AI in a New Millennium: Obstacles & Opportunities

Aaron Sloman

School of Computer Science, University of Birmingham, UK

http://www.cs.bham.ac.uk/~axs/

Last revised: October 6, 2013

Contents

1 INTRODUCTION

This paper was originally prepared as part of the booklet for the two day tutorial on 'Representation and learning in animals and robots' at IJCAI05[1]

The Cognitive Systems initiative within the EC's Sixth Framework Programme (FP6) described in a presentation by Colette Maloney in October 2004[2], sponsors long term research on very difficult questions about intelligent integrated systems combining many aspects of human competence. The primary emphasis is on science rather than applications. The first wave of 8 projects started in September 2004[3]. One of the requirements for the funded projects is to engage in training of researchers working on the projects. Another objective is 'community building' — one of the goals of the IJCAI tutorial.

There is growing interest world-wide in research on cognitive systems, including DARPA's 'Cognitive Systems' initiative[4], the UK's Foresight Cognitive Systems project[5] and no doubt others unknown to the author. Some of these initiatives attempt to select problems that have a good chance of being solved within a few years or at most a decade. Others take a longer term view, like some of the Research Grand Challenges sponsored by the UKCRC[6], one of which is GC5: 'Architecture of Brain and Mind'[7]. In the spirit of these grand challenges, questions to be addressed at the tutorial include how to make progress and how to assess progress towards answering questions or designing systems that are at present far beyond the state of the art.

However, it is one thing to identify long term ambitions but taking steps to fulfil them is fraught with difficulties, expressed well by someone at a recent workshop on future prospects for robotics, who said that every few years he heard researchers in AI say what they were going to do for the next few years, and every time there was little change in what they were going to do. This paper attempts to diagnose some of the causes of this phenomenon and suggest remedies. To that extent it is extremely ambitious and tendentious. But the author is always ready to learn, so comments and criticisms are invited.

2 AI in a New Millennium: Obstacles & Opportunities

2.1 AI as both Science and Engineering

AI has always had two overlapping, mutually-supporting strands, namely *science* (concerned with understanding what is and isn't possible in natural and artificial intelligent systems and how, and why some of those possibilities are realised when they are) and *engineering* (concerned mainly with producing new useful kinds of machines). The ma-

[1] See http://www.cs.bham.ac.uk/research/projects/cosy/conferences/edinburgh-05.html

[2] Available at ftp://ftp.cordis.lu/pub/ist/docs/directorate_e/cogn/colette-bled-version_2_en.pdf

[3] See URL ftp://ftp.cordis.lu/pub/ist/docs/dir_e/cognition/overview_of_the_first_batch_of_fp6_cognitive_systems_projects.ppt

[4] http://www.darpa.mil/ipto/ E.g see the 'learning locomotion' solicitation: http://www.darpa.mil/ipto/solicitations/open/05-25_PIP.htm

[5] http://www.foresight.gov.uk/Previous_Projects/Cognitive_Systems/index.html

[6] The initially selected computing research grand challenges are summarised in a booklet *Grand Challenges in Computing Research*, edited by Tony Hoare and Robin Milner, available at http://www.ukcrc.org.uk/gcresearch.pdf

[7] described more fully here http://www.cs.bham.ac.uk/research/cogaff/gc/

jority of the funding has understandably always 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 (including, for example, Turing, McCarthy, Minsky, Simon, and Newell).[8]

There are many textbooks, journal or magazine articles and research papers which summarise the state of the art, including both science and engineering, in various subfields of AI, computational cognitive science, or related disciplines.

The presumption of the EC Cognitive Systems initiative and the CoSy project which led to the IJCAI tutorial is 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 summary of the Cognitive Systems Initiative in 2002[9] begins as follows:

> Objective: To construct physically instantiated or embodied systems that can perceive, understand (the semantics of information conveyed through their perceptual input) and interact with their environment, and evolve in order to achieve human-like performance in activities requiring context-(situation and task) specific knowledge.

That may sound like a purely practical objective, but other documents and presentations by EC officials[10], have made it clear that the primary aim is to advance *scientific* understanding. The online abstract[11] for the CoSy project (extracted from the grant proposal) comments:

> We assume that this is far beyond the current state of the art and will remain so for many years. However we have devised a set of intermediate targets based on that vision. Achieving these targets will provide a launch pad for further work towards the long term vision. In particular we aim to advance the science of cognitive systems through a multi-disciplinary investigation of *requirements*, *design options* and *trade-offs* for human-like, autonomous, integrated, physical (e.g. robot) systems, including requirements for architectures, for forms of representation, for perceptual mechanisms, for learning, planning, reasoning, motivation, action, and communication. The results of the investigation will provide the basis for a succession of increasingly ambitious working robot systems to test and demonstrate the ideas. Devising demanding but achievable test scenarios, including scenarios in which a machine not only *performs* some task but shows that it *understands* what it has done, and why, is one of the challenges to be addressed in the project. Preliminary scenarios have been proposed. Further scenarios, designs and

[8] See, for example, McCarthy (2004) Minsky (2005)

[9] recent information about related EC calls, meetings and funded projects can be found at the website of the 'Cognition Unit': http://www.cordis.lu/ist/directorate_e/cognition/index.htm

[10] E.g. this presentation by Colette Maloney in October 2004
ftp://ftp.cordis.lu/pub/ist/docs/directorate_e/cogn/colette-bled-version_2_en.pdf

[11] http://www.cognitivesystems.org/abstract.asp

implementations will be developed on the basis of (a) their potential contribution to the long term vision, (b) their achievability (which may not be obvious in advance) and (c) the possibility of practical applications. Tools will be developed to support this exploration. The work will use an 'open' framework facilitating collaboration with a variety of international projects with related objectives.

The tutorial was intended as part of the process of engaging with other researchers interested in long term research on cognitive systems of the sort referred to in Section 1, namely research on integrated systems combining different kinds of functionality, and based on research in several disciplines or sub-disciplines. We are not proposing that everyone working on a specific highly focused research problem should immediately switch to research on integrated systems, though we hope that a few more people will be inspired to think about a wider range of related problems by this tutorial.

2.2 Difficulties and how to address them.

There are several features of the research environment and features of the research itself that make the tasks difficult. We attempt to identify some of the obstacles to progress, and suggest at least partial remedies in the following subsections. This methodological discussion is tentative and open to criticism and suggestions for improvement.

Some of the difficulties relate to the difficulty in being clear about what the hard problems actually are – including the problem of 'ontological blindness' discussed below. Some obstacles relate to tendencies among rival research groups to argue about whether X or Y or Z is best, instead of trying to understand the structure of the space containing X, Y, and Z among other possibilities, and the tradeoffs between options. Some relate to paying excessive attention to normal and obvious adult human, capabilities, whilst ignoring the deep genetic heritage which we share with many other animals, and the variations in human capabilities across different ages, cultures, personality types, etc., including the many puzzling and illuminating consequences of brain damage or deterioration.

