Phenomenology, Representations and Complexity

Achim G. Hoffmann
Technische Universität Berlin, Department of Computer Science, FR 5-11
Franklinstr.28/29, W-1000 Berlin 10, Germany, E-mail: achim@cs.tu-berlin.de

Abstract. The paper refutes the general phenomenological argument that knowledge cannot be completely represented by symbols and, hence, symbolic AI does not work. Moreover, the viewpoint of algorithmic information theory is introduced for considering the limitations of artificial intelligence. Based on the formal considerations of the notion of algorithm we will come to new conclusions from the phenomenological critic at the possibility of (symbolic) AI.
Keywords: Phenomenology, Representations, Algorithmic information theory, Physical symbol systems

1 Introduction

In this paper we consider some fundamental issues of artificial intelligence. Among these issues will be problems of formal limitations of AI, i.e. discussing the *physical symbol system* hypothesis [8]. We will respond to philosophical critics as these of Dreyfus (see e.g. [1]) or these of Winograd and Flores (see [12]). These critics are closely connected to the problem of knowledge representation.

But most importantly, in section 4.3, we will advocate a new consequence from the phenomenological arguments against AI.

The paper is organized as follows. In the following section the formal framework for our viewpoint of AI is given. Section 3 discusses the implications of our viewpoint to philosophical critics from a phenomenological point of view. In section 4, the implications of different philosophical views of the representation problem are discussed. Section 5 contains the conclusions.

2 Formal Foundations

Traditional approaches to AI are explicitly or implicitly based on the claim that the human mind is just some kind of symbol manipulating entity or at least can accurately simulated by a such. This claim is also called the *physical symbol system hypothesis* [8]. and has been heavily attacked by, e.g. Dreyfus [1] and Winograd & Flores [12] from a phenomenological point of view. We identify the physical symbol system hypothesis with the claim that the human mind can be adequately simulated by an appropriate Turing machine. In the following we will take a closer look to the Turing machine in connection with the *Invariance theorem* [6] of algorithmic information theory.

2.1 Universal Turing machines

Following the introduction of the Turing machine by A.M. Turing [10] it has been recognized that there also exist various kinds of *universal Turing machines.* A universal Turing machine U is a Turing machine that is able to simulate any particular Turing machine whose Turing machine table is provided on U's input tape. Since a universal Turing machine U can execute arbitrary algorithms, one may wonder how long the input to U has to be in order that U computes a particular output string s.

2.2 Algorithmic information theory

The minimal length of an input string to U (program) to turn U to print an output string s, is measured as the algorithmic information of s. For technical convenience, we consider only finite or infinite **binary** strings consisting of '0's and '1's not containing any blanks or other symbols. The length of a binary encoded program p is denoted by $|p|$.

Definition 1 *The length of the shortest program for constructing s is called its* **Kolmogorov complexity** *$K(s)$.*

According to the Invariance Theorem (see e.g. [6]) the particular kind of considered universal Turing machine U makes only a difference of a certain constant c. Moreover, there are 2^n different binary strings of length n. Since each string requires another program there are strings s of length n whose Kolmogorov complexity $K(s)$ is at least n. Note: One program produces exactly one string - there is no input to the program. Moreover, *most* strings of length n have Kolmogorov complexity $K(s)$ of n.magnitude

European Conference on Artificial Intelligence 1992 pp. 610-614

Examples: Strings as '11111111111111' or '00000000000' or '1010101010101010' etc. are simple strings, since their description by encoding a program which outputs the string under consideration is short. Mainly, only the length of the string is required, i.e. $\log_2(length\ of\ string)$ bits are sufficient. In addition, an indication whether '0's , '1's or alternating '1's and '0's should be printed is necessary.
In contrast to the strings above, strings like '1011001001110110010111101010' are more complex, i.e. require a longer program for getting printed, since the structure needs a particular description.

2.3 Algorithmic information for *describing* the mind

According to Turing's paper *Computing machinery and intelligence* [11], we assume that for considering intelligent behavior of agents our considerations can be restricted to agents which communicate with their environment through finite strings of symbols from a finite alphabet.[1] In other words, we can say that an agent behaves intelligent, if it shows a certain appropriate output to the input supplied to it. Furthermore, all observable behavior - intelligent or not - is a mapping from some finite input to some finite output. It appears to be no restriction, if it is assumed that the length of the input to an agent through its lifetime is finite. However, then there is a finite number of possible different inputs to such an agent. Thus, the required I/O-function that models the *intelligence* of any particular agent is simply some function from a finite number of input symbols to a finite number of output symbols. Hence, there definitely exists some TM that models intelligent behavior. Furthermore, this I/O function can be encoded as a binary string $s(f_{Int})$, e.g. as the encoding of an appropriate TM table. Therewith f_{Int} has a certain Kolmogorov complexity $K(s(f_{Int}))$.
In other words, the goal of AI (at least for the engineering approach) is to develop a physical implementation of such a function f_{Int}. It is clear that f_{Int} can be represented, e.g. as a binary string for feeding some universal Turing machine U. Assume the number of possible binary input symbols n_i is upper bounded by $n_i \leq 10^{15}$. The number of output symbols through the agents lifetime n_o may be upper bounded by $n_o \leq 10^{10}$. Then the length $|s(f_{Int})|$ of the binary encoded function f_{Int} is upper bounded as follows.

