ON THE ADEQUATENESS OF AI-SYSTEMS

STEFFEN HÖLLDOBLER

Fachgebiet Wissensverarbeitung, Fakultät Informatik, TU Dresden
D-01062 Dresden, Germany

and

MICHAEL THIELSCHER

Fachgebiet Intellektik, Fachbereich Informatik, TH Darmstadt
Alexanderstraße 10, D-64283 Darmstadt, Germany

In: I. Plander, ed., Proc. of the Int. Conf. on AIICSR, Smolenice Castle, Slovakia, Sep. 1994.

Research in Intellectics, ie. Artificial Intelligence (AI) and Cognition, has always focussed on two primary goals, viz. to understand how humans behave intelligently and to make systems behave intelligently. Unfortunately, many researchers had and still have the habit of following only one of these goals although it is not at all clear that one can get closer towards one of these goals without getting closer to the other one at the same time. Can we expect to build a robot which applies common sense reasoning to solve a given task in a previously unknown environment if we do not understand how the human nervous system perceives, processes, and acts in a similar scenario? Can we expect to fully understand the human nervous system if we cannot build a machine which behaves similarly? Machines already outperform humans in many simple manual and some simple intellectual activities. We do not want to give up these advantages. On the contrary, we want our machines to perform intellectually more demanding activities like planning, diagnosing, or common sense reasoning. What is the kind of similarity between a machine and a human that we are aiming at?

Computational models have widely been used in modern psychology to explain perception, thought, and behavior. Moreover, many researchers in the Neurosciences believe that the activities of our brains and nervous systems are computations in the very same sense as computations are the kind of activities performed by computers (see eg. [3]). One should observe that this believe has various consequences. If we additionally assume that all computations performed by a computer are formalized — and this assumption is widely accepted in Computer Science and AI – then it follows on the one hand that anything a machine could possibly achieve in simulating human intelligence can be formally described [10] and, on the other hand, that intelligent behavior can be formally described [9].

A computational model usually consists of three levels, viz. the abstract or specification level, the algorithmic level, and the physical level. This is independent of whether we follow a top–down or bottom–up approach. For example, in AI one often starts with a specification, eg. a certain logic, develops an algorithm for the specification, eg. a calculus together with appropriate tactics and strategies, and finally implements the algorithm on a computer. On the other hand, in Cognitive Science one often starts with a biological system, eg. the human, analyses the hardware, ie. the receptors, the neurons, their connections, etc., tries to understand the "algorithms" which are "executed" on this hardware, and finally attempts to specify the behavior of the biological system.

From Computer Science we learn that theoretically the three levels are independent and, in particular, that algorithms can be designed independently of the underlying physical system. As a typical computer as well as a neuronal system are universal one could argue, that what really matters are the specification and the algorithmic level, and if an algorithm is correct and complete with respect to the given specification, then it does not matter whether it is run on a silicon–based computer or an a neuronal system. Is this really true?

Let us have a look at an example. Suppose we want to design a system which recognizes persons and, in particular, recognizes our grandmother. From an AI point of view we would specify the visual properties of our grandmother in, say, some logic language and, whenever a visual scene is given, we match the persons in the visual scene against the stored visual properties of our grandmother. The matching task can be encoded such that it is essentially a validity problem in the logic used and, hence, some calculus can be used to compute the answer. One should note that the task is quite tricky as we know more persons than just our grandmother and we do not only consider standardized pictures of our grandmother but want the system to recognize grandma independently of whether she is sitting in the church, riding a bicycle, or cleaning up her apartment. We should not be surprised to discover that the time to find an answer is exponential with respect to the size of the knowledge base. On the other hand experience tells us that it takes usually only a few milliseconds to recognize our grandmother in the real life even if she is standing in the middle of a crowd of people.

More generally, many tasks which are specified naively in some formal language turn out to be exponential in time whereas humans solve these tasks in the millisecond range. As real neurons are quite slow computational devices we can conclude that only few steps may have taken place until our nervous system has reached a conclusion (see eg. [4]). Although the nervous system is massively parallel the results from Theoretical Computer Science tell us that parallelization alone cannot solve the apparent gap between the performance of a naivly designed AI system and a human.

