Creativity explained by Computational Cognitive Neuroscience

Frédéric Alexandre

Creativity explained by Computational Cognitive Neuroscience

Frederic Alexandre
Inria Bordeaux Sud-Ouest; Labri UMR 5800; IMN UMR 5293
146 rue Leo Saignat; 33076 Bordeaux, France
Frederic.Alexandre@inria.fr

Abstract

Recently, models in Computational Cognitive Neuroscience (CCN) have gained a renewed interest because they could help analyze current limitations in Artificial Intelligence (AI) and propose operational ways to address them. These limitations are related to difficulties in giving a semantic grounding to manipulated concepts, in coping with high dimensionality and in managing uncertainty. In this paper, we describe the main principles and mechanisms of these models and explain that they can be directly transferred to Computational Creativity (CC), to propose operational mechanisms but also a better understanding of what creativity is.

Introduction

Artificial Intelligence (AI) has developed approaches based on knowledge manipulation and others based on data processing. The latter ones have made tremendous progresses recently, mostly due to technological improvements. Nevertheless, the development of AI remains constrained by a series of limitations, which are present from the birth of this domain and still unsolved. This is also the case for CC, seen as a subfield of AI (Colton and Wiggins 2012).

The three limitations of Artificial Intelligence

These limitations are mostly methodological, which means that they do not prevent the design of projects of AI or CC but they render difficult their development up to a realistic size and they miss a realistic description of certain characteristics, particularly when targeting their autonomy.

Giving a semantic grounding Probably, one of the most fundamental limitations of AI is the difficulty to make sense and exploit the meaning of the objects they are manipulating. This limitation is also important in CC where a novel point of view can come from such a semantic analysis. This is often related to the incapacity to ground these objects in the real world (Harnard 1990) and to propose solutions related to the intrinsic meaning of the objects in the task. This problem is generally addressed by developping ontologies of the task, but these ontologies are often built explicitly from knowledge, whereas in humans, intrinsic meanings are generally acquired implicitly and difficult to express explicitly.

Coping with high dimensionality Recently, the increase of computational resources allowed AI to explore tasks in high dimensional spaces but this progress remains negligible as compared to the virtually infinite size of state spaces in realistic tasks. Similarly, (Colton and Wiggins 2012) underline the growing tendency in CC to exploit web and other large resources. A strategy is to use heuristics and to explore a limited part of the state space. A side effect of long lasting exploration in high dimensional state spaces is that learning requires many examples and long training times, which is not realistic, as compared to the ability of animals and especially of humans to learn quickly with few examples.

Managing uncertainty Tasks in the real world can be stochastic, marring observations with noise; they can be volatile, implying that a relation learned one day can suddenly be no longer valid. A related important problem, when an intelligent system receives an error signal, is to decide if this is due to stochasticity (in the case, the best strategy is to insist) or to volatility (requiring to look for a new response). This is related to the balance between exploration and exploitation, questioned in many models in AI. Exploration can be purely random, but this is not consistent with observation of a certain degree of flexibility in changing environments, where the appropriate behavior is immediately reinstated each time the same context is revisited. Choosing between exploration and exploitation and going beyond random processes are also important issues in CC. All these problems are generally tackled in various probabilistic frameworks (like bayesian techniques). Some of them propose solid and mathematically grounded methods but they are computationally prohibitive, thus adding to the problem of high dimensionality defined in the previous item.

The CCN approach

CCN models are primarily developed as a way to operationalize proposed cognitive concepts, together with their brain implementation. They can sometimes give a biological basis to classical models in AI as it is the case in Reinforcement Learning. They can also help study how the brain solves some problems that remain difficult with classical mathematical approaches. This is specifically the case with the three limitations mentioned above, as we describe now in more details.

