Preprint (author's cut) version of a contribution to a
Handbook of Computational Psychology

Putting the Pieces Together Again

Aaron Sloman
School of Computer Science
University of Birmingham, UK
http://www.cs.bham.ac.uk/~axs/

December 8, 2007

Contents

1 Introduction

Instead of attempting to sum up the *achievements* of the field, this chapter complements the review of requirements for work on integrated systems in Chapter 1, by presenting a personal view of some of the major unsolved problems, and obstacles to solving them. In principle, this book should soon be out of date, as a result of world-wide growth in research on cognitive systems. However, it is relatively easy to identify long-term ambitions in vague terms, e.g. the aim of modelling human flexibility, human learning, human cognitive development, human language understanding or human creativity; but taking steps to achieve those goals is fraught with difficulties. So progress in modelling human and animal cognition is very slow despite many impressive narrow-focus successes, including those reported in earlier chapters.

An attempt is made to explain why progress in producing realistic models of human and animal competences is slow, namely, (1) the great difficulty of the problems; (2) failure to understand the breadth, depth and diversity of the problems; (3) the fragmentation of the research community; and (4) social and institutional pressures against risky multidisciplinary, long-term research. Advances in computing power, theory and techniques will not suffice to overcome these difficulties. Partial remedies will be offered, namely identifying some of the unrecognised problems and suggesting how to plan research on the basis of 'backward-chaining' from long term goals, in ways that may, perhaps, help warring factions to collaborate and provide new ways to select targets and assess progress.

1.1 The scope of cognitive modelling

Although Artificial Intelligence (AI) and Cognitive Science have different aims, AI has always had two overlapping, mutually-supporting strands, namely *science* (concerned with understanding what is and is not possible in natural and artificial intelligent systems), and *engineering* (concerned mainly with producing new useful kinds of machines). 'Cognitive science' and 'cognitive modelling' overlap significantly with the science strand of AI (as documented in great detail in M. Boden (2006)). However, the majority of AI researchers have a strong engineering orientation. In contrast, those who were responsible for many of the key ideas in AI were all interested in AI primarily as a contribution to the general 'science of mind', including, for example, Turing (1950), McCarthy (2004), Minsky (2006), Simon (1997), and Newell (1990). Some people wish to restrict cognitive science to a *subset* of AI as science, namely the study of 'natural' systems. Many are even more narrow and study only *human cognition*. Such restrictions reduce opportunities to understand how systems produced by evolution relate to the space of *possible* behaving systems – as if physicists restricted their studies to naturally occurring physical entities (e.g. plants and planets) while ignoring physical phenomena in artefacts like prisms, plasma lamps and power stations. So, a full understanding of how minds work, deep enough to support realistic modelling, requires a broad multi-disciplinary approach combining neuroscience, psychiatry, developmental psychology, linguistics, animal behaviour studies (including insects and microbes), biology, computer science, mathematics, robotics, linguistics, and philosophy. It may also require physics and chemistry for reasons explained later. We need all those disciplines in order to understand the variety of possible ways of combining different kinds of competence within an integrated, embodied, organism-like agent, able to learn and develop as ants, antelopes and humans do.

We cannot expect everyone working on highly focused research problems to switch immediately to long term research on *integrated* biologically inspired systems combining multiple competences. But more researchers should think about these issues. In particular, we should inspire some of the brightest young researchers to do so, despite all the institutional and financial pressures to focus on narrow, more practical, goals, and despite the deep intellectual difficulties discussed in the following sections.

1.2 Levels of analysis and explanation

Different kinds of commonality link natural and artificial systems, some concerned with physical mechanisms, some with high level (virtual machine) design features, and some with task requirements. Organisms use many different physical mechanisms in their sensors, effectors and internal information-processing systems. Not all of them have brains or nervous systems, though all acquire and use information from the environment in controlling behaviour, e.g. determining direction of movement in microbes, of growth of shoots or roots, or orientation of leaves in plants. Likewise, artificial behaving systems differ in processor design and materials used, and at present use totally different physical mechanisms from those used in organisms, although those differences may be reduced in future.

Progress will depend not only on analysis of the variety of physical substrates and their trade-offs, but also investigation of types of form of representation, types of algorithm, types of dynamical system, and types of architectural decomposition – independently of how they are implemented in physical substrates. Many researchers ignore details in the hope of capturing important features of biological systems. For instance, people developing neural net models tend to ignore the roles of the many chemicals involved in biological information-processing, such as neurotransmitters and hormones. It could turn out that such omissions seriously undermine the long-term significance of their models, so more needs to be understood about the roles of physical and chemical information-processing in organisms. (Compare the chapter by Stellan Ohlsson, this volume – Chapter 13).

In addition to research on underlying mechanisms, and research on high level virtual machine specifications, there is also a type of research that involves identifying the diverse *requirements* that the mechanisms and designs need to satisfy. This is sometimes called 'task analysis', though organisms do not merely perform one type of task. Tasks that cognitive scientists need to analyse are more like the collection of requirements for travel to remote planets, compounded with use of diverse bodily forms in diverse environments. Requirements for a whole organism include a specification of the niche, or sequence of niches, in which it evolved or developed.

2 Difficulties and how to address them.

Previous chapters have said much about explanatory mechanisms and architectures. Most of this chapter is about requirements (i.e. niches) and how to learn about them: a much harder task than many suppose. Contributing to a field like this requires several years of multidisciplinary post-graduate study in order to understand the problems. Unfortunately, both institutional features of current academic research environments and intrinsic features of the research problems make this difficult.

2.1 Institutional obstacles

Severe institutional and financial deterrents obstruct multidisciplinary research, even when lip-service is paid to the idea. It is especially risky for untenured researchers worried about getting research grants and building up publication lists – achieved more easily in work that focuses only on minor extensions or new applications of old techniques. That pressure prevents researchers from taking time to acquire the multidisciplinary knowledge required to investigate integrated systems (whole animals, whole robots) and makes them choose less risky research strategies. It also produces factions working on different sub-problems who compete for funds, students, and attention, instead of collaborating, and whose teaching produces only partly educated students who believe myths such as that symbolic AI 'failed', or, at the other extreme, that neural mechanisms are merely aspects of implementation. Although many funding agencies promote research that integrates subfields and disciplines (Hendler 2005), because of the career risks, research on integrated multifunctional ('whole-animal') systems remains impossible for younger researchers. (That is in addition to the intrinsic difficulties described in Section 2.2.) Reducing institutional pressures causing individuals to focus narrowly will require the academic community to teach politicians and managers that assessment by *measurable* results is no way to foster deep, high-calibre research or teaching. Sections 7 and 8 present a new research framework that may help to integrate future research communities, although that will require exceptionally able researchers.

2.2 Intrinsic difficulties in making progress

Many researchers have identified detailed obstacles to progress. Some very high-level obstacles, however, appear not to have received