# AI in a New Millennium: Obstacles and Opportunities [1]


## Aaron Sloman, University of Birmingham, UK
**http://www.cs.bham.ac.uk/˜axs/**


AI has always had two overlapping, mutually-supporting strands: science, which is mainly concerned with understanding what is and isn't possible in natural and artificial intelligent systems, and engineering, which is concerned mainly with producing new useful kinds of machines. The majority of funding has understandably been available for the engineering strand, and the majority of researchers have a strong engineering orientation. However, the intellectual giants responsible for many of the key ideas in AI and for setting up the leading research centres were all interested primarily in AI as science-for example, Turing, McCarthy, Minsky, Simon, and Newell.[2]

I am not alone in thinking that, despite substantial progress in many subfields of AI, not enough is being done to address the scientific problems related to combining many different kinds of competence within an integrated, embodied, human-like agent. The EC's Cognitive Systems Initiative[3] is just one program that has made it clear its primary aim to advance scientific understanding.

## 1 Ontological blindness

Several features of the research environment and the research itself make the tasks difficult. Some of the difficulties relate to the difficulty in being clear about what the hard problems actually are. My colleague Ron Chrisley and I (2005) have described the problem of *ontological blindness*. It arises out of misidentifying what organisms are doing or the tasks that robots may need to accomplish and, as a result, failing to identify the various subfunctions that need to be modelled or explained.

For example, most research on vision is concerned with how to extract from the optic array information about objects, properties, relationships, and processes that exist in a scene. This ignores the role of vision in seeing what does not yet exist but could exist-the possibilities for action and the constraints on those possibilities, described by J.J. Gibson (1979) as *affordances*. A more complete analysis would have to discuss ways of representing these possibilities and constraints-how an animal or robot learns to see them and how the ontology for describing them grows.

Similarly, a vast amount of AI work on vision seems to be concerned with recognition of objects. But that fails to address what goes on when we see things we do

[1]Summarised by Linda World, Senior Editor IEEE Computer Society, on the basis of Aaron Sloman's introductory notes for the IJCAI-05 Tutorial on Representation and Learning in Robots and Animals http://www.cs.bham.ac.uk/research/projects/cosy/conferences/ijcai-booklet/,

[2]McCarthy (2004), Minsky (2005)

[3]http://fp6.cordis.lu/fp6/areas.cfm?CALL_ID=74#744

not recognize. Do we know what the results of perceiving something unrecognisable should be?

Similar kinds of ontological blindness can afflict students of language. The popular view, that language exists to enable communication, at first seems obviously correct. But it ignores the deeper problem of how it is possible to have any meaning to communicate. Unifying the study of language with the study of other aspects of intelligence might help us model both of them as outgrowths of rich forms of syntactic and semantic competence used in purely internal information processing in other animals and in prelinguistic children.

## 2 Scaling up and scaling out

Another kind of ontological blindness involves varieties of complexity. From the earliest days of AI it was obvious that combinatorial explosions threatened progress. For this reason, an AI system must scale up, performing with reasonable space and time requirements as the task complexity increases.

But another kind of complexity requirement often goes unnoticed, which requires what we'll call *scaling out* in integrated systems with multiple functions. Vision and language provide good examples of capabilities that cannot be understood fully except insofar as they relate to other capabilities with which they combine. Missing the functions of vision that relate to requirements for action and thought can lead to impoverished theories of vision. Focuses the study of language entirely on linguistic phenomena such as morphology and syntax can fail to address how language is used for reasoning, how it relates to and builds on capabilities that exist in young children or other animals that cannot use language, and so on.

We can specify a general form of requirement for a model or theory of how vision and language work, how plans are made and executed, how mathematical or other reasoning operate-namely, the proposed mechanisms should be able to form a usefully functioning part of an integrated complete agent combining many different capabilities.

The kinds of combination required can vary, of course. In the simplest cases, submodules are given tasks or other input, run for a while (as "black boxes"), then produce results that other modules can use. Many AI architectures assume this sort of sense-decide-act cycle, but that sort of model fails to account for the variety of extended, concurrent, interacting processes humans and many other animals-and even some robots-support.

