w_{i,\mathrm{key}} = p_i \times p_{\mathrm{key}} - h_{i,\mathrm{key}} \times p_{\mathrm{key}} - h_{\mathrm{key},i} \times p_i$;

    Set $\Delta w_{\mathrm{key},i} = p_{\mathrm{key}} \times p_i - h_{\mathrm{key},i} \times p_i - h_{i,\mathrm{key}} \times p_{\mathrm{key}}$;

  end

  w=w+Δw;

end

Algorithm 1: Pseudocode for training glocal palimpsest Hopfield net learning algorithm

A.4 Figure 4a

Memories learned by the Hopfield network.

```
{
       1111        ,
       1111        ,
     11111111      ,
     11111111      ,
   1111    1111    ,
   1111    1111    ,
   1111    1111    ,
 1111        1111  ,
 1111111111111111  ,
 1111        1111  ,
1111           1111,
1111           1111
},
{
11111111111111     ,
1111       1111    ,
1111       1111    ,
1111       1111    ,
1111     1111      ,
11111111111111     ,
1111       1111    ,
1111         1111  ,
1111         1111  ,
1111         1111  ,
1111       1111    ,
11111111111111
},
{
     1111111111      ,
   1111      1111    ,
 1111          1111  ,
1111                 ,
1111                 ,
1111                 ,
1111                 ,
1111                 ,
1111                 ,
 1111          1111  ,
   1111      1111    ,
     1111111111
},
{
   1111              ,
   1111              ,
   1111              ,
   1111              ,
   1111              ,
   1111              ,
   1111              ,
   1111              ,
   1111              ,
   1111              ,
   1111              ,
   1111
},
{
        1111         ,
        1111         ,
        1111         ,
        1111         ,
        1111         ,
        1111         ,
        1111         ,
        1111         ,
        1111         ,
 1111   1111         ,
 1111   1111         ,
  11111111
},
{
111111111111         ,
1111     1111        ,
1111       1111      ,
1111         1111    ,
1111         1111    ,
1111         1111    ,
1111         1111    ,
1111         1111    ,
1111         1111    ,
1111       1111      ,
1111     1111        ,
111111111111
},
{
     1111111111      ,
   1111      1111    ,
 1111          1111  ,
1111           1111  ,
1111                 ,
1111                 ,
1111     1111111111  ,
1111           1111  ,
1111           1111  ,
 1111        111111  ,
   1111    11111111  ,
     11111111  1111
},
{
1111         1111    ,
1111         1111    ,
1111         1111    ,
1111         1111    ,
1111         1111    ,
11111111111111111    ,
1111         1111    ,
1111         1111    ,
1111         1111    ,
1111         1111    ,
1111         1111    ,
1111         1111
},
{
111111111111111111   ,
1111                 ,
1111                 ,
1111                 ,
1111                 ,
1111111111111111     ,
1111                 ,
1111                 ,
1111                 ,
1111                 ,
1111                 ,
111111111111111111
},
```

