GRAND CHALLENGE 5:
The Architecture of Brain and Mind

Integrating Low-Level Neuronal Brain Processes with High-Level Cognitive Behaviours, in a Functioning Robot

Web site: http://www.cs.bham.ac.uk/research/cogaff/gc/

Moderator: Aaron Sloman[1]

Introduction

What is the most powerful and most complicated computer on the planet? Wrong! it's not a machine you can buy for millions of dollars, it's the amazing system that we all own, the few kilos of grey and white mush in our heads.....

Brains are the most impressive products of biological evolution. We are still far from understanding what they do and how they do it, although neuroscience is advancing rapidly. Biological evolution, in combination with social evolution and individual learning, also produced our minds, which are intimately connected with our brains. Despite much progress in psychology, we are also far from understanding many of the things minds do and how they do them.

- *Brains*, the contents of our skulls, are composed of extraordinarily intricate physical structures, performing many tasks in parallel at many scales, from individual molecules to large collections of cooperating neurones or chemical transport systems. In some ways brains are like computers, insofar as they have vast numbers of components that can change their states, and this allows them to perform a huge variety of different tasks. In other ways brains are very different from computers — though the study of brains has already begun to inspire the design of computers of the future.
- *Minds* are abstract machines whose existence and ability to do things depend on all the 'wetware' components that make up brains. But, whereas brains contain visible and tangible things like nerve cells and blood vessels, minds contain invisible, intangible things like ideas, perceptions, thoughts, desires, emotions, memories, knowledge and skills of many kinds. Minds and their contents cannot be seen by opening up skulls, though their effects are visible all around us in many physical behaviours of humans and other animals and in enormous changes in the physical world brought about by processes in minds, including the development of computers.

Gaps in our knowledge

Many processes in brains and minds are not yet understood, including how we:

- see many kinds of things around us,
- understand language (like the language you are now reading),
- learn new concepts,
- decide what to do,
- control our actions,
- remember things,
- enjoy or dislike things,
- become aware of our thoughts and emotions,
- learn about and take account of the mental states of others,

[1] School of Computer Science, The University of Birmingham. http://www.cs.bham.ac.uk/˜axs/. Several others contributed, especially Mike Denham. Jim Cunningham suggested several improvements on a first draft.
(Last revised November 14, 2004.)

- appreciate music and jokes,
- sense the passage of time.

In this project we aim to understand both brains and minds well enough to produce robots with a large collection of human-like capabilities, unlike all current robots, which are very limited.

The historical roots of the project

This research addresses two centuries-old quests that have fired the imagination of many inventors, scientists, philosophers and story-tellers: the attempt to understand what we are, and the attempt to make artificial human-like systems, whether entertaining toys, surrogate humans to work in inhospitable environments or intelligent robot helpers for the aged and the infirm.

Minds are machines: virtual machines

Minds are in some ways like software systems running on computers: since both minds and software systems depend on complex physical machinery to enable them to exist and to perform their tasks. Yet neither minds nor running software systems can be observed or measured by opening up the physical machinery and using methods of the physical sciences, and in both cases the mappings between the abstract components and the physical components are very subtle, complex, and constantly changing.

This 'strange' relationship between invisible, intangible mental things and our solid flesh and bones has intrigued and mystified mankind for centuries. Despite the efforts of many philosophers and scientists it remained unexplained for centuries. However, around 1842 Ada Lovelace[2] wrote some notes on the 'analytical engine' designed by the British mathematician and inventor Charles Babbage[3] in which she showed a clear recognition of something that only began to become clear to others more than a century later. She noted that the Analytical Engine, like other calculating machines, acted on numbers, which are abstract entities, and observed that it

> *might act upon other things besides number, were objects found whose mutual fundamental relations could be expressed by those of the abstract science of operations, and which should be also susceptible of adaptations to the action of the operating notation and mechanism of the engine . . .*

She suggested, for instance, that using this "abstract science of operations .... *the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent.*" This idea of a *virtual machine* running on a *physical machine* and performing many abstract operations was expanded and refined through developments in computer science and software engineering in the 20th Century, following pioneering work of Alan Turing[4] and many others. We now know how to design, make, debug, maintain and sell(!) many kinds of virtual machines including word-processors, email systems, internet browsers, operating systems, planning systems, spelling correctors, teaching packages, data-mining packages, plant control systems, and many more. Many of them outperform humans on specific tasks (e.g. playing chess, statistical analysis), yet there is nothing that combines the capabilities of three year old child, or even a squirrel, or a nest-building bird. The most powerful chess machines cannot discuss or explain their strategies.

Animal brains in general and human brains in particular have many capabilities as yet unmatched by anything we know how to build. Can we ever replicate all that functionality?

[2] See **http://www-gap.dcs.st-and.ac.uk/˜history/Mathematicians/Lovelace.html** and **http://www.sdsc.edu/ScienceWomen/lovelace.html**,

[3] See **http://ei.cs.vt.edu/˜history/Babbage.html** Babbage's design was partly influenced by the role of punched cards in Jacquard's automated loom. **http://www.csc.liv.ac.uk/˜ped/teachadmin/histsci/htmlform/lect4.html**

[4] **http://www.turing.org.uk/turing/**

The availability of more powerful computers than ever before, advances in many areas of Artificial Intelligence and Computer Science, and an ever-growing body of research results from Neuroscience and Psychology provide the inspiration for a new concerted attempt at a major leap forward, by developing new deep theories, tested in a succession of increasingly ambitious working models, namely robots combining more and more human abilities, e.g. able to see, manipulate objects, move around, learn, talk about what they are doing, develop self-understanding, think creatively, engage in social interactions and provide both advice and physical help to others when appropriate.

