

The Hippocampus According to the Ouroboros Model, the "Expanding Memory Index Hypothesis"

Knud Thomsen PSI

Paul Scherrer Institut

NUM Forschung mit Neutronen und Muonen

CH-5232 Villigen PSI, Switzerland

e-mail: knud.thomsen@psi.ch

Abstract—Understanding information processing in the brain demands more than one approach or method. Combining well-established and diverse novel experimental findings from neuroscience, connectionist approaches, and artificial intelligence, the perspective of the recently proposed cognitive architecture of the Ouroboros Model offers an integrative view on the functional role of the mammalian hippocampus: it is hypothesized to implement a rapidly laid down index, first establishing, contributing to binding, and then retrieving together memory entries, thus iteratively expanding the overall cognitive system in an autocatalytic process.

Keywords - Cognition; schema; iterative processing; discrepancy monitoring; hypercycle; hash table; extreme learning machine.

I. INTRODUCTION

The hippocampus is a central neural structure on top of the processing hierarchy in mammalian brains. Homologous structures have been found in reptiles and birds. Widely and reciprocally connected to cortical areas, in mammals the hippocampus is comprised of several distinct main regions: Dentate Gyrus (DG), CA3, and CA1; see Fig. 1. These appear to implement successive stages of processing: DG as the input stage has been implicated in pattern separation whereas CA3 is best understood as a content-addressable auto-association memory providing pattern completion, as explained in some detail below [1]. A hippocampal memory indexing theory has been proposed 30 years ago, and it has fared quite well over time [2][3].

The paper is structured as follows. In section II, the gist of the Ouroboros Model is presented in just a few words, and in section III, a coarse conceptual sketch of an expanded dynamic indexing view is outlined. Conclusions and future work are indicated in Section IV.

II. THE OUROBOROS MODEL IN TWO WORDS

The Ouroboros Model offers a cognitive architecture aiming at an encompassing account [4]. Cognition in general is explained as resulting from two fundamental building blocks: a memory structured into cohesive chunks called schemata, and a cyclic process termed 'consumption analysis', which "cultivates" consistency by monitoring for discrepancies and thereupon directing attention and also triggering memory storage according to demand.

A key tenet of the Ouroboros Model, in particular, is that novel memory entries are laid down as kind of "snapshots" linking all prevalent cortical activations at occasions marked as important by the outcome of consumption analysis [5]; this most efficiently includes distinct index records.

III. THE EXPANDING MEMORY INDEX HYPOTHESIS

Mapping structures of the Ouroboros Model to living brains, it has been hypothesized in various previous accounts that the hippocampus embodies an index to features in distributed representations over widely spread specialized cerebral cortex areas [2][3][6].

Arguing that there are diverse requirements on different forms of human memory it has been proposed that these can best be addressed with two memory systems: a limited fast, one-shot, component based in the hippocampus, and another with vast capacity and with some essential features slower but gradually improving, in the cerebral cortex [7].

Memorizing new episodes in the form of complete activation-images, demands fast and encompassing storage, exactly as has been described for the hippocampus [5][7]. (Almost) all activity in the entire brain is effectively bound together because the hippocampus sits on top of many diverse processing areas [8].

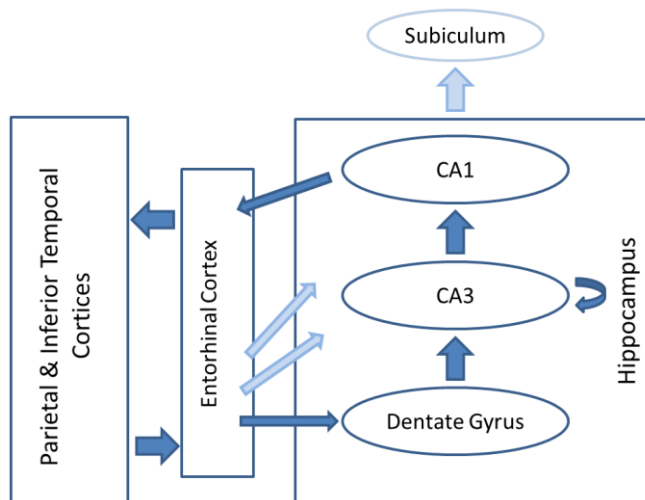


Figure 1. Principal stages for the hippocampus, with dark arrows indicating the links mentioned in the text [8].

