Introduction to Symposium GC5: Architecture of Brain and Mind Integrating high level cognitive processes with brain mechanisms and functions in a working robot

Aaron Sloman[⋆†]

[⋆]Computer Science, University of Birmingham UK
http://www.cs.bham.ac.uk/~axs/
A.Sloman@cs.bham.ac.uk

Abstract

This symposium is inspired by UKCRC Research Grand Challenge 5: Architecture of Brain and Mind. The aim is to provoke unified discussion of long term research goals in AI, Cognitive Science, and related disciplines, especially goals concerned with giving computers a useful and general subset of human capabilities, implemented in a biologically inspired fashion. The symposium can also be seen as part of a series of related events attempting to promote a high-level long-term vision of achievable scientific goals of AI/Cognitive Science, including The DAM (Designing an Mind) Symposium at AISB'00 (Davis, 2005), the Tutorial on Philosophical Foundations of AI at IJCAI'01 (Sloman and Scheutz, 2001), the St. Thomas symposium in 2002 (Minsky et al., 2004), and the IJCAI'05 Tutorial on Learning and Representation in Animals and Robots (Sloman and Schiele, 2005). It presents themes central to the EC-funded Cognitive Systems initiative,[*] including the CoSy project[†] which is part of that initiative, whose members have helped to organise this symposium, and the euCognition project[‡] which is funding this meeting. A common feature is the focus on *scientific* goals rather than useful *applications* though implementation of working systems is central to the proposed methodology. This introduction to the symposium provides some background and highlights some of the major problems to be overcome.

1 Introduction

In October 2002, under the auspices of The UK Computing Research Committee (UKCRC)[1], Tony Hoare and Robin Milner initiated discussions of "grand challenge" research projects in computing[2]. Seven grand challenge proposals emerged, listed in the booklet available on the UKCRC web site. One of them was "GC-5: Architecture of Brain and Mind — Integrating high level cognitive processes with brain mechanisms and functions in a working robot."[3]

It is concerned with the attempt to understand and model natural intelligence at various levels of abstraction, demonstrating results of our improved understanding in a succession of working robots, along with a succession of increasingly realistic implementations of models of brain mechanisms capable of implementing the competences to be explained.

Robots produced within this grand challenge project should have an interesting and challenging subset of the capabilities of a child aged somewhere between 2 and 5, including the ability to go on learning, and the ability (some of the time) to understand what they are doing and why. One way for such a robot to demonstrate all of that functionality would be being capable of helping and conversing with a disabled person who wishes to avoid being dependent on other humans, at least around the house, without the robot first having to be programmed explicitly with knowledge about that house and its contents, and that person's needs and preferences.

However, it is not enough to produce something that works: we can already do that, thanks to biological evolution. The deep problem that makes this a scientific challenge is explaining how this is possible.

[*] http://www.cordis.lu/ist/cognition/projects.htm
[†] http://www.cognitivesystems.org
[‡] http://www.eucognition.org/
[1] http://www.ukcrc.org.uk/
[2] http://www.ukcrc.org.uk/grand_challenges
[3] http://www.cs.bham.ac.uk/research/cogaff/gc

2 The need to re-integrate AI

Achieving this scientific understanding requires us to bring together work in neuroscience, cognitive science, various areas of AI, linguistics, and other relevant disciplines, to produce an integrated theory of how a functioning system can combine many human capabilities, including various kinds and levels of perception, different kinds of reasoning, planning, problem solving, wondering about, varieties of learning (including grasping new abstract concepts and developing new fluent skills), many kinds of actions of varying complexity, different uses of language, varieties of affect including motivation and emotions, social interaction, and various forms of creativity.

Current robots perform many tasks (some practical, some merely for entertainment) but usually they do not combine their perceptual and manipulative skills with the ability to communicate and cooperate, and they do not know what they are doing, why they are doing it, what difference it would make if they did things in a different way, or how they would have had to change their actions if circumstances had changed, etc., and they cannot give help or advice to another robot or a person performing such tasks.

The most advanced chess playing programs could be installed in a robot, but that would not enable the robot to detect the need to adjust the level of its play to help a beginner, as even a not very advanced human chess player might.

Most current robots need to be given goals because they have no internally generated motives or concerns of their own (although if they include AI planning mechanisms they can generate subgoals of externally provided goals). In particular, they lack the playful, exploratory, curiosity-driven activities that seem to enable human children and some other animals to learn much about their environment in a relatively short time.

