

Essay

On Gödel's Incompleteness Theorem, Artificial Intelligence & Human Mind

Victor Christianto*

Abstract

In this essay, I discuss Gödel's incompleteness theorem and plausible implications to Artificial Intelligence/Life and human mind. Perhaps we should agree with Sullins III, that the value of this finding is not to discourage certain types of research in AI, but rather to help move us in a direction where we can more clearly define the results of that research. Gödel's incompleteness theorem has its own limitations, but so do Artificial Life systems. Based on our experiences, human mind has incredible abilities to interact with other part of human body including heart, which makes it difficult to simulate in AI/AL systems. However, it remains an open question to predict whether the future of AI including robotics science can bring this gap closer or not.

Key Words: Gödel, Incompleteness Theorem, artificial intelligence, artificial life, human mind.

Introduction

In 1931 a German mathematician named Gödel published a paper [9] which included a theorem which was to become known as his Incompleteness Theorem. This theorem stated that:

To every w -consistent recursive class k of formulae there correspond recursive class-signs r , such that neither $v \text{ Gen } r$ nor $\text{Neg } (v \text{ Gen } r)$ belongs to $\text{Flg}(k)$ (where v is the free variable of r)

In more common mathematical terms, this means that "all consistent axiomatic formulations of number theory include undecidable propositions" [9].

Another perspective on Gödel's incompleteness theorem can be found using polynomial equations [10]. It can be shown that Gödel's analysis does not reveal any essential incompleteness in formal reasoning systems nor any barrier to prove the consistency of such systems by ordinary mathematical means [10]. Further, Beklemishev discusses the limits of applicability of Gödel's incompleteness theorems in [11].

Does Gödel's incompleteness theorem limit Artificial Intelligence?

In the 1950s and 1960s, researchers predicted that when human knowledge could be expressed using logic with mathematical notation, it would be possible to create a machine that reasons

*Correspondence: Victor Christianto, <http://www.sciprint.org>, <http://independent.academia.edu/VChristianto>
Email: victorchristianto@gmail.com . Phone: (62) 341-403205.

known as artificial intelligence. This turned out to be more difficult than expected because of the complexity of human reasoning [12].

Nowadays, it is widely accepted that general purpose of artificial intelligence (AI) is to develop (1) conceptual models; (2) formal rewriting processes of these models; and (3) programming strategies and physical machines to reproduce as efficiently and thoroughly as possible the most authentic, cognitive, scientific and technical tasks of biological systems that we have labeled Intelligent [5, p.66].

According to Gelgi [1], Penrose claims that results of Gödel's theorem established that human understanding and insight cannot be reduced to any set of computational rules. Gelgi goes on to say that:

Gödel's theorem states that in any sufficiently complex formal system there exists at least one statement that cannot be proven to be true or false. Penrose believes that this would limit the ability of any AI system in its reasoning. He argues that there will always be a statement that can be constructed which is unprovable by the AI system.

The above question is very interesting to ponder, considering recent achievements in modern AI research. There are ongoing debates on this subject in many online forums, see for instance [5-9]. Here I give a summary of those articles and papers in simple words. Hopefully this effort will shed some light on this debatable subject. Those arguments basically stand either on the optimistic side (that Gödel's theorems do not limit AI) or on the pessimistic side (that Gödel's theorems limit AI).

Mechanism and reductionism in biology and implications to AI/AL

It is known that mechanistic or closely related reductionist's theories have been part of theoretical biology in one form or another at least since Descartes [8]. The various mechanistic and reductionist's theories are historically opposed to the much older and mostly debunked theories of vitalism. These theories (the former more than the latter), along with formism, contextualism, organicism and a number of other "isms" mark the major centers of thought in the modern theoretical biology debate [8].

The mechanistic and reductionist's view of the world were criticized by F. Capra in his book, *The Turning Point* [13].

According to Sullins III, Artificial Life (AL) falls curiously on many sides of these debates in the philosophy of biology [8]. For instance, AL uses the tools of complete mechanization, namely the computer, while at the same time it acknowledges the existence of emergent phenomena. Neither mechanism nor reductionism is usually thought to be persuaded by arguments appealing to emergence. Facts like this should make our discussion interesting. It may turn out that AL is hopelessly contradictory on this point, or it may provide an escape route for AL if we find that Gödel's incompleteness theorems do pose a theoretical road block to the mechanistic-reductionist's theories in biology.

Sullins III also writes that most theorists have outgrown the idea that life can be explained wholly in terms of classical mechanics [8]. Instead, what is usually meant is the following:

- 1) Living systems can and/or should be viewed as physico-chemical systems.
- 2) Living systems can and/or should be viewed as machines (This kind of mechanism is also known as the machine theory of life).
- 3) Living systems can be formally described. There are natural laws which fully describe living systems.