For studies of natural intelligence, the requirement to scale up may be far less important than the requirement to scale out. Humans, for instance, do not scale up! Suitably programmed computers can do complex numerical calculations that would defeat all or most humans, but this does not enable them to explain what a number is or why arithmetic is useful.

# 3 Information processing architectures

Solving the deep integration problems of cognitive systems with multiple functions may prove much more difficult than anyone anticipates. It is certainly conceivable that biological evolution discovered powerful forms of information processing long ago that scientists and engineers do not yet understand. We need a deep theory about nature's information-processing architectures and the capabilities they do and do not support.

Biologists make a distinction among animals that is relevant to this question:

- *precocial species*, which comprise the vast majority, seem to have all their main competences determined genetically-for example, grazing mammals that can run with the herd shortly after birth; and
- *altricial species*, the small subset that are born helpless, physiologically underdeveloped, and apparently cognitively incompetent but end up with capabilities that appear to be far more cognitively complex-for example, building nests in trees, hunting other mammals, manipulating various kinds of tools.

The distinction actually occurs along a spectrum (Sloman and Chappell, 2005), but some altricial species, especially humans, seem to learn to do things rapidly, almost effortlessly, in a wide range of environments, giving them competences as adults-or even as young children-that none of their ancestors had. At present the mechanisms supporting such learning are not well understood, and AI has no learning mechanisms or self-constructing architectures that can account for it.

Meta-semantic competence is another important distinction: to perceive and have intentions involving not merely physical things but also semantic states representing entities, states, and processes that themselves have semantic content, such as your own thoughts or those of others. Humans are not alone in having meta-semantic competence, but the richness of their capabilities-whether inwardly or outwardly directed-does seem unusual. Many disciplines-philosophy, sociology, anthropology, psychology, ethology-study how one intelligent individual can think about others, communicate with them, and engage in various kinds of shared activities. Philosophers know that relevant theories must solve deep problems, such as the breakdown of normal modes of reasoning because things referred to in beliefs, desires, intentions, and so on need not exist. Moreover, a stone or tree cannot be correct or mistaken: it just exists, but a thought or belief can be true or false.

Many disciplines look at questions such as when and why a meta-semantic capability evolved and how much it depends on learning as opposed to genetically determined competence. But hardly anyone discusses the architectural and representational requirements for an organism or machine to represent or reason about semantic contents. John McCarthy (1995) is an exception.

# 4 Methodological lifting

AI researchers must make many choices in their work: forms of representation, algorithms, architectures, kinds of information to be used, types of hardware, design and testing procedures, programming languages, development environments, and other software tools. Too often the proponents of one or another design option get into silly squabbles about which one is right or best.

By shifting the questions to a higher level, former opponents can become collaborators in a deeper research project. Instead of arguing over whether to use neural or "symbolic" forms of representations, we can instead explore the space of possible forms of representation. We need more research addressing metalevel questions to clarify the design options and trade-offs on the basis of detailed task requirements instead of fashion or prejudice. For example, Marvin Minsky depicts the trade-offs between symbolic and neural mechanisms in his paper, "Future of AI Technology" (1992).

The metalevel analysis of a space of possibilities can help to end fruitless debates over such questions as to whether representations are needed in intelligent systems, or which sorts of representations are best.

The need for this move to a higher level is particularly clear in relation to the current state of teaching AI. Students are often introduced to the choice between "symbolic" representation and tools or artificial neural nets and other numerical-statistical formalisms through their teachers' prejudices. In some cases, they do not even learn the existence of alternatives to the approach they are taught. A generation of researchers trained with blinkered vision is hardly likely to achieve the major advances required to solve our hard problems.

As a scientific research community, in addition to identifying specific, somewhat arbitrary, target systems, we should attempt to identify a structured set of scientific goals that advance our knowledge-not just our capabilities, however important that may be. In a field as complex as the study of intelligence, we cannot expect anything as simple and clear as Hilbert's list of unsolved mathematical problems at the start of the 20th century. But perhaps we can move in the direction of identifying important questions we should try to answer.

