On the Computational Complexity of Cognition

Cosmo Harrigan[*]

Department of Mathematics, University of Washington

1   Introduction

In his seminal paper *Computing Machinery and Intelligence* [1], Alan Turing proposes to ask, *"Can Machines Think?"* Cognition is often defined as the ability of an embodied agent to process information intelligently. The fields of computationalism and connectionism in cognitive science offer models to explain the mechanisms of cognition. John Searle famously argued in [2] that even if machines were able to outwardly resemble humans, they would not truly harbor understanding. In this paper, I will argue that this definition of cognition is not sufficient to resolve this debate, and I will present an outline of an alternative definition that analyzes cognition in terms of the *computational complexity* of a system as a way of measuring its degree of understanding.

I will begin with an overview of the paradigms of computationalism and connectionism in cognitive science, and then I will present an analysis of constraints on mind design, and a brief survey of formal measures of complexity and efficiency. I will then use these ideas to provide a formal characterization of the degree of understanding present in a cognitive system, and an analysis of possible objections.

2   Background

2.1   Computationalism

The paradigm of *computationalism* is concerned with modeling cognitive systems as symbol-processing systems. In [3], Newell and Simon famously presented the "Physical Symbol System Hypothesis", which describes a system that performs manipulations on abstract tokens which refer to elements of a physical reality, and posits that a physical symbol system is both necessary and sufficient as a means for cognition. Around the same time, Fodor argued that minds possessed a symbolic "Language of Thought" [4].

---

[*]cosmoh@uw.edu

2.2 Connectionism

An alternate interpretation of the mechanisms of cognition, *connectionism*, was advocated for by Rumelhart and McClelland [5, 6]. In connectionism, the mind is understood as a massively parallel architecture, composed of simple elementary units which form associations with each other and represent concepts by way of intricate interconnections between individual elements.

2.3 Levels of Analysis

Marr had a lasting influence on the field of cognitive science with the "Tri-Level Hypothesis", which argued for the analysis of cognitive systems at three different levels of functionality: hardware implementation, representation and algorithm, and computational theory [7]. Later, Poggio proposed to extend this scaffolding with a fourth level encompassing learning [7]. We will see that this argument for a hierarchical structure in the study of minds is compatible with the spirit of the thesis which will be presented in detail later, while it is not particularly compatible with the alternative "lookup table" argument which we mention next.

2.4 Lookup Tables and Understanding

In *Why Philosophers Should Care About Computational Complexity* [8], Aaronson asks us to imagine a finite but very large "lookup table", filled with all possible question histories along with the appropriate next response, which could be utilized by a computer tasked with taking the Turing Test. By consulting the table every time it needed to answer a question, the computer would be able to successfully navigate the test without needing any complex internal representations which we might have expected would be necessary for cognition. This is a very clear way to express an argument which is similar in spirit to Searle's "Chinese Room" argument [2]. Later in this paper, I will analyze this argument in more detail.

3 Constraints on Mind Design

I will now describe three types of design constraints for minds: size, response time and capacity. Then, in Section 4, I will review possible measures of efficiency and complexity. In Section 5, I will present a unification of these themes: how to measure the degree of understanding present in a mind.

3.1 Size of Minds

In order to be physically instantiated in reality, a mind must "fit". Let us take a human brain, for example. The average brain has roughly 100 billion neurons and 10^{26} atoms and is able to perform its incredible cognitive feats with this amount of machinery. The aforementioned lookup table, on the other hand,

would require more storage space than is available in the entire observable universe [8]. Hence, it is clearly not a physically realizable mechanism for cognition; it is merely a philosophical fiction.

3.2   Mind Response Times

Furthermore, in order to be effective as an embodied agent, a mind must be able to respond to its environment within a sufficiently small timeframe for its responses to be useful. As a result, whatever mechanisms are used to implement cognition, they must be capabable of processing information sufficiently quickly.

3.3   Capacity of Minds

In order to be effective, a mind needs to possess the ability to remember salient aspects of the past, and to output adequate responses to a complicated array of stimuli. In order to satisfy these constraints, the mind must have sufficient storage capacity and sufficient information processing capacity.