According to Sullins III, reductionism is related to mechanism in biology in that mechanists wish to reduce living systems to a mechanical description [8]. Reductionism is also the name of a more general world view or scientific strategy. In this world view, we explain phenomena around us by reducing them to their most basic and simple parts. Once we have an understanding of the components, it is then thought that we have an understanding of the whole. There are many types of reductionist strategies [8].

According to Sullins III, reductionism is a tool or strategy for solving complex problems [8]. There does not seem to be any reason that one has to be a mechanist to use these tools. For instance one could imagine a causal reductionistic vitalist who would believe that life is reducible to the *elan vital* or some other vital essence. And, conversely, one could imagine a mechanist who might believe that living systems can be described metaphorically as machines but that life was not reducible to being only a property of mechanics.

Sullins III also asserts that the strong variety of AL does not believe that living systems should only be viewed as physico-chemical systems [8]. AL is life-as-it-could-be, not life-as-we-know-it, and this statement suggests that AL is not overly concerned with modeling only physico-chemical systems. Postulates 2 and 3 seem to hold, though, as strong AL theories clearly state, that the machine or formal theory of life is valid and that simple laws underlie the complex and nonlinear behavior of living systems.

Sullins III goes on with his argument, saying that at least one of the basic qualities of our reality will always be missing from any conceivable artificial reality, namely, a complete formal system of mathematics [8]. This argument tends to make more sense when applied to strong AI claims about intelligent systems understanding concepts. He also concludes that it is impossible to completely formalize an artificial reality that is equal to the one we experience, so AL systems entirely resident in a computer must remain, for anyone persuaded by the mathematical realism posited by Gödel, a science which can only be capable of potentially creating extremely robust simulations of living systems but never one that can become a complete instantiation of a living system [8].

However, Sullins III also writes that the value of this finding is not to discourage certain types of research in AL, but rather to help move us in a direction where we can more clearly define the results of that research [8]. In fact, since one of the above arguments rests on the assumption that the universe is infinite and that some form of mathematical realism is true, if we are someday

able to complete the goal advanced in strong AL, it would seem to cast doubt on the validity of the assumptions made above.

For a debate on this issue in the context of fuzzy logic, see for instance Yalciner et al. [5]. The debates on the possibility of thinking machines or the limitations of AI research have never stopped. According to Yalciner et al., these debates on AI have been focused on three claims:

- (1) An AI system is in principle an axiomatic system.
- (2) The problem solving process of an AI system is equivalent to a Turing machine.
- (3) An AI system is formal, and only gets meaning according to model theoretic semantic [16].

More than other new sciences, AI and philosophy have things to say to one to another: any attempt to create and understand minds must be of philosophical interest [5]. May be we will never manage to build real artificial intelligence. The problem could be too difficult for human brain over to solve. Yalciner et al. also write that a fundamental problem in artificial intelligence is that nobody really knows what intelligence is [5]. The problem is especially acute when we need to consider artificial systems which are significantly different to humans.

Human mind is beyond machine capabilities

According to Gelgi, it follows that no machine can be a complete or adequate model of the mind, that minds are essentially different from machines [1]. This does not mean that a machine cannot simulate any piece of mind; it only says that there is no machine that can simulate every piece of mind. Lucas says that there may be deeper objections. Gödel's theorem applies to deductive systems, and human beings are not confined to making only deductive inferences. Gödel's theorem applies only to consistent systems, and one may have doubts about how far it is permissible to assume that human beings are consistent [1].

Therefore, it appears that there are some characteristics of human mind which go beyond machine capabilities. For example there are human capabilities as follows:

- a. To synchronize with heart, i.e., to love and to comprehend love;
- b. To fear God and to acknowledge God: "The fear of the LORD is the beginning of knowledge" (Proverbs 1:7);
- c. To admit own mistakes and sins;
- d. To repent and to do repentance; and
- e. To consider things from ethical perspectives.

All of the above capabilities are beyond the scope of present day AI machines, i.e., it seems that there is far distance between human mind capabilities and machine capabilities. However, we can predict that there will be much progress by AI research. For instance, by improving AI-based

chess programs (such as new generations of *Deep Blue*), one could see how far the machine can go.

Furthermore, there are other philosophical arguments concerning the distinction between human mind and machine intelligence. Dreyfus contends that it is impossible to create intelligent computer programs analogous to the human brain because the workings of human intelligence are entirely different from that of computing machines [5]. For Dreyfus, the human mind functions intuitively and not formally. Dreyfus's critique on AI proceeds from his critique on rationalist epistemological assumptions about human intelligence. Dreyfus's major attack targets the rationalist conception that human understanding or intelligence can be "formalized" [5, p.67].