There are a variety of reasons which may explain this gap. The used specification language may be too expressive. The given specification may be faulty, the selected algorithms may not be appropriate, the demanded properties of the algorithm like, for example, completeness may be too severe, the applied tactics and strategies may be clumsy, the chosen data structures and the machine architecture may be unsuitable, etc. Obviously, these causes do not only concern the implementation level but the other levels as well. We do not want to analyse each of these causes independently and in detail as we believe that we have to consider all of them in order to get closer to the ulimate goal of Intellectics. Rather we would like to discuss whether there is a kind of abstract or general property that AI–systems and methods should satisfy. Clearly, AI–systems and methods should be correct or, at least, should approximate a correct solution arbitrarily accurate. But what about other properties?

In this talk we will discuss the adequateness of AI–systems and methods. Following [2] a (proof) method is *adequate* if — roughly speaking — for any given knowledge base, the method solves simpler problems faster than more difficult ones. Thereby simplicity is measured under consideration of all (general) formalisms available to capture the problem.

Before we discuss the notion of adequateness by means of four example, we will make some general remarks. The systems, methods, knowledge bases, and problems referred to in the previous characterization are assumed to be formalized objects as we believe that all computations done by a computer are formalized. Although adequateness is only defined for proof methods in [2] this is by no means a restriction and the definition applies to other formalized methods as well. As already stated in [2], the definition hinges on the given knowledge base as one could always solve a problem in one step if its solution is added to the knowledge base. Last but not least, one should observe that if we accept the hypothesis that our brain and nervous system computes then our brain and nervous system is itself a formalism and has to be taken into consideration.

Specification Languages may be Inadequate. Our first example, which is concerned with image interpretations and that is taken from [13], illustrates inadequteness on the most abstract level, namely the specification language itself that describes the problem.

The authors of [12] present a logical specification language for the problem of interpreting simple hand-drawn sketch maps consisting of arbitrary chains. On condition that some system has numbered the chains and abstracted form the image a set of relations between theses chains such as c_1 *and* c_2 *cross each other* or c_1 *meets* c_2 etc., the task is to determine whether a chain denotes a road, a river, or a shore. In [12] this task is formalized as a problem to generate models of an appropriate propositional formula which contains the various relations which characterize a particular image along with a set of general axioms — i.e. constraints such as *rivers cannot cross each other* etc. — restricting interpretations to realistic scenarios.

In [13] it is shown that the problem of generating models for such a propositional formula is intractable in general as it is NP-complete. On the other hand, [13] also elaborates the reason for this intractability, namely the presence of unnatural and unexpected interpretations, viz. the unlikely coincidence of the source of a river with some point of a road. It is shown that whenever such an unintuitive coincidence is ignored, by providing an additional constraint, then the task to compute models of a corrsponding propositional formula becomes linear wrt the number of chains detected in the image. Hence, it turned out that the unrestricted specification language is inadequate as it causes the problem to be intractable by the consideration of unnatural solutions.

Calculi may be Inadequate. The inadequateness of a calculus regarding a specification shall be illustrated by a well-known calculus designed for reasoning about actions and change in dynamic systems.

One application of this kind of problem solving consists in determining the goal state of such a system given an initial state along with a particular sequence of actions whose execution causes changes of system states. This task is usually called *temporal projection*. A difficulty when trying to formalize this kind of reasoning using a logical specification is the dynamical aspect of state transition. In the *Situation Calculus* [7, 8] a state is described by a number of facts which are represented by atomic formulas, i.e. by a number of properties which hold forever and, thus, have to be restricted to the particular state in which they are assumed to hold by employing an additional argument. This representation leads to the famous *frame problem*, i.e. to the question how to formalize the *inertia assumption* stating that facts which are not affected by the execution of an action keep their validity. Each concrete implementation of the Situation Calculus has to include additional axioms to express this assumption, e.g. *successor state axioms* as in [11].

In [6] we illustrate by a simple example that the Situation Calculus and especially the method presented in [11] is inadequate. Informally, the example is as follows: Let each cell of some array be initialized with an arbitrary integer number, and let an action sequence be given which increases the value of each cell by 1. The temporal projection problem is to determine the goal state of the array. It is obvious that the time complexity of this task is linear wrt the size of the array.

However, the application of the approach [11] to this scenario requires quadratic time as it is necessary for *each* single increment operation to apply a particular instance of a successor state axiom to *each* fact describing the contents of a cell which is not affected by the operation. As the Situation Calculus in general is based on the technique to associate an additional state argument to each fact, this justifies the claim that this calculus is not an adequate computational mechansim and, hence, implementations which are based on this method cannot constitute adequate AI-systems.