However impressively current robots may perform specific tasks on the factory floor or in some demonstration, they do not have the variety of competences, the integration, or the self-understanding of a 3 or 4 year old child and they cannot learn most of the things a child can learn.

3 Why not? What is missing?

There are many reasons for these limitations in the current state of the art, but some of the main ones are:

- We lack deep, comprehensive characterisations of the competences of typical children, at any age, including their visual and other perceptual competences, their manipulative, deliberative, problem-solving, communicative competences, what sort of ontology they use and how it changes, what their motivations and affective states are, what kinds of development and learning go on.
- We do not know what sorts of architectures, forms of representation, virtual machines, brain mechanisms, are capable of producing those competences, including the division between innate, genetically determined, mechanisms and information structures, and what is, grown, developed or learnt as a result of interacting with the physical and cultural environment (Sloman and Chappell, 2005).

It is also fair to say that, although we can specify in very general terms what is required of a useful domestic robot, e.g. the ability to learn to find its way around the building, using vision and other sensors, the ability to manipulate domestic objects of various sorts, the ability to communicate in natural language well enough to obey instructions, answer questions, accept advice, and offer help, we don't really have any clear and detailed set of requirements that are both worth aiming for and eventually achievable.

For instance, although there is often reference to the need for AI systems to be able to 'scale up' to 'human-level' competence, it is generally forgotten that humans do not scale up: as tasks become more complex we can degrade quite drastically, and in many narrowly defined tasks machines are already better than most humans, for instance doing arithmetic, playing chess and controlling sophisticated modern aeroplanes. Although machines often scale up better in specific tasks they do not do what I've called 'scaling out', namely combining old competences in novel ways as circumstances demand.

Another thing that is often said, following (Gibson, 1986), is that robots need to be able to perceive and understand *affordances*, but people have different views of what affordances are, and as far as I know there is no representative list of affordances a good domestic robot will need to perceive and use.

4 Developing a 'roadmap'

So among the main tasks involved in the Grand Challenge are specifying long term requirements and explaining what sorts of mechanisms are capable of satisfying those requirements. Both tasks are so difficult and will last so long into the future that we cannot hope to get them right soon. AI 'prophets' of all fashions and factions have been notoriously over-optimistic in the past, bringing the whole field into disrepute as a result. Can we avoid this mistake?

In the booklet for the IJCAI'05 Tutorial on Learning and Representation in Animals and Robots[4], a scenario-based strategy for achieving greater realism in goals and timescales was proposed. This required collaboratively producing a partially ordered network of scenarios of various kinds and degrees of complexity, ordered according to dependency relationships and difficulty. If such a graph ended with a wide range of scenarios representing, for example, performances of a child-like but very useful and congenial autonomous domestic robot able to help a partially disabled (e.g. blind) person in a typical house, and included very many intermediate scenarios, with relatively small steps between the different stages, all shown to be ultimately dependent on scenarios that might be achieved in the next few years, then that could define a long term 'roadmap' or collection of roadmaps that could be used both to guide research plans and to measure progress.

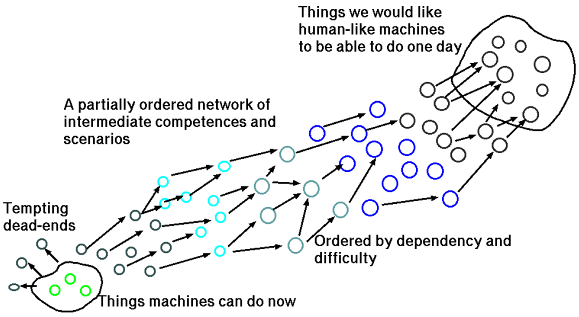

Figure 1

We would know at any time which bits of the graph of scenarios had been achieved and what remained to be done. People could agree on that without necessarily agreeing on which methods, architectures, forms of representations, algorithms, mechanisms, etc. should be used. But the agreed roadmap could provide a common way evaluating progress which is now lacking except for very narrow, specialised, and often arbitrary benchmarks (e.g. sets of images to be classified – a task that may not have much to do with the use of vision in action).

Of course our understanding of what should and should not go into the graph would continue to develop, so the graph would not be fixed permanently from the start: one effect of research inspired by it would be to rebuild the roadmap as we learn more both about requirements and about usable mechanisms and designs.

[4] http://www.cs.bham.ac.uk/research/projects/cosy/conferences

5 Methods and tools to help build roadmaps

A problem with this proposal is that many people find it very difficult to think up a systematic and comprehensive collection of future scenarios of the kind required. (That was the experience of members of the CoSy project team at Birmingham, for instance.) So we have been working on a methodology and some (initially simple) supporting tools to help with development of this network of roadmaps.