Some of the difficulties of both researchers and funding agencies relate to the problems of selecting good ways of making progress towards very distant goals when we don't yet have a clear vision of the intermediate research stages that can lead to those goals.

And of course many difficulties arise from the fact that, after less than a century of intense investigation of mechanisable varieties of information-processing, what we still understand is only a tiny fragment of what evolution produced over billions of years. The same can be said about how far current mechanical or mechatronic engineers are from emulating the amazing materials and mechanisms found in many animals. The last point may have deep implications regarding still unknown requirements for varieties of information processing underlying animal intelligence.

2.3 Institutional and financial obstacles

Quite apart from the great intellectual difficulty of the problems, there are severe institutional and financial deterrents to such research: it can be especially risky for young untenured researchers worried about building up their publication lists, for it is obvious that

publishable results and research grants can be achieved more easily in work that focuses only on new applications of old techniques, or minor extensions to those techniques. On the other hand, not only the EC, but many other funding agencies are beginning to favour research that integrates subfields and disciplines, as noted recently in Hendler (2005). Nevertheless because of the risks and lack of funding, research on ways of achieving more integrated systems may be possible for only a small subset of the research community. Of course, the narrowly focused but deep work on specific problems of vision, of language, of learning, of reasoning, of motor control, etc. has to go on in any case. However, a new framework for that research could help to integrate the research community in future, if even a small number of people of sufficiently high quality address the hard problems of integration.

2.4 Failing to see problems: ontological blindness

One of the most subtle obstacles to progress can be described as 'ontological blindness' (discussed in Sloman and Chrisley (2005)). This arises out of mis-identifying what organisms are doing, or the tasks that robots may need to accomplish, and as a result failing to identify the various sub-functions that need to be modelled or explained. We'll illustrate this in connection with vision and language.

2.4.1 What are the functions of vision?

In his very influential book, David Marr (1982) suggested that the function of vision was to provide information about shape, size, location, motion, and colour of objects. In contrast J.J. Gibson (1979) pointed out that there is a far more subtle function of perception, namely to provide information about "affordances" which are abstract properties of the environment related to possible actions and goals of the perceiver.[12]

On the first view, vision would be the same for a lion and a lamb surveying the same terrain, whereas, from a Gibsonian viewpoint, it would be argued that the biological requirements and action capabilities are so different in hunting and grazing mammals that they need to perceive very different affordances, for which they have different genetically determined or learnt visual mechanisms and capabilities — despite similarities in overall body structure. For instance the requirements for catching and eating meat are very different from the requirements for grazing: vegetable matter does not attempt to escape, and grass not require peeling, breaking open, etc. Similar differences may exist between birds that build nests from twigs and those that build nests using only mud, or between birds that fly and birds that don't, etc.

One test for whether a vision system perceives affordances is how it sees empty 3-D space, or empty 2-D surfaces. Since there are no objects in empty space, if all the functions of a visual system are concerned with perception of objects, then empty space cannot be seen. Yet we can see an empty 2-D or 3-D space as full of potential for various kinds of actions, depending on where the space is, how big it is, how close we are to it, what other things are in the vicinity, and what our current capabilities and concerns

[12] See Mark Steedman's tutorial notes.

are. Someone like Picasso can see potential in a blank surface that most people cannot.[13] A mathematician wondering how to calculate the area of a circle may see the potential for inscribing and circumscribing an unending succession of regular polygons with with ever-increasing numbers of sides just inside and just outside the circle. A bird holding a twig to be added to a partially built nest may need to see a place where something like that twig would be useful, and a route by which it can be inserted.

How many vision researchers have treated seeing empty space as one of the functions of vision, and, if they have, how have they defined what is to be seen in empty space? (Is there any relevant work, apart from work on path-planning, which requires recognition of regions of empty space that afford traversal?)

One way to sum up ontological blindness regarding the functions of vision is this: most research on vision is concerned with how to extract from the optic array information about objects, properties, relationships and processes *that exist in the scene*. This ignores the role of vision in seeing *what does not yet exist but could exist*, i.e. the possibilities for action and the constraints on those possibilities: the positive and negative affordances. A more complete analysis of these problems would have to discuss ways of representing such possibilities and constraints, how an animal or robot learns to see them, including how the ontology available for describing them grows.

2.4.2 Seeing without recognising objects

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 not recognise. Seeing involves acquiring information about spatial structure and relationships which does not depend on recognition of objects, though it may be the basis of recognition, and may sometimes be facilitated by recognition. But systems based entirely around recognition of objects and their properties and relationships must fail when confronted with things that cannot be recognised because they are new.

Do we know what the results of perceiving something unrecognisable should be? A good theory of vision might explain how different kinds of brain damage can differentially affect the ability to recognise different classes of objects, without removing the ability to see[14].

Besides the need for intelligent machines to be able see affordances related to their action capabilities, needs and goals there are other aspects of the functions of vision that appear not to have been investigated by vision researchers in recent years. Nearly 30 years ago Barrow and Tenenbaum (1978) drew attention to aspects of perception of shape properties and spatial relations of 3-D surface fragments that seem to be independent of object recognition, e.g. seeing the shape of the portion of the surface where a cup's handle meets the bowl, or seeing how the 3-D orientation of parts of the rim of a cup or jug vary around the rim, including the pouring lip if there is one.[15]

[13]The AARON program, by Harold Cohen, accessible online at http://www.kurzweilcyberart.com/ (click on 'Watch AARON paint now'), also seems to have some grasp of 2-D affordances and they way they change as a painting grows, along with at least a primitive understanding of 3-D structure.

[14]As illustrated graphically by this web site on *prosopagnosia* an affliction in which the ability recognize faces is lost: http://www.prosopagnosia.com/main/stones/index.asp

[15]There has been work using range-finders to obtain 3-D structure, including, for example

Many animals appear to be able to see and make use of surface structure and shapes of fragments of objects they do not necessarily recognise (e.g. consider a carnivore's task in eating its prey).[16]

Young children seem to spend much of their first year or two developing competences related to many different aspects of shape perception, including competences such as pushing, pulling, picking up, putting down, throwing, inserting, stacking, bending, twisting, breaking, assembling, disassembling, opening, shutting, etc., much of which precedes learning to talk. Some of this is related to seeing their own hands and the hands of others, whose shapes need to change during the performance of many actions.

So perhaps there is something deep and general that we need to understand in order to be understand evolutionarily old and important aspects of vision on which many other competences build. We may be tempted to assume that all human visual competence depends on having hands that can manipulate things. But we must remember that babies born without arms, e.g. after the thalidomide tragedy in the 1960s, can grow up into intelligent adults. This may depend on a powerful mixture of genetic endowments shared with normal humans, including kind of *vicarious* learning capability that we use when watching others do things we cannot do ourselves. Is that shared with other animals? How many robots can do this? The current interest in imitation may lead to new insights, provided that it is based on mechanisms, architectures, and forms of representation that support perception of affordances.