$$|s(f_{Int})| \leq 2^{(10^{15})} \times 10^{10} \approx 10^{100000000000000}$$

[1]See the conclusions for remarks on these restrictions.

Certainly $|s(f_{Int})|$ is extremely huge. However, the function f_{Int} has not to be represented explicitly as a binary string but may be *compressed* as well as possible by describing the function as rules for entire classes of input strings (substrings). The size of the most compressed form of *any* representation is *lower bounded* by the Kolmogorov complexity $K(s(f_{Int}))$.
In the following section, the universal Turing machine together with complexity considerations will be opposed to phenomenological arguments against the physical symbol system hypothesis.

3 Phenomenology and Complexity

Arguments which indicate that human intelligence does not essentially rest on representations, has been put against the physical symbol system hypothesis by e.g. Dreyfus [1] or Winograd & Flores [12]. Phenomenology argues that human cognition and understanding is basically connected with social interaction. Moreover, the human mind only starts reflecting about objects and their relations in an outer world in situations of breakdowns. That is, in situations where the usual acting cannot proceed as expected. E.g. only when I cannot open the door, I remind the door 'explicitly' (consciously) in its physical appearance. Moreover, phenomenology argues that human values and goals, and similarly ontological considerations, are *implicitly* contained in the way people act within their environment as well as in the way people look at the world, i.e. which objects exist for them in which situation and which properties the objects have. In contrast to that, one has argued, computers *have* to represent 'values' and 'goals' as well as their 'view' of the world *explicitly* by symbols and therewith computers cannot behave like humans. Similarly, Frixione et al. [3] claim that intelligent systems must have a *subsymbolic* level of activity.

The implicitness of symbol manipulation

However, let us take a closer look to the Turing machine model: The way a particular Turing machine T works is only *implicitly* contained in the way T manipulates its symbols on its tape. This way of manipulating symbols is encoded in T's Turing table instead of being represented *explicitly* by symbols on T's tape. However, the Turing table has yet to be interpreted. And an interpreting agent has to 'know', how to interpret such a table. But a correct interpretation - even an interpretation at all - is only possible with some *implicit* 'knowledge'. about the way how to interpret the symbols that constitute the representation of a Turing table. Moreover, this implicit 'knowledge' is given in the way a universal Turing machine manipulates

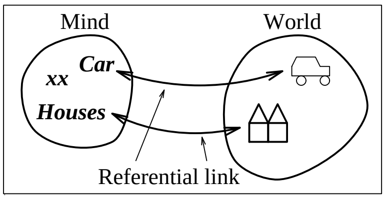

Figure 1: The basic idea of representation. Symbols in the mind represent objects of the world.

the symbols on its tape. Due to the infinite regress one would encounter, it follows that not all necessary 'knowledge' can be represented by symbols.

The complexity of thoughts

When one makes *complexity* an issue, one recognizes that the *conscious* human thought processes appear of rather limited complexity. Simply because our consciousness can evidently only cover a very limited number of items simultaneously. As a consequence, the subconscious processes (i.e. the implicit knowledge) has to be of large complexity in terms of algorithmic information theory. Thus, in fact these subconscious processes bear a great impact on the emerging conscious thoughts, which has been pointed out by phenomenology. The universal Turing machine is - so to say - a minimal kernel (which bears a minimal complexity) of 'implicit knowledge' from which we can simulate any 'implicit knowledge' of arbitrary complexity. Due to the large difference in complexity of the universal Turing machine and subconscious processes, the possible correspondence of the symbol system and human thoughts is so less obvious.

Moreover, in contrast to the rather convincing arguments of phenomenology the discovery of the universal Turing machine shows that there exists some 'universal implicit knowledge' that allows to represent arbitrary knowledge explicitly - anyway whether it *appears* to be implicit or explicit knowledge.

4 Representations and Complexity

Representation appears to be a central notion in cognitive science as well as in artificial intelligence. For a very simple version, see Figure 1.

4.1 Representations in the model-theoretic sense

The classical approach considers the real world as a logical model in which logical expressions may be interpreted. The standard model-theoretic semantics are given by an interpretation function of the atomic syntactical symbols of a given language L into a domain of individuals D. The most important point is, that for each compound syntactical expression its interpretation in the domain D is defined by rules that correspond one-to-one to the syntactical rules for building the compound expressions.
However, this approach shows a convenient way for building AI systems. Of course, this way is only applicable in domains where a clear-cut ontology is available and deductive reasoning suffices for the tasks that are supposed to be accomplished by the system.