Later, during remembering, when trying to reestablish an overall activation in the cortex, hippocampus acts as driver providing effective index-entries and bringing to bear efficient pointers to the distributed cortical representations.

The hippocampus thus can be understood as a top "convergence-divergence zone", which receives convergent projections from the sites whose activity is to be recorded, and which sends divergent projections to these same areas, similar to a proposal by Damasio [9]. The here proposed account is distinct from more standard conceptualizations of memory stressing clearly separated components, i.e., short- and long-term stores or more constricted situation-specific control processes like in the Atkinson–Shiffrin model [10].

Following the ideas of [2] and [7], memories according to the Ouroboros Model are stored in a partly redundant way: linked via the hippocampus and also, definitively after some consolidation over time (and slow wave sleep), permanently bound cortical representations. No 'transfer', in the sense of completely erasing hippocampal traces, normally takes place; cortex entries just become more independent with use.

Some type of two-pronged memory implementation appears indispensable for meta-cognition and meta-memory, which naturally enable an agent to assess her status of memory (-retrieval). Prominent examples in this context are the tip-of-the-tongue phenomenon and *déjà vu*.

The hippocampal complex is hypothesized to work as an effective hash-table; index values are established from the activity distributions they code for in a given context, and, in turn, their activation later allows addressing and combining the detailed entries spread over widely dispersed cortex areas. At both storage sites, pattern completion is effective, enabling full retrieval from only a fraction of the features, efficiently implementing content-addressable memory.

During encoding, as well as for retrieval, the two repositories reinforce each other mutually. This constellation can be described as a minimum-hypercycle comprising only two main sub-stages [11][12]. These are distributed over the cerebral cortex and, on the other end, the CA3 section of the hippocampus. (For very important memories, e.g., ones containing strong (aversive) emotions, an additional structure is regularly involved, i.e., the amygdala, which allows for particularly fast reactions in already encoded contexts.)

Entries in the hippocampus, like in local neocortex areas, are laid down in patches of orderly maps. Sufficiently different contexts have been shown to lead to remapping in the hippocampus [13]. Specificity is taken to be ensured by well-separated attractor states in CA3 due to its prominent recurrent connections, while CA1 recoding for the output can preserve matched correspondence with cortical areas; see Fig.1. Different schemata as generalizations of place-cell maps would correspond to specific hippocampal mappings. Dynamically, these associations are seen as manifest in theta oscillations, which originate mainly in the hippocampus and cingulate cortex, and phase-locked high gamma oscillations in neocortex [14].

Due to the auto-association capabilities of the CA3 region, full activations can be provoked already when only part of their constituents are available first. In order to avoid disturbing overlap between distinct memories, a random

component in the assignment from the DG, which effects very powerful input to CA3 storage, has been proposed [1].

An efficient additional way of minimizing collisions, i.e., preventing the overloading of a neural associative network, would be to add new neurons in the distributing stage as new (index) entries are required. After decades, in which the dogma "no new neurons in the adult brain" was accepted as valid, it has been found that the dentate gyrus in the hippocampal formation is one of two regions in the adult human brain where new neurons are continually developing and functionally integrated into working brain circuitry [15]. Contributions to better pattern separation and memory resolution have been suggested as main function of these added neurons [16][17]. There is ample evidence for their enhanced generation and survival when demand for new representational resources is presumably high, i.e., they are boosted by physical exercise in enriched environments, accompanied by significant improvement of previously impaired hippocampal long-term potentiation and cognitive performance in a mouse model [18].