Research on these topics is extremely difficult. Perhaps that explains why the tasks identified by Barrow and Tenenbaum have largely been forgotten while most researchers work on other tasks that do not involve detailed understanding of spatial structure and affordances. Great progress has been made on tasks like object recognition or classification, object tracking, trajectory prediction, grasping or pushing simple objects, and path traversal – all of which are worthy research topics, of course, but form only a relatively subset of functions of vision. Other functions, not discussed here include the role of vision in fine-grained control of actions like grasping, posture-control, perceiving varieties of motion, developing many kinds of athletic capabilities using vision, parking a car, perceiving causal relationships, understanding the operation of a machine, perceiving social interactions, aesthetic appreciation of natural and artificial scenes and objects, communication, learning to read, then later fluently reading, text, sight-reading music, and many more. Some distinct visual capabilities can be exercised in parallel, e.g. when walking on difficult terrain whilst enjoying the view, or judging how to hit a moving tennis ball while

---

http://www.nrl.navy.mil/techtransfer/fs.php?fs_id=AI02
http://vicos.fri.uni-lj.si/view_publication.asp?id=6
http://homepages.inf.ed.ac.uk/cgi/rbf/CVONLINE/entries.pl?TAG182
http://eia.udg.es/~jpages/examples/examples.html
http://doi.ieeecomputersociety.org/10.1109/ISCV.1995.476995
but often the aim of such work is to produce only the kind of mathematically precise 3-D information that suffices for generating images of the scene from multiple viewpoints, rather than the kind of 'qualitative' information about surface shape and structure that supports perception of affordances. It is very likely that there is relevant work that I am unaware of. While writing these notes I came across this useful 1996 survey by Varady, Martin and Cox: http://ralph.cs.cf.ac.uk/papers/Geometry/RE.pdf

[16] A domain of objects we call 'polyflaps' designed for research on learning to perceive affordances is described in http://www.cs.bham.ac.uk/~axs/polyflaps — Jackie Chappell found that parakeets presented with cardboard examples played with, manipulated, and chewed them, despite never having seen them previously.

seeing what the opponent is doing.

Perhaps it is time for a collaborative multi-disciplinary project to expand our ontology for thinking about vision by attempting to develop a comprehensive taxonomy of varieties of functions of vision, along with a first draft analysis of requirements for mechanisms, forms of representation, types of learning and architectures to support such functions, especially under the constraint of having only one or two eyes that have to be used to serve multiple concurrently active processes that perform different tasks while sharing lower level resources. Such a project will benefit from the scenario-driven research described below.

2.4.3 Is language for communication?

Similar kinds of ontological blindness can afflict students of language. The popular view, which at first sight seems obviously correct, is that language exists in order to enable communication. From that viewpoint meanings are assumed to exist and language is seen as a solution to the problem of conveying meaning. But this ignores the deeper problem of how it is possible for person, or a chimp taught to use sign language or a button board, to have any meaning to communicate. Thoughts, percepts, memories, suppositions, desires, intentions, cannot exist in an animal or machine unless there is something that encodes or expresses their content. Very young children clearly have intentions, desires, information gained from the environment, and even things they want to communicate before they have learnt how to communicate in language. They are often very creative, for example moving an adult's head to face in the direction where there is something the child wishes the adult to attend to. Moreover, it seems very clear that many other animals can be attentive, afraid, puzzled, surprised, or repeatedly trying to do something, all of which involve states with semantic content. So they must have some internal states, processes or structures that express or encode that semantic content, and which allow the specific content to have consequences for internal and external behaviour. The internal and external encodings need not use the same forms of representation, though there will be considerable overlap in their requirements.

If we define 'language' as whatever is used to express semantic content, whether for oneself or another agent, then it seems that many animals have languages; and languages capable of expressing meanings with complex structures must have evolved before what we now call language, since they often exist without human language capabilities. But those older languages are used *within* individual animals, not for communication *between* animals. (This might be described as an animal communicating with itself.)

Of course, it is (or used to be!) a commonplace in AI that intelligent systems must use representations. But there are many general requirements for representations that are not only met by the forms of representation used in AI work on high level vision, reasoning, planning, learning, problem solving, but are also met by what we call language. These common requirements include the ability to construct novel meanings, the ability to make inferences, the ability to combine fragments of information to form more complex structures, the variety of meaning structures and the potentially unbounded complexity of meanings that can express facts or questions, or can control actions, all apparently necessitating the use of compositional semantics to achieve the required ability to cope

with novelty.[17]

That suggests that in order to unify the study of language with the study of other aspects of intelligence we need to stop treating languages and linguistic meanings as *sui generis* and see to what extent we can model them as outgrowths of rich forms of syntactic and semantic competence used in purely internal information processing in other animals and in pre-linguistic children.

Of course that does not contradict the claim that addition of external languages (including pictorial and other forms of communication) allowed rapid acceleration of both learning in individuals and cultural evolution that is part of the explanation of the unique capabilities of human beings. (Discussed below.)

2.4.4 Varieties of complexity: '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. If the solution to a problem involves n actions and for every action there are at least k options, then the space of possible action sequences to be searched for a solution has size at least k^n, which grows exponentially with n. For this reason it is frequently noted that a test for an AI system is whether it 'scales up', namely continues to perform with reasonable space and time requirements as the complexity of the task increases. But another kind of complexity requirement often goes unnoticed, which requires what we'll call 'scaling out' in integrated systems with multiple functions.

The examples of vision and language illustrates this: often a particular capability cannot be understood fully except insofar as it is related to other capabilities with which it can be combined. We have seen how missing the fact that the functions of vision relate to requirements for action and thought can lead to impoverished theories of vision. Similarly work on language that focuses entirely on linguistic phenomena, e.g. phonemics, morphology, syntax, semantics, may fail to address the problem of how language is used for many purposes (including thinking and reasoning, as well as communicating), how it is learnt, 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 language works, how vision works, how plans are made and executed, how mathematical or other reasoning works, how learning works, etc., 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 there are sub-modules that are given tasks or other input, and which run for a while (as 'black boxes') then produce results that can be used by other modules. Many AI architectures assume that sort of structure: they are represented by diagrams with arrows showing unidirectional flow of information between modules. Often they assume that there is a *sense-decide-act* cycle, in which a chunk of input comes in via the senses, is processed by sending packets of derived information through various modules (some of which may be changed as a result) until some external behaviour is produced, and then the cycle repeats.

[17] This sort of point was argued in Sloman (1979) but appeared to convince nobody.

This is clearly wrong. At least in humans there seems to be a deeper integration: as remarked previously different competences can interact while they are running in parallel and before specific tasks are complete. For humans, many other animals, and for robots with complex bodies and multiple sensors acting in a fast changing environment, the *sense-decide-act* model fails to account for the variety of extended, concurrent, interacting, processes that are capable of mutual support and mutual modulation.[18]

For instance, while you are looking for an object you have dropped, if you hear someone say 'Further to the left' that can help you recognise what you were looking for. Likewise while you are trying to work out what someone means by saying 'Put the bigger box on the shelf where there is more room, after making space for it' you may notice three shelves one of which is less cluttered than the others, and work out which shelf is being referred to and what might be meant by 'making space for it' in the light of what you can see of the size of the bigger box. This need not be done by fully analysing the sentence, deciding it is ambiguous, then setting up and acting on a goal to find more information to disambiguate it. What you see can help the interpretation of a sentence even before the sentence is complete.

