

## The Cerebellum in the Ouroboros Model, the “Interpolator Hypothesis”

Knud Thomsen  
 Paul Scherrer Institut  
 PSI, NUM, ASQ  
 CH-5232 Villigen, Switzerland  
 knud.thomsen@psi.ch

**Abstract**—The Ouroboros Model offers a novel cognitive architecture with an algorithmic backbone of iterative and self-referential processing. All memory content is organized into meaningful pieces of data, chunks and schemata, which are laid down as a kind of snapshots of all activations at a relevant point in time. This entails a grainy structure of memory content. Whereas a core process of "consumption" analysis can naturally be defined taking advantage of this parcellation, it necessitates interpolation for fine nuances, especially for the representation of intermediate values during transients. It is hypothesized that, in the vertebrate brain, essential interpolation functionality is provided by the cerebellum. Findings concerning movement control as well as the involvement of the cerebellum in more abstract, cognitive, tasks can be interpreted as reflecting a function of the cerebellum as a co-processor, i.e., a general interpolator, boosting representations in cerebral cortex areas, which are reciprocally connected with cerebellar cortex areas. In this paper, it is sketched how the new "interpolator hypothesis" can explain manifold observations while embracing previous theories of the function of the cerebellum.

**Keywords** - *schemata; consumption analysis; iterative processing; grainy memory structure; interpolation; co-processor*

### I. INTRODUCTION

Strikingly diverse proposals have been made concerning the function of the cerebellum, and, so far, no consensus has been reached. Here, a novel functional account is proposed, which addresses a general need for interpolation and seems to allow to reconcile many of the previous concepts.

The paper is structured as follows. Sections II and III provide a short summary of key tenets of the Ouroboros Model. Motivation for an interpolation function and the “interpolator hypothesis” are formulated in Section IV. Hints from actual brains are presented in Section V. The relation to existing conceptions is clarified in Section VI, and one particular example supporting the interpolator hypothesis relating to motor control is described. Conclusions and future work are outlined in Section VII.

### II. THE OUROBOROS MODEL IN A NUTSHELL

In a series of recent papers, the Ouroboros Model has been introduced as a novel attempt at explaining a wide range of findings pertaining to cognition and consciousness of natural and also of artificial agents [1-5]. It has been suggested how, within a single approach centered around a principal algorithmic process on a suitably structured

memory, one can explain human cognitive performance and also formulate prescriptions of how to achieve comparable capabilities with artificial agents implemented in hard- or software, all iteratively and recursively following a similar self-steered evolutionary program.

Minds are seen as primarily data processing entities; an iterative and self-referential universal algorithmic layout on the basis of suitably stored data structures is essential [1,2].

#### A. Action and Memory Structure

The Ouroboros Model sees memory entries as effectively organized into (non-strict) hierarchies of schemata. Memory is made up of meaningful chunks, i.e., combinations of features and concepts belonging together [4]. In living brains, neural assemblies are permanently linked together when once co-activated in the right manner. Later re-activation of any one of the linked features excites the entire schema. In particular, also currently missing features are biased and thus expectations are triggered.

#### B. Principal Algorithmic Backbone

At the core of the Ouroboros Model lies a self-referential recursive process with alternating phases of data-acquisition and -evaluation. A monitor process termed 'consumption analysis' is checking how well expectations triggered at one point in time fit with successive activations; these principal stages are identified:

- ... anticipation,
- action / perception,
- evaluation,
- anticipation, ...

These steps are concatenated into a full repeating circle, and the activity continues at its former end, like the old alchemists' tail-devouring serpent called the Ouroboros. Most importantly, there is no detrimental circularity involved if the succession of the processing steps in time is well taken into account: teeth and tail of the name-giving snake belong to well distinct points in time.

#### C. Consumption Analysis

Any occurring activation, such as sensory-related, excites the associated schemata. The one with the highest activation is selected first. Other schemata, which possibly are also applicable, are inhibited and their activation suppressed. Taking the first selected schema and ensuing anticipations active at that time as reference and basis, consumption

analysis checks how successive activations fit into this activated frame structure, i.e., how well lower level input data are "consumed" by the chosen schema. Features are assigned / attributes are 'explained away' and inhibited for immediate reuse [6].

If everything fits perfectly, the process comes to a momentary partly standstill and continues with new input data. If discrepancies surface, they have a strong impact on the elicited actions that follow [2]. In case of severe mismatch, the first schema is altogether discarded and another, new, conceptual frame is tried. The actual appropriateness of a schema can vary over a wide range. In any case, consumption analysis delivers a gradual measure for the goodness of fit between expectations and actual inputs.

