# *Reductio ad Absurdum:* On Oversimplification in Computer Science and its Pernicious Effect on Artificial Intelligence Research

Kristinn R. Thórisson[1,2]

[1]Center for Analysis and Design of Intelligent Agents / School of Computer Science, Reykjavik University, Venus, Menntavegur 1, Reykjavik, Iceland

[2]Icelandic Institute for Intelligent Machines, 2.h. Uranus, Menntavegur 1, Reykjavik, Iceland

thorisson@gmail.com

**Abstract.** The Turing Machine model of computation captures only one of its fundamental tenets – the manipulation of symbols. Through this simplification it has relegated two important aspects of computation – *time* and *energy* – to the sidelines of computer science. This is unfortunate, because time and energy harbor the largest challenges to life as we know it, and are therefore key reasons why intelligence exists. As a result, time and energy must be an integral part of any serious analysis and theory of mind and thought. Following Turing's tradition, in a misguided effort to strengthen computer science as a science, an overemphasis on mathematical formalization continued as an accepted approach, eventually becoming the norm. The side effects include artificial intelligence research largely losing its focus, and a significant slowdown in progress towards understanding intelligence as a phenomenon. In this position paper I briefly present the arguments behind these claims.

## 1 Introduction

A common conception is that the field of computer science provides an obvious and close to ideal foundation for research in artificial intelligence (AI). Unfortunately some fundamental derailments prevent computer science – as practiced today – from providing the perfectly fertile ground necessary for the two to have the happy marriage everybody is hoping for. Here we will look at two major such derailments.

## 2 Derailment #1: The Turing Machine

Alan Turing's characterization of computation – as the sequential reading and writing of symbols by a simple device we now know as a Turing Machine (Turing 1948) – is generally considered to be a cornerstone of computer science. Turing's influential paper *On Computing Machinery and Intelligence* (Turing 1950) took notable steps towards considering intelligence as a computational system. The foundation of what came to be called *artificial intelligence* – the quest for making machines capable of what we commonly refer to as "thought" and "intelligent action" – was laid a few years after his papers were written. The main inspiration for the field came of course from nature – this is where we still find the best (some might say *only*) examples of intelligence. The original idea of artificial intelligence, exploration of which already started with cybernetics (cf. Heylighen & Josly 2001), was to apply the tools of modern science and engineering to the creation of generally intelligent machines that could be assigned any task: that could wash dishes and skyscraper windows to writing research reports and discovering new laws of nature; that could invent new things and solve difficult problems requiring imagination and creativity.

Before we go further on this historical path, let's look at two natural phenomena that play a large role in science and engineering: time and energy. Time and energy are directly relevant to the *nature of intelligence* on two levels. First, because every computation must take place in a medium, and every medium requires some amount of time and energy to act, there are limits on the number of computations that can be produced a given timeframe. This level of detail is important to AI because intelligence must be judged in the context of the world – including the computing medium – in which it occurs: If the mind of an intelligent agent cannot support a sufficient computation speed for it to act and adapt appropriately in its environment, we would hardly say that the agent is "dumb" because it would be *physically incapable* of acting intelligently. The physical properties of environments present time and energy constraints; the "hardware" of a thinking agent must meet some minimum specification to support thought *at sufficient speeds for survival*. Unless we study the role time and energy play at this level of the *cognitive computing medium* we are neither likely to understand the origins of intelligence nor its operating principles.

At the cognitive level time must in fact occupy part of the *content* of any intelligent mind: Every real-world intelligent agent must be able to understand and think about time, because everything they do happens in time. The situation is similar with respect to energy at this level (although in the case of humans it used to be more relevant the past than it is now, as foraging and farming occupied more time in our ancestors' minds than ours). In either case, *a key role of intelligence* from moment to moment remains in large part to help us handle the ticking of a real-world clock by *thinking* about time: To make *better use of time*, to be able to *meet deadlines* and understand the implications of *missing* them, to *shorten* the path from a present state to a new state, to *speed up* decision time by using past experiences and decision aids, and so on. Having unbounded time means that any problem can be solved by a complete search of all possibilities and outcomes. But if this is the case, intelligence is essentially *not needed:* Disregarding time renders intelligence essentially *irrelevant*. And so the very subject of our study has been removed. Therein lies the rub: Unlike the paths taken (so far) in some of the subdomains of computer science, the field of AI is fundamentally dependent on time and energy – these are two of its main raison d'être – and therefore *must be an integral part of its theoretical foundation*.

