Concept Formation in the Ouroboros Model

Knud Thomsen

Paul Scherrer Institute, CH-5232 Villigen PSI, Switzerland

Abstract

According to the Ouroboros Model several occasions can be distinguished over the course of the general autonomous cyclic activity in which new concepts are established and associated memories are preferentially laid down. Whereas a rather standard habituation process can lead to the extraction of statistical regularities from repeated common (consecutive) excitation, two specific instances are peculiar to the Ouroboros Model; the consumption analysis process marks events when especially successful or the contrary. In addition, new concepts can be assembled very quickly by combining previously existing building blocks.

Relations of these theoretical considerations to supporting recent experimental findings are briefly outlined.

The Ouroboros Model in a Nutshell

The Ouroboros Model proposes a novel algorithmic architecture for efficient data processing in living brains and for artificial agents [1]. At its core lies a general repetitive loop where one iteration cycle sets the stage for the next. All concepts of an agent are claimed to be organized into hierarchical data structures called schemata, also known as frames, scripts or the like.

During perception, any sensory input activates schemata with equal or similar constituents encountered before, thus expectations for the whole unit and especially for its parts are kindled. This corresponds to the highlighting of empty slots in the selected schema as biasing anticipated features facilitates their actual excitation. These predictions are subsequently compared to the factually encountered input. Depending on the outcome of this "consumption analysis" different next steps are taken. In case of partial fit search for further data continues; a reset, i.e. a new attempt employing another schema, is triggered if the occurring discrepancies are too big. Basically the same process applies for all other actions like memory search, active movements of the agent or language production.

Self-referential monitoring of the whole process, and in particular of the flow of activation directed by the consumption analysis, yields valuable feedback for the optimum allocation of attention and resources including the selective establishment of useful new concepts.

According to the Ouroboros Model basically four different situations in which novel concepts are formed and corresponding fresh memory entries are first created and shaped can be distinguished.

Ways to Concept Formation

Two types of occasions are directly marked in the Ouroboros Model as interesting by the outcome of the consumption analysis, and preferentially for them new records are laid down:

- Events, when everything fits perfectly; i.e. associated neural representations are stored as kind of snapshots of all concurrent activity, making them available for guidance in the future as they have proved useful once.

- Constellations, which led to an impasse, are worthwhile remembering, too; in this case for future avoidance.

These new memories stand for junks, i.e. concepts, again as schemata, frames or scripts. Their building blocks include whatever representations are active at the time when the “snapshot” is taken, including sensory signals, abstractions, previously laid down concepts, and emotions. They might but need not include / correspond to a direct representation unit like a word. At later occasions they will serve for controlling behavior, by guiding action to or away from the marked tracks.

Knowledge thus forms the very basis for the data processing steps, and its meaningful expansion is a prime outcome of its use as well; the available data base of concepts / schemata is steadily enlarged and completed, especially in areas where the need for this surfaced and is felt most strongly.

Even without the strong motivation by an acute alert signal from consumption analysis novel categories and concepts are assembled on the spot:

- New concepts are built from existing structures

We can quickly establish new compound concepts, whole scenes, from previously existing building blocks, i.e. by combining (parts of) other concepts; here is an example:

Let us assume that we hear about “the lady in the fur coat”. Even without any further specification a figure is defined to a certain extent including many implicit details. Also in case we heard this expression for the first time the concept is established well enough for immediate use in a subsequent cycle of consumption analysis, expectations are effectively triggered. When we now see a woman in this context, we are surprised if she is naked on her feet (…unless she is walking on a beach). Fur coats imply warm shoes or boots, unless the wider frame already deviates from defaults.

In parallel to the above described instant establishing of concepts and the recording of at least short time episodic memory entries there exists a slower and rather independent process:

- Associations and categorizations are gradually distilled from the statistics of co-occurrences.

In the sense, that completely disadvantageous or fatal activity would not be repeated many times, also this grinding-in of associations can be understood as a result of successful or even rewarded activations.

Activity, which forms the basis of this comparatively slow process can pertain to many different representations starting from low level sensory signals to the most abstracts data structures already available, and of course, their combination.

Relation to Recent Experimental Findings

The most important ingredient in the Ouroboros Model is the repetitive and self-reflective consumption analysis process. A key conjecture derived from this algorithmic structure is the highlighting of interesting occasions and the quick recording of corresponding memories for advantageous future use. The Ouroboros Model proposes to distinguish "index-entries" as pointers to the "main text" containing more details. On the basis of a wealth of experimental data, a similar general division of work in the mammalian brain has been proposed some time ago [2]. Hippocampal structures are well suited for fast recording of distinct entries, they are thought to link memories spread widely over the cortex, where minute details are memorized on longer time scales.

Dopamine signals are widely considered to act as highlighting behaviorally important events, midbrain dopamine neurons code discrepancies between expected and actual rewards [3].

If dopamine now is the best established marker for discrepancies and if associated constellations should lead to the immediate recording of new concepts, at least of their specific index entry, one would expect that dopamine release has a profound impact on hippocampal long term potentiation, generally accepted as a decisive substrate for memories. This now is exactly what has been found just recently: dopaminergic modulation significantly increases sensitivity at hippocampal synapses [4]. In addition, temporal contrast is lost, i.e. not only consecutive activations lead to enhancement, but also activations in reverse order, which normally result in an attenuation of a connection. Thus, a memory entry is established, which connects in an encompassing snapshot all activity associated with the occurrence of a dopamine burst.

Along with the enhanced storage of "index entries", the preferential establishment of traces in the "text" occurs. In the cortex, several neuromodulator systems, in particular widespread cholinergic innervation, have been conjectured to control attention and associative learning under the control of error driven learning mechanisms [5].

The Ouroboros Model holds that in the brain often several mechanisms working to the same end are implemented in parallel.

Given the demanding boundary conditions, in particular, the stringent time constraints, for any actor in the real world, not all memories are of equal value or even sorted out to the same degree. Incompletely processed information has been claimed to be discarded off-line in living brains while sleeping and dreaming [6].

During the process of clearing the brain of not (yet) useful remainders, sleeping and dreaming might still serve to prime associative networks [7]. Unassociated but otherwise well-established information has been found to be integrated into associative networks, i.e. schema structures, after REM (rapid eye movement) sleep.

Obviously, much work is still required to establish detailed relations as suggested by the Ouroboros Model.

References

[1] K. Thomsen, "The Ouroboros Model", BICA 08, Technical Report FS-08-04, Menlo Park, California: AAAI Press, 2008.

[2] R. C. O'Reilly and J. W. Rudy, "Computational Principles of Learning in the Neocrotex and Hippocampus", Hippocampus 10, 389-397, 2000.

[3] W. Schultz, "The Reward Signal of Midbrain Dopamine Neurons", News Physiol. Sci. 14, 249-255, 1999.

[4] J-C. Zhang, P.M Lau, and G.Q Bi, "Gain in sensitivity and loss in temporal contrast of STDP by dopaminergic modulation at hippocampal synapses", PNAS 106, 13028-13033, 2009.

[5] W. M. Pauli and R. C. O'Reilly, "Attentional control of associative learning – A possible role of the cholinergic system", Brain Research 1202, 43-53, 2008.

[6] K. Thomsen, "The Ouroboros Model", Cogprints 6081, http://cogprints.org/6081/, 2008.

[7] D. J. Cai, S. A. Mednick, E. M. Harrison, J. C. Kanady, and S. C. Mednick, "REM, not incubation, improves creativity by priming associative networks", PNAS 106, 10130-10134, 2009.