4   Formalizing Efficiency

Before we can properly discuss the efficiency of minds, we need to discuss some measures of complexity that can be used as a quantitative measure of the efficiency of a system. I will present several measures from the fields of computational complexity theory and algorithmic information theory which I believe will be useful as a starting point, and describe them in the context of cognitive science.

4.1   Computational Complexity Theory

Computational complexity theory is concerned with the mathematical analysis of the complexity of algorithms and information processing systems. The *time* and *space* requirements of an algorithm are taken into account; in particular, the way the requirements vary in response to changes in the size of the inputs to the system is considered.

A core distinction is made between algorithms which run in *polynomial time* and algorithms which run in *exponential time* relative to the size of their input. They are divided into separate *complexity classes*, and only algorithms which run in polynomial time are considered to be feasible.

4.2   Algorithmic Information Theory

In algorithmic information theory, the descriptive complexity of concepts and objects is studied, and the notion of efficiency is formalized.

It is standard to discuss the complexity of strings, and in particular, binary strings. However, this should not be an issue, since we can consider a mind to be made up of a collection of discrete elements, and if it became necessary, we

could agree on some formal mapping between the configuration of a mind and a binary string that describes it.

4.3   Kolmogorov Complexity

The motivation behind Kolmogorov complexity is to describe the complexity of an object in terms of the length of the most concise description of the object. The *Kolmogorov complexity* [9, 10, 11] [12, Chapter 2] of a binary string is defined as the length of the shortest computer program which, when executed on a reference machine, produces the string as output. The reference machine can be taken to be a Turing machine, and the choice of reference machine is arbitrary and produces equivalent results up to a constant factor.

Hence, it is straightforward to imagine that we could quantify the Kolmogorov complexity of a mind as the length of the shortest computer program which, when executed on a reference machine, produces a complete description of that mind as its output.

Interestingly, we could take an alternative route, and choose to quantify the Kolmogorov complexity of the cognitive behavior produced by a mind instead. I will say more on this in the next section.

4.4   Resource-bounded Kolmogorov Complexity

The initial formulation of Kolmogorov complexity does not take into account the runtime of the program in producing the output, nor the amount of space it needs for its computations. This is clearly not sufficient for certain purposes; and, indeed, there exist variants of the measure, called *resource-bounded Kolmogorov complexity* [12, Chapter 7], which study the Kolmogorov complexity of an object conditional on specified bounds on space and time complexity.

If we adopt the alternative approach described in the previous section and measure the resource-bounded Kolmogorov complexity of the behavior produced by a mind, then it seems natural to set the space and time bounds analogously to the constraints on size and response time in mind design which we defined previously in Section 3.

4.5   Logical Depth

Another measure of descriptive complexity, the *logical depth* of an object, was proposed in [13] and is defined in [12, Chapter 7] as "the necessary number of steps in the deductive or causal path connecting an object with its plausible origin", where the origin is a more compressed description of the object.

As before, we could describe the logical depth of a mind itself; but, for our purposes, it will be more useful to consider the logical depth of the behavior produced by a mind, as a quantification of the complexity of that behavior.

An important consideration in these types of complexity measures is that, intuitively, we would like to assign a high measure of complexity to objects that are richly structured, while penalizing objects which are hard to predict but are

essentially random. This was one of the goals in the formulation of the measure of logical depth.

5 Measuring Cognition

With the preliminary background behind us, I propose the following conjecture quantifying the amount of *understanding* present in a cognitive system:

> The amount of understanding present in an intelligent system is directly proportional to the logical depth of its behavior and inversely proportional to the Kolmogorov complexity of the mind responsible for the behavior.

By requiring the mind to have low Kolmogorov complexity in order to have high understanding, scenarios such as the lookup table and the "Chinese Room" would no longer qualify as possessing high understanding. Furthermore, by requiring high logical depth on the part of the behavior that is produced, the system must arrive at complex, structured results requiring extensive deliberation in order to be said to harbor understanding. Aaronson proposed a similar concept in [8].

