The Science of a “New AI”

Erich Prem
Austrian Research Institute for Artificial Intelligence
Schottengasse 3, A-1010 Wien, Austria
Email erich@ai.univie.ac.at

December 15, 1992

Abstract

This paper seeks to describe the science of a “New AI”. It is explained why this new development is belonging to Artificial Intelligence and why it should be called “new”. We establish a list of criteria, which a new AI must satisfy. Several existing mainstreams of AI are revisited in the light of this explanation. Finally, we describe one bottom-up approach to symbol grounding in more detail and point out implications of such a model, which can be seen as a step towards “new” production systems or heuristics.

1 Introduction

A brief glimpse at those publications which regard themselves as belonging to Artificial Intelligence (AI) or which are classified as such by AI researchers makes it clear that AI is a wide-spread discipline covering many different subjects. Large AI conferences consist of symposia that cover a wide thematic range from mere formal logics to machine learning. It may thus be surprising that there are scientists who—in spite of the above mentioned variety—try to introduce a “new AI”. The question arises, as to whether there is space for a new AI besides the old one. This will be subject of our investigation in this paper together with the clarification of problems, methods and goals of a “new AI”. We also seek to clarify relationships between this and other scientific fields of research. The existing different approaches to a new AI are described in order to make clear what they have in common and what they possibly can achieve.

Of course, we ought to begin with a few words on "old" (or classic) AI. Old AI—and here it does not differ from its new counterpart—follows the goal of *making computers smart.* Therefore, it is a classical engineering science and constructive in its character. AI tries to enable computers to do whatever humans do, with the *restriction to certain specific domains of human activity.* Classic AI teaches us that such systems must be knowledge-based. Consequently, they must possess a structure for knowledge representation and a mechanism for the generation of new knowledge (or the desired behavior). This knowledge, according to old AI, is to be partitioned into small, understandable units, which work together in a linguisticly describable way in order to yield the desired results. Thus the method to a large extent consists in problem analysis and system synthesis by means of necessarily reductive constructs. Seemingly, the most appropriate of such constructs is the *rule*, which can be used for problem analysis and for the constructive description of a solution. The rule itself mostly connects two predicates, which again are mainly made up from symbols. Closely related constructs are frames, scripts, semantic networks etc.

However, this is not the right place to repeat all the critiques of such a method. Criticism especially arises whenever the old AI makes efforts to use its systems as explantions for human cognitive phenomena[1]. Instead, we want to point out what is new about new AI, what is not and how the differences arose.

2 Is the "New AI" AI?

The new AI mainly uses new *methods* (namely neural networks) to *model* cognitive and, hence, *intelligent* systems. Therefore, it is certainly *AI*. "New" AI becomes active in a constructive way to yield a model of cognition. By virtue of the specific goal- or behavior-oriented view of computer science this means: to yield a model of complex behavior which is difficult to simulate

[1]See e.g. H.Dreyfus *Why Computers Can't Think.* Macmillan 1984. Also: P.Smolensky "Connectionist AI, Symbolic AI, and the Brain." in *Artificial Intelligence Review 1* 1987, 95–109.

and to predict, but nevertheless behaves according to some prespecified goals. If goals would be of no importance at all, new AI could hardly be part of a technological endeavour. However, it should be noted that the "new" AI stresses the *cognitive* modeling aspect and thus becomes a part of Cognitive Science. Instead of being oriented on a classical representation hypothesis new AI's backbone is biology and evolutionary theory together with their hypothesis that human intelligence is historically based on those of animals.

3 But what is new in "New AI"?

As briefly mentioned above, from the methodological assumptions of old AI assumptions concerning the subject of AI were derived. The very classical, natural-science like reduction of the problem down to a correct set of rules was desirable for the sake of technical realizability (von Neumann computer). But this reduction also implied a reduction in its substance to so-called *domains* of human intelligence, e.g. automatic theorem prooving, playing chess, natural language processing, etc. These areas were the ones that seemed most likely to be approachable by a classical theoretic and objective solution. These areas are also subjects of "intelligence tests" and therefore said to be "objective" in their very nature. A solution in these domains seemed to be commonly valid and therefore posess a theory-like character. This theory is to be a constructive theory of general problem solving with understandable explanations gained from its atoms. The orientation at problems that had thus become "objective" needed to lead to an orientation at mathematics, logics and linguistics. All these disciplines traditionally tend to exclude the subject from their investigations. Especially logic promises to guarantee truth, if the designer correctly puts a set of predicates and rules together. The goal of this science soon was an objective, understandable, and reductive theory of *parts* of cognition and of human behavior. The *parts*, however, left a bitter taste for old AI, but the importance of this problem for the AI community decreased with a reference to forthcoming computer generations and future capacities.