There are well documented examples of close interaction between vision and spoken language comprehension, including the 'McGurk effect' reported in McGurk and MacDonald (1976) in which the same recorded utterance is heard to include different words when played with videos of speakers making different mouth movements. Other kinds of interaction can occur between active and currently suspended processes: e.g. something you see or think of while doing one task may give you an idea about how to finish another task on which you are stuck: a common phenomenon in scientific and mathematical discovery. In some cases, that sort of thing can even cause the current task to be dropped, with attention switching to a much more important previously suspended task. Requirements for 'anytime planning', which involve planning modules that can take account of time pressures and deliver partial results on request are another well-studied example. There is growing interest in 'incremental' processing in natural language, which may help to support such deep interactions between linguistic and non-linguistic capabilities.[19]

This requirement for competences to be capable of being integrated with other competences in flexible ways might be named the 'scaling out' requirement, in contrast with the more widely used criterion of adequacy, namely 'scaling up'. The latter requires a mechanism to be able to cope with increasingly complex inputs without, for example, being defeated by a combinatorial explosion.

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! Although there are many human capabilities that are nowhere near being matched by current machines, all of them seem to be limited in the complexity they can cope with, a point related to what Donald Michie (1991) called 'the human window'. As a result there are already many specialised forms of competence where machines far outperform most, or all, humans. Such AI systems *scale up*, but they do not *scale out* insofar as they have only very narrowly focused competence. For example, suitably programmed computers can do complex numerical calculations that would defeat all or most humans, but that does not enable them to explain what a number is or why it is useful to be able to do arithmetic.

[18] For an early discussion of this see Sloman (1978).

[19] For example this 2004 workshop http://homepages.inf.ed.ac.uk/keller/acl04_workshop/

Deep Blue can beat the vast majority of humans at playing chess, but cannot (as far as I know) teach a child to play chess, help a beginner think about his mistakes, modify its play so as to encourage a weaker player by losing sometimes, explain why it did not capture a piece, explain what its strategy is, or discuss the similarities and differences between playing chess and building something out of meccano.

Is there any artificial chess system that is capable of being puzzled as to why its opponent did not make an obviously strong move? What are the requirements for being puzzled? Compare being surprised.

Not seeing how different competences need to interact if we are to achieve systems with human-like capabilities is another form of ontological blindness: occurrences of such interactions are part of our everyday life, but we don't necessarily notice them or remember them when planning our research.

It is possible that solving the problems of deep integration of cognitive systems with multiple functions will be much more difficult than anyone anticipates. For example, it is at least conceivable that there are powerful forms of information processing discovered and used long ago by biological evolution that have not yet been understood by human scientists and engineers. Investigation of this issue is included in one of the UK Computing Research grand challenges on new forms of computation, summarised here http://www.cs.york.ac.uk/nature/gc7/.

2.5 Are humans unique?

One of the curious facts about this question is that even among scientists who are supposed to be dispassionate seekers after knowledge there are both passionate claims that humans are unique, e.g. because of their use of language, their self-consciousness, their ability to produce and appreciate art, or some other characteristics, and also equally passionate claims (some of them from champions of animal rights) that the continuity of evolution implies that we are not unique, merely slightly different from other animals, such as chimpanzees, or foxes. I fear that sometimes either kind of passion comes from some unscientific commitment, e.g. to religious reasons for *wanting* to think of humans as unique, or a concern for animal welfare that uses Darwinian theory as a basis for claims that the similarity of other animals to humans gives them similar rights. (Confusion of ethical and scientific questions can arise in many different contexts, of course.)

The debate is silly because the correct answer is obviously "Yes and No".

- Yes: humans are unique because there are things humans do that no other (known) animals can do, such as prove theorems about infinite structures, compose poems, utter communications using subjunctive conditionals, send people and machines to the moon and outer space, or make tools to make tools to make tools to make tools .... to make things we use for their own sake.
- No: humans are not unique because there are huge numbers of facts about their bodies, their behaviour, their needs, their modes of reproduction and development, and how they process information that are common to other animals.

Further discussion of the question seems to be pointless until we know a lot more about how humans and other animals work, and what the similarities and differences actually are.

Unfortunately, we still understand relatively little about how they work, partly because, as mentioned above, we don't have clear and accurate knowledge about what their capabilities, especially their information-processing capabilities, actually are, and partly because many of the mechanisms and architectures supporting such capabilities are still unknown. Instead of wasting effort on spurious debates we should try to deepen our understanding of the facts. In the mean time ethical decisions, such as decisions about hunting of animals, or about abortion will have to be based on ethical premises, not spurious scientific arguments. (The question whether 'ought' can be derived from 'is' goes back to the philosopher David Hume, and earlier philosophers. This is not the place to pursue it.)

If we had a deep theory of the variety of types of information-processing architectures in nature and what capabilities they do and do not support (including internal information-processing capabilities), and if we knew which animals have which sorts, then such emotion-charged debates might give way to reasoned analysis and collection of relevant evidence to settle questions, or acknowledgement that some of the questions use concepts that are partly indeterminate (e.g. 'cluster concepts') so that there are no answers. Similar comments can be made about the question whether a foetus is conscious or feels pain, whether various kinds of animals suffer, etc.

2.5.1 Altricial and precocial skills in animals and robots

One thing that is not generally noticed that is relevant to this is the contrast between

- the vast majority of species that seem to have all their main competences determined genetically (so-called 'precocial' species) including grazing mammals that can run with the herd shortly after birth
- the small subset that are born helpless, physiologically under-developed and apparently cognitively incompetent but end up with capabilities (e.g. nest-building in trees, hunting other mammals, use of hands to pick berries, and various kinds of tool use) that appear to be far more cognitively complex than those achieved by the former group.

The latter are labelled 'altricial' by biologists, though there is not a sharp distinction but a spectrum of cases in which different mixtures of altricial and precocial skills are combined.[20]

Some of the altricial species, especially humans, seem to be capable of relatively rapidly and almost effortlessly, learning, in a wide range of environments, to do things that are appropriate – so that as adults (or even as young children) they may have competences none of their ancestors had. In contrast, the competences of precocial species (e.g. deer, chickens) may be shaped to a small degree by the environment in which they live, or altered by slow and laborious training (e.g. circus training) that is unlike the spontaneous and rapid learning through play found in humans and other primates. It seems that at present the mechanisms supporting that rapid, spontaneous learning are not well understood, and there are no learning mechanisms or self-constructing architectures in AI that can account for this.

[20]The altricial-precocial spectrum for robots is discussed in Sloman and Chappell (2005) in this conference.

The learning by doing outlined in Philipona et al. (2003) may be an example of a type of mechanism that is relevant. Another important topic may be selection of actions and percepts as 'interesting' as discussed in Colton et al. (2000). We have yet to understand the varieties of animal motivation related to cognition as opposed to motivation related to physical and reproductive needs.