Giving a semantic grounding It is generally considered that access to the semantic meaning of objects is due to the fact that we have a body that we can perceive externally (sense of exteroception, including proprioception) and internally (sense of interoception (Craig 2009)). Considering the consequences of taking these sensations into account onto cognition is studied in what is called embodied AI (Pfeifer, Bongard, and Grand 2007).

Internal representation of the body is built in the insular cortex. On the one hand, rich signals of pain and pleasure allow to elaborate reinforcement signals much more complex than the unique scalar generally used in reinforcement learning. Learning to anticipate these signals (pavlovian conditioning, see a model in (Carrere and Alexandre 2015)) elicits what is called emotions and plays a pivotal role in decision making by allowing to compare different signals under a common currency, as summarized by the principle of the somatic markers (Damasio, Everitt, and Bishop 1996).

On the other hand, other important internal signals form motivations and can be related to physiological needs (extrinsic motivations) as well as needs for certain kinds of information (intrinsic motivations). Such signals are particularly important to generate internal goals and to prevent from obeying only stimuli in the environment. This is an important dimension for goal-directed behaviors, see computational frameworks in (Pezzulo and Castelfranchi 2009).

Coping with high dimensionality One important contribution has arisen from CCN about the way humans can learn in high dimensionality with the Complementary Learning System (CLS) framework (McClelland, McNaughton, and O'Reilly 1995). In this framework, it is proposed that humans can, at the same time, learn slowly concepts in semantic memory, in the cortex, and store quickly episodes in the hippocampus. Later on, by a phenomenon called consolidation, the hippocampus will send episodes back to the cortex to train it off-line and ensure good properties to the learning process, for example to avoid mix-up, also called catastrophic forgetting. This kind of principles has been used in Machine Learning to propose ways of training large architectures (for example deep networks) with "not so big data", cf a model in (Drumond, Viéville, and Alexandre 2019).

Recently, progresses in neuroscience about the hippocampus and in the corresponding models led to more precise explanation about information encoding in the hippocampus, and particularly about place cells and grid cells (model in (Stachenfeld, Botvinick, and Gershman 2017)) and their proposed contribution in the goal-oriented exploration of high-dimensional spaces. It has particularly been proposed that replays of episodes for consolidation are not random but obey subtle strategies (implemented in (Mattar and Daw 2018)) and that they can be organized in time to participate to the elaboration of knowledge-based strategies in the frontal cortex (cf computational framework in (Daw 2018)). It has also been experimentally shown (Derdikman and Moser 2010) that the hippocampus not only replays stored episodes but is also able to create new episodes combining pieces of old episodes, thus promoting an imaginative training with virtual episodes.

These concepts and models are today transfered to the domain of AI to elaborate what is called Episodic Reinforcement Learning (Gershman and Daw 2017) and participate to the design of more realistic and faster learning in high dimensional data spaces.

Managing uncertainty Defining the stochastic or volatile nature of the environment has been related to the role of neuromodulation and exploited in models (Alexandre and Carrere 2016), adapting a general framework of behavior selection to the estimated kind and level of uncertainty.

As for the flexible adaptation of behavior to a volatile environment, the role of the prefrontal cortex has been mentioned for a long time in neuroscience, observing that patients with a frontal lesion demonstrate perseveration and are unable to adapt to changes (Nauta 1971). Accordingly, it has been proposed that the prefrontal cortex is the place where nondominant behaviors are defined (Wise 2008) or, to tell it differently, the place where top-down modulations are sent to other regions of the brain to insert internally generated priorities and to help resist to immediate responses driven by stimuli (Mesulam 2008), thus defining two important mechanisms in the prefrontal cortex: the inhibition of dominant and occasionally non-adapted behaviors and the triggering and maintenance by working memory of attentional focus towards characteristics supposedly important to generate the presently appropriate behavior.