Nowadays it is commonplace in AI that there exist classes of problems for this theory that cannot be solved satisfactorily. Problems of context, of recognizing "gestalt", situativity, etc. resisted all persistent efforts of reducing them to predicates and rules. Soon these constructs were recognized by and blamed in a new AI as causing the trouble. Especially the idea of *designing* was replaced by *learning.*

Another original idea in AI—to be abolished by its new pendant—was that the artificially cognitive system, in order to behave properly, had to *mirror nature* in a reasonable way (and reasonable here again includes understandable, explainable, and linguistic). An engineering science could not allow intuitively or randomly driven structures as a basis of a system's behavior. It had to rely on disposition designing ones which needed to be linguistic in its atoms. The technical solution by means of symbols, predicates and rules promised to be the only viable way with respect to technical realizability for the following reasons.

The main quality of an AI-problem always was combinatorial explosion and the diversity of possible actions or solutions, and often also the manifoldness of inputs. This is why an effective mechanism for reducing the complexity of the problem is badly needed. This reduction should comprise the manifoldness into cases, correctly decide for cases and enable that all cases are covered. *Rules*—by means of using predicates—promised to achieve this comprehension with respect to a specific goal and thus to reduce the complexity of the problem. Note that the same is said to be achieved by words, in which things are set identical to themselves[2]. (I.e., two different dogs are nevertheless “dogs”. Thus in the word “dog” two different things are set identical.) In classical AI the same is achieved in the predicates, so that an already predicated input to the system can be successfully managed by a proper set of rules. But the manifoldness of input has then been thouroughly reduced.

As opposed to this, the program of “new” AI cannot be satisfied in this way. It expects a *non-reductive treatment of all aspects of human cognition without any predicating pre-processing of the manifoldness of sensory input.* It excludes conventional rules, because it does not assume that a designer can anticipate all necessary aspects.[3] It excludes symbols and predicates (at least in the classical sense) because it has recognized the reductive process in old AI of setting things identical in the predicates as too restrictive, especially if such a comprehension is designed. Note that we face two different prob-

[2]From a commentary on J.G.Herder *Abhandlung über den Ursprung der Sprache.* Reclam 1975, p.143.

[3]For a motiviation of this point see H.Dreyfus *Being-in-the-world.* MIT Press, Cambridge, MA, 1990, p.115–127.

lems here. Firstly, the reductive comprehension of a manifoldness happens through a symbol or a predicate. Secondly, the design of the appropriate symbols is achieved by an interpreting designer. Thus, in order to stick to the above mentioned richness of behavior, from the comprehending symbol as the atomic unit we must arrive at the single sensory input as the atom of new AI. New AI does not try a priori to exclude the possibility that a single sensory input can be decisive for the overall behavior or that all inputs at every time-step are important. This is what one termes "holistic".

Nevertheless, some comprehensive structure will be needed. But instead of having this structure designed through an already interpreting system engineer, it must now be constructed in a way that ensures the above mentioned qualities. The only remaining approach now is learning. Consequently, new AI must reaffirm *history* and *subjectivity*[4]. It not only welcomes them, but regards both as necessary conditions of intelligence. Thus *humans*—and in as far as they are its product *nature*—are placed into the foreground. The new AI is not only theoretical simulation, engineer-like reaching of goals, or formal computer science, but also an epistemic endeavour. It talks about the subject's experiences and about their contents separately. Depending on how much the new AI is oriented towards humans, it poses the question of how to construct the most effective system structures in terms of an epistemological problem.

Recapitulating, a "new AI"—as opposed to the old one—should posess the following features:

- non-reductive treatment of all aspects of human cognition
- including the subject and not expelling it
- replacing the design of linguistic atoms by constructive learning
- no mirroring of nature
- no prespecifing "setting equal"
- affirming history
- atomic entity is the sensory feature

[4]A similar point has been made by C.Lischka: Über die Blindheit des Wissensingenieurs, die Geworfenheit kognitiver Systeme und anderes ... KI 4/87, pp.15–19.

- be anthropocentric and allow epistemological endeavours

The variety of approaches to new AI is briefly summarized in the next section.

4 Main directions in "New AI"

Depending on which of the above aspects is paid most attention to, several distinct possibilities for a new AI arise. One that mainly tries to reduce the importance of the preinterpreting system designer is *Artificial Life*. The idea is to define only the conditions for survival for an agent and generate several possible ones by means of a simulated evolution. Goals of the designer and of the agent are designed as goals of the simulated evolution. In cases of obvious success a fitting structure will be attributed to the agent. And the best way to explain its behavior will be to interprete it as goal-driven and intentional. It is obvious that such an approach is necessarily slow, restricted in the possible complexity of its agents and that it mainly proceeds by chance.