Perhaps we can make significant progress on these topics if we look more closely at a variety of phenomena found in the animal world, including recent work on animal tool-making and use e.g. Chappell and Kacelnik (2002) and Chappell and Kacelnik (2004). Related discussions and empirical data can be found in Cummins and Cummins (2005), Csibra and Gergely (in press) , and Tomasello *et al.* (200?). Perhaps future work in AI on altricial robots will enable us to rewrite Piaget's theories, e.g. Piaget (1954).

2.5.2 Meta-semantic competence

In addition to the plasticity and rapidity of learning in altricial species there is another factor that is important, namely meta-semantic competence: the ability not merely to perceive, think about or have intentions involving physical things such as rocks, trees, routes, food, and the bodies of animals (including one's own), but also to have semantic states that represent entities, states and processes that themselves have semantic content, such as one's own thoughts, intentions or planning strategies, or those of others. The label 'meta-management' for an architectural layer with meta-semantic competence applied to the system itself was coined by Luc Beaudoin in his PhD thesis. (The word 'reflective' is sometimes used in that way but also often used in other ways – another example of confused terminology in the study of architectures.) Closely related ideas have been developed by Marvin Minsky and Push Singh, focusing mainly on attempts to model human competence.[21]

It seems to be clear that humans are not alone in having meta-semantic competence, but the richness of their meta-semantic competence, whether directed inwardly or outwardly does seem to be unusual. We still do not know what sorts of forms of representation, mechanisms and architectures support this, nor how far they are genetically determined and how far a product of the environment, e.g. cultural learning.[22]

There is much discussion in many disciplines (e.g. philosophy, sociology, anthropology, psychology, ethology) of the ability of one individual to think about other intelligent individuals, to communicate with them, to engage with them in various kinds of shared activities. Philosophers know that there are deep problems concerned with referential opacity that need to be solved by such theories: for instance the problem of breakdown of normal modes of reasoning because things referred to in beliefs, desires, intentions, etc. need not exist. You cannot kick or eat something that does not exist, but you can think about it, talk about it or run away from it. Moreover, a stone or tree cannot be correct or mistaken: it just exists, but a thought or belief can be true or false. The growth of understanding of these matters is a subject of concern for developmental psychologists. Other questions often discussed in various disciplines include: when a meta-semantic capability evolved, why it evolved, how much it depends on learning as opposed to genetically determined competence, whether and how it is influenced by a culture, and what its

[21] For more on this see Sloman and Chrisley (2003).

[22] See Mark Steedman's notes for the tutorial.

consequences are.

But hardly anyone discusses what the architectural and representational requirements are for an organism or machine to represent, refer to, or reason about, semantic contents. One of the exceptions is McCarthy (1995).

This is one of many topics requiring further study by researchers thinking about integrated cognitive systems.

2.6 Using detailed scenarios to sharpen vision

One way of helping to overcome 'ontological blindness', i.e. failure to notice some of the functions of natural cognition, is to formulate some of the design goals in terms of *very detailed* scenarios[23]. If the scenarios are described in minute detail, e.g. using imaginary 'film-scripts' for future demonstrations of competence, then close attention to individual steps in the scenario can generate questions of the form: 'How could it do that?' which might not be noticed if a competence is described at too general a level. An example illustrating some of the previous points might be a scenario in which a robot learns to paint pictures of various sorts on a blank canvas. Such work may explain how a three year old child who is well able to hold a pencil and make spirals and other things on a sheet of paper cannot copy a square drawn on the paper.

The need for 'fine grain' in scenario specifications is not always appreciated. For example, merely specifying that a robot will help infirm humans in their own homes does not generate as many questions as specifying that the robot will be able to see wine-glasses on a table after a meal and put the used ones into a dishwasher without breaking them. How will it tell which have been used? Compare the differences between red and white wine. Will it also be able to do that for coffee cups? How will it control its movements in picking up the glasses? What difference does the design of its hand make? E.g. does the task require force feedback? Will it pick up only one glass or cup at a time or more than one in the same hand? How will it avoid bumping a glass against other objects in a partly loaded dishwasher? Under what conditions will it make a mistake and break a glass, and why? Can it improve its competence by practice, and if so how will that happen, and what sorts of improvement will occur? Will it be able to modify its behaviour appropriately if the lights are dimmed, or if its vision becomes blurred through camera damage, or if part of its hand is not functioning? Will it be able to explain why it picked up only two glasses at a time and not more? Can it explain how it would have changed its behaviour if the glasses had been twice as big, or if they had had wine left in them?

Each question leads to bifurcations in the possible scenarios to be addressed, depending on whether the answer is "yes" or "no".

Merely specifying a form of behaviour to be demonstrated does not specify research goals, for it may be that the behaviour is largely pre-programmed by genetic mechanisms in an animal or explicit programming in a robot (as in so-called 'precocial' species) or it may be a result of a process of learning and development that is capable of producing a wide variety of end results depending on the environment in which it occurs (as in so-called 'altricial' species). Instead of arguing over which is the better target scenario

[23] As tentatively illustrated here: http://www.cs.bham.ac.uk/research/cogaff/gc/targets.html, http://www.cs.bham.ac.uk/research/projects/cosy/scenarios/, http://www.cs.bham.ac.uk/~axs/polyflaps/

or which is the better design, the proposed scenario-based methodology can allow both extremes and also a variety of intermediate cases (including a combination of abilities that are precocial and some that are altricial) and investigate the detailed requirements for all the combinations, and their trade-offs.

Another way of generating task requirements is to bring people from different disciplines together to discuss one another's problems and results. So AI robotics researchers looking to the long term future should pay attention to discoveries of psychologists, students of animal behaviour, neuroscientists, and clinicians who identify failures of competence arising out of various kinds of brain damage or deterioration. Better still, they should engage in collaborative research and frequent discussions of the issues. A recent example of a non-AI research paper with rich implications for AI research is Csibra & Gergely (in press). Related observations can be found in Tomasello *et al.*

2.7 Resolving fruitless disputes by methodological 'lifting'

There are many choices to be made when doing AI research. These include selecting 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, and, in recent years, debating whether robots should or should not have emotions.[24] Too often the proponents of one or other option get into what can best be described as silly squabbles about which design option is right or best. They are silly if the terms used are ill defined, or if there is no *best* option, only a collection of *tradeoffs*.

2.7.1 Analyse before you choose

As illustrated in the previous subsection, instead of continuing with these debates we can learn more by shifting the questions to a higher level, which enables former opponents to become collaborators in a deeper research project. E.g. instead of arguing over whether neural or 'symbolic' forms of representations should be used, we can instead explore the space of possible forms of representation (i.e. all the significantly different ways of encoding and manipulating information in an information-using system), trying to understand the various dimensions in which the formalisms differ, and trying to understand what the individual types are and are not good for, what mechanisms they require, how they differ in relation to a range of criteria such as speed, accuracy, reliability, generality, etc. In many cases the answers are not obvious, so if the options and trade-offs can be made clear by research addressing the meta-level questions, then future researchers can choose options wisely on the basis of detailed task requirements, instead of following fashions or prejudice.