Corresponding models describe the prefrontal cortex as a region where tasks in specific contexts are represented, thus defining the notion of Task Sets (Domenech and Koechlin 2015), learning in certain contexts to inhibit the default behavior and to suggest new behaviors by an attentional process (cf model in (O'Reilly et al. 2002)) biasing the current perception. In order to select the pertinent nondominant behavior, some interesting mathematical frameworks have been proposed (Collins and Koechlin 2012) to tame the combinatorial exploration and generate a limited number of hypotheses. Other interesting and bio-inspired computational mechanisms have been proposed to control and monitor the execution of such behaviors, particularly in the case of hierarchical planning (Pezzulo and Castelfranchi 2009).

These models and concepts are also presently transferred to AI in so-called Meta Reinforcement Learning (Wang et al. 2018), considering that, to adapt to the changing world, a meta learner must quickly learn to select a specialized (and slowly learning) learner, depending on the context.

Another important limitation in AI: Creativity

It is often mentioned though disputable (Boden 2009) that creativity remains one of the rare cognitive phenomena that AI cannot replicate (with other phenomena like sense of humor, not to say consciouness). Remarkably, studying solutions proposed by CCN to remedy limitations in AI (and consequently in CC), we argue (and will establish below) that they massively rely on cognitive characteristics related to creativity. We consequently propose that CCN could be pivotal to fuel CC with fresh ideas and particularly add a global view often missing in AI.

Defining creativity

Whether dealing with creativity in specific domains (like musical improvisation or scientific creativity) or from a general point of view (Beaty et al. 2016), authors often mention that creativity has two main steps: creating a novel idea and verifying that this idea is useful or appropriate to the task (Dietrich 2004); else the idea is adapted or rejected and another novel idea is elaborated. These two steps are respectively termed divergent and convergent thinking in human creativity research (Jung et al. 2013). Corresponding mechanisms in CC are called "generate and evaluate". The first step can be associated to insight (Kounios and Beeman 2014) and originate from an emotional or a more cognitive process (Dietrich 2004). As for the second step, it is generally proposed that the assessment of the value of what has been proposed is carried out by brain circuits associated to executive functions. The brain circuits responsible for these two steps have been studied in different ways as it will be discussed in more details below.

The two steps of creativity

Concerning the divergent thinking step, it is reported that self generated thoughts can be spontaneous or goal directed (to meet specific task demands) (Dietrich 2004). This can correspond to the reinterpretation of a situation to produce a nondominant interpretation (Kounios and Beeman 2014) or to using generative models to try to explicitly build a novel solution, which has been qualified as a more restricted but more efficient solution (Dietrich 2004). (Boden 2009) proposes that novel ideas can be produced by combination, exploration or by transformation. (Dietrich 2004) proposes four types of creativity: In the first step, novelty can come from emotional or cognitive structures and the processing can be deliberate or spontaneous.

Among neuronal mechanisms evoked above, it can be remarked that the hippocampus is a good candidate to generate spontaneous novel ideas and that the mechanisms in the prefrontal cortex, responsible for selecting a new Task Set, might participate to the explicit elaboration of new thoughts. Let us also underline that the model proposed in (Collins and Koechlin 2012) explicitly mentions that if no existing Task Set is appropriate to the task, a new Task Set might be created by the combination between two existing Task Sets or by the random generation of a new one.

The convergent thinking step is often related to cognitive control, the role of which is also to ensure appropriateness in the execution of a behavior. Appropriateness can be syntactic (checking that procedural constraints are respected, which is easy to do with a generative model) or semantic (checking that manipulated objects have consistent values, which is probably less easy (Dietrich 2004; Boden 2009)). In all cases, it is proposed that these verifications are made by the PFC, with the corresponding information set in working memory for their conscious evaluation.

The two steps of creativity in brain circuitry

The functional description of brain circuits is often made through networks of brain regions (Mesulam 2008), highlighting some regions as critical hubs for the spreading and processing of information. The attentional, semantic, default, cognitive control and salience networks are the major ones and play a central role in creativity (Jung et al. 2013).