Just as in mathematics, we can show that answering some questions will enable others to be answered-or at least take us nearer to answering them. So we should try to identify relations between unsolved problems in AI. For example, perhaps if we can describe in detail, with the help of psychologists, some of the competences displayed by young children at different stages of development in different cultures, and if we analyse the architectural and representational requirements for those competences in detail, we will gain insight into the variety of developmental paths available to humans. That, in turn, may give us clues regarding the mechanisms capable of generating such patterns of learning and development.

Looking at typical interactions between these kinds of learning and other things such as varieties of play, growth of ontologies, kinds of social interaction, and kinds of self-understanding might help us overcome the difficulty of identifying what needs to be explained. It can also address the further difficulty of different subcommunities disagreeing about what is important or interesting-perhaps partly because of competition for limited funds.

# 5 Scenario-based backward chaining

Instead of trying to propose specific design goals, over which there is likely to be strong disagreement regarding priorities, perhaps we may agree on a principled methodology for generating and analysing relations between structured collections of goals that can provide milestones and criteria for success.

One such method is based on the use of detailed scenarios. Suppose we describe in great detail a variety of scenarios involving human-like or animal-like behavior that far exceed what the current state of the art can achieve. If we then analyse requirements for producing the detailed behaviours, we might be able to generate "precursor scenarios" for those scenarios, and precursors for the precursors, where a precursor to a distant scenario at least prima facie involves competences that are likely to play a role in that scenario.

In this way, by careful analysis of long- and intermediate-term goals, we can work backwards from them to identify a partially ordered set of scenarios. We can annotate those scenarios with hypotheses to be tested, regarding kinds of knowledge, learning, representations, mechanisms, and architectures that might realize the scenarios.

The scenarios can also determine a collection of milestones to measure progress. The "measure" will not be a number, but a location in a partially ordered collection of initially unexplained capabilities.

We can also work forwards from the current state of the art identifying new competences selected on the basis of their apparent relevance to the more remote scenarios, but we are likely to make better choices when we have mapped at least some of the terrain a long way ahead.[4]

The key point here is that we need to two kinds of metalevel tasks in planning research:

- describing and analysing research problems, their relationships to other problems, the evidence required to determine whether they have been solved, the methods that might be relevant to solving them, the possible consequences of solving them-both scientific and engineering; and prioritising research problems.

People can collaborate and reach agreement on the former while disagreeing about the latter. By making the construction, analysis, and ordering of possible scenarios an

[4]The analysis of the role of ordered scenarios in defining research milestones arose from discussions with John Salasin and Push Singh in connection with the DARPA Cognitive Systems project. See also http://www.cs.bham.ac.uk/research/cogaff/gc/targets.html

explicit community-wide task, we separate the identification and analysis of research problems-a task that can be done collaboratively-from projects aiming to solve the problems or test specific rival hypotheses, which may be done competitively.

# 6 References


Gibson, J. (1986). The Ecological Approach to Visual Perception. Lawrence Erlbaum Associates, Hillsdale, NJ. (originally published in 1979).

McCarthy, J. (1995). Making robots conscious of their mental states. In AAAI Spring Symposium on Representing Mental States and Mechanisms. Accessible via http://www-formal.stanford.edu/jmc/consciousness.html.

McCarthy, J. (2004). What is Artificial Intelligence? http://www-formal.stanford.edu/jmc/whatisai/whatisai.html.

Minsky, M. L. (1992). Future of AI Technology. Toshiba Review, 47(7). http://web.media.mit.edu/ minsky/papers/CausalDiversity.html.

Minsky, M. L. (2005). The Emotion Machine (draft). http://web.media.mit.edu/?minsky/.

Sloman, A. and Chappell, J. (2005). The Altricial-Precocial Spectrum for Robots. In Proceedings IJCAI'05, page In press, Edinburgh. Available at http://www.cs.bham.ac.uk/research/cogaff/altricial-precocial.pdf.

Sloman, A. and Chrisley, R. L. (2005). More things than are dreamt of in your biology: Information-processing in biologically-inspired robots. Cognitive Systems Research, 6(2):145-174. http://www.cs.bham.ac.uk/research/cogaff/04.html#cogsys