Shortcomings

Unfortunately, many important potential AI applications are of a different nature; at least some inductive or analogical reasoning, nonmonotonic reasoning or reasoning under uncertainty and vagueness is required.

For example in inductive reasoning, usually there is an ambiguity among different hypotheses, when generalizing inferences should be performed. Thus, there is a need for expressing some *preference relation* among competing hypotheses.
As indicated in, e.g. [4] for symbolic approaches and in [5] for non-symbolic approaches, there is no uniform inductive inference procedure, which describes human inductive reasoning behavior. The human inductive reasoning behavior contains some complex (implicit) preference relation on the competing hypotheses. However, any preference relation among a finite number of hypotheses can be represented by symbols, although these symbols do not refer to some 'natural' ontological entities, i.e., to the ontological entities of the domain of inductive reasoning. Moreover, it is quite clear that such a preference relation contains a large amount of information - otherwise a uniform inductive inference procedure would exist.
The same problem exists for reasoning under uncertainty: Human reasoning incorporates a lot of interdependencies of subjective probabilities, when probability estimates of combined uncertain events are made. However, the general applicability of this model-theoretic approach has been rejected by phenomenology as well as by the late Wittgenstein [14] himself; once being a strong advocate of the model-theoretic idea [13].

4.2 Alternative approaches

Auto-organizative systems

Another rather popular approach today, is the idea of emphasizing the auto-organizative capacities of biological systems, e.g. the human mind. Particularly, Maturana and Varela [7] propell this auto-organizative tradition exploring the capability of biological systems to create *their own world of experience*. In their view,

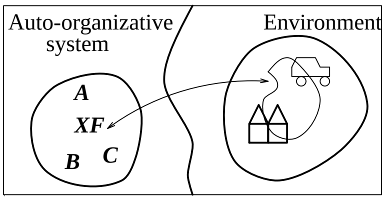

Figure 2: The schema of auto-organizative systems. The auto-organizative system is conceived as completely autonomous. I.e., there is no direct coupling between the cognitive structures in the system and the world with which the system interacts. Thus, the ontological entities of the world are <u>not</u> one-to-one reflected by the cognitive concepts of the system.

mental representations do not replicate or even capture information from some external reality, but rather emerge by means of environmental interactions. Here, these interactions have just a 'triggering role' and not an 'instructive' one. No informative exchange followed by a similarity or structural matching occurs, but rather a simple coupling between environmental perturbations and the mind's generation of novel ways of being. The mind is autonomous and finds alone, in an infinite repertoire of possible behaviors, the one which is simply associated with an environmental perturbation. Association does not mean correspondence. No structural matching is required between an environmental signal and any mental structure. See Figure 2. Roughly spoken, the semantic of the subject's representation turns out to be committed in a larger degree to the subject than to the object. First results on the complexity of possibly acquired mental structures in auto-organizative systems are obtained in [2]. These results indicate that simple auto-organizative principles are rather limited in producing meaningful mental structures.

Quine's ontological relativism

The logician and philosopher W.v.O. Quine rejected the model-theoretic absolute ontology in 'real world applications'. He advocates much more a relativistic viewpoint [9]: *'To be is to be the value of a variable.'* Which means, that ontological entities are determined by the *structure of the language* that is *used* for describing phenomena in a certain domain. However, there is always a choice which particular structure of language is used. The choice depends on the culture and the conditions of usefulness. Hence, ontology can change, when the needs of the respective community change. This is what has been emphasized by phenomenology.

Moreover Quine recognized, that the reference of linguistic expressions cannot be determined by considering single expressions. It is rather necessary, to consider the use of the entire linguistic framework around a particular expression in order to determine the actual meaning of an expression like: 'It is not allowed to smoke in the aircraft.' E.g., in order to determine, what an *aicraft* is, one has to know what 'to fly' means, etc.

4.3 Consequences for physical symbol systems

Phenomenology as well as the auto-organizative systems approach as well as Quine's ontological relativism indicate that there is yet a significant amount of auxiliary information necessary for physical symbol systems in order to behave comparable to the human mind.

Basically, the choice for setting up a particular ontology, respectively the appropriate language in a particular situation respectively the development of autonomous subjective concepts has to be guided by additional information.

If the observation of phenomenology holds, that these processes appear not to be rule-guided, this merely implies that the required information is of rather large Kolmogorov complexity. This contrasts the standard implication, that is drawn from these philosophical arguments, namely that physical symbol systems are not appropriate for modeling the mind.