The default network (activated by default, for spontaneous thinking, including parietal cortex, medial prefrontal cortex and hippocampal regions) is much involved in creativity and is known for its links to episodic memory in the hippocampus, spontaneous retrieval, replay activity and simulations based on personal past experiences (Dietrich 2004).

To evaluate the semantic and syntactic appropriateness of candidate ideas, a reference is made toward an emotional system for biological significance of events (involving the semantic network, associated to the insula and other limbic regions) and an information processing system to perform detailed feature analysis involving attentional control networks (Beaty et al. 2016). As a synthesis, pertinence, adaptation or rejection of candidate ideas is made by the cognitive control network, involving the lateral prefrontal cortex and the anterior cingulate cortex, performing top-down modulation of self generated information for efficacy evaluation, selection and adaptation to the task (Beaty et al. 2016).

Dynamics of inhibition and excitation in hubs and networks are observed with the analysis of EEG activity in behavioral and psychological tasks (Kounios and Beeman 2014). It must be remarked that corresponding tests of creativity are in fact relying on measures of fluency, flexibility, originality and elaboration, all measures also related to problem solving. This is also a strong argument to propose that the same brain circuits and cognitive mechanisms are used for higher brain functions and for creativity (Dietrich 2004).

Discussion

CCN models presented in this paper insist on the need to build, inspired from neuroscience, different kinds of representations in different regions of the brain (Alexandre, Carrere, and Kassab 2014) and to organize the selection of behavior through the interactions between different kinds of memories (Alexandre 2000).

Understanding these mechanisms and corresponding brain circuitry is particularly important to address current limitations in AI. As a central claim for this paper, we propose that these mechanisms and cognitive processes are also central to understand creativity as it is carried out in the brain. Perhaps this is not so surprising if we consider that among the mechanisms the brain has developed to circumvent the tedious and systematic analysis of some situtations, creativity might be a choice mechanism to efficiently explore novel approaches.

The present paper mentions operational models that could be directly exploited to implement CC. It also proposes paths of research to implement specific mechanisms. Spontaneous processing could be related to the hippocampus and its role in the default network, whereas deliberate processing would be more frontal and related to the combination of Task Sets. The more difficult aspect of combinational creativity could be linked to the limitation of semantic grounding, thus pleading for a closer look to the semantic network. Using CCN models to implement CC could be also an easier way to choose the type of creativity to generate or even

to study fundamental differences between human and non-human creativity (Colton and Wiggins 2012).

The very nature of creativity itself can be questioned from the elements reported here. Creativity is often associated to the generation of completly novel ideas. What is reported here about convergent but also divergent thinking is that most of the time, old memories are used to create new ideas. Here also, it is the old that makes the new. What is also reported here (in line with (Boden 2009) claiming that creativity is not magic) is that CC, indeed considered as a scientific questioning, suffers mainly from fragmentation in AI research (Colton, de Mantaras, and Stock 2009) and could highly benefit from the global framework proposed by cognitive neuroscience.

References

Alexandre, F., and Carrere, M. 2016. Modeling Neuromodulation as a Framework to Integrate Uncertainty in General Cognitive Architectures. In *The Ninth Conference on Artificial General Intelligence*.

Alexandre, F.; Carrere, M.; and Kassab, R. 2014. Feature, Configuration, History : a bio-inspired framework for information representation in neural networks. In *International Conference on Neural Computation Theory and Applications*.

Alexandre, F. 2000. Biological Inspiration for Multiple Memories Implementation and Cooperation. In *In V. Kvasnicka P. Sincak, J. Vascak and R. Mesiar, editors, International Conference on Computational Intelligence*.

Beaty, R. E.; Benedek, M.; Silvia, P. J.; and Schacter, D. L. 2016. Creative Cognition and Brain Network Dynamics. *Trends in Cognitive Sciences* 20(2):87–95.

Boden, M. A. 2009. Computer Models of Creativity. *AI Magazine* 30(3):23–23. Number: 3.