It is important to note that index values are most useful if distinct. "Continuous" versions seem adequate within limits, but requirements for very fine-scaled representations and smooth transitions might more efficiently be taken care of by some general-purpose interpolator as has been argued to be the function of the cerebellum [19].

For the hippocampus, it is further proposed to investigate the idea how the observed addition of newborn neurons in living brains can be understood as a version of an extreme learning machine (ELM), a recent neural network lay-out featuring the random addition of hidden neurons, which yields spectacular improvements in learning rate compared to standard learning routines [20]. It has been shown that randomly introducing neurons and then selecting the ones which work best can further enhance the efficiency of ELMs compared to versions without pruning [21].

Following the Ouroboros Model, newborn neurons are inserted in living brains in a manner "better than random", i.e., their number and timing being controlled by the demand for finer differentiation in the already existing network of schemata [5][18]. This would actually fit nicely with hopscotch hashing, where additional entries are added demand-oriented for resolving hash collisions and inserted locally in the relevant neighborhood ensuring quick retrieval.

New and uncommitted neurons are most useful in the separation stage (DG) with initial high and rather localized sensitivity, and their weights constrained by neighbors, possibly not only for better discriminating and stabilizing but later also for condensing and compactifying (sparse) activity, which was before distributed more widely and implementing some type of population-code, all in tight interaction with the linked cortex representations and also inducing similar tuning, and likely also including pruning, there.

IV. CONCLUSION AND FUTURE WORK

It is claimed that the Ouroboros Model sheds some light on the requirement and implementation of a continually expanding index to representations distributed over the

cortex in mammalian brains: new entries are rapidly laid down as "snapshots" concatenating all concurrent activity upon a trigger by consumption analysis. The mammalian hippocampus is thus hypothesized to serve a fundamental role for relatively quickly establishing and also reliably retrieving distinct entries in episodic memory as well as for episodic simulation employing exactly that schema-type, which is claimed to form the ever-expanding basis for efficient cognition. Request-oriented storage of index entries in tight interplay with detailed cortical records thus efficiently implements autocatalytic learning and growth. This appears to be a significant step towards explaining how information is processed in vertebrate brains.

The Expanding Memory Index Hypothesis of the Hippocampus:

- The hippocampus provides an index to wide-spread cortical representations.
- Content-addressable at both repositories, memories can most efficiently be retrieved from partial keys.
- Entries in the hippocampus and in the cerebral cortex mutually endorse each other and thus form a hypercycle.
- Memory separation/orthogonalization capability for the unique indexing of novel episodes is greatly enhanced by adding adult-born neurons in the DG.
- This is somewhat similar to the addition of hidden units in Extreme Learning Machines (ELMs).
- All of this fits nicely with the Ouroboros Model.

A further piece of the puzzle will be to elucidate the link of the hippocampus with another central (control) structure, i.e., anterior cingulate cortex, ACC, which has been shown recently to possess direct monosynaptic connections to CA3 and CA1 regions [22]. With the hippocampus and the ACC in the center together, and including general cerebral cortex as well as subcortical structures, the neuronal basis for extensive cyclic iterative processing can be outlined.

Work on the Ouroboros Model in general and also on the role of the hippocampus and its connections is in progress. Following a fundamental self-reflective and self-consistent approach, a first schematic sketch will be filled-in with more fine-grained and quantitative details in subsequent iterations; collaborations to this end are most welcome.

ACKNOWLEDGMENT

A number of insightful comments and very helpful suggestions for improvements by several reviewers are gratefully acknowledged.

REFERENCES

[1] E. T. Rolls, "The mechanisms for pattern completion and pattern separation in the hippocampus," *Frontiers in Systems Neuroscience*, vol. 7, pp. 1-21, 2013.