As an example, see Minsky's 'Causal diversity' depiction of trade-offs between symbolic and neural mechanisms Minsky (1992). His much older paper Minsky (1963) also includes many relevant observations about trade-offs between design alternatives. Another once well known meta-level paper is McCarthy and Hayes (1969) which produced

[24] See Sloman and Croucher (1981); Arbib and Fellous (2005) and this online presentation http://www.cs.bham.ac.uk/research/cogaff/talks/#cafe04

a first draft list of criteria for adequacy of forms of representation.[25] That paper's emphasis on logic provoked a charge of narrowness in Sloman (1971), and a rebuttal in Hayes (1984). One of the most recent developments of this thread is a PhD thesis on proofs using continuous diagrams Winterstein (2005). Some preliminary steps towards a more general overview of the space of possible forms of representation are in Sloman (1993, 1996).[26]

Earlier discussions of requirements for forms of representation or of mechanisms usually failed to take account of requirements for an integrated, embodied, agent with a complex body embedded in a partially unknown and and continuously changing environment. Such an agent will typically have concurrently active processes concerned with managing the state of the body, including controlling ongoing actions and continuously sensing the environment in parallel with other internal processes such as reminiscing, deliberating, thinking about what someone is saying, planning a response, etc., as well as aesthetic and emotional responses. More recent work on requirements for complete architectures in robots and other systems interacting with a rich dynamic environment begins to address this complexity, but such work is still in its infancy, and the gaps in our knowledge are easily revealed by analysis of requirements for detailed scenarios.

2.7.2 The need to survey spaces of possibilities

The 'meta-level' analysis of a space of possibilities (for forms of representation, for mechanisms, for architectures, etc.) 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. Some of these debates turn out to be inherently muddled because what one faction offers as an *alternative* to using representations another will describe as merely using a different *sort* of representation. If we have a deep understanding of the structure of the space of possibilities containing the proposed alternatives, and their tradeoffs, then how we *label* the options is of lesser consequence. However, we'll have a better chance of agreeing on labels when we agree on what variety of things we are labelling. (Compare the importance of the periodic table of the elements in the history of the physical sciences.)

The need for this move to a higher level is particularly clear in relation to the current state of teaching of AI as regards whether to use more 'symbolic' forms of representation and tools or artificial neural nets and other numerical/statistical formalisms and methods. Learners are often introduced to these issues through the prejudices of their teachers. In some cases this can lead to kinds of teaching were students do not even learn about the existence of alternatives to the approach they are taught. This became very clear when we were attempting to select candidates for a robotics research position: many applicants claimed to have MSc or PhD degrees in AI yet had never encountered a symbolic parser, or problem solver, and had apparently never heard of STRIPS or any other planning system.[27] Similarly, people who learn to use a particular sort of architecture (e.g. ACT-R,

[25] E.g. metaphysical adequacy, epistemological adequacy, and heuristic adequacy, to which we could add things like learnability and evolvability – in a particular environment.

[26] And further work on implicit and explicit forms of representation in organisms can be found in this draft incomplete paper http://www.cs.bham.ac.uk/research/cogaff/sloman-vis-affordances.pdf

[27] An excellent introduction is Ghallab et al. (2004)

CLARION, ICARUS, PRS, SOAR, Subsumption, ...) may never learn about very different alternative possible architectures.[28] A generation of researchers trained with blinkered vision is hardly likely to achieve the major advances required to solve our hard problems, even if different subgroups have different blinkers.

We can summarise as follows:

- Before choosing the best X it may be best to understand the space of possible Xs
- Often there is no best X, but a collection of tradeoffs
- Instead of trying to decide what the precise boundaries are between Xs and non-Xs it may be better to investigate varieties of X-like things, the dimensions in which they vary and the tradeoffs: a result of this may be that the X/non-X distinction evaporates and is replaced by a rich taxonomy of cases.

2.7.3 Towards an ontology for types of architectures

There are many other questions whose resolution can have a sound scientific basis only after progress in analytical research at a higher level. For example, over the last two decades there has been a major shift of emphasis in AI research from investigations of *algorithms* and *representations* needed for specific tasks to the study of *architectures* in which many components performing different tasks can be combined. Various specific architectures have been proposed, some of them surveyed in Langley and Laird (2006).[29] However it is not clear that the research community has so far developed an adequate analysis of possible requirements for different sorts of architectures nor an adequate ontology for describing and comparing alternative architectures. Moreover the existing terminology that is widely used for labelling components, e.g. as 'reactive', 'deliberative', 'reflective', 'affective', 'symbolic', 'sub-symbolic', is not based on well-defined, clearly specified categories. For example, some researchers describe as *deliberative*, a reactive system in which sensory input can activate two possible behaviours thereby triggering a competitive process to select only one of them; whereas others have called that *proto-deliberative*, reserving the label *deliberative* for mechanisms that can manipulate representations of variable structure using compositional semantics. A richer meta-level ontology for types of architectures would allow a variety of intermediate cases. Some researchers reserve the label 'reactive' for systems whose internal state does not change, whereas others allow reactive systems to learn and have changing goals, whilst lacking deliberative mechanisms for constructing and comparing hypothetical alternatives. As indicated above the word 'reflective' is also used with different meanings when describing architectures or components of architectures.[30]

[28] For a useful partial survey see Langley and Laird (2006).

[29] Occasionally architectures are confused with tools used for implementing them. For instance 'SOAR' can refer to an abstract specification of an architecture defined in terms of a collection of competences relevant to certain kinds of reasoning and learning, or it can refer to a toolkit that supports the development of instances of the SOAR architecture. But the abstract architecture could be implemented in many other tools, or even in different programming languages. This section is not concerned with tools.

[30] Papers in the Cognition and Affect project http://www.cs.bham.ac.uk/research/cogaff/ present the CogAff schema as part of an ongoing attempt to provide a more principled ontology for possible architectures. Types of architectures need to be related to the space of possible sets of requirements: i.e. *niche space*.

Going beyond the present terminological morass will benefit from the realisation that biological evolution produced a large set of intermediate cases which we do not yet understand, some of which occur during early stages of human infant and child development – though observing developmental processes in virtual machines that bootstrap themselves is a task fraught with difficulties (Sloman and Chappell, 2005). We may need to understand many of the intermediate cases if we are to match our designs for applied AI systems to a much richer variety of practical problems, as evolution has already done with its designs.

2.8 Assessing scientific progress

A well known psychologist once commented that whenever he heard AI researchers giving seminars they talked about what they were going to do, and occasionally what they had done, but rarely about what they had *discovered*. Is there any way in which we can, as a research community, map out varieties of progress and research milestones for the next five or ten years, with the aim of *advancing knowledge* – as opposed to merely *doing things*, however worthwhile? A partial answer is given by the suggestions in the previous subsection: there is scientific work to be done producing systematic meta-level theories about varieties of forms of representation, mechanisms, architectures, functions, requirements, etc., which define the spaces from which we can choose components of designs and explanatory theories. But that does not answer substantive questions about how human vision works, or how crows build nests, or how children learn language, or how capabilities found in nature may be replicated or improved on in artificial systems. Neither do we explain such things by merely building systems that work, if we don't understand how they work and why they are better or worse than other possible designs, etc., or why they are better in some contexts and worse in others.