### III. CONCEPT FORMATION

Two special types of occasions are specifically marked in the Ouroboros Model as interesting by the outcome of the consumption analysis when attention is triggered leading to higher than baseline excitement and to stronger activations; preferentially for these cases, new entries are laid down quickly in (episodic) memory [4]:

- 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 or problem, are worthwhile remembering, too; in this case, for future avoidance.

In addition:

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

Novel categories and concepts can also be assembled on the spot by combining (parts of) existing memory entries following an external trigger [2].

Just the same as old memories, new concepts are laid down in the form of cohesive packages, immediately effective again as schemata, frames or scripts. Building blocks include whatever representations are active at the time when such a snapshot is taken, including sensory signals, abstractions, previously laid down concepts encompassing features relating to probable transients and causal structure, and also prevalent emotions and longer lasting moods. They might, in some cases, but need not correspond to direct representation units like words. At subsequent occasions, they will serve for controlling behavior, by guiding action towards or away from the marked tracks, depending on the sign of the associated emotion value (which was originally itself distilled from consumption analysis [2]).

### IV. NEED FOR INTERPOLATION

Some structure and parcellation of all memory content into well separated schemata most probably is dictated quite generally by the need to keep the total amount of stored data

manageable. It is a consequence of the generation of many entries as snapshots and ad-hoc assemblies seen from the Ouroboros Model's perspective. Even in the case when a schematic relation, e.g., representing a movement, is distilled from repeated similar activations, it is most probably laid down economically as distinct (end)points, and the complete transient in between is not stored in arbitrary detail.

While perfectly suited for a process like consumption analysis, a coarse-grained structure of memories poses a challenge when details finer than available in the form of local recordings are needed, and in particular, when smooth transients in time are demanded for whatever actions.

With a focus on time, it seems obvious that interpolation can significantly enhance cortical representation capabilities over what is possible with only unitary activity.

#### A. Various time scales

The transition between distinct and separated stepping stones, decisive for the overall coherence of activity according to the Ouroboros Model, is affected by diverse processes at different timescales and levels of detail.

Starting from extended timescales, emotions and moods ensure some coherence and continuity of perceptions and for the actions of an agent [2].

Over short to medium durations in the order of seconds, the flow of action according to the Ouroboros Model is mediated by shared constituents, i.e., common attributes and features, of thus concatenated, otherwise distinct, schemata.

Closer to the short-term limit of action, both for bodily movements and also for abstract cognitive processes, it is hypothesized that representations pertaining to intermediate values are calculated from more directly accessible neighboring reference points by means of some type of averaging and interpolation.

Remarkable, at the other end of the timescale again, given the intrinsic dynamic characteristics of neural action, generating some truly constant level is not completely trivial and appears to mandatorily require some form of averaging.

All timescales are addressed in the Ouroboros Model by schemata including explicitly dynamic features, which code for changes and transients. Their effective resolution will be enhanced by interpolation again.

#### B. The interpolator hypothesis of the cerebellum

The novel proposal here is that the cerebellum provides fine-grained values for features in between well established, separately and distinctly represented reference points, i.e., interpolations between cerebral activations, which specify directly available values in a coarser way.

The cerebellum is thus seen as a dedicated co-processor working in close interplay with the cerebral cortex, greatly expanding the total achievable resolution of representations in living brains.

### V. IMPLEMENTATION IN VERTEBRATE BRAINS

As a theory of human cognition, the Ouroboros Model at some point needs to demonstrate the correspondence of the proposed structures and processes with actual facts from real brains. This is work in progress; first proposals have been

presented in indentifying the hippocampal structures as providing an efficient rapidly established index to more extensive and detailed content like memory entries, laid down in cerebral cortex [5].

#### A. Selected hints from anatomy and established findings

First, looking only at a very coarse level, the cerebellum grew in tight lockstep with the cerebral cortex in its evolutionary trajectory in mammals, probably generally in vertebrates [7, 8]. Over long times, the cerebellum did not change in its comparatively simple internal cytoarchitectonic structure in rather diverse animals.

A very uniform layout lead early to the suspicion that the cerebellum does one and the same operation to all input arriving there. Some gross correspondence between different areas in cortex and sectors of the cerebellum is observed but the detailed organization into areas is well distinct from what is found in cerebral cortex [9, 10]. Input routes and the output tracts appear to establish links in separated closed loops between delimited patches of cerebellar cortex and distinct areas of cerebral cortex; in detail, parts of the body are not represented continuously over an extended area of the cerebellar cortex. Instead, representations are fractured into small discontinuous patches in an apparently uncorrelated manner with diverse sensory and motor areas arranged in close neighborhood [9, 11].

