A quest to understand Mind

V. Kabir Veitas[1,2,3]

[1] *SingularityNET Foundation*
[2] *Center Leo Apostel - VUB*
[3] *Evolution, Complexity & Cognition Group - VUB*

October 13, 2018

Abstract

Draft section from Veitas, V. K. (2019). *Synthetic Cognitive Development in decentralized self-organizing systems (PhD Thesis, forthcoming)*. Brussels, Belgium.

Produced for HIVE MINDS Collab, Fall 2018, The New School[1].

1 AI research perspectives

The modern quest for creating machines capable of thinking like humans has started around the middle of the nineteenth century, undoubtedly triggered by the advent of digital computers. Two major events are usually mentioned as signifying the beginning of the Artificial Intelligence research programme: Alan Turing's article "Computing Machinery and Intelligence" (Turing, 1950) and Dartmouth Summer Research Project on Artificial Intelligence in 1956.

In a course of six decade long research, scientific and popular discourse around AI, a number of informal and formal descriptive terms have emerged aiming at signifying different aspects or types of intelligence, as well as research perspectives. We review the major concepts and contexts in which they are used in order to pave the ground for further discussion and illustrate the precariousness and controversy around the concept of intelligence at large.

1.1 General Artificial Intelligence

General Artificial Intelligence (AGI) is considered by its proponents a return to the roots (Pennachin, Goertzel, and Geisweiller, 2014, p. 1) of the original AI research program formulated by the organizers of the Dartmouth Workshop in 1956 as "an attempt [...] to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves" (McCarthy et al., 2006).

Aiming at coming up with a standard definition of *universal intelligence*, Legg and Hutter (2007) have collected over 70 definitions of intelligence and distinguished them into three broad categories: *collective* (found in encyclopaedias and dictionaries), *psychologist* and *AI researcher* definitions. They distilled their own account – ***intelligence measures an agent's ability to achieve goals in a wide range of environments*** – by observing the most common features used for describing intelligence:

[1] https://courses.newschool.edu/courses/PSAM5550/2092/

Embodiment : it is a property that an individual agent has as it interacts with its environment or environments;

Goal directness : it is related to the agent's ability to succeed or profit with respect to some goal or objective;

Efficiency : it depends on how able the agent is to adapt to different objectives and environments.

Likewise, Goertzel (2009) proposes a notion of *efficient pragmatic general intelligence* and defines it as ***the capability of a system to choose actions maximizing its goal-achievement, based on its perceptions and memories, and making reasonably efficient use of its computational resources.***

1.2 Narrow Artificial Intelligence

As defined by Kurzweil (2005), narrow AI refers to machines and algorithms which perform a specific function that once required human intelligence to perform, and does that at human levels or better. The difference between the concepts of 'general' and 'narrow AI' seems to be 'only' in that the latter emphasizes different functions considered intelligent, while the former – the integration of these functions into a single system capable of more than the sum of its parts. Interestingly, the definition of 'narrow AI' is mostly used by AGI researchers to distinguish their research programme from the multitude of research areas in machine perception, natural language processing, machine learning, sensor fusion, neural networks, and many others, which, for arguably historical reasons, got under the umbrella of the notion of AI research (Pennachin, Goertzel, and Geisweiller, 2014, p. 2).

Yet, while the historical separation is obvious, conceptual borderline is not that straightforward, especially considering latest achievements of the 'narrow AI'[2]. As a research field, narrow AI holds a potential and ambition to reach the 'original AI agenda', albeit this has started to be vocalized only lately. For example, Jürgen Schmidhuber, one of the the pioneers of 'deep learning' techniques, powering the current success of 'narrow AI', envisions how Artificial General Intelligence could grow out of current specialized pattern recognition networks using the principles of reinforcement learning (Herman Kahrs, 2017).

1.3 Global Brain

The Global Brain is a metaphor for an emerging intelligent network that is formed by all people with the computers, knowledge bases and communication links that connect them together (Heylighen, 2002). Following the metaphor, the Global Brain is the nervous system of the organism of the human society. The origin of the concept can be tracked at least to the middle of the twentieth century, yet gained a scientific and technical perspective with the rise of computer networks, Internet and social networks.

Apart from social, organismic, philosophical, utopian, technical, cybernetic and many other perspectives, the Global Brain metaphor first of all emphasizes the concept of *distributed intelligence* which emerges from a network of interacting heterogeneous agents of lower capabilities. The relation of intelligence and network structure is strongly grounded in the neurophysiology and is very well reflected in AI research and practical applications. A network of interconnected

[2] AlphaGo – an AI Go player; Libratus – an AI Poker player

and interacting processes is largely regarded as a correct image of an intelligence machinery by pioneer as well as contemporary AI researchers (Minsky, 1988; Goertzel, 2002) and forms a basis of a *connectionist* approach.