Fast forward to the present. The field we know as 'computer science' has been going strong for decades. But it gives *time* and *energy* short shrift as subjects of importance. To be sure, progress continues on these topics, e.g. in distributed systems theory and concurrency theory, among others. But it is a far cry from what is needed, and does not change historical facts: Few if any programming languages exist where time is a first-class citizen. Programming tools and theories that can deal properly with time are sorely lacking, and few if any real methods exist to build systems for realtime performance without resorting to hardware construction. Good support for the design and implementation of energy-constrained and temporally-dependent systems (read: *all* software systems) is largely relegated to the field of "embedded systems" (cf. Sifakis 2011) – a field that limits its focus to systems vastly simpler than any intelligent system and most biological process found in nature, thus bringing little additional value to AI. As a result, much of the work in computer science practitioners – operating systems, databases, programming tools, desktop applications, mathematics – are rendered irrelevant to a serious study of intelligence.

What caused this path to be taken, over the numerous others possibilities suggested by cybernetics, psychology, engineering, or neurology? Finding an explanation takes us back to Turing's simplified model of computation: When he proposed his definition of computation Turing branched off from computer engineering through a *dirty trick:* His model of computation is completely mute on the aspects of *time* and *energy.* Yet mind exists in living bodies because time is a complicating factor in a world where energy is scarce. These are not some take-it-or-leave-it variables that we are free to include or exclude in our scientific models, these are inseparable aspects of reality.

As an incremental improvement on past treatments, some might counter, Turing's ideas were an acceptable next step, in a similar way that Newton's contributions in physics were before Einstein (they were not as thoroughly temporally grounded). But if time and energy are not needed in our *theories* of computation we are saying that they are irrelevant in the *study of computation*, implying that it does not matter whether the computations we are studying take no time or infinite time: The two extremes would be *equivalent*. Such reductio ad absurdum, in the literal meaning of the phrase, might possibly be true in some fields of computer science – as they happen to have evolved so far – but it certainly is not true for AI. If thinking is computation we have in this case rendered time irrelevant to the study of thought. Which is obviously wrong.

An oversimplification such as this would hardly have been tolerated in engineering, which builds its foundations on physics. Physicists take pride in making their theories actually match reality; would a theory that ignores significant parts of reality have been made a *cornerstone* of the field? Would the theory of relativity have received the attention it did had Einstein not grounded it with a reference to the speed of light? Somehow $E = m$ is not so impressive. The situation in computer science is even worse, in fact, because with Turing's oversimplification – assuming infinite time and energy – nothing in Einstein's equation would remain.

## 4 Derailment #2: Premature Formalization

The inventors of the flying machine did not sit around and wait for the theory of aerodynamics to mature. Had the Wright brothers waited for the "right mathematics", or focused on some isolated part of the problem simply because the available mathematics could address it, they would certainly not be listed in history as the pioneers of aviation. Numerous other major discoveries and inventions – electricity, wireless communications, genetics – tell a similar story, providing equally strong examples of how scientific progress is made without any requirement for strict formal description or analysis.

In addition to relegating time and energy to a status of little importance in AI, rubbing shoulders with computer science for virtually all of its 60-year existence has brought with it a general disinterest in natural phenomena and a pernicious obsession with formalization. Some say this shows that AI suffers from *physics envy* – envy of the beauty and simplicity found in many physics equations – and the hope of finding something equivalent for intelligence. I would call it a *propensity for premature formalization*. One manifestation of this is researchers limiting themselves to questions that have a clear hope of being addressed with today's mathematics – putting the tools in the driver's seat. Defining research topics in that way – by exclusion, through the limitations of current tools – is a sure way to lose touch with the important aspects of an unexplained natural phenomenon.