The above argument can be seen as stronger than Penrose's. However, one should admit the fundamental differences between human intelligence and machine intelligence. Human intelligence is very good in identifying patterns and subjective matters. However, it is usually not very good in handling large amounts of data and doing massive computations. Nor can it process and solve complex problems with large number of constraints. This is especially true when real time processing of data and information is required. For these types of issues, machine intelligence is an excellent substitute [5].

Concluding remarks

In this essay, I discuss Gödel's incompleteness theorem and its plausible implications to artificial intelligence/life and human mind.

Perhaps, we should agree with Sullins III, that the value of this finding is not to discourage certain types of research in AL, but rather to help move us in a direction where we can more clearly define the results of that research [8]. Gödel's incompleteness theorems have their own limitations, but so do Artificial Life/Intelligence systems. Based on our experiences, human mind has incredible abilities to interact with other part of human body including heart, which makes it so difficult to simulate in AL/AI. However, it remains an open question to predict whether the future of AI including robotics science can bring this gap closer or not. In this regard, fuzzy logic may offer a way to improve significantly AL/AI research [15].

References

- [1] Gelgi, F. (2004) Implications of Gödel's Incompleteness Theorem on A.I. vs. Mind, *NeuroQuantology*, (3), 186-189. URL: <http://www.neuroquantology.com/index.php/journal/article/view/46>
- [2] Chalmers, DJ. (1995) Minds, Machines, and Mathematics. Psyche Symposium on Roger Penrose's book *Shadows of the Mind*. URL: <http://consc.net/papers/penrose.html>
- [3] Gödel, K. (1931) On Formally Undecidable Propositions of Principia Mathematica and related systems I. *Monatshefte für Mathematik und Physik*, vol. 38 (1931), pp. 173-198. <http://www.ddc.net/ygg/etext/godel/>
- [4] Pelletier, F.J. (2000) Metamathematics of fuzzy logic, *The Bulletin of Symbolic Logic*, Vol. 6, No.3, 342-346, URL: <http://www.sfu.ca/~jeffpell/papers/ReviewHajek.pdf>

- [5] Yalciner, A.Y., Denizhan, B., Taskin, H. (2010) c. TJFS: *Turkish Journal of Fuzzy Systems* Vol.1, No.1, pp. 55-79. URL: http://www.tjfs-journal.org/TJFS_v1n1_pp55_79.pdf
- [6] Karavasileiadis, C. & O'Bryan, S., (2008) *Philosophy of Logic and Artificial Intelligence*.
- [7] Collins, J.C. (2001) On the compatibility between physics and intelligent organisms, arXiv:physics/0102024. URL: <http://arxiv.org/abs/physics/0102024>
- [8] Sullins III, J.P. (1997) Gödel's Incompleteness Theorems and Artificial Life - Digital Library & archives, Society for Philosophy and Technology, <http://scholar.lib.vt.edu/ejournals/SPT/v2n3n4/sullins.html>
- [9] Makey, J. (1995) Gödel's Incompleteness Theorem is Not an Obstacle to Artificial Intelligence, URL: http://www.sdsc.edu/~jeff/Godel_vs_AI.html
- [10] Norman, J.W. (2011) Resolving Gödel's incompleteness myth: Polynomial Equations and Dynamical Systems for Algebraic Logic, arXiv:1112.2141 [math.GM]. URL: <http://arxiv.org/abs/1112.2141>
- [11] Beklemishev, L.D. (2010) Gödel incompleteness theorems and the limits of their applicability. I, *Russian Math. Surv.* 65 857, URL: <http://iopscience.iop.org/0036-0279/65/5/R03>; or <http://www.mi.ras.ru/~bekl/Papers/goedel-en.pdf>
- [12] <http://www.wikipedia.org/Logic>
- [13] Capra, F. (1982) *The Turning Point*. Bantam Books. See URL: <http://www.juwing.sp.ru/Capra/CONTENTS.htm>
- [14] Chaitin, G.J. (1999) A century of controversy over the foundations of mathematics, arXiv: chaos-dyn/9909001. URL: <http://arxiv.org/abs/chaos-dyn/9909001>
- [15] Schumann, A. & Smarandache, F. (2007) *Neutrality and many-valued logics*. American Research Press. 121 p. URL: <http://arxiv.org/abs/0707.3205>
- [16] Wang, P. (2006) Three Fundamental Misconceptions of Artificial Intelligence, http://www.cis.temple.edu/~pwang/Publication/AI_Misconceptions.pdf
- [17] http://www.wikipedia.org/Philosophy_of_artificial_intelligence
- [18] Straccia, U. (2000) On the relationship between fuzzy logic and four-valued relevance logic, arXiv:cs/0010037. URL: <http://arxiv.org/abs/cs/0010037>
- [19] Born, R.P. (2004) Epistemological Investigations into the Foundations of Artificial Intelligence, http://www.iwp.jku.at/born/mpwfst/04/0401Turing_engl_1p.pdf