Building on many disciplines

Much research in Computer Science departments is already strongly interdisciplinary, but like many major breakthroughs this project will require new ways of combining results from researchers in several disciplines: including

- *neuroscientists* studying brain mechanisms and architectures,
- *psychologists, linguists, social scientists, ethologists and philosophers* studying what minds can and cannot do, and
- *researchers in computer science and AI* developing techniques for *specifying* and *implementing* many kinds of abstract mechanisms and processes in present and future physical machines.
- *researchers in mechanical engineering, materials science, electronic engineering* extending materials and mechanisms available for robot bodies and brains.

Inspired by both past and future advances in neuroscience the project will attempt to build machines that simulate as much as is known about how brain mechanisms work. In parallel with that, the project will attempt to implement many kinds of abstract mental processes and mechanisms in physical machines, initially without requiring biological realism in the implementation mechanisms, but gradually adding more realism as our understanding increases.

How will it be done?

Several mutually-informing tasks will be pursued concurrently:

Task 1 *Bottom-up* specification, design, and construction of a succession of computational models of brain function, at various levels of abstraction, designed to support as many as possible of the higher level functions identified in other tasks.

Task 2 Codification and analysis, partly from a software engineering viewpoint, many typical, widely-shared, human capabilities, for instance those shared by young children, including perceptual, motor, communicative, emotional and learning capabilities, and using them:

(a) to specify a succession of increasingly ambitious design goals for a fully functioning (partially) human-like system,

(b) to generate questions for researchers studying humans and other animals which may generate new empirical research leading to new design goals

Task 3 *Top down* development of a new theory of the kinds of *architectures* capable of combining all the many information-processing mechanisms operating at different levels of abstraction, and testing of the theory by designing and implementing a succession of increasingly sophisticated working models, each version adding more detail.

Different research groups will study different parts of the problem, but always in the context of the need to put the pieces together in working systems. Analysing ways in which the models produced in task 3 succeed and fail, and why, will feed back information to the other two tasks.

Targets, evaluation and applications

As a 15 to 20 year target we propose demonstration of a robot with some of the general intelligence of a young child, able to learn to navigate a typical home and perform a subset of domestic tasks, including some collaborative and communicative tasks. Unlike current robots it should know what it is doing and why, and be able to cooperate with or help others, including discussing alternative ways of doing things. Linguistic skills should include understanding and discussing simple narratives about things that can happen in its world, and their implications, including some events involving capabilities, motives and feelings of humans. The robot could be tested in various practical tasks, including helping a seriously disabled or blind person cope without human help.

This long-term target will be broken down into a large collection of sub-goals, with different subsets used to define intermediate milestones for judging progress. Achieving all this will require major scientific advances in the aforementioned disciplines, especially if one of the sub-goals is production of biologically plausible mechanisms capable of supporting the robot's functionality.

Success could also provide the foundation for a variety of practical applications in many industries, in unmanned space exploration, in education, and in the ever-growing problem of caring for disabled or blind persons wishing to lead an active life without being totally dependent on human helpers. Perhaps some people reading this will welcome such a helper one day.

Although many practical applications are possible, the primary goal is to increase our understanding of the nature and variety of natural and artificial information-processing systems. This is likely to influence many areas of research including psychology, psychiatry, ethology, linguistics, social science and philosophy. It could transform ideas about education.

There is no guarantee of success: The project could fail, if it turns out, for instance, that the problems of understanding how brains work are far deeper than anyone imagines, or if replicating their capabilities turns out to require much greater computer power than will be available in the next few decades. However, there is a distinct possibility that this research will eventually lead to the design of new kinds of computers that are far more like brains than current machines are.

International collaboration

Several international research programmes, including "Cognitive Systems" initiatives in Europe, in the USA and in Japan are now supporting related research: international collaboration is essential for success in such a demanding project. Work inspired by the UK Foresight Cognitive Systems initiative will help with the foundations of the project. Several projects funded by the 25M euro EC Framework 6 "Cognitive Systems" initiative starting in 2004,[5] will provide significant steps towards the achievement of the grand challenge.

Sources of further information

- **http://www.nesc.ac.uk/esi/events/Grand_Challenges/**
  General information about the UK Computing Research Grand Challenges initiative.
- **http://www.cs.bham.ac.uk/research/cogaff/gc/**
  More detailed information about the "Architecture of Brain and Mind" proposal.
- **http://archives.nesc.ac.uk/gcproposal-5/**
  Email discussion archives, leading up to the development of this proposal.
- Enquiries may be sent by email to: Aaron Sloman <A.Sloman@cs.bham.ac.uk>

[5] E.g. this FP6 project includes a UK partner: **http://www.cs.bham.ac.uk/research/projects/cosy/**