Those who dare to go one step further into the agent, design its environment, its possible interactions (sensors and effectors), its connectionist structure and, of course, its goals. This mainstream of research is usually labelled *Autonomous Agents*. The difference to Artificial Life is that in this case more direct influence on the agent's structure is allowed. The designer specifies architectures in order to achieve the desired behavior of the agent, which is mainly based on goal-seeking and interaction with its environment. Of course, the above list of criteria must be fullfilled. The agent must possess a structure that allows to act effectively, which means for us—as the interpreting observers—that in similar situations similar actions are to be undertaken by the agent. Therefore, there must be a comprehensive structure causing this behavior, but without having been designed so that the holistic aspect is lost. The only way to achieve this is incremental construction of this structure through a mechanism that ensures that in principle all sensory inputs can get most important in each case of action. As we are mainly interested in actions of the agent, it will be useful to decide upon the quality of the behavior-causing structure depending on the effectiveness of its actions. The credo of interaction with the environment is central for Autonomous Agents, but it is mainly the consequence of our wish to see the agent *act* and a technical consideration. Such a system can only be interpreted (i.e. explained) through observation again. In case of fitting behavior we will attribute motivation, intentionality, or even intelligence.

One of the main problems of such an approach to new AI lies in the design of the goals. It may be dangerous to assume that a rich and complex behaviour of the agent can be achieved by a set of primitive goals. But the design of a large set of goals involves problems similar to the design of rules. The question of what could be the right set of goals for a human being to behave may be equally unanswerable as is the question for the correct and complete set of rules to describe its behavior. A possible practical answer to this problem could again be learning. We see Marvin Minsky[5] and Hubert Dreyfus[6] (and also Wittgenstein and Heidegger[7]) united in the demand that goals may have to be taken up from an ongoing culture.

In order to accelerate the development of correctly behaving agents, it may be desirable to perform one step further into the internal structures of the agent. Here, new AI must begin to replace the constructs of classical AI by its own ones in a reasonable way and, of course, sticking to its principles. It then seeks alternatives to symbols, predicates and obviously also to inference. Therefore, we must distinguish between purely evolutionary, more behavior- and motivation-oriented, and mainly technically interested approaches. The former stress the aspect of (inter-) action, the latter try to proceed faster through designing, whilst seeking to avoid the constructs of classical AI. In the final section we give an ouline for such a bottom-up approach to new AI, developed by Georg Dorffner[8].

5 "New" production systems

In a first step the "symbol" of classic AI is to be replaced. In order to stick to the principles of new AI this must happen so that a rich input maps onto

[5] M.Minsky *The Society of Mind*, Simon & Schuster, New York, 1985, p.175.

[6] H.Dreyfus, S.Dreyfus *Making a Mind Versus Modelling the Brain: Artificial Intelligence Back at a Branchpoint.* in Daedalus, Winter 1988, p.39.

[7] Ibd.

[8] G.Dorffner "A step toward sub-symbolic language models without linguistic representations" in R.Reilly, N.Sharkey (Eds.) *Connectionist Approaches to Language Processing (Vol.I)* Lawrence, Erlbaum, Hove (in press).

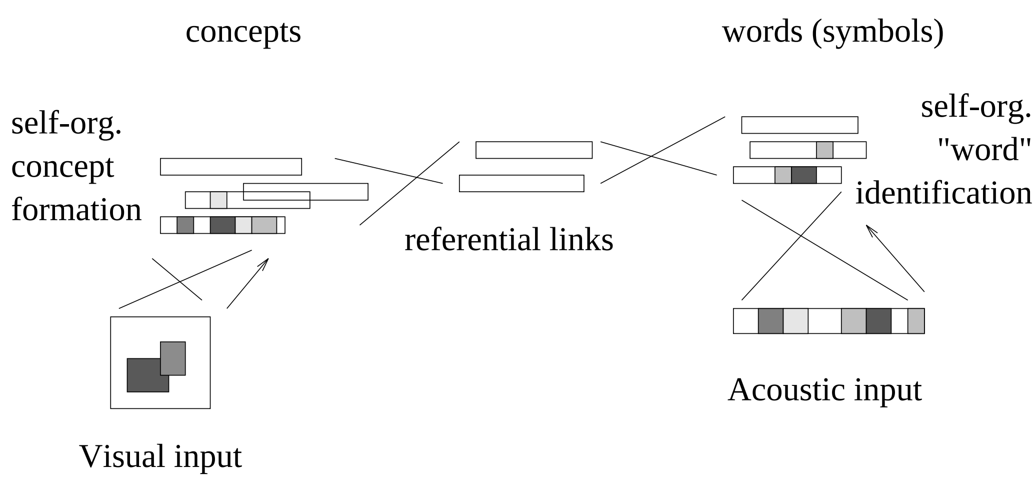

Figure 1: The symbol grounding architecture

a rich output in a way that does not restrict the manifoldness of meaning by means of a reductive structure or technique. The main achievement of a cognitively relevant symbol is that one sensory input can stand for another but with the additonal possibility that the necessary categorization of inputs (the "setting identical") happens through self-organization of the inputs.