Relations in a network naturally represent dynamic, seemingly chaotic, evolving and not fitting into clear logical and semantic structures interactions and interrelations among many heterogeneous agents. Yet the main message of the Global Brain metaphor lies in its emphasis on the importance of *self-organized coordination of decentralized processes*. Actually, all intelligence is essentially decentralized, albeit to a different degree: brains and minds are products of interactions of neurons and coordination of neural activity in brain areas; collective intelligence of eusocial insects originates from their interaction via pheromone trails; the intelligence of civilizations and societies emerges from coordination among individual humans.

1.4 Universal Intelligence

The theory of Universal Intelligence provides a formal definition of a universally intelligent agent (AIXI), capable "to achieve goals in a wide range of environments", which conforms to the definition of Artificial General Intelligence given above. Direct practical application of this theoretical and mathematical abstraction of an intelligent agent requires specific optimization mechanisms and algorithms for achieving useful down-scaling of otherwise incomputable model (Legg, 2008). A "practically universal intelligent agent" optimizes its behaviour with respect to a given environment (or set of environments) by running iterative cycles of observation, learning, prediction, decision, action and reward (Hutter, 2012; Hutter, 2013). Agent-environment interactions are modelled by formalizing both agent and environment as probabilistic functions each feeding its output to the other's input (Legg, 2008). Remarkably, general agents, built using this formalism and techniques, are able to learn different (albeit rather simple as of now) environments without any context related adjustments.

1.5 Freedom and constraint

In the context of the quest for creating synthetic intelligence, concepts of General Artificial Intelligence, Narrow Artificial Intelligence, Global Brain and Universal Intelligence represent complementary research perspectives rather than competing theories. These perspectives more often interact by enriching rather than denying each other's results or theoretical approaches. Even if one of them (and 'narrow AI' is currently seems to be in the lead) provides an essential breakthrough for AI research program, it is never isolated from important influences from and repercussions to other perspectives.

One of the ways to see how the AI research perspectives interact is to identify conceptual axes along which they take distinct approaches. The following axes explain well their differences and similarities:

Environmental interaction : how much the behaviour of a theoretical AI agent and its implementation allows for a change – depending on the influence of an environment; a usually overlooked aspect of this axis is the ***degree of agent's influence upon environment***.

Goal directedness : how much the agent's behaviour is guided or can be explained by *a-priori* defined goals and values and how much these goals can change or adjust to circumstances.

Efficiency : how important are efficiency and optimal behaviour considerations for a given perspective.

Obviously, these axes are not orthogonal – efficiency can be considered only with respect to the goal of the behaviour, while development of goals and values is often related to the degree of environmental interaction. Furthermore, they are not exhaustive – i.e. there are many more and will always be axes according to which the intelligent behaviour can be discriminated and analysed. Yet we can distinguish one overarching principle which grasps well the positioning of any perspective on any conceptual axis depending on how much constraint is imposed upon an agent. The principle is build upon observation that each axis is useful as much as it allows to grasp how much AI architecture allows for its own development in terms of interactions with environment, goal directness and efficiency / resource utilization. This principle is henceforth called **freedom and constraint**.

We can informally map all AI research perspectives in the continuum between complete freedom and ultimate constraint. The quest for synthetic intelligence can be understood as a search for a balance between freedom and constraint in terms of specific manifestations of intelligent behaviour (agents, organisms, AI architectures, etc.) with respect to specific context (goals, environments, etc.).

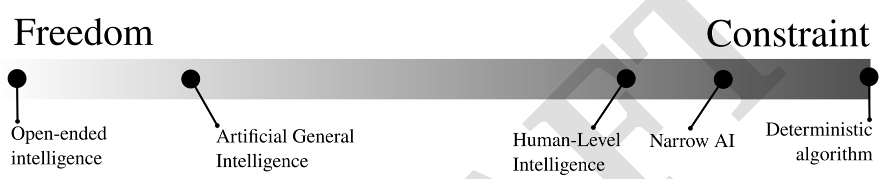

Figure 1: Research perspectives as positioned on freedom and constraint axis.

1.6 Open-ended intelligence

Open-ended intelligence is a novel theoretical approach to general intelligence proposed by Weinbaum and Veitas (2017) and is: "a process where a distributed population of interacting heterogeneous agents achieves progressively higher levels of coordination. In coordination, here we mean the local resolution of disparities by means of reciprocal determination that brings forth new individuals in the form of integrated groups of agents (assemblages) that exchange meaningful information and spontaneously differentiate (dynamically and structurally) from their surrounding milieu" (Weinbaum and Veitas, 2017, p. 14). It is a philosophical concept which first, ***allows for maximum freedom*** (see Figure 1) and second, defines intelligence not in terms of its specific manifestations or features, but in terms of non-linear process of bringing about ***precarious balances between the freedom and constraint*** invited and supported by specific contexts.

In contrast with open-ended intelligence, all of the aforementioned types of intelligence are examples of *goal-oriented intelligence* which is characterized by (1) more or less sharp agent-environment distinction where environment is independent of the agents' behaviour and is objectively knowable; (2) agents having a priori given goals while interacting in a knowable environment and (3) reward driven behaviour of agents with respect to their goals.

