CURRENT SCIENCE

Volume 64 Number 6 25 March 1993

Artificial intelligence (AI) as science and AI as engineering

R. Narasimhan

Artificial intelligence (AI) as science is concerned with psychologically viable computational modelling of agentive behaviour. An agent is characterized by an *action-repertoire* which it can deploy intentionally to bring about desired-for changes in the world (including itself). Systematic study of agentive behaviour, then, involves articulating computational architectures and computational processes relating to one or more of the several aspects of agents. In AI as engineering, on the other hand, the objective is to design a *working* AI system that exhibits some well-defined agentive behaviour, without necessarily worrying about homologies between the artificial system and the natural system. In this article we develop in some detail the above theses and outline problems and assessment procedures that confront research workers in AI.

Artificial intelligence as science

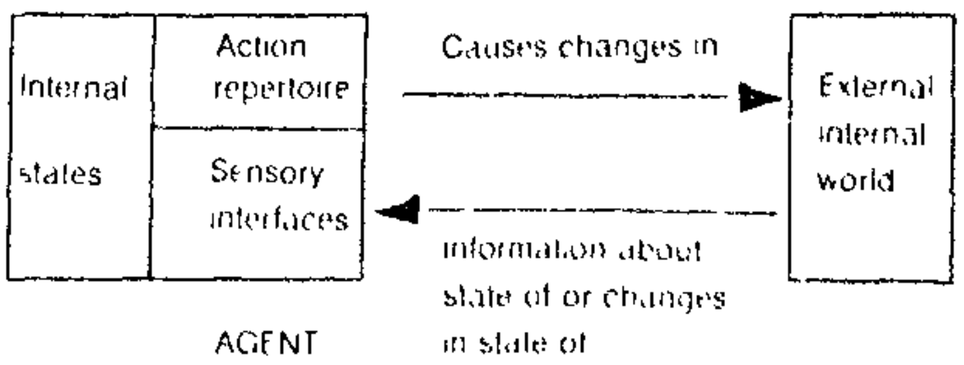
IF there were to be a science of artificial intelligence (AI), what would it be concerned with?

In a recent paper with the intriguing title 'How to get a PhD in AI', Bundy *et al.*[1] provide an insightful characterization of AI as follows: 'AI can be regarded as analogous to applied mathematics, with the role of pure mathematics being played by the cognitive sciences; psychology, linguistics, and so on. Under this analogy, the pupose of AI is to provide and explore *computational techniques* for cognitive modelling.'

This is an unusually productive approach to characterizing the domain of AI. But the definition as given above is unsatisfactory in that it does not quite get down to first principles. The physical sciences are concerned with the study of physical phenomena: that is, the occurrences and properties of physical objects (i.e. matter), and physical processes involving such objects (i.e. physical events). The methodology of the physical sciences centrally involves the construction of mathematical models (or mathematical theories) to account for physical phenomena. Mathematics (i.e. pure and applied mathematics formalisms) provides the theoretical framework for constructing such models or theories. Figure 1 schematizes this approach of the physical sciences. Opposed to physical objects are *agents*. An agent is characterized by an *action repertoire* which it can deploy intentionally to bring about desired-for changes in the world (including itself). The study of agentive behaviour is concerned with articulating computational (i.e. information-processing) models to account for the action capabilities of specific agents, or classes of agents. Computational formalisms underpin this model-construction activity. Computer science is the domain of the totality of computational formalisms. The domain of AI, as science, is delimited to those classes of computational models to account for agents and agentive behaviour. Figure 2 is a schematization that is a direct analogue of Figure 1.