Carrere, M., and Alexandre, F. 2015. A pavlovian model of the amygdala and its influence within the medial temporal lobe. *Frontiers in Systems Neuroscience* 9(41).

Collins, A., and Koechlin, E. 2012. Reasoning, Learning, and Creativity: Frontal Lobe Function and Human Decision-Making. *PLOS Biology* 10(3):e1001293+.

Colton, S., and Wiggins, G. A. 2012. Computational Creativity: The Final Frontier? In *ECAI'12*.

Colton, S.; de Mantaras, R. L.; and Stock, O. 2009. Computational Creativity: Coming of Age. *AI Magazine* 30(3):4pp.

Craig, A. 2009. How do you feel – now? the anterior insula and human awareness. *Nat. Rev. Neurosci.* 10:59–70.

Damasio, A. R.; Everitt, B. J.; and Bishop, D. 1996. The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 351(1346):1413–1420.

Daw, N. D. 2018. Are we of two minds? *Nature Neuroscience* 21(11):1497–1499.

Derdikman, D., and Moser, M.-B. 2010. A Dual Role for Hippocampal Replay. *Neuron* 65(5):582–584. Publisher: Elsevier.

Dietrich, A. 2004. The cognitive neuroscience of creativity. *Psychonomic Bulletin & Review* 11(6):1011–1026.

Domenech, P., and Koechlin, E. 2015. Executive control and decision-making in the prefrontal cortex. *Current Opinion in Behavioral Sciences* 1:101–106.

Drumond, T. F.; Viéville, T.; and Alexandre, F. 2019. Bio-inspired Analysis of Deep Learning on Not-So-Big Data Using Data-Prototypes. *Frontiers in Computational Neuroscience* 12.

Gershman, S. J., and Daw, N. D. 2017. Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology* 68(1):101–128.

Harnard, S. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena* 42:335–346.

Jung, R. E.; Mead, B. S.; Carrasco, J.; and Flores, R. A. 2013. The structure of creative cognition in the human brain. *Frontiers in Human Neuroscience* 7.

Kounios, J., and Beeman, M. 2014. The cognitive neuroscience of insight. *Annual Review of Psychology* 65:71–93.

Mattar, M. G., and Daw, N. D. 2018. Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience* 21(11):1609–1617.

McClelland, J. L.; McNaughton, B. L.; and O'Reilly, R. C. 1995. 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(3):419–457.

Mesulam, M. 2008. Representation, inference, and transcendent encoding in neurocognitive networks of the human brain. *Annals of Neurology* 64(5):367–78.

Nauta, W. J. 1971. The problem of the frontal lobe: A reinterpretation. *Journal of Psychiatric Research* 8(3-4):167–187.

O'Reilly, R. C.; Noelle, D. C.; Braver, T. S.; and Cohen, J. D. 2002. Prefrontal cortex and dynamic categorization tasks: representational organization and neuromodulatory control. *Cereb Cortex* 12(3):246–257.

Pezzulo, G., and Castelfranchi, C. 2009. Thinking as the control of imagination: a conceptual framework for goal-directed systems. *Psychological Research* 73(4):559–577.

Pfeifer, R.; Bongard, J.; and Grand, S. 2007. *How the body shapes the way we think: a new view of intelligence*. Bradford Books. MIT Press.

Stachenfeld, K. L.; Botvinick, M. M.; and Gershman, S. J. 2017. The hippocampus as a predictive map. *Nature Neuroscience* 20(11):1643–1653.

Wang, J. X.; Kurth-Nelson, Z.; Kumaran, D.; Tirumala, D.; Soyer, H.; Leibo, J. Z.; Hassabis, D.; and Botvinick, M. 2018. Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience* 21(6):860–868. Number: 6 Publisher: Nature Publishing Group.

Wise, S. P. 2008. Forward Frontal Fields: Phylogeny and Fundamental Function. *Trends in neurosciences* 31(12):599–608.