The idea is to think in terms of a three-dimensional grid of competences. One dimension represented by columns in Figure 2,[5] is concerned with types of *entity* to which competences can be applied (e.g. 2-D and 3-D spatial locations, regions, routes, inert objects, mobile objects, objects with goals, perception, and action, and various kinds of more abstract objects such as beliefs, proofs, numbers, plans, concepts).

Entity-types	E1	E2	E3	E4	E5	E6	E7	E8
Competences								
C1								
C2								
C3								
C4								
C5								
C6								
C7								

Figure 2

Another dimension illustrated in the rows in Figure 2, is concerned with types of *competence* that can be applied to instances of some or all of the types of entities; for instance competences like perceiving, manipulating, referring to in thought, referring to in language, constructing, destroying, explaining, wondering about, and many more. These two dimensions determine a grid of possible sets of requirements that could form targets for a robot project. Some of the boxes in the grid would, of course, be empty. For instance some of the abstract entities, such as numbers, beliefs, and plans cannot acted on physically. So there would be no scenario examples of such actions.

The third dimension of the grid could be thought of as the depth of the boxes. Within a category some scenarios would be very complex, very difficult and

[5] And at http://www.cs.bham.ac.uk/research/projects/cosy/matrix, and described in
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0602 'Towards a Requirements Grid: A Conceptual Framework and Draft Tool for Generating Requirements and Scenarios'

a long way into the future, whereas others might be much more easily and quickly attainable. An example of the former might involve a robot reliably able to clear away things from a dinner table and put dirty objects into a dishwasher, including glasses, plates, cups sauces, knives, forks, etc., while putting other things into drawers and cupboards, or the refrigerator. A much simpler task could involve the robot using a special-purpose device to carry a book from a shelf to its owner.

When large numbers of examples of competence/object pairs have been described, and their dependencies analysed, it would be possible to define scenarios involving different subsets of entries in the grid.

A new kind of research project could be concerned almost entirely with producing entries for the grid and analysing their dependencies, with a view to identifying promising short term targets that are very likely to be intermediate steps towards very difficult long term goals. Both collaborating and competing groups of researchers could use sets of entries from the grid to specify agreed benchmarks and milestones.

The idea of a rectangular grid of rows and columns was introduced here merely for the sake of exposition. In fact, as already implied by the claim that some boxes will be empty, a more complex structure for the grid will be needed. For instance some of the boxes may have many more sub-divisions than others. There is also a need to be able to refer conveniently to combinations of competences in different parts of the grid as forming a new competence, for example using what is seen to disambiguate a spoken sentence while it is being uttered, or using what someone is saying to facilitate seeing a complex structure. These refinements and elaborations of the grid are topics for future research.

6 Forward-chaining vs backward-chaining research

The 'backward chaining' research methodology described here, contrasts with the 'forward chaining' that most people do, which is to ask what improvements can be added to their current techniques and systems. The problem is that there is no guarantee that those improvements will take us nearer to achieving the long term scientific goals, even if they usefully help to solve immediate engineering problems, such as recognising faulty parts on a production line. An early presentation of this methodology, arising out of a DARPA cognitive systems consultation is here http://www.cs.bham.ac.uk/research/cogaff/gc/targets.html

Unfortunately experience shows that most people find it very difficult to develop the distant scenarios in sufficient detail for them to provide a basis for the backward chaining. This is one of the reasons so much research uses the forward chaining methodology based on trying to think of ways to improve current systems rather than trying to derive new steps by working back from distant coals.

7 Implementation in biological mechanisms

One strand of the research programme as described in the previous sections involves the top-down specification of requirements and exploration of possible ways of meeting those requirements using whatever tools and mechanisms seem to be up to the task. However it is possible that there are deep, not yet understood, difficulties in doing this that mean that we shall eventually learn that only mechanisms with many of the important properties of biological brains are capable of supporting such working human-like performances. However, it is totally implausible that existing computational models of brain mechanisms are anywhere near being adequate to the task.

So a major feature of this grand challenge is that in parallel with the top down research into requirements and design possibilities there should be bottom up (and middle out) research both investigating what the biological mechanisms are and how they work and also attempting to produce artificial systems based on the same principles. It is not likely that merely applying empirical techniques of psychology and neuroscience will suffice to unravel these mysteries, without thinking about designs for virtual machines able to support the many competences that can be observed both in natural settings and in laboratory experiments. If researchers do not know which high level capabilities the brain mechanisms need to support they may not notice or investigate subtle features of the mechanisms that are required for such support, just as a physicist or electronic engineer studying computers without knowing anything about operating systems, compilers, virtual memory, privileges, security, recursive languages, and so on, will probably not come up with a good description of what a computer, considered as an electronic machine, is and does.