Studying agentive behaviour

For the purpose of computational modelling of agentive behaviour, an agent can be schematized as follows:



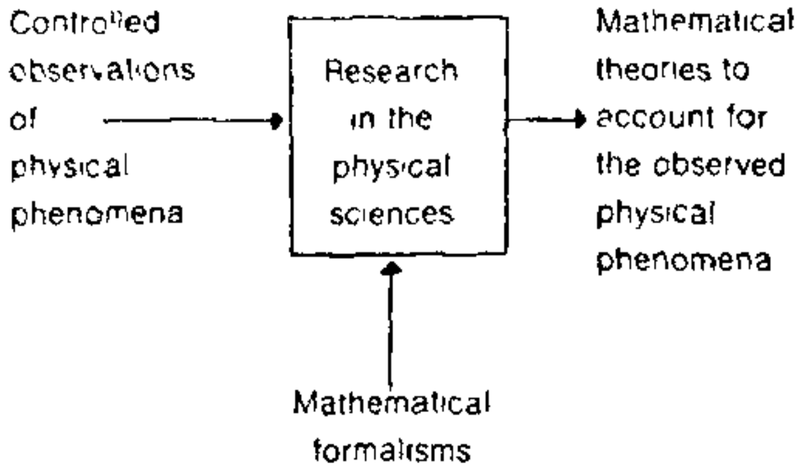


Figure 1.

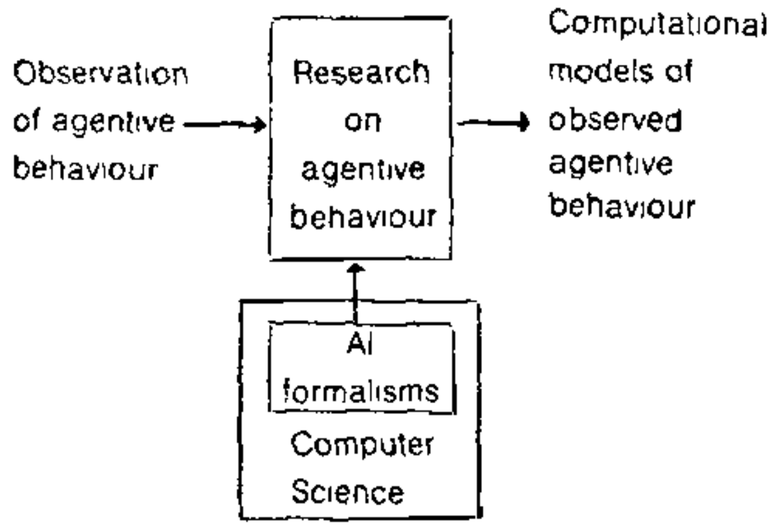


Figure 2.

Systematic study of agentive behaviour, then, involves articulating computational architectures and computational processes relating to one or more of the following aspects of agents. It is to be noted that the term 'agent' is not restricted to 'human' agents. In the behavioural sense intended here all (biological) organisms are agents.

Nature of information abstracted about the world and about the self

- Varieties of sensory modalities
- Kinds of information abstracted in each modality
- Integration of information across modalities
- Internal representation of the integrated information.

Action repertoire

- Varieties of effector organs
- Action repertoires available.

Control of behaviour

- Direct stimulus control
  * Simple reflexes
  * More elaborate fixed-action-patterns.
- Planning and experimenting
  * Trial and error
  * Knowledge-based.

Extension of behavioural capabilities

- Use of tools
  * To extend action repertoire
  * To extend sensory capabilities
- Fabrication of tools
- Communication (of knowledge/know-how)
  * Using signals/vocalizing
  * Using signs: speech, language behaviour
  * Through writing-reading (literacy).

- *Social structures and organizations*
  * Based on communication using signals and fixed-action patterns
  * Based on language behaviour and speech
  * Based on written material (literacy).

It is immediately clear from the above enumeration that the computational modelling of agentive behaviour in this comprehensive sense has a very wide overlap with other specialized disciplines such as ethology, psychology, sociology, education, neurophysiology and anatomy, to name only the more important ones. AI, as science, can make significant contributions to computational modelling of agentive behaviour only if AI researchers work with specialists from the above disciplines in closely coordinated interdisciplinary research teams.

