



HAL
open science

Higher Cognitive Functions in Bio-Inspired Artificial Intelligence

Frédéric Alexandre, Xavier Hinaut, Nicolas P. Rougier, Thierry Viéville

► **To cite this version:**

Frédéric Alexandre, Xavier Hinaut, Nicolas P. Rougier, Thierry Viéville. Higher Cognitive Functions in Bio-Inspired Artificial Intelligence. ERCIM News, 2021, Special topic "Brain inspired computing", 125. ⟨hal-03189215⟩

HAL Id: hal-03189215

<https://inria.hal.science/hal-03189215v1>

Submitted on 2 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Higher Cognitive Functions in Bio-Inspired Artificial Intelligence

by Frédéric Alexandre, Xavier Hinaut, Nicolas Rougier and Thierry Viéville (Inria, France)

Major algorithms from artificial intelligence (AI) lack higher cognitive functions such as problem solving and reasoning. By studying how these functions operate in the brain, we can develop a biologically informed cognitive computing; transferring our knowledge about architectural and learning principles in the brain to AI.

Digital techniques in artificial intelligence (AI) have been making enormous progress and offer impressive performance for the cognitive functions they model. Deep learning has been primarily developed for pattern matching, and extensions like Long Short Term Memory (LSTM) networks can identify and predict temporal sequences. Adaptations to other domains, such as deep reinforcement learning, allow complex strategies of decision-making to be learnt to optimise cumulated rewards.

Contrary to human performance, these techniques require large training corpora and long training times. These challenges can be partly addressed by increasing data storage and computing power capacities. But another major flaw is less often mentioned: all the underlying cognitive functions are rather elementary and do not correspond to what is generally considered as human intelligence. The questions are thus: how is it possible to implement prospective reasoning and planning with these techniques? What about problem solving and creativity? Together with neuroscientists, we are studying the principles of brain organisation and memory network interactions (Figure 1). The main goal of our work, which takes place in the Mnemosyne Inria lab at the Bordeaux Neurocampus, is to transfer major experimental findings to new models of AI, capable of such higher cognitive functions.

We started by modelling the loops between the frontal cortex and the basal ganglia, known to play a major role in decision-making and in the acquisition of skills [1]. This has led us to propose that a major distinction should be made between two kinds of loops. Loops involving agranular areas of the frontal cortex, which have been well studied in rodents, are responsible for learning sensorimotor skills and stimulus-driven decision-making. These loops can be related to elementary cognitive functions described above, providing immediate responses to external cues, and internal reinforcement. In contrast, granular frontal regions present in primates, which have not been so well studied, are generally associated with what is called meta-cognition. Here, the same learning principles are applied to the history of performance of the elementary functions to decide which of these elementary functions are triggered, inhibited or updated, in a given context. This process yields contextual flexibility and rapid updating, as observed in higher cognitive functions.

From a modelling point of view, this organisation of architecturally similar loops is very interesting because it implies that similar computing principles (implemented in the circuitry of cortico-basal loops) are exploited on different kinds of information to implement different kinds of behaviour (reflexive and reflective). We are, for example, investigating the possibility of basing decisions on the mechanisms of working memory [2], which represents a history of past activity, instead of decisions being made based on the activity itself. Likewise, decisions between different strategies can be made from levels of confidence estimated from previous experiences, instead of simply deciding from actual rewards. Still relying on architectural similarities, future work could be dedicated to the processing of temporal sequences: Projections of agranular frontal areas to the parietal cortex implement the temporal chaining of sensorimotor skills and homologous projections from the granular frontal areas are reported to play a role in sequential subgoal decomposition.

In addition, studying the projections from the hippocampus to the cortico-basal loops has given us important insights that can help us address another major problem related to prohibitive computational times for learning skills and decision criteria. The hippocampus is reported to be the place of episodic learning, storing past episodes and learning to replay them to train semantic and procedural memory. In neuroscience, this has been shown to be a major way to accelerate learning for the latter kinds of memory. We are currently working on computational models in this area, which are adapted to both improve learning time and decrease the size of the training corpora needed. In addition, the hippocampus is known to perform rapid and arbitrary binding of information. It has been proposed that the hippocampus in association with the frontal cortex can build prospective memory, also called imagination, which is fundamental for planning in the future. This is another area that we intend to study.

