Epistemic Computation and Artificial Intelligence

Jiří Wiedermann[1] and **Jan van Leeuwen**[2]

Abstract

AI research is continually challenged to explain cognitive processes as being computational. Whereas existing notions of computing seem to have their limits for it, we contend that the recent, epistemic approach to computations may hold the key to understanding cognition from this perspective. Here, computations are seen as processes generating knowledge over a suitable knowledge domain, within the framework of a suitable knowledge theory. This, machine-independent, understanding of computations allows us to explain a variety of higher cognitive functions such as accountability, self-awareness, introspection, knowledge understanding, free will, creativity, anticipation, curiosity in computational terms, as well as to understand the mechanisms behind the development of intelligence. The argumentation does not depend on any technological analogies.

Extended abstract

Computation has proved to be a powerful metaphor for understanding cognitive processes. The all important question is how it can explain and model them, and whether our understanding of computation is sufficient for it. We argue that the epistemic approach to computation provides the proper abstraction for it, pushing its application to new frontiers.

Epistemic approach

Traditionally, computations are seen as processes performed by computers. This focuses attention on HOW computations are performed, and makes their understanding technology-dependent. However, the primary question should be WHAT computations do. In our view, the only reasonable answer is that computations produce knowledge. This view is the starting point of the *epistemic theory of computations*, developed by us since 2013. In this approach, knowledge is defined in the (observer-dependent) framework of a suitable *knowledge domain* over which a computation operates. All knowledge, about some subset of the domain, is captured by a corresponding *knowledge theory* which can be more or less formal, or completely informal. *Axioms* describe the *elementary knowledge* of (representations of) objects in the domain. The ways in which new, derived knowledge can be obtained are described by *inference* or *derivation rules.* Computational processes are related to the theory via the following condition: *whatever can be derived within the given theory must be supported by the corresponding computational process.* If this condition holds then, what knowledge can or cannot be generated over the given knowledge domain and the "quality" of this knowledge (e.g., its agreement with an observation), depend solely on the properties of the underlying knowledge theory and, hence, of the computation. The approach has proved to be very versatile in apprehending a great variety of computational phenomena.

Cognitive functions in AI

Within the epistemic framework one can naturally define and explain a number of cognitive functions computationally, which would be cumbersome otherwise.

a) Accountability: an ability of a system to generate knowledge justifying its own decisions.

b) Awareness: this means that a system has knowledge about the problem being solved and that it can deliver explanations if asked to do so.

c) Introspection: an ability of a system to recall knowledge about its previous actions and their derivation.

d) Epistemic understanding. Accountability, self-awareness and introspection together give rise to understanding the knowledge domain over which a system operates. A system is able to explain the meaning of terms it works with and, based on its previous experience (recorded in its knowledge base), to apply them in new contexts.

e) Free will. We say that cognitive system A has free will with respect to cognitive system B if and only if, based solely on the observation of A's actions, B is not able to always generate knowledge about A's future actions in concrete situations. This definition differs from numerous definitions of free will that see the concept from an inner view of a system.

f) Creativity: a manifestation of a creative process, which is any process generating a solution to a problem (in the form of knowledge) that is new for the given cognitive system. Its counterpart is a routine process, which solves a known problem with the help of known procedures. In general, a creative process seeks explicit knowledge that is given implicitly, via a set of conditions that the knowledge to be found must satisfy. In our approach, the basic strategy for a creative process is the systematic examination of all knowledge that can be generated in the framework of a given theory. This initially inefficient, but universal process of knowledge discovery is cultivated in the course of its repeated use. Knowledge discovery is then seen as a potentially never ending evolutionary self-improving learning process whose goal is to improve its creative abilities. We describe several basic techniques to be used in a creative process: interactive refinement, automatic extraction of user preferences, and guided interaction with the environment.

g) Anticipation: an ability to generate knowledge in the form of predictions about the future occurrence of events or conditions in an epistemic domain. It is seen as the result of a "wired" creativity, a

[1] Institute of Computer Science of Czech Academy of Sciences and Czech Institute of Informatics, Robotics and Cybernetics of Czech Technical University, Prague, Czech Republic email: jiri.wiedermann@cs.cas.cz