Planning, scheduling and reasoning about action

The externally observable behaviour of an agent consists of the complex of actions it engages in. At the most primitive level an agent engages in an action to bring about a desired-for change in the state of the external world or of its internal world. The desired-for change is the *goal* (objective, aim, purpose or intent) of the action. Goals, in general, may require the execution of a complex *program* of actions. (*Program* is used in the information processing sense. A *reflex* would thus be a pre-wired program.) Habits are already built-up programs that get executed more or less autonomously. In other cases, achieving a goal may require the deliberate formulation of a *plan* of actions. That is, a program to achieve a goal may not be available ready-made but may have to be built-up, either on the basis of theoretical considerations or through exploratory experiments. In either case, an agent must have available to it *knowledge* or a *knowledge-base* to construct a plan of actions. Execution of a plan of actions would in general be guided by the assessment of outcomes of already executed actions. Assessing the state of an environment is based on *judgement*, that is, the capability to gather relevant evidence, to evaluate and to arrive at conclusions. *Understanding* a situation involves the successful utilization of available knowledge

to *assimilate* the situation, or of enlarging the knowledge-base (i.e. the available knowledge) to *accommodate* to the situation. It would seem necessary to distinguish between *knowledge* and *belief* at the behavioural level. Not all available knowledge may be forthcoming in a situation to base one's judgement on. The issues involved here are complex and systematic studies are unavailable to assist behaviour modelling. Belief-structures that come into play in psychopathic conditions have been studied. It is clear that what is needed here is a more systematic understanding of the way motivational states condition and influence the availability of knowledge in any given circumstance.

Judgement is, thus, an essential aspect of understanding. The knowledge-base directly involved in understanding (i.e. in arriving at judgements) may be called the beliefs or the belief-structure of an agent. The notion 'desired-for change' involves a valuation process. Many alternative goals may be potentially desirable but it may be possible to strive for only one at a time, or only a few of them may be accessible in a given circumstance. So the goals, would have to be ordered (or ranked or weighted) on the basis of their *value* as evaluated by the agent. Of course, the values assigned to goals may change from time to time, or be based on the prevailing circumstance. It any case, one must clearly predicate a *value-system* as underlying the functioning of an agent. At the most primitive level this value-system could only be based on the innate motivational states of the agent. Subsequently the motivational states underlying agentive behaviour should be assumed as augmented or modified on the basis of the past behavioural interactions of the agent with the world and their outcomes.

This is an ideal state of affairs. In actual situations, planning, scheduling and reasoning about actions, on the part of an agent, fall far short of this ideal. The principal question arising at this stage is as follows: What strategies are available to an agent to plan and schedule its actions in this manner? In the case of human beings the availability of language makes a qualitative difference in this respect.

Although communication-capability is something humans share with other animals, language behaviour would seem to be available only to human beings. Language behaviour enables humans to engage in qualitatively different modes of communication. *Describing, specifying, and instructing*, are modes of *interpersonal interaction which would seem to be* impossible without the availability of the language modality of behaviour.

Language enables the *articulation* of the aspects of the world being dealt with. This ability to articulate the situational aspects—aspects of objects, agents, and events and their inter-relationships—is an essential prerequisite to describing, specifying and instructing. Through specification and instruction one is enabled to plan and programme activities in this world. Articulated self-expression plays a crucial role in the complex programming of one's own behaviour. Through language behaviour human beings are able to deal with not only the world that is *immediately* available for interaction, but with worlds distanced from them in *space* and *time*. Moreover, they are able to deal not only with the actual world out there that is given, but with imagined (possible) worlds and counterfactual situations.

Using language in this mode is characteristic of *'reflective behaviour'. John Dewey*[2] *a long-time ago* argued that the central aim of the educational process should be fostering reflective thinking in children. He emphasized three principal values of thinking, i.e. reflective thinking.