The approach to intelligence as a goal directed behaviour is a well established and is a prevailing mode of thinking – not surprisingly so, given its practical value in many domains, including psychology, robotics and AI research. Already James (1890) in his seminal study of human mind has chosen to follow the principle that "[t]he pursuance of future ends and the choice of means for their attainment are the mark and criterion of the presence of mentality in a phenomenon" (James, 1890, p. 8).

While goal-oriented intelligence is a measure of an agents' competence to match actions to observations such that it will achieve optimal results in a variety of environments, open-ended intelligence is the *process of emergence of intelligence* itself, including goal-directed intelligence and its manifestations. Open-ended intelligence therefore considers maximally fluid environmental interaction by encompassing processes of agent - environment differentiation and formation of agent's identity at the first place. Moreover, open-ended intelligence includes the processes of goal and value formation as well as determination of problematic situations which lead to goal formation. Finally, agents do not have a priori goals or values and interact with other similar agents in an environment shaped by the interaction itself.

The goal of this chapter is to introduce and shortly describe philosophical and theoretical concepts which are essential for conceiving actual mechanisms of the *process of becoming intelligent*. Designing and building such mechanisms based on the concept of open-ended intelligence is the direction of this work. The metaphysical framework of open-ended intelligence itself is developed in depth and breath by Weinbaum (2018).

If we ask ourselves, which the domain of science and philosophy, which process observable in nature and theory most closely embraces the notion of maximum freedom, the answer is quite obvious: *evolution*. It is the most general manifestation of the open-ended intelligence process and the best known exemplar. Other domains closely related to the concept are complex adaptive systems, complexity science and network science. We will next attend to these and related domains from the perspective of open-ended intelligence.

References

Goertzel, Ben (2002). *Creating Internet Intelligence: Wild Computing, Distributed Digital Consciousness, and the Emerging Global Brain*. en. Google-Books-ID: w5nbBwAAQBAJ. Springer Science & Business Media. ISBN: 978-1-4615-0561-7 (cit. on p. 3).

— (2009). "Toward a Formal Characterization of Real-World General Intelligence". In: *Proceedings of the 3rd Conference on Artificial General Intelligence*. AGI, pp. 19–24 (cit. on p. 2).

Herman Kahrs (2017). *Creating Human Level AI: How and When / Ray Kurzweil* (cit. on p. 2).

Heylighen, Francis (2002). "The global brain as a new utopia". In: *Zukunftsfiguren, Suhrkamp, Frankurt* (cit. on p. 2).

Hutter, Marcus (Oct. 2012). *Universal Artificial Intelligence*. Presentation by Marcus Hutter at Singularity Summit Australia 2012. Singularity Summit Australia 2012 (cit. on p. 3).

— (2013). "To Create a Super-Intelligent Machine, Start with an Equation". In: *The Conversation* November.29, pp. 1–5 (cit. on p. 3).

James, William (1890). *The Principles of Psychology*. English. Vol. 1. American science series. New York: Henry Holt and Company (cit. on pp. 4, 5).

Kurzweil, Ray (2005). *The Singularity is Near: When Humans Transcend Biology*. en. Viking. ISBN: 978-0-670-03384-3 (cit. on p. 2).

Legg, Shane (2008). "Machine Super Intelligence". Engels. PhD thesis. University of Lugano (cit. on p. 3).

Legg, Shane and Marcus Hutter (2007). "A collection of definitions of intelligence". In: *Advances in artificial general intelligence: Concepts, archtectures and algorithms*. Ed. by Ben Goertzel

and Pei Wang. Vol. 157. Frontiers in Artificial Intelligence and applications. Amsterdam: IOS Press, pp. 17–24 (cit. on p. 1).
McCarthy, John et al. (2006). "A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955". In: *AI magazine* 27.4, p. 12 (cit. on p. 1).
Minsky, Marvin (Mar. 1988). *Society Of Mind*. en. Simon & Schuster. ISBN: 978-0-671-65713-0 (cit. on p. 3).
Pennachin, Cassio, Ben Goertzel, and Neil Geisweiller (Feb. 2014). *I. Engineering General Intelligence, Part 1: A Path to Advanced AGI via Embodied Learning and Cognitive Synergy*. English. 2014 edition. New York: Atlantis Press. ISBN: 978-94-6239-026-3 (cit. on pp. 1, 2).
Turing, A. M. (Oct. 1950). "Computing Machinery and Intelligence". In: *Mind*. New Series 59.236, pp. 433–460. ISSN: 0026-4423 (cit. on p. 1).
Weinbaum, David R. (Weaver) (Mar. 2018). "Open-ended intelligence". English. PhD Thesis. Brussels: Vrije Universiteit Brussel (cit. on p. 5).
Weinbaum, David (Weaver) and Viktoras Veitas (Mar. 2017). "Open ended intelligence: the individuation of intelligent agents". In: *Journal of Experimental & Theoretical Artificial Intelligence* 29.2, pp. 371–396. ISSN: 0952-813X (cit. on p. 4).