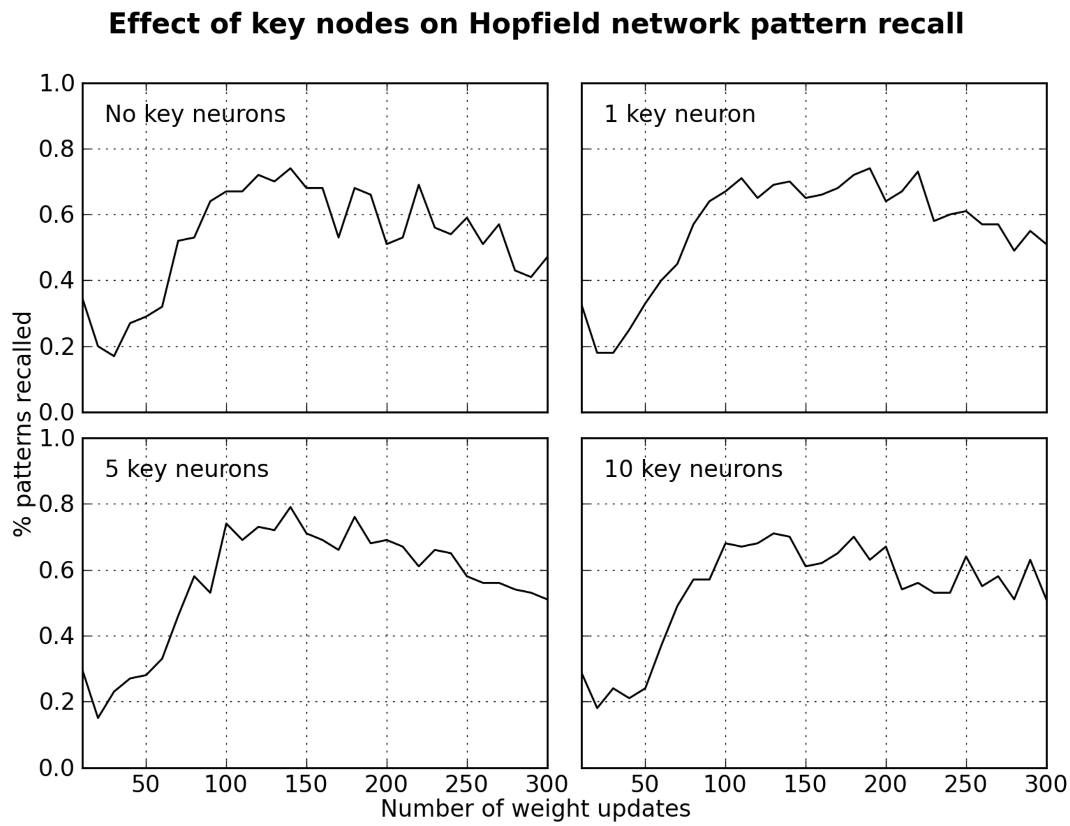
```
{
 111111111111111111   ,
 1111                 ,
 1111                 ,
 1111                 ,
 1111                 ,
 1111111111111111     ,
 1111                 ,
 1111                 ,
 1111                 ,
 1111                 ,
 1111                 ,
 1111
}
```

A.5 Figure 4b

Example test patterns used to probe Hopfield network retrieval behavior.