1. Thinking makes possible action with a conscious *aim, i.e. it converts action that is merely appetitive,* blind and impulsive into intelligent action.
2. Thinking makes possible systematic preparations and inventions: i.e. science and technology are consequences of reflective thought.
3. Thinking enriches things with meaning: i.e. reflective thought enables us to provide theoretical underpinnings *to the objects and events of this world.*

Dewey notes that the first two values increase one's power of *control* over oneself and the world, while the *third increases our understanding of ourselves and the* world. But he cautions that the values described 'do not automatically realize themselves. For anything approaching their adequate realization, thought needs careful and attentive educational direction. Nor is that the whole story. Thinking may develop in positively wrong ways and lead to false and harmful beliefs···. The *history of scientific belief shows that when a wrong* theory once gets general acceptance, men will expend ingenuity of thought in buttressing it with additional errors rather than surrender it and start in a new direction···.' Capacity to engage in reflective thinking should be the principal focus of attention in school education. What the schools need to concentrate on is not to teach students how to solve some pregiven types of problems but how to equip themselves with general problem-solving skills.

Of all nonhuman animals, chimpanzees are genetically closest to human beings. Recently there has been much interest in teaching chimpanzees language skills. For references to critical analyses of three of the more important projects, see ref. 3. For additional information about the controversies that these studies have given rise to, see ref. 4. Most standard animal training experiments (e.g. circus animals) are concerned with

extending the behavioural repertoire of the animals in the sensori-motor modalities. But in these chimpanzee experiments, an attempt has been made for the first time, and with some noteworthy success, to extend the communication capabilities of these animals.

However, the language behaviour of these chimpanzees, did differ in significant ways from that of human children of comparable age. The language taught was hardly ever used as a tool for exploration; it was hardly ever used to describe on going activities—either one's own or other people's. It was hardly ever used to express anything more than their own *immediate* physical wants, and to a limited extent their *current affects (impatience at delay, rejection, appeal, appeasement, etc).*

Köhler[5], in his classic experiments discussed in his *The Mentality of Apes*, also notes that the group of chimpanzees he experimented with over a period of several years, never *planned their actions keeping in mind future* contingencies. Contrasting this with human behaviour, he remarks: 'Even the most primitive man makes ready his digging stick, though he is not going to dig right away; i.e. when the objective conditions for the use of his tool are not at hand'.

AI is centrally concerned with replicating the planning, scheduling and reasoning behaviour of human beings rather than the more restricted problem-solving behaviour of infra-human animals. In the latter case, the skills would seem to be restricted to the here-and-now, i.e. to the world directly available to their sensori-modalities.

AI and common sense

A general criticism acknowledged by all—practitioners of AI, as well as critics of AI—is that expert systems (i.e. domain-knowledge-based problem-solving systems) lack common sense. But there is no general consensus on what constitutes common sense.

We can attempt a solution to this puzzle in the case of human beings as follows. From functional and pragmatic considerations, we can group the spectrum of observed behaviour thus:

1. Perceptual-motor behaviour as evidenced in speech, vision, locomotion and manipulation
2. The aptitude to judge, understand, and learn from past experience (i.e. the primitives of cognition in some *sense)*
3. Natural language usage to communicate with other human beings.

Now we can take the following as the defining features of common sense.

1. Sensori-motor behaviour as evidenced in vision, speech, locomotion and manipulation, and also *evidenced in skilled-behaviour in general*
2. Communication based on natural-language behaviour
3. Cognition, consisting of
- Problem-solving and planning in articulated problem-situations
- Articulation restricted to natural language usage and simple visual schemata such as charts, lists, maps, tables, etc.

With this definition, to say that an AI expert system lacks common sense is to say the following: tacit knowledge underpins our behaviour in the sensori-motor domain and also, plausibly, much of our communication competence in natural language. Naive *pictorial mode representation is within our naive* cognitive competence. All this knowledge is implicitly taken for granted when human beings plan and solve problems in naive behavioural situations. This fund of implicit knowledge forms the 'context' that we bring to bear on any more formalized knowledge that we are *told or taught, or that we learn, or that is given to us* through instruction. This tacit knowledge-base that acts as a 'context' is totally missing in the case of all AI systems. And if any part of it is needed for planning and problem-solving, it has to be fully articulated and explicitly represented *in the knowledge-base that* underpins the AI system performance.