If we adopt this conjecture as a tool for understanding cognition, then it has interesting consequences for the debate between computationalism and connectionism as the most fruitful model for cognition. Rather than forcing us to decide between the two paradigms, it shifts the focus onto measuring the resulting system and the behavior it produces. As a result, we are free to combine computationalist and connectionist models of the mind into integrated cognitive architectures in whatever way is deemed most fruitful for producing complex behaviors from simple architectures.

6 Objections

One possible objection to the conjecture that I have presented is that Kolmogorov complexity is not a computable quantity: we can exhibit a short description of an object, thereby placing an upper bound on its Kolmogorov complexity, but we can never be certain that a yet shorter description doesn't exist. However, due to the fact that we can design algorithms to place ever tighter bounds on the quantity, in technical terms the Kolmogorov complexity is said to be *upper semi-computable*, and compression algorithms are useful method for approximating it. As a result, I think this objection should not be a serious obstacle.

Another objection that might be raised is that we are not really capturing the essence of understanding in this definition, but rather analyzing a peripheral issue. It may be the case that the conditions I have outlined are necessary, but not sufficient, conditions for understanding to be present in a cognitive system. This merits further consideration, but I believe that the arguments

presented herein would still serve a useful purpose in providing a framework for analyzing understanding, and in order to strengthen the objection, a more rigorous definition of what is meant by understanding would need to be provided.

7   Conclusion

In this paper, I discussed the issue of measuring the degree of understanding present in a cognitive system and considered a methodology to refute the popular argument from Searle who attempted to eliminate the possibility of machines harboring true understanding.

I began with an overview of the field of artificial intelligence and cognitive science, starting with Alan Turing and the paradigms of computationalism and connectionism, along with an overview of the tri-level hypothesis and the basic argument from Searle that I set out to refute. I then presented a concise catalogue of constraints on mind design, and an overview of formal methods for measuring efficiency and complexity selected from the fields of computational complexity and algorithmic information theory.

Building upon these concepts, I presented a conjecture for a more rigorous definition of understanding in a cognitive system that takes into account resource constraints and analyzes both the cognitive system itself and the behavior that it produces. Finally, I analyzed and responded to possible counterarguments regarding computability and the meaning of understanding.

It is my hope that the conjecture I have presented and the analysis of issues from cognitive science in the language of computational complexity and algorithmic information theory could provide inspiration and provoke new results taking advantage of the cross-pollination of ideas between the disciplines.

References

[1] Alan M Turing. "Computing machinery and intelligence". In: *Mind* 59.236 (1950), pp. 433–460.

[2] John R Searle. "Minds, brains, and programs". In: *Behavioral and brain sciences* 3.03 (1980), pp. 417–424.

[3] Allen Newell and Herbert A Simon. "Computer science as empirical inquiry: Symbols and search". In: *Communications of the ACM* 19.3 (1976), pp. 113–126.

[4] Jerry A Fodor. *The language of thought*. Vol. 5. Harvard University Press, 1975.

[5] David E Rumelhart, James L McClelland, PDP Research Group, et al. *Parallel distributed processing*. Vol. 1. IEEE, 1988.

[6] David E Rumelhart. "The architecture of mind: A connectionist approach". In: *Mind readings* (1998), pp. 207–238.

[7] David Marr. "Vision: A computational investigation into the human representation and processing of visual information". In: *Inc., New York, NY* 2 (1982).

[8] Scott Aaronson. "Why philosophers should care about computational complexity". In: *In Computability: Gödel, Turing, Church, and beyond (eds.* Citeseer. 2012.

[9] AN Kolmogorov. "Three approaches to the quantitative definition of information". In: *Problems of Information Transmission* 1.1 (1965), pp. 1–7.

[10] Ray J Solomonoff. "A formal theory of inductive inference. Part I". In: *Information and control* 7.1 (1964), pp. 1–22.

[11] Gregory J. Chaitin. "On the Length of Programs for Computing Finite Binary Sequences". In: *Journal of the ACM* 13.4 (1966), pp. 547–569. ISSN: 00045411.

[12] Ming Li and Paul Vitányi. *An introduction to Kolmogorov complexity and its applications*. Springer Science & Business Media, 2013.

[13] Charles H Bennett. "Logical depth and physical complexity". In: *The Universal Turing Machine A Half-Century Survey* (1995), pp. 207–235.