Much funded research in AI is defined in terms of specific goals that look like practical achievements, e.g. producing a system that will do something that no machine has done before, whether it be attending a conference and giving a talk Simmons et al. (2003), performing well at soccer (www.robocup.org), helping with rescue operations after a disaster (www.rescuesystem.org), or identifying potential terrorists at airports, etc. As a scientific research community, in addition to identifying specific, somewhat arbitrary, target systems, however interesting and important, we should attempt to identify a structured set of scientific goals which advance our knowledge, as opposed to merely advancing our capabilities (however important that may be). We cannot expect there to be anything as simple and clear as Hilbert's list of unsolved mathematical problems in a field as complex and diverse as the study of intelligence, which has not yet achieved the clarity and rigour of mathematics at the start of the 20th Century (and perhaps never will because of the nature of AI as science). 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 should we 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 in detail the architectural and representational requirements for those competences, that will give

us insight into the variety of developmental paths available to humans. That in turn may give us clues regarding the mechanisms that are capable of generating such patterns of learning and development. In particular, instead of projects that focus only on language learning, or visual learning, or development of motor control, we can look at typical interactions between these kinds of learning and other things such as varieties of play, growth of ontologies, kinds of enjoyment, kinds of social interaction, and kinds of self-understanding.

This strategy may help us overcome the difficulty of identifying what needs to be explained, referred to as 'ontological blindness' above. It can also address a further difficulty, namely that different sub-communities disagree as to what is important or interesting: partly because they are in competition for limited funds, or simply because of limitations in what they have learnt.

So instead of trying to propose specific scientific 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, allowing new goals to be set as we continue to apply the method. One such method is based on the use of detailed scenarios mentioned above.

2.8.1 Scenario-based backward chaining research

Suppose we describe *in great detail* a variety of scenarios involving various kinds of human-like or animal-like behaviour whose achievement is far beyond the current state of the art. The dishwasher-loading scenario in 2.6, above, is one example, but we could produce hundreds more, relating to everyday competences of humans of different sorts and as many again involving competences of other animals.

If we then analyse requirements for producing the detailed behaviours, this may enable us 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.[31]

2.8.2 An example collection of scenarios

Using attached knobs a young child may be able to lift 'cut-out' pictures of various animals (a cat, an elephant, a cow, etc.) from a sheet of plywood, and then put them back in their appropriate recesses. Analysis of requirements for doing that reveals at least the following intermediate competences each of which can be achieved without going on to the next stage

- Being able to lift a picture from its recess (using its attached knob).
- Being able to put down a picture.
- Being able lift a picture from its recess and put it somewhere else
- Being able to lift a picture from the table and put it on the plywood sheet
- Being able to put the picture down in the general location of its recess

[31] A provisional template for specifying such scenarios is provided here http://www.cs.bham.ac.uk/research/projects/cosy/scenarios/. This is still under development.

- Being able to see that the picture is not yet in its recess
- Being able to move the picture about at random until the picture drops in
- Seeing that the reason why the picture will not go into its recess is that its boundary is not aligned with the boundary of the recess
- Being able to use the perceived mismatch in the boundaries to slide and rotate the picture till it drops into the recess
- Being able to say which picture should go into which recess
- Being able to explain why the non-aligned picture will not fit into its recess
- Being able to help a younger child understand how to get the pictures back into their recesses.

This crude sketch of an ordered collection of competences leaves out much of the fine detail in the progression, but indicates possible stages about which we can ask: what mechanisms, forms of representation, etc. can account for this competence? How needs to be added to the child's ontology to enable competence to improve? (E.g. boundary of a shape, alignment and misalignment of two boundaries.) What mechanisms can account for the development of the competence from precursor competences? What mechanisms can enable successor competences to develop from this competence? We should not start with the *assumption* that some uniform learning mechanism is involved at all stages. Moreover, we need not assume that all required forms of learning are present from the start: some kinds of learning may themselves be learnt. This fact enables us to ask questions about kinds of meta-competence and whether they also develop, and if so how?

The last example, the ability to help a younger child, has many precursors which would need to be unpacked as part of a detailed analysis, including being able to see and think about another individual has having goals, as perceiving objects, as performing intentional actions, as making mistakes, as not knowing something, etc.

2.8.3 Assessing (measuring?) progress

In this sort way, by careful analysis of long-term and intermediate-term goals, and working backwards from them, we can expect to identify a *partially* ordered set of scenarios. Those scenarios can be annotated with hypotheses to be tested, regarding kinds of knowledge, kinds of learning, forms of representation, mechanisms and architectures that may enable the scenarios to be achieved. They can also determine a collection of milestones that will measure progress. The 'measure' will not be a number, but a location in a partially ordered collection of initially unexplained capabilities. Of course, as the research proceeds, the collection of scenarios, the presupposition/precursor links, and the hypothesised components of adequate models and explanations will change.

Sometimes rival hypotheses will be proposed, and that will help to sharpen some of the research goals associated with the scenarios, by suggesting variants of the scenarios, or constraints on implementation, that may lead to tests that will show which hypothesis is better – or whether each is better only for a subset of cases.

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.[32]

2.8.4 Replacing rivalry with collaboration

The key point here is that we need to separate two kinds of meta-level tasks involved in planning research:

- The task of *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 (including both scientific and engineering consequences).
- The *prioritising, justification, or selection* of research problems, e.g. deciding what is important and should be funded.

People can collaborate and reach agreement on the former while disagreeing about the latter. But the process should lead researchers to be less intensely committed to answers to the second question: questions about what is important are not usually themselves important in the grand scheme of advancing knowledge. (The philosopher, J.L. Austin, once dealt with an objector by saying 'Truth is more important than importance'.)

Moreover, when we understand the science better we shall be in a better position to discuss the benefits of different ways of allocating scarce research resources. For instance, work on clarifying and analysing a problem can contribute to a decision to postpone research on solving the problem e.g. by showing that there is a hard prior problem that should be addressed first, or by pointing out that the costs would be higher or the benefits lower than work on some other project. The meta-level theoretical work on the problem can be a significant contribution to knowledge in revealing good routes to intermediate goals. Theoretical analysis of which mechanisms, formalisms, architectures, knowledge systems, will or will not be sufficient to support particular types of scenarios, and why some approaches will fail is a contribution to science. (Compare the role of complexity theory in software engineering. Time and space complexity are important, but they are not the only obstacles to progress in AI. The need to 'scale out' in the sense explained above may be a more serious obstacle.)

By making the construction, analysis and ordering of possible scenarios, along with analysis of corresponding design options and tradeoffs, an explicit community-wide task (as the Genome project was a community-wide task among biologists), we separate the task of identifying and analysing research problems and their relationships, a task that can be done collaboratively, from projects aiming to solve the problems or aiming to test specific rival hypotheses, which may be done competitively.