Two serious problems arise at this stage. Firstly, we do not know the nature of all this tacit knowledge and, therefore, are unable to articulate it effectively. Secondly, at the representation level we are unable to *unify this tacit knowledge successfully with problem* domain-knowledge which is usually expressed in some form of logical formalism. For a counter-view asserting the possibility of such a unification of all knowledge, expert as well as common sense, see ref. 6.

A deep issue we seem to have very little understanding *of—and which, consequently, does not figure at all in* AI expert system efforts—is the relationship between knowledge and affect. In the case of human beings affect seems to play a crucial role in the dynamics of knowledge acquisition and use. Is this role a purely motivational one, or is it consequential at the information processing level? Even partial answers to this question would alter our approaches to the structuring of artificial intelligences in quite significant ways; see ref. 7 also.

AI as engineering

In AI as engineering the objective is to design a working AI system that exhibits some well-defined agentive behaviour (or some delimited aspects of such behaviour). The primary concern is with realizing an

efficiently working system satisfying the functional requirements and not with establishing (or even investigating) homologies between the artificial system and the natural system. The following are some illustrative task environments for AI to cope with:

- Make typescripts of handwritten documents
- Produce typescripts from dictation
- Perform medical diagnosis and advise
- Assist in equipment maintenance and repair
- Manipulate objects as instructed
- Assemble equipment from kits.

Recently there has been much interest in integrating database and AI technologies. Interaction between these two major subareas of computer science should offer much scope for innovative research. (See ref. 8 for more discussion on the possibilities for research.)

Expert systems are perfect exemplars of engineered AI systems. As demonstrated by the more successful of the expert systems, it is as engineering and technology that AI has the most tangible benefits to offer in the foreseeable future.

AI as engineering and technology provides a practical approach to answering three major questions that a would-be AI researcher has to contend with:

1. How does one choose a significant research problem?
2. How does one equip oneself to tackle such problems?
3. How does one evaluate the research achievement?

All engineered systems are necessarily closed-world systems. But within this overall limitation many opportunities are available for innovation. How to cope with 'noise' and ambiguity in the input; how to cope with information left unspecified; how to improve performance based on feedback from the past; how to be more friendly where the task environment is interactive; and similar issues offer possibilities for research. Ultimately, how many such design issues are tackled, and how systematically (i.e. in how principled a manner) they are tackled, distinguish a good piece of research from a not-so-good one in AI.

1. Bundy, A., du Boulay, B., Howe, J. and Plotkin, G., How to get a Ph.D. in AI, in *AI: Tools. Techniques and Applications* (eds. O'Shea, T. and Eisenstadt M.), Harper & Row, 1984.
2. Dewey, J., How We Think: A Restatement of the Relation of Reflective Thinking to the Educational Process, (Selections included in: *John Dewey on Education* (ed. Archambault, R. D.)), Chicago University Press, Chicago, 1964.
3. Narasimhan, R., *Modelling Language Behaviour*, Springer-Verlag, Heidelberg, 1981.
4. Marx, J. L., *Science*, 1980, **207**, 1330–1333.
5. Köhler, W., *The Mentality of Apes*, Vintage Book, a division of Random House, 1956.
6. Lenat, D. B., Guha, R. V., Pithman, K., Pratt, D. and Shephard, M., CYC: Toward Programming with Common Sense, *Comm. ACM*, 1990, **33**(8), 30–49.
7. Narasimhan, R., Human intelligence and artificial intelligence: How close are we to bridging the gap?, *Vivek*, Jan. 1990; also *IEEE Expert*. April 1990.
8. Narasimhan, R., Integrating database and AI technologies, Inaugural Address *Comad-89*; *CSI Communication*, Jan. 1990.