Even if everything can be done using conventional computers, the aim of this grand challenge, is not *merely* to understand how such diverse functions can be integrated in single system at a high level of abstraction which might be modelled on computers, or future artificial information-processing machines, but also to explain how they can be implemented in actual biological mechanisms. So an aim of the project is to continue developing our understanding of brain mechanisms (e.g. chemical, neural, etc. mechanisms) including showing how those mechanisms are able to support the high level functionality required by a child or human-like robot robot.

For this purpose, natural minds can be viewed as virtual machines implemented in brains. Since human minds surpass artificial minds in many ways at present, we may discover that this is partly due to using a different kind of physical implementation from current computers. There could be other reasons: it may be that our current designs for AI systems are simply far too simple because we have not yet understood what kinds of functionality they need nor what kinds of architectures, forms of representation and algorithms can provide those kinds of functionality in an integrated system.

Another Grand Challenge being pursued in parallel with this one is GC7: Journeys in Non-Classical Computation, investigation forms of computation that are different from those expressed in conventional programs using conventional computers.[6]. It may be that as both GC5 and GC7 make progress they will have to be more closely intertwined.

8 GC5 and the EC Cognitive Systems initiative

By coincidence, at the same time as this Grand Challenge project was being discussed in the UK in late 2002 and 2003, the European Commission was formulating a very closely related initiative, the *Cognitive Systems* initiative of Framework Programme 6. Since then that initiative has begun to fund a variety of projects, as described on its website[7], including the recently initiated euCognition project[8] which is funding this symposium – for which we are very grateful. Several of the projects already being funded include people who were involved in discussions of the Grand Challenge.

9 Themes for GC5 and the symposium

This grand challenge is far too complex and ambitious to be covered exhaustively in a two-day symposium. The following collection of themes was set out as possible topics for discussion when the symposium was proposed. However it is not possible to cover more than a small subset in the time available, so the list can serve as background to the presentations and help to determine what sorts of comments and questions are relevant. Additional context for the symposium is provided by the enduring symposium website, along with other events, past[9] and future, will help to stimulate discussion long after the conference is over.

A guiding principle in formulating the themes is to replace destructive factional debates about which are the right goals, methods and theories with constructive collaborative analysis of the alternatives and their tradeoffs; and to provide agreed ways of determining whether progress has been made towards the long term goals, avoiding the criticism made by H.L. Dreyfus that results achieved in AI are no more progress towards its goals than climbing trees is progress towards travel to the moon.

Theme 1. Requirements

What needs to be explained/modelled and how can we check that we have good requirements specifications?

This is the core question that drives everything else. There are many things humans can do and selecting a set of competences to be explained and modelled requires great care. The focus on abilities, including learning abilities of young children, arose from the observation that adult competences are typically based on a vast amount of individual learning and idiosyncratic history, whereas the common competences of young children are the basis for many kinds of future development in different cultures and different physical environments, from cave-dwellings to homes in skyscrapers.

But that still leaves open what those competences are and that requires extremely careful observation. (Notice that finding what sorts of competences children are capable of developing is different from doing research on the precise age at which they occur or whether particular environments can accelerate or retard their development.)

Often it is not clear what some human competence is until many examples have been analysed. E.g. some people think that what needs to be explained about vision is simply how a depth-map is computed – distance to contact in all directions. Others believe that the function of a visual system is to segment reti-

[6] See http://www.cs.york.ac.uk/nature/gc7/

[7] http://www.cordis.lu/ist/cognition/projects.htm

[8] http://www.eucognition.org/

[9] E.g. the UKCRC grand challenge conferences in 2004 and 2006, the EC Cognitive Systems 'Kickoff' Conference http://www.cognitivesystems.org and also the 2005 IJCAI tutorial on learning and representation in animals and robots http://www.cs.bham.ac.uk/research/projects/cosy/conferences

nal images and recognise objects. Three more very different alternative requirements specifications are found in the work of D. Marr, in J.J. Gibson and people who emphasise dynamical systems. It is arguable they have all done only a *partial* requirements analysis for vision (or more generally perception).