The architecture of the network is depicted in Fig. 1. The model consists of two sensory inputs, two components for object recognition and a set of layers for building the *referential links*. One part of the model is used for perceiving and clearly identifying words (the symbols), the other for forming concepts about the perceived input. Word and object recognition are based on self-organizing categorization. Both types of concepts are connected through *referential links*. These achieve the identification of the conceptual states in the "word" and " object"-parts and establish a connection between them. Importantly, this construction leads to a discreteness of referential links with the practical effect that blendings between external embodiments (acoustic inputs) do not imply blendings of associated concepts.

Self-organisation in this model is regarded as a necessary condition of conceptualisation but also for recognition of the object and its *name*. Concepts are contructed and the conceptual symbol is equipped with rich sensory input and thus with semantics. If we replace "concept" by "object" we can say that actually the object itself is constructed through the system. The symbol (the name of the object) then is a symbol with respect to that individual cognitive agent. The step from "concept" to "object" is a mere question of point of view. It is the concept within the agent, in which objects are set equal to themselves. It is the symbol which achieves the identification of objects as objects.[9] When we noted before that new AI reaffirms subjectivity we neglected that it is still desirable to have some "objective" influence on the model. Words (or symbols) are the possibility to achieve this. While the interaction of the agent with the object only depends on external reality and internal construction of the agent, the symbol can also depend on the internal construction of *others*.[10] The identification of something as something happens always without a guarantee, but reality and social context provide the necessary control. This control is important if we want the agent to not only live its own solipsistic version of reality. The symbols ensure an "objective" interface between the constructions of different cognitive agents.

The proximity of the whole model to contructivist epistemics is obvious. The important necessary question is, as to whether more attention should be paid to this similarity than to the usual phenomenon that philosophy uses the leading discipline as the main source for examples and methaphors. Kant used physics for this purpose, Freud's epistemics was psychology, the version of Konrad Lorenz is biology. Today it is philosophers like Paul Churchland[11] who use the new AI[12] for epistemic purposes. There is indeed a new quality in this line of argumentation. New AI differs from the first three disciplines in the fact that now for the first time systems are constructed according to the theoretic principles and can thus be tested empirically with respect to their behavior and also to biological foundations. The model of symbol grounding is a model of the possibility of object constitution and of conceptualisation. Note that all we could achieve so far is the replacement of symbols in classical AI by their cognitively relevant counterpart. Our next steps must include a construction through interaction, and mechanisms to generate more complex symbols, e.g. part-whole hierarchies[13]. Then, it will be necessary to develop

[9]These ideas have been strongly influenced by W.Zeidler: *Grundriß der transzendentalen Logik.* Junghans, Cuxhaven, 1992.

[10]We could also say "on the deliberate usage" or "on the achievements of a language-using society".

[11]P.Churchland *A Neurocomputational Perspective.* Bradford, Cambridge, 1989.

[12]or its ancestor connectionism

[13]A first approach can be found in: G.Dorffner *Taxonomies and Part-Whole Hierarchies*

new-AI pendants to predicates and inference. At this point new AI turns into new *heuristics*, or an *ars inveniendi*, which is searching for a methodology to generate *new* knowledge. Here it differs from a large part of classic AI, which—through deductive reasoning—only derives what is already prespecified. (Of course, with the advantage of deriving only thruths.) A system that works in accordance to new AI will have to take other forms of reasoning into account, at least something similar to induction and abduction. Obviously, this bottom-up approach (from the structures towards the goal) still has a long way to go.

6 Conclusions

In this paper we have argued that "new AI" constitutes a field of research that is as well AI as it is "new". A list of criteria has been established, which must be fulfilled by a research program in order to justifiably belong to this science. Several existing research endeavours meet these requirements, among which we find *Artificial Life, Autonomous Agents* and *Symbol Grounding.* The latter is a program which consists in a constructive bottom-up approach to the problem of AI. The possible advantage of such a methodology is to proceed faster than, for example, the construction of Autonomous Agents or to achieve more complex models. Symbol grounding can be seen as an initial step to the development of "new" production systems and finally, truly cognitive models which—by means of their symbolic character—also provide a method for creating (or production of) *new knowledge*, or an *ars inveniendi.*

7 Acknowledgments

The author wishes to thank Prof.Robert Trappl for having made this research possible and Georg Dorffner for many valuable discussions.
The Austrian Research Institute for AI is sponsored by the Austrian Federal Ministery of Science and Research.

in the Acquisition of Word Meaning., Proc. of the 14th Annual Conference of the Cognitive Science Society, Erlbaum, 1992.