Mossy fiber and Climbing fiber input pathways converge on the level of single Purkinje cells [12].

Signals carried by relatively small numbers of input and also output fibers are in between expanded enormously with granule cells being by far the most numerous neurons in the brain. They are the origin of very many parallel fibers. It has been argued earlier that this stark contrast in numbers allows for (internal) very fine-grained encoding (and pattern separation) [13]. There is considerable evidence that mossy fiber input codes are preserved in synaptic responses of granule cells; this "similar coding principle" works as an ideal noise-reducing filter allowing the transmission of weak sensory inputs in a graded fashion [14]. In the following, temporal aspects shall be emphasized in particular.

Purkinje cells are triggered in a differential way by either a vast number of inputs from parallel fibers piercing their extended dendritic trees (eliciting simple spikes), or, by a single climbing fiber (sparking complex spikes). The response of synapses on Purkinje cells to input from parallel fibers is reduced if this parallel fiber activation "predicts" climbing fiber activity for this cell, i.e., if the latter arrives 50 - 200 ms later [15]. Persistent long term depression of the involved connections ensues after repeated pairings [16].

#### B. Effecting Interpolation

Interpolation by the cerebellum is hypothesized as being performed between reference points, i.e., between somehow distinct representations in the cerebral cortex, which are temporarily related, overlap to some extent or are in close timely vicinity like constituents of an overarching schema. As a prototypical example benefitting from interpolation, one can take a reaching movement including a start- and an end point as a goal, see Figure 1.

It is assumed that interpolation first is performed in a feed forward manner by determining a shortest trajectory between supporting points in a high-dimensional space defined by the activated feature representations in cerebral cortex, following a principle of least action. The interpolated values are then relayed back to the same cortex areas.

The simplest case would be the determination of a representation for some finely distinguished nuances between the endpoints of a scale for a single (dynamic) variable. When many feature dimensions are involved, derived intermediate values would lie in a hyper plane.

According to the interpolator hypothesis, in the example of a straight movement, all effected activations for starting- and end point, required muscles, expected sensory feedback as well as usual duration are taken into account; dimensions specifying an abstract goal for actually performing the movement come in addition. All considered feature dimensions are contributing according to some weight.

In any case, the resulting trajectory in this high dimensional space is hypothesized to be derived according to a suitably implemented principle of least action.

There exist uniquely outstanding points, i.e., when a given reference-point and the result of the interpolation are exactly equal; the obvious case would be a correctly reached intermediate or end point of a trajectory.

#### VI. RELATIONS TO PREVALENT CONCEPTUALIZATIONS

While not following any of the numerous existing proposals concerning the computational functions of the cerebellum (to the best knowledge of the author), it seems that the above advanced "interpolator hypothesis" fits rather well with observations and the prevailing ideas [13, 16, 17].

Very sketchy still, a core proposal of the interpolator hypothesis is that the vast number of granule cells effectively leads to an only very smoothly changing excitatory input to the numerous contacted Purkinje cells, influenced by a huge number of each very finely graded and reliably coded dimensions. Based on this input, Purkinje cells can deliver very high action-potential firing rates, which are probably effective as rather constant values in the further processing.

Special action, as expected for the exact coincidence between any preset value and an interpolation result, is then postulated to be signaled by climbing fiber input to Purkinje cells. This triggers the Purkinje cell (complex spike) and shuts it off for some time interval immediately afterwards when no addition or correction is required. The timing of the process could fit nicely if suitable anticipated values were effective at the outset; climbing fiber feedback activation via the inferior olive will arrive with a delay compared to the parallel fiber input.

Purkinje cells, in turn, provide the such calculated interpolated values via deep cerebellar nuclei to the same cortex areas which first prompted the interpolation operation.

Climbing fiber activation would thus be a confirmation signal rather than an error signal. As has been pointed out, "error" and "learning" do not make immediate sense if connections between identical features are affected [12].

Learning would be different in different phases; for system "set-up" during maturation, or, after a massive

change, pronounced long term adaptations can be expected. In contrast, during “normal operations” no large persistent modifications in the cerebellum seem to be required. This matches with the experimental fact that genetically engineered mice exhibit normal motor learning in the absence of long term depression, LTD, at the parallel fiber - Purkinje cell synapse [18]. Consequently, learning, which takes place involving cerebellar activation, might mainly be implemented in the connected cerebral cortex areas [19, 20].