This will help to counter the current tendency for research groups or sub-communities who do the research to specify their own criteria for evaluation independently of what others are doing, a symptom of an immature and fragmented science.

[32]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

An important side-effect of this work is that it can also provide a means of evaluating research *proposals*. It is sometimes said that every wave of AI researchers proposes to do what previous researchers had proposed to do, but failed to do, posing the question: why should the new proposals be taken seriously? One answer may be that new proposals are too often 'forward chaining' proposals regarding how known techniques, formalisms, architectures, etc. will be used to solve hard problems: a well-tried recipe for failure. Instead we need more work based on the kind of detailed 'backward-chaining' analysis of long term task requirements for integrated systems, as proposed here. Perhaps that will lead to a major change in the fortunes of AI research projects.

3 CONCLUSION

The preceding discussion has been very vague in places, somewhat polemical and lacking in detailed examples. Nevertheless it should at least provoke some people to pay attention to classes of problems and ways of thinking about them that are now too rarely found in AI research and teaching. There seems to be growing awareness that it is time to consider the task of integration as a long term challenge. Perhaps events like the IJCAI tutorial and documents like this will provoke some very bright young researchers to strike out in new directions that in future years will be seen to have transformed the research landscape, leading to deep new scientific understanding and many new applications that are now far beyond the state of the art.

NOTE: The references that follow should not be treated as a comprehensive bibliography, or even a representative sample. They are merely the publications that occurred to the author while this document was being written. Additional references occur in some of the footnotes.

References

Arbib, M. and Fellous, J.-M., editors (2005). *Who Needs Emotions?: The Brain Meets the Robot*. Oxford University Press, Oxford, New York.

Barrow, H. and Tenenbaum, J. (1978). Recovering intrinsic scene characteristics from images. In Hanson, A. and Riseman, E., editors, *Computer Vision Systems*, pages 3–26. Academic Press, New York.

Beaudoin, L. (1994). *Goal processing in autonomous agents*. PhD thesis, School of Computer Science, The University of Birmingham, Birmingham, UK.

Chappell, J. and Kacelnik, A. (2002). Tool selectivity in a non-mammal, the New Caledonian crow (Corvus moneduloides). *Animal Cognition*, 5:71–78.

Chappell, J. and Kacelnik, A. (2004). New Caledonian crows manufacture tools with a suitable diameter for a novel task. *Animal Cognition*, 7:121–127.

Colton, S., Bundy, A., and Walsh, T. (2000). On the notion of interestingness in automated mathematical discovery. *Int. Journ. of Human-Computer Studies*, 53(3):351–375.

Csibra, G. and Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Johnson, M. H. and Munakata, Y., editors, *Processes of Change in Brain and Cognitive Development. Attention and Performance XXI*, pages 249–274. Oxford University Press, Oxford.
http://www.cbcd.bbk.ac.uk/people/gergo/a&p_pedagogy.pdf.

Cummins, D. and Cummins, R. (2005). Innate modules vs innate learning biases. *Cognitive Processing: International Quarterly Journal of Cognitive Science*, 3(3-4):19–30.

Ghallab, M., Nau, D., and Traverso, P. (2004). *Automated Planning, Theory and Practice*. Elsevier, Morgan Kaufmann Publishers, San Francisco, CA.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, MA.

Hayes, P. J. (1984). Some problems and non-problems in representation theory. In Brachman, R. and Levesque, H., editors, *Readings in knowledge representation*. Morgan Kaufmann, Los Altos, California.

Hendler, J. (2005). A Letter from the Editor: Fathoming Funding. *IEEE Intelligent Systems*, pages 2–3. Accessible at http://www.computer.org/intelligent.

Langley, P. and Laird, J. (2006). Cognitive architectures: Research issues and challenges. Technical report, Institute for the Study of Learning and Expertise, Palo Alto, CA. http://cll.stanford.edu/~langley/papers/archrev.ps.

Marr, D. (1982). *Vision*. W.H.Freeman, San Francisco.

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

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

McCarthy, J. and Hayes, P. (1969). Some philosophical problems from the standpoint of AI. In Meltzer, B. and Michie, D., editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, Edinburgh, Scotland. http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264:746–748.

Michie, D. (1991). Machine intelligence and the human window. *Applied Artificial Intelligence*, 5(1):1–10.

Minsky, M. L. (1963). Steps towards artificial intelligence. In Feigenbaum, E. and Feldman, J., editors, *Computers and Thought*, pages 406–450. McGraw-Hill, New York.

Minsky, M. L. (1992). Future of AI Technology. *Toshiba Review*, 47(7).

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

Philipona, D., J.K.O'Regan, and Nadal, J.-P. (2003). Is there something out there? Inferring space from sensorimotor dependencies. *Neural Computation*, 15(9). http://nivea.psycho.univ-paris5.fr/Philipona/space.pdf.

Piaget, J. (1954). *The Construction of Reality in the Child*. Ballantine Books, New York. Last chapter online http://www.marxists.org/reference/subject/philosophy/works/fr/piaget2.htm.

Simmons, R. G. et al. (2003). Grace: An autonomous robot for the aaai robot challenge. *AI Magazine*, 24(2):51–72.

Sloman, A. (1971). Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence. In *Proc 2nd IJCAI*, pages 209–226, London. William Kaufmann.

Sloman, A. (1978). *The Computer Revolution in Philosophy*. Harvester Press (and Humanities Press), Hassocks, Sussex.

Sloman, A. (1979). The primacy of non-communicative language. In MacCafferty, M. and Gray, K., editors, *The analysis of Meaning: Informatics 5 Proceedings ASLIB/BCS Conference, Oxford, March 1979*, pages 1–15, London. Aslib.

Sloman, A. (1993). Varieties of formalisms for knowledge representation. *Computational Intelligence*, 9(4):413–423. (Special issue on Computational Imagery).

Sloman, A. (1996). Towards a general theory of representations. In D.M.Peterson, editor, *Forms of representation: an interdisciplinary theme for cognitive science*, pages 118–140. Intellect Books, Exeter, U.K.

Sloman, A. and Chappell, J. (2005). The Altricial-Precocial Spectrum for Robots. In *Proceedings IJCAI'05*, pages 1187–1192, Edinburgh. IJCAI. http://www.cs.bham.ac.uk/research/cogaff/05.html#200502.

Sloman, A. and Chrisley, R. (2003). Virtual machines and consciousness. *Journal of Consciousness Studies*, 10(4-5):113–172.

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.

Sloman, A. and Croucher, M. (1981). Why robots will have emotions. In *Proc 7th Int. Joint Conference on AI*, pages 197–202, Vancouver. IJCAI.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5):675–735.

Winterstein, D. (2005). *Using Diagrammatic Reasoning for Theorem Proving in a Continuous Domain*. PhD thesis, University of Edinburgh, School of Informatics. http://www.era.lib.ed.ac.uk/handle/1842/642.