[2] T. J. Teyler and P. DiScenna, "The hippocampal memory indexing theory," *Behavioral Neuroscience*, vol 100(2), pp. 147-154, 1986.

[3] T. J. Teyler and J. W. Rudy, "The hippocampal indexing theory and episodic memory: updating the index," *Hippocampus*, vol. 17, pp. 1158-1169, 2007.

[4] K. Thomsen, "The Ouroboros Model in the light of venerable criteria," *Neurocomputing*, vol. 74, pp. 121-128, 2010.

[5] K. Thomsen, "Concept formation in the Ouroboros Model," *Third Conference on Artificial General Intelligence*, Lugano, Switzerland, Proceedings of AGI 2010, pp. 194-195.

[6] K. Thomsen, "Ouroboros Model mapped to Brain," *8th IBRO World Congress of Neuroscience*, Florence, Italy, 2011.

[7] J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly, "Why there are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory," *Psychological Review* 102, pp. 419-457, 1995.

[8] T. Nakashiba, J. Z. Young, T. J. McHugh, D. L. Buhl, and S. Tonegawa, "Transgenic Inhibition of Synaptic Transmission Reveals Role of CA3 Output in Hippocampal Learning," *Science* vol. 319, 1260-1264, 2008.

[9] A. Damasio, "Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition," *Cognition* 33, pp. 25-62, 1989.

[10] R. C. Atkinson and R. M. Shiffrin, "Human memory: A proposed system and its control processes," in K. W. Spence and J. T. Spence, *The psychology of learning and motivation* (Volume 2). New York: Academic Press, pp. 89-195, 1968.

[11] M. Eigen and P. Schuster, "The Hypercycle, A Principle of Natural Self-Organization, Part A, Emergence of the Hypercycle," *Naturwissenschaften*, vol. 64, pp. 541-565, 1977.

[12] M. Eigen and P. Schuster, "The Hypercycle, A Principle of Natural Self-Organization, Part B, The Abstract Hypercycle," *Naturwissenschaften*, vol. 65, pp. 7-41, 1978.

[13] J. K. Leutgeb, S. Leutgeb, M.-B. Moser, and E. I. Moser, "Pattern Separation in the Dentate Gyrus and DCA3 of the Hippocampus," *Science*, vol. 315, pp. 961-966, 2007.

[14] R. T. Canolty et al., "High Gamma Power is Phase-Locked to Theta Oscillations in Human Neocortex," *Science*, vol. 313, pp. 1626-1628, 2006.

[15] P. S. Eriksson et al., "Neurogenesis in the adult human hippocampus," *Nature Medicine*, vol. 11, pp. 1313-1317, 1998.

[16] A. Sahay, D. A. Wilson, and R. Hen, "Pattern separation: a common function for new neurons in hippocampus and olfactory bulb," *Neuron*, vol. 70, pp. 582-588, 2011.

[17] J. B. Aimone, W. Deng, and F. H. Gage, "Resolving new memories: a critical look at the dentate gyrus, adult neurogenesis, and pattern separation," *Neuron*, vol. 70, pp. 589-596, 2011.

[18] M. E. Sakalem et al., "Environmental Enrichment and Physical Exercise Revert Behavioral and Electrophysiological Impairments Caused by Reduced Adult Neurogenesis," *Hippocampus* 27, pp. 36-51, 2017.

[19] K. Thomsen, "The Cerebellum according to the Ouroboros Model, the 'Interpolator Hypothesis,'" *Journal of Communication and Computer*, vol. 11, pp. 239-254, 2014.

[20] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, pp. 489-501, 2006.

[21] G. B. Huang and L. Chen, "Enhanced random search based incremental extreme learning machine," *Neurocomputing*, vol. 71, pp. 3460-3468, 2008.

[22] P. Rajasethupathy et al., "Projections from the neocortex mediate top-down control of memory retrieval," *Nature* 526, 653-659, 2015.