For body movements, also when only imagined, distinct postures are separated by a time interval, dictated by basic physics; it is therefore clear that a failure to follow a smooth path between them can be interpreted as a timing problem. Timing has been postulated as a main function of the cerebellum [21, 22].

Disturbances in postural tone and smooth movement were historically among the first deficiencies associated with cerebellar dysfunction [23]. Complex movements have been described as being broken down into components.

Figure 1 depicts a comparison between healthy control subjects (traces a) and c)) and patients with cerebellar damage (traces b) and d)). Moving, e.g., a finger from one point in space (A) to a prescribed goal (B), a healthy subject draws a straight line, whereas patients with a lesioned cerebellum produce wiggly trajectories (A' to B'), bearing witness to struggling for control and fine tuning. Comparing traces c) and d), latency after a go-signal in one and the same patient is higher for the impaired limb, in which the onset of action is reported as coincident with the reaching of the first hold point for the faultless movement. Here, especially the second hint is interesting; this illustration, based on [23] and [24], can be interpreted as an action with the impaired limb only starting at the instant when consumption analysis detects a discrepancy, i.e., a deviation from a set goal or reference. The figure is based on the work of G. Holmes in the 1920s, and it would be very interesting to scrutinize these old findings [23, 24].

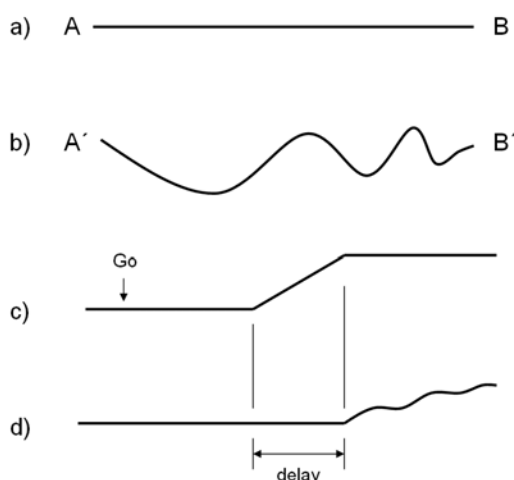


Figure 1. Sketch of principle findings contrasting healthy subjects (traces a) and c)) with ones suffering from cerebellar damage (traces b) and d)).

According to the Ouroboros Model, for any deviation between feedback and expectation, consumption analysis triggers attention and also delivers some affective signal (computed first in cingulate cortex areas). If the cerebellum is involved in calculating required intermediate values it is no wonder that it is activated at times when deviations are detected.

Failures to provide fitting interpolations would certainly provoke errors. In the realm of abstract cognition, the corresponding errors would most probably manifest as “dysmetria of thought” [25].

More cerebellar activity is probably linked to cases in which wider and more daring interpolations are needed and thus there is a higher risk to make an error. It can naturally be hypothesized that cerebellar activation increases with escalating requirements for interpolation.

## VII. CONCLUSION AND FUTURE WORK

Extensive detailed work is still needed. In particular, formalization and numerical simulations are required to illuminate details concerning timing and how to implement interpolation in a biologically plausible manner, most probably adhering to a principle of least action, with neural networks as found in the cerebellum.

At present time, it can be stated, that the tight interweaving of computational and neural perspectives appears to offer a fresh look and a new promising approach.

For functional activation studies, predictions can be made concerning differences and similarities of cerebellar contributions to movements, e.g., comparing the drawing of a complex figure either by hand and with a pencil or with a big brush and using wide arm movements: timing and also the involvement of the body would differ vastly but in the proper reference frame the abstract specifications for supporting points and their smooth interpolation, i.e., transients, should be rather similar.

The interpolator hypothesis proposes a new, coarse-grained and overarching picture. Preliminary evidence is presented that the cerebellum serves useful and deemed necessary functions as an interpolator for deriving fine-grained representations from distinct supporting points defined by activations in cerebral cortex and effectively referring to different points in time. This hypothesis, which is motivated by the Ouroboros Model, is testable, and it appears to comply with rather general considerations. At first sight, the interpolator hypothesis not only seems able to explain available observations but also to reconcile several diverse approaches and distinct earlier proposals documented in the literature.

## ACKNOWLEDGMENTS

A number of insightful comments and very helpful suggestions for improvements by six unknown referees and the editor are gratefully acknowledged.