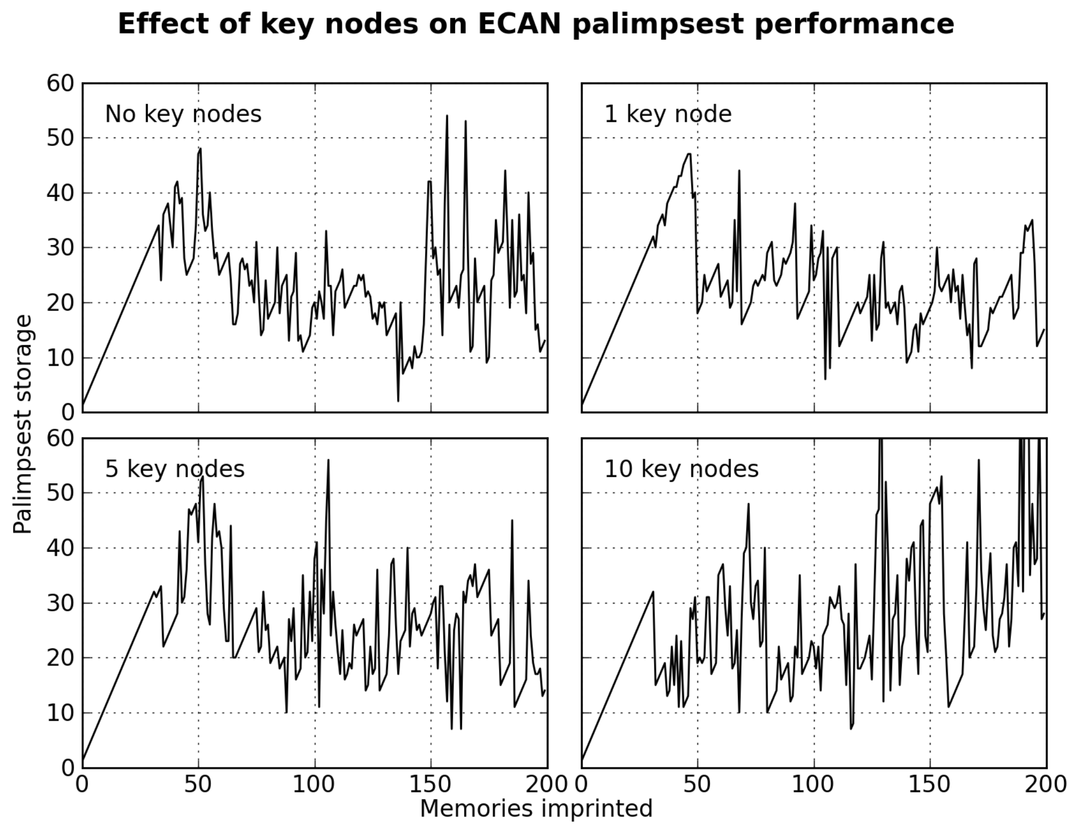
```
{
       1111          ,
       1111          ,
     11111111        ,
    111111111        ,
   1111    1111      ,
   1111    1111      ,
   1111    11111     ,
 1111        111     ,
 1111111111111111    ,
 11111      111111   ,
11111          1111  ,
1111           1111
},
{
1111111111111111     ,
1111       11111     ,
1111       11111     ,
1111       1111      ,
1111     1111        ,
11111111111111       ,
1111       1111      ,
1111        11111    ,
1111         11111   ,
1111         1111    ,
1111       1111      ,
11111111111111
},
{
     1111111111      ,
   1111      1111    ,
 1111         11111  ,
1111              1  ,
1111                 ,
1111                 ,
1111                 ,
1111                 ,
 1111             1  ,
  1111         1111  ,
   1111      1111    ,
     1111111111
},
{
 11111111            ,
 11111111            ,
   1111              ,
   1111              ,
   1111              ,
   1111              ,
   1111              ,
   1111              ,
   1111              ,
   1111              ,
 11111111            ,
 11111111
},
{
      11111111       ,
      11111111       ,
        1111         ,
        1111         ,
        1111         ,
        1111         ,
        1111         ,
        1111         ,
 11     1111         ,
 1111   1111         ,
 11111 11111         ,
  11111111
}
```

A.6 Figure 5



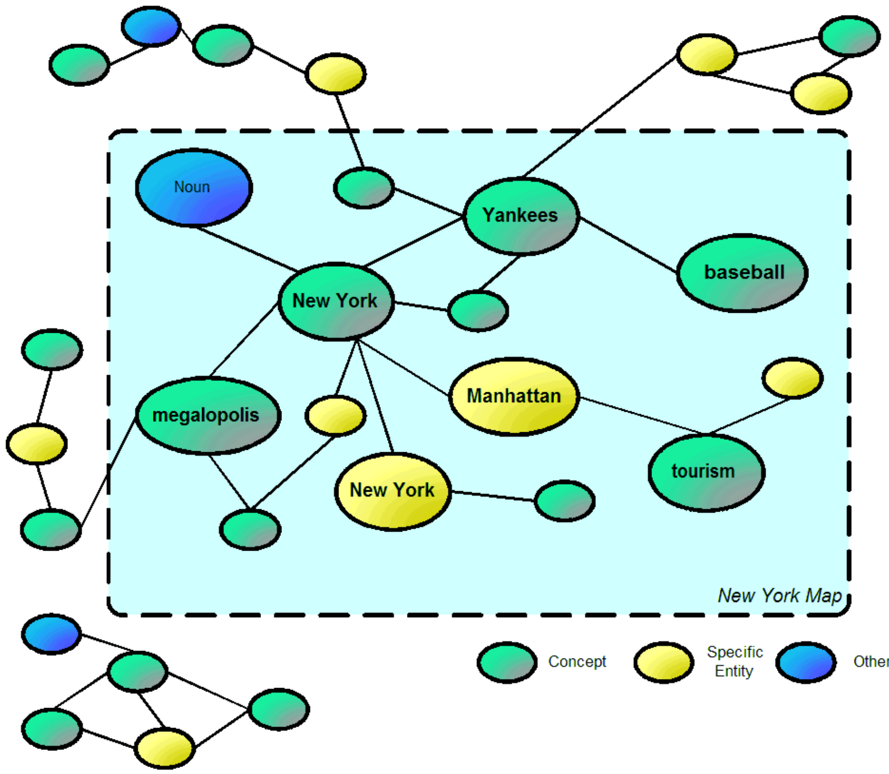
The memory recall for a modified Hopfield network, using Storkey & Valabregue's (1999) update rule and comparing the effect of including key nodes to facilitate the storage of "glocal" memories.