Here we have sketched a brief overview of the cerebral architecture responsible for higher cognitive functions and have explained that knowledge in neuroscience is already available to transfer computing principles to AI. Interestingly, this cerebral architecture is tightly associated with brain regions responsible for elementary cognitive functions related to sensorimotor processing and reinforcement learning. This research program can consequently be seen as an extension of existing models in AI. A more general reference to the nervous system [3] provides the anchoring of these models in the perceptual, bodily, emotional and motivational dimensions of behaviour, thus providing direct links for the robotic implementation of autonomous agents, which we are also investigating. On a different note, accessing higher cognitive functions also corresponds to manipulating structured knowledge representations and mechanisms of abstract reasoning, which are classically studied in symbolic AI. This opens unique perspectives into the study of the emergence of meaning and internal thought, which apply to the domains of natural language processing and educational science.

Links:

<http://www.imn-bordeaux.org/en/teams/mnemosyne-mnemonic-synergy/>

References:

- [1]: Boraud, T., Leblois, A., & Rougier, N. P. (2018). A natural history of skills. *Progress in Neurobiology*, 171, 114-124. <https://doi.org/10.1016/j.pneurobio.2018.08.003>
- [2]: Strock, A., Hinaut, X., & Rougier, N. P. (2019). A Robust Model of Gated Working Memory. *Neural Computation*, 32(1), 153-181. https://doi.org/10.1162/neco_a_01249
- [3]: Alexandre, F. A global framework for a systemic view of brain modeling. *Brain Inf.* 8, 3 (2021). <https://doi.org/10.1186/s40708-021-00126-4>

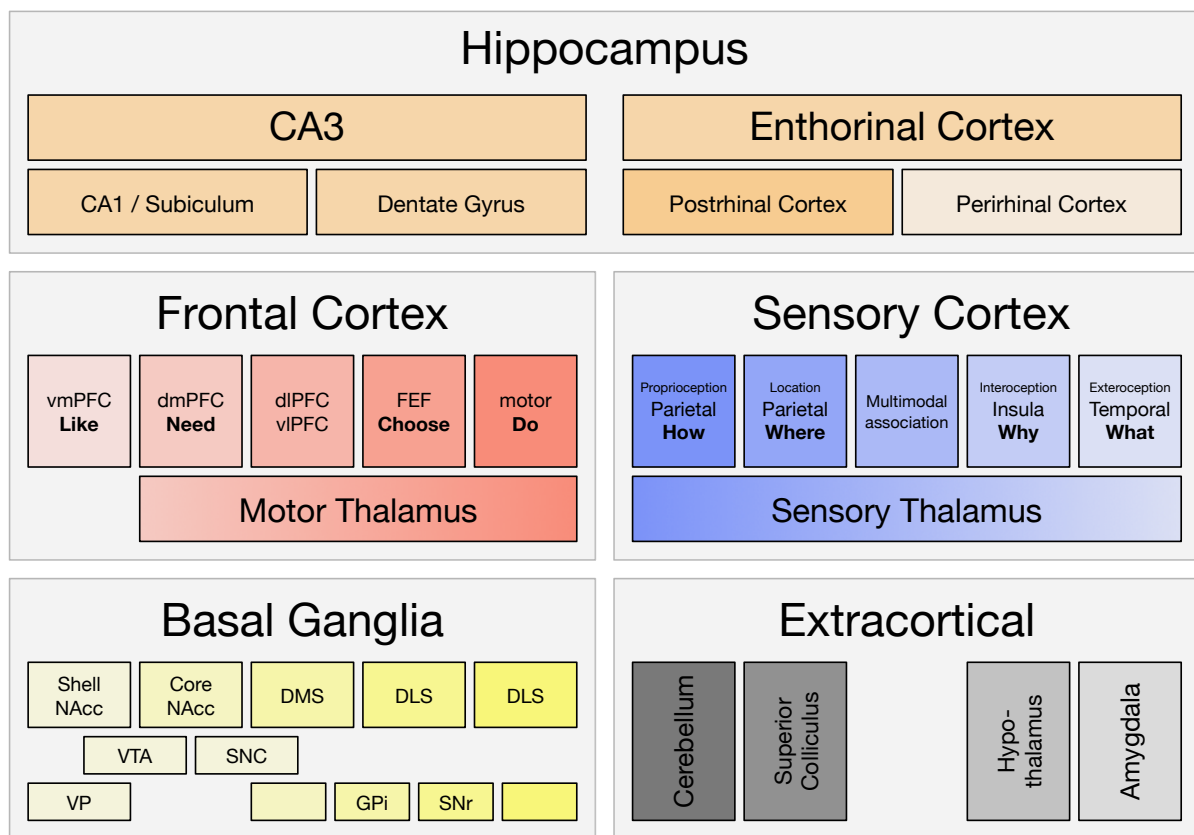


Figure 1:

This global architecture of the brain can be considered to be made of five cortico-basal loops and their main afferents from hippocampus and other extra-cortical structures. The three most central loops, which implement higher cognitive functions, include granular frontal areas.