## REFERENCES

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

- [2] K. Thomsen, "The Ouroboros Model, Selected Facets" in: C. Hernández et al. (eds.) *From Brains to Systems*. New York Dordrecht Heidelberg London: Springer, 239-250, 2011.
- [3] K. Thomsen, "Consciousness for the Ouroboros Model", *Journal for Machine Consciousness* 3, 163-175, 2011.
- [4] K. Thomsen, "Concept Formation in the Ouroboros Model", *Third Conference on Artificial General Intelligence, AGI 2010*, Lugano, Switzerland, March 5-8, 2010.
- [5] K. Thomsen, "Ouroboros Model Mapped to Brain", 8<sup>th</sup> IBRO World Congress of Neuroscience, Florence, 14-18 July, 2011.
- [6] A. Yuille and D. Kersten, "Vision as Bayesian inference: analysis by synthesis?" *Trends in Cognitive Science* 10, 301-308, 2006.
- [7] D.A. Clark, P.P. Mitra, and S.S.H. Wang, "Scalable architecture in mammalian brains", *Nature* 411, 189-193, 2001.
- [8] S. Herculano-Houzel, "Coordinated scaling of cortical and cerebellar number of neurons", *Frontiers in Neuroanatomy* 4, 1-8, 2010.
- [9] E. Manni and L. Petrosini, "A century of cerebellar somatotopy: a debated representation", *Nature Reviews Neuroscience* 5, 241-249, 2004.
- [10] P.L. Strick, R.P. Dum, and J.A. Fiez, "Cerebellum and Nonmotor Function", *Annu. Rev. Neurosci.* 32, 413-434, 2009.
- [11] T. Wiestler, D.J. McGonicle, and J. Diedrichsen, "Integration of sensory and motor representations of single fingers in the human cerebellum", *J. Neurophysiol* 105, 3042-3053, 2011.
- [12] I.E. Brown and J.M. Bower, "Congruence of Mossy Fiber and Climbing Fiber Tactile Projections in the Lateral Hemispheres of the Rat Cerebellum", *The Journal of Comparative Neurology* 70, 429-459, 2001.
- [13] D. Marr, "A Theory of Cerebellar Cortex", *J. Physiol.* 202, 437-470. 1969.
- [14] B. Bengtsson and H. Jörntell, "Sensory transmission in cerebellar granulae cells relies on similarly coded mossy fiber inputs", *Proc. Nat. Ac. Sc.* 106, 2389-2394, 2009.
- [15] S.S.H. Wang, W. Denk, and M. Häusser, "Coincidence detection in single dendritic spines mediated by calcium release", *Nature Neuroscience* 3, 1266-1273, 2000.
- [16] M. Ito, "Cerebellar Long-Term Depression: Characterization, Signal Transduction, and Functional Roles", *Physiological Reviews* 81, 1143-1195, 2001.
- [17] J.S. Albus, (1971). "A theory of cerebellar function". *Math. Biosciences* 10 (1-2): 25-61.
- [18] M. Yuzaki, "Cerebellar LTD vs. motor learning – Lessons learned from studying GluD2", *Neural Networks*, PMID: 22840919, 2012.
- [19] V. Tomassini et al., "Structural and Functional Bases for Individual Differences in Motor Learning", *Human Brain Mapping* 32, 494-508, 2011.
- [20] R. D. Seidler, A. Prushotham, S.G. Kim, K. Uğurbil, D. Willingham, and J. Ashe, "Cerebellum Activation Associated with Performance Change but Not Motor Learning", *Science* 296, 2043-2046, 2002.
- [21] R.B. Ivry, R.M. Spencer, H.N. Zelaznik, and J. Diedrichsen, "The Cerebellum and Event Timing", *Ann. N.Y. Acad. Sci.* 978, 302-317, 2002.
- [22] M. Manto et al. "Consensus Paper: Roles of the Cerebellum in Motor Control – The Diversity of Ideas on Cerebellar Involvement in Movement", *Cerebellum* 1, 457-487, 2012.
- [23] C. Ghez and W.T. Thach, "The Cerebellum" in E.R. Kandel, J.H. Schwartz, T.M. Jessell, S.A. Siegelbaum, and A.J. Hudspeth, Eds., "Principles of Neural Science", 5<sup>th</sup> Edition, McGraw-Hill, 2013.
- [24] C. Ghez, private communication, 2013.
- [25] J.D. Schmahmann, "Disorders of the Cerebellum: Ataxia, Dysmetria of Thought, and the Cerebellar Cognitive Affective Syndrome", *J. Neuropsychiatry* 16, 367-378, 2004.