Adequateness Implies Massive Parallelism. Driven by the observation that humans can draw a variety of inferences effortlessly, spontaneously, and with remarkable efficiency, Lokendra Shastri's and Venkat Ajjanagadde's goal was to identify a class of problems and to specify a computational model such that it is biologically plausible, matches psychological data, and answers queries efficiently [14]. Efficiency is defined with respect to a knowledge base, which is assumed to be quite large. The number of processors is bound by the size of the knowledge base and the time to answer a query should be much smaller than — or even independent of — the size of the knowledge base.

The class of problems considered in [14] is a class of definite formulae which is queried by a goal clause. The class is restricted such that all branches in the search tree can be investigated in parallel. The computational model is a massively parallel, connectionist one and much effort has been put into the model to make it biologically plausible. It is an open question whether the problems defined by Shastri and Ajjanagadde are problems which humans can solve efficiently and effortlessly and, henceforth, the question whether the model matches psychological data remains to be tested.

In [1] a formal semantics for Shastri's and Ajjanagadde's computational model was defined by showing that reasoning in this model is nothing but reasoning by reduction in a standard first–order calculus. But, in order to meet the time constraints this first–order calculus has to be implemented in parallel, which can be done along the lines outlined in [5]. In other words, adequateness implies massive parallelism.

Seeking Adequate Computational Models. One of the most active research areas in Cognition is Vision. How does our nervous system perceive and identify the thousands of objects encountered each day? The favored model among psychologists, physiologists and reasearchers in AI consists of two or more layers. In the first layer features like color, orientation, or size are extracted from patterns of light and in the higher layers these features are combined to form objects, figures, ground, etc. (eg. [15]). Two major constraints govern the research in this field. Our nervous system has only a limited number of neurons and, thus, we cannot have a neuron for each possible combination of features. Experiments can be made to time subjects on certain visual tasks and computational models should meet these time constraints. For example, a person is asked to find a certain object within a set of distractors. If the object differs in a simple feature from all distractors, then the object can be found almost immediately and independently from the number of distractors. Thus, a kind of parallel search must have been taken place. If, however, the object differs in a conjunction of features, then the time to find the object often depends on the number of distractors. Hence, a kind of serial search must have been performed. This is clearly a simplified analysis and conjunction search is much more difficult (eg. [16]). In any case, the goal is to develop adequate computational models.

1. A. Beringer and S. Hölldobler. On the adequateness of the connection method. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, pages 9–14, 1993.
2. W. Bibel. Perspectives on automated deduction. In R. S. Boyer, editor, *Automated Reasoning: Essays in Honor of Woody Bledsoe*, pages 77–104. Kluwer Academic, Utrecht, 1991.
3. P. S. Churchland and T. J. Sejnowski. *The Computational Brain.* Bradford Book, The MIT Press, 1992.
4. J. A. Feldman and D. H. Ballard. Connectionist models and their properties. *Cognitive Science*, 6(3):205–254, 1982.
5. S. Hölldobler. Automated inferencing and connectionist models. Technical Report AIDA–93–06, Intellektik, Informatik, TH Darmstadt, 1993. (Postdoctoral

Thesis).
6. S. Hölldobler and M. Thielscher. Properties vs. resources — Solving simple frame problems, 1994. (submitted).
7. J. McCarthy. Situations and actions and causal laws. Stanford Artificial Intelligence Project: Memo 2, 1963.
8. J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of Artificial Intelligence. In B. Meltzer and D. Michie, editors, *Machine Intelligence 4*, pages 463 – 502. Edinburgh University Press, 1969.
9. A. Newell. Physical symbol systems. *Cognitive Science*, 4:135–183, 1980.
10. Z. W. Pylyshyn. *Computation and Cognition — Towards a Foundation of Cognitive Science*. The MIT Press, 1986.
11. R. Reiter. The frame problem in the situation calculus: A simple solution (sometimes) and a completeness result for goal regression. In V. Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation — Papers in Honor of John McCarthy*, pages 359–380. Academic Press, 1991.
12. R. Reiter and A. K. Mackworth. A logical framework for depiction and image processing. *Artificial Intelligence*, 41:125–155, 1989.
13. B. Selman. Domain-specific complexity tradeoffs. In *Proceedings of the European Conference on Artificial Intelligence*, 1994.
14. L. Shastri and V. Ajjanagadde. From associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioural and Brain Sciences*, 16(3):417–494, September 1993.
15. A. Treisman. Properties, parts, and objects. In K. Boff, L. Kaufman, and J. Thomas, editors, *Handbook of Perception and Human Performance*, chapter 35, pages 1–70. Wiley, 1986.
16. A. Treisman and S. Sato. Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, 16:459–478, 1990.