Similar things can be said about what needs to be explained/modelled regarding: learning, motivation, emotions, affect in general, linguistic ability, reasoning, action control, mathematical abilities, creativity, exploration-based learning, aesthetic capabilities, humour, consciousness, etc.

Work on the requirements grid may help to clarify many of these issues though there is still a vast amount to be done.

Theme 2. Empirical evidence and theories
Research on the grand challenge needs to be informed by research in other empirical disciplines, including biology (e.g. animal behaviour, evolutionary theories), neuroscience, linguistics, cognitive/clinical/developmental psychology, social sciences, that helps either to refine the requirements or support/contradict proposed models and explanations (theme 3).

In particular work on empirical evidence from very young children, unusual humans (e.g. people with brain damage) or other animals can be useful in helping to avoid narrow thinking based on what normal adult humans (in our culture) can do.

What other animals do may provide evidence about evolutionary precursors and about unnoticed subsystems in human capabilities.

Theme 3. Designs for Models/Explanations
This includes:
3.1. High-level (virtual machine) designs
proposed as meeting some subset of requirements (whether for explanation or for applications). [The 'top-down' approach.]
3.2. Implementation designs
3.2.a. Natural/biological mechanisms
3.2.b. Artificial mechanisms for neural/ chemical/developmental mechanisms proposed as capable of supporting the high level designs. [The 'bottom-up' approach.]
Suitable topics for discussion would include reports on current work in progress, and reports on current systems with analyses of how they are inadequate (they are ALL inadequate, in relation to the long term goals of GC5), and how those inadequacies may be overcome or reduced.

I.e. *merely* reporting on what some system can do and how it does it is not appropriate for a GC5 symposium.

Theme 4. Philosophical/conceptual issues
Discussion of what is meant by ascribing various kinds of capabilities and processes to animals or machines, and whether conceptual categories currently in use (e.g. cognition, learning, intentionality, sub-symbolic, emotion, information-processing, consciousness) are confused and in need of refinement.

Eg. do our current concepts allow us to discuss adequately which organisms do or do not process information, or use representations, or have motives or beliefs. Instead of focusing so much on particular capabilities and how to implement them we also need research into a good taxonomy (or other framework) of varieties of types of information-processing capability.

Theme 5. Recommendations for managing GC5
There is an important need for a project like this to have some sort of roadmap, even if it is regularly revised, and criteria for assessing progress.

Developing the three-dimensional requirements grid described above (and improving its structure) could be part of the process of producing such a roadmap. It is also important to resist fragmentation of effort and destructive rivalries.[10]

Theme 6. Social/ethical implications
Many people think social and ethical implications should always be discussed as part of such a project. Often such discussions tend to be either very shallow or wildly speculative or based on some strong ethical bias against advances in AI, or some combination of all of those. Care will be needed to ensure that we do not fall into such traps.

There are potentially profound applications that could follow from significant progress with GC5, not just in the obvious areas of building new smart robots and other machines, but in connection with applications arising from a deeper understanding of how humans work, including new ways of doing education, counselling, therapy, diagnosis of brain disorders and other kinds of mental disorders, and perhaps new forms of treatment.

I am especially interested in ways in which understanding better what enables a toddler to grow up to be mathematician. Such understanding could revolutionise mathematics education at all levels, including

[10] For more on this see http://www.cs.bham.ac.uk/research/cogaff/gc/targets.html (discussion of targets, milestones, in the context of a scenario-based research methodology), and http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0503 (extract from the IJCAI'05 tutorial booklet.)

primary schools. But discussing all this may still be premature.

Acknowledgements

Thanks to the euCognition network for funding this symposium, and for members of the EU-funded CoSy project for help with ideas in this introduction and with arrangements for the symposium.

References

D. N. Davis, editor. *Visions of Mind: Architectures for Cognition and Affect*. Idea Group Inc., Hershey, PA, USA, 2005.

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

M. Minsky, P. Singh, and A. Sloman. The St. Thomas common sense symposium: designing architectures for human-level intelligence. *AI Magazine*, 25(2):113–124, 2004. http://web.media.mit.edu/˜push/StThomas-AIMag.pdf.

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

A. Sloman and M. Scheutz. Tutorial on philosophical foundations: Some key questions. In *Proceedings IJCAI-01*, pages 1–133, Menlo Park, California, 2001. AAAI. http://www.cs.bham.ac.uk/˜axs/ijcai01.

A. Sloman and B. Schiele, editors. *Tutorial on Learning and Representation in Animals and Robots*. IJCAI'05, Edinburgh, 2005. Available at www.cs.bham.ac.uk/research/projects/cosy/conferences.