A.7 Figure 6



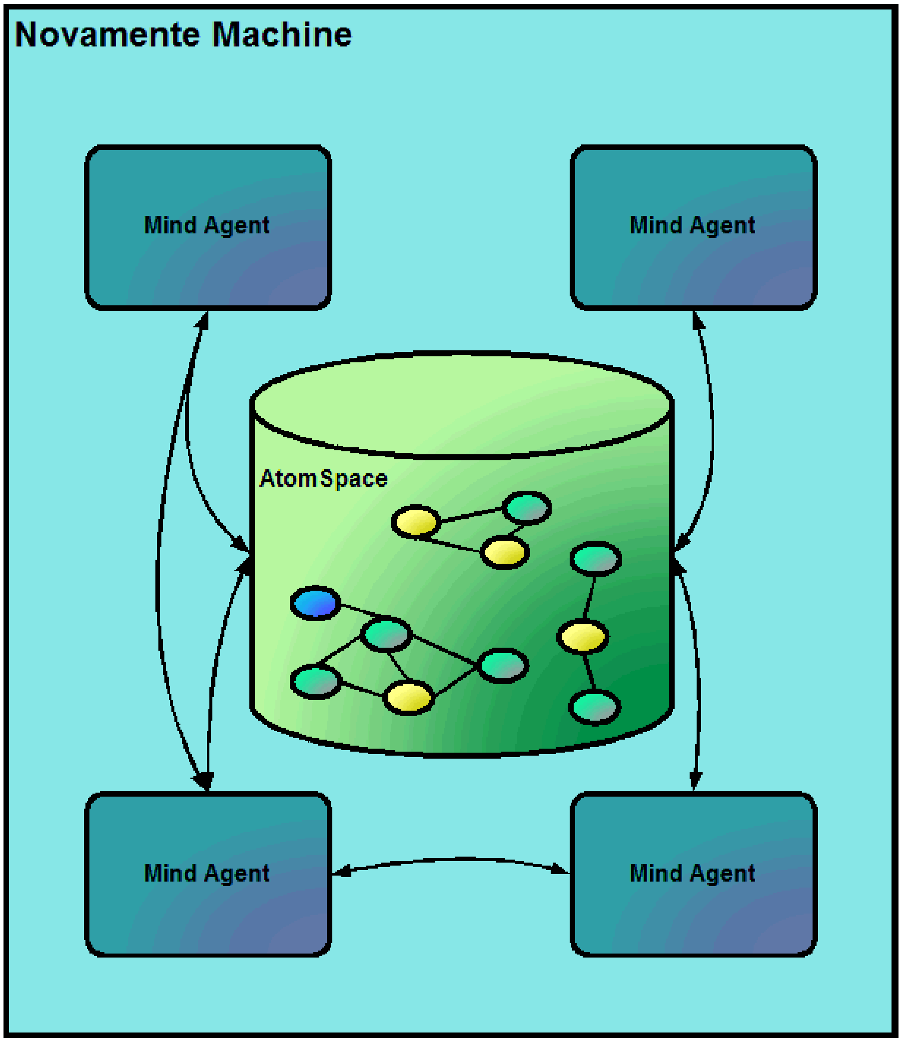
Using a palimpsest testing scheme applied to a Hopfield network equivalent using ECAN dynamics within the OpenCog framework. The network has a link density of 0.3, with 1% of links forgotten before each new pattern is imprinted. Although the palimpsest storage from one pattern to the next is noisy, the average palimpsest storage is greater when key nodes are included.

A.8 Figure 7



An example hypergraph, demonstrating different kinds of nodes and links, and a map, or global attractor, containing nodes and links related to the concept of New York.

A.9 Figure 8



The composition of a Novamente "unit", consisting of a number of MindAgents that all access and update an AtomSpace which allows acces to the underlying hypergraph.