Mathematical formalization does not work without clear definitions of terms. Definition requires specifics. Such specification, should the mathematics invented to date not be good for expressing the full breadth of the phenomena to be defined (which for complex systems is invariably the case), can only be achieved through *simplification* of the concepts involved. There is nothing wrong with simplification in and of itself – it is after all a principle of science. But it matters *how* such simplification is done. Complex systems implement intricate causal chains, with multiple negative and positive feedback loops, at many levels of detail. Such systems are highly sensitive to changes in topology. Early simplifications are highly likely to leave out *key aspects* of the phenomena to be defined. The effects can be highly unpredictable; the act will likely result in devastating *oversimplification*.

*General* intelligence is capable of learning *new tasks* and adapting to *novel environments*. The field of AI has, for the most part, lost its ambition towards this general part of the intelligence spectrum, and focused instead on the making of specialized machines that only slightly push the boundaries of what traditional computer science tackles every day. Part of the explanation is an over-reliance on Turing's model of computation, to the exclusion of alternatives, and a trust in the power of formalization that borders on the irrational. As concepts get simplified to fit available tools, their correspondence with the real world is reduced, and the value of subsequent work is diminished. In the quest for a stronger scientific foundation for computer science, by threading research through the narrow eye of formalization, exactly the opposite of what was intended has been achieved: The field has been made *less scientific*.

## 5 What Must Be Done

In science the questions are in the driver seat: A good question comes first, everything else follows. Letting the tools decide which research questions to pursue is not the right way to do science. We should study more deeply the many principles of cognition that are *difficult* to express in today's formalisms, system architectures implementing multiple feedback loops at many levels of detail, for instance; only this way can we simultaneously address the self-organizing hierarchical complexity and networked nature of intelligent systems. Temporal latency is of course of central importance in feedback loops and information dissemination in a large system. All this calls for greater levels of system understanding than achieved to date (cf. Sifakis 2011), and an understanding of how time and energy affect operational semantics.

The very nature of AI – and especially artificial *general* intelligence (AGI) – calls for a study of *systems*. But systems theory is immature (cf. Lee 2006) and computer science textbooks typically give system architecture short shrift. The rift between computer science and artificial intelligence is not a problem in principle – computer science could easily encompass the numerous key subjects typically shunned in AI today, such as non-axiomatic reasoning, existential autonomy, fault-tolerance, graceful degradation, automatic prioritization of tasks and goals, and deep handling of time, to name some basic ones. Creativity, insight and intuition, curiosity, perceptual sophistication, and inventiveness are examples of more exotic, but no less important, candidates that are currently being ignored. Studying these with current formalisms is a sure bet on slow or no progress. We don't primarily need formalizations of cognitive functions per se, first and foremost we need *more powerful tools:* New formalisms that don't leave out key aspects of the real world; methods that can address its dynamic complexity head-on, and be used for representing, analyzing, and ultimately understanding, the operation of large complex systems *in toto*.

*Acknowledgments*. Thanks to Eric Nivel, Pei Wang, Hrafn Th. Thórisson, Helgi Páll Helgason, Luca Aceto, Jacky Mallett, and the anonymous reviewers for comments on the paper. This work has been supported in part by the EU-funded project HUMANOBS: Humanoids That Learn Socio-Communicative Skills by Observation, contract no. FP7-STReP-231453 (www.humanobs.org), by a Centres of Excellence project grant (www.iiim.is) from the Icelandic Council for Science and Technology Policy, and by grants from Rannís, Iceland.

## References

Heylighen, F. & C. Joslyn (2001). Cybernetics and Second-Order Cybernetics. *Encyclopedia of Physical Science & Technology*, 3rd ed. New York: Academic Press.

Lee, E. E. (2006). Cyber-Physical Systems - Are Computing Foundations Adequate? Position Paper for *NSF Workshop On Cyber-Physical Systems: Research Motivation, Techniques and Roadmap*.

Sifakis, J. (2011). A Vision for Computer Science – the System Perspective. *Cent. Eur. J. Comp. Sci.*, **1**:1, 108-116.

Turing, A. (1948). Intelligent Machinery. Reprinted in C. R. Evans and A. D. J. Robertson (eds.), *Cybernetics: Key Papers,* Baltimore: University Park Press, 1968.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, **59**:236, 433-460.