The consequence of the objections in this paper is actually the following:

The additional information has to be encoded in a rather comprehensive way, in order to allow an expert or a knowledge engineer to implement the required information. However, the symbols that can carry the required information will not refer to the ontological entities of the respective domains. But these symbols will rather refer to some abstract entities like certain preference relations among competing hypotheses for inductive reasoning. Thus, perhaps the human expert or at least the knowledge engineer should learn to think in abstract entities of this kind. If the knowledge engineer succeeds with that attempt, he/she would greatly enhance the current scope of knowledge transfer, leading to the implementation of much more capable physical symbol systems.

Thus phenomenology indicates, that AI should develop appropriate frameworks for encoding the mentioned additional information.

5 Conclusions

In this paper, we have considered the notion of algorithmic information within the discussion on the limitations of AI. Instead of arguing qualitatively about what computers can't do, we proposed a quantitative consideration. Thus, the question whether machines can behave like an intelligent human being should be replaced by the following question:

> Can a given task (that can be accomplished by an intelligent human) be accomplished by a machine whose program is at most of length k ? (For some particular k.)

Moreover, the phenomenological critic at the physical symbol system hypothesis has been rejected in favour of considering the quantitative aspects of intelligent systems.

In section 4, we discussed the consequences of various critics at the classical representational view. It turned out, that there is yet a need for encoding a significant amount of auxiliary information. The symbolic encoding of this information does not refer to any ontological entities of the respective domain, but to *abstract* entities like preference relations. This fact indicates the need in artificial intelligence for developing appropriate frameworks in which the encoding of such necessary auxiliary information can be done. Within these frameworks, it should be possible to encode the required knowledge comprehensively. Moreover, we can say that phenomenology shows that AI needs frameworks for encoding the auxiliary information for (non-deductive) guiding reasoning processes. This conclusion contrasts the manyfold claims, that there is a need for so-called 'sub-symbolic' approaches like neuronal models of computation, which appear more or less incomprehensible in what will be finally computed.

Possible limitations of the presented considerations

One may argue that this model is too restrictive since it does not cover asynchronous or analogue signals. When we consider systems with analogue receptors and anlogue effectors we cannot immediately apply algorithmic information theory. However, we can observe the following if we distinguish different ways of behavior: For each particular actual behavior of such a (possibly self-organizing) system, there has to be a particluar system design. Hence, we may count the number of different system behaviors and obtain a corresponding number of different designs. I.e. we encounter very similar lines of thought as in algorithmic information theory.

Moreover, the assumption that a design of an *intelligent* system is rather simple implies that all the many different and more complex designs result in dumb system behavior !

Acknowlegdement I would like to thank the VUB AI-group at the Vrije Universteit Brussel for fruitful discussions on this paper.

References

[1] H. L. Dreyfus and S. E. Dreyfus. *Mind over Machine - The Power of Human Intuition and Expertise in the Era of the Computer.* Free Press, 1986.

[2] R. Freivalds and A. G. Hoffmann. An inductive inference approach to classification. In *Proceedings of the 3^{rd} Workshop on Analogical and Inductive Inference*, Schloß Dagstuhl, Germany, October 1992. Springer-Verlag. (to appear.)

[3] M. Frixione, G. Spinelli, and S. Gaglio. Symbols and subsymbols for representing knowledge: a catalogue raisonné. In *Proceedings of the 11^{th} International Joint Conference of Artificial Intelligence*, pages 3–7, 1989.

[4] A. G. Hoffmann. General limitations on machine learning. In *Proceedings of the 9^{th} European Conference on Artificial Intelligence*, pages 345–347, Stockholm, Sweden, August 1990.

[5] A. G. Hoffmann. On computational limitations of neural network architectures. In *Proceedings of the 2^{nd} IEEE Symposium on Parallel and Distributed Processing*, pages 818–825, Dallas, Texas, USA, December 1990. IEEE.

[6] M. Li and P. M. B. Vitanyi. Two decades of applied Kolmogorov complexity. In *Proceedings of the 3^{rd} Annual Conference on Structure in Complexity Theory*, pages 80–101, 1988.

[7] H. R. Maturana and F. J. Varela. *The Tree of Knowledge. The Biological Roots of Human Understanding.* New Science Library, 1987.

[8] A. Newell. Physical symbol systems. *Cognitive Science*, 4:135–183, 1980.

[9] W. v. O. Quine. *Ontological Relativity and Other Essays.* Columbia University Press, New York, 1969.

[10] A. M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 2(42):230–265 and (43) 544–546, 1937.

[11] A. M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950.

[12] T. Winograd and F. Flores. *Understanding Computers and Cognition: A new Foundation for Design.* Norwood Publisher, 1986.

[13] L. Wittgenstein. *Tractatus logico-philosophicus.* 1922.

[14] L. Wittgenstein. *Philosophical Investigations.* Macmillan, Oxford, 1953.