[2] Center for Philosophy of Computer Science, Utrecht University, the Netherlands email: J.vanLeeuwen1@uu.nl

limiting result of creativity cultivation. Consequently, anticipation becomes a routine process that works as an efficient substitute of an originally creative process.

h) Epistemic curiosity: a perpetual need to discover new knowledge. It is intimately related to creativity and anticipation. Similar to creativity, it is a life-long learning process whose cultivation causes that not everything is explored and exploration is not made randomly. Curiosity is invoked when anticipation fails.

i) Epistemic self-improvement: an ability of artificial knowledge systems to improve their knowledge theory. To achieve it, such systems have mechanisms for discovering and repairing logical inconsistencies in their theory. As long as there exist contradictory facts within their theory and the systems at hand can find them, and as long as there exist unexplored objects and phenomena in the underlying knowledge domain, epistemic self-improving systems keep increasing their intelligence. Such systems can, at least in some domains, overcome human intelligence. Unlike the popular idea of software self-improvement that aims at streamlining derivation procedures in a cognitive system, self-improvement of knowledge theories aims at the heart of the intelligence — viz. the quality and quantity of the epistemic data.

Conclusions

Our paper uses the epistemic approach to computations as we developed it and extends it by new results and new apprehensions. We presents them uniformly from the viewpoint of AI. In the full paper, more detailed descriptions and discussions of the studied cognitive functions are given. Algorithmic knowledge-processing mechanisms that enable their realization within the given computational model of cognitive systems will also be identified.

Viewing computations as knowledge generation processes has great potential for AI. It enables us to understand and define nontrivial cognitive functions in a natural way, and brings insight into algorithmic mechanisms behind the development of intelligence. We intend to extend it to a more complete, Jungian scheme of functionalities.This would presents an essential contribution to the philosophy and theory of computational cognitive systems.

ACKNOWLEDGEMENTS

The research of the first author was partially supported by ICS AS CR fund RVO 67985807 and the Czech National Foundation Grant No. 15-04960S.

References

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies.* Oxford University Press

Kroll, J.A., et al. (2016). Accountable algorithms (March 2, 2016). University of Pennsylvania Law Review, Vol. 165, 2017 Forthcoming; Fordham Law Legal Studies Research Paper No. 2765268. Available at SSRN: https://ssrn.com/abstract=2765268

Piccinini, G., Scarantini, A. (2011). Information processing, computation and cognition. *J. Biol. Phys.* 37:1 (2011): 1-38.

Samsonovich, A.V. (2010). Toward a Unified Catalog of Implemented Cognitive Architectures. BICA 221 (2010): 195-244, online catalogue http://bicasociety.org/cogarch/.

van Leeuwen, J., Wiedermann, J. (2017). Knowledge, representation and the dynamics of computation. In: G. Dodig-Crnkovic, R. Giovagnoli (Eds): *Representation and Reality: Humans, Animals and Machines.* Berlin: Springer

Wiedermann, J., van Leeuwen, J.(2013). Rethinking computation. In: *Proc. 6th AISB Symp. on Computing and Philosophy: The Scandal of Computation - What is Computation?*, AISB Convention 2013 (Exeter, UK), AISB, pp. 6-10

Wiedermann, J., van Leeuwen, J. (2014). Computation as knowledge generation, with application to the observer-relativity problem. In: *Proc. 7th AISB Symposium on Computing and Philosophy: Is Computation Observer-Relative?*, AISB Convention 2014 (Goldsmiths, University of London), AISB, 2014

Wiedermann, J., van Leeuwen, J. (2015a). What is Computation: An Epistemic Approach. In: *Italiano, G. et al., (eds.). SOFSEM 2015: Theory and Practice of Computer Science.* LNCS 8939, Berlin: Springer, pp. 1-13

Wiedermann, J., van Leeuwen, J. (2015b). Towards a computational theory of epistemic creativity. In: *Proc. 41st Annual Convention of AISB 2015.* London, pp. 235-242

Wiedermann, J., van Leeuwen, J. (2017). Understanding and controlling Artificial General Intelligence Systems. In: *Proc. 10th AISB Symposium on Computing and Philosophy: Language, Cognition and Philosophy*, AISB Convention 2017, (University of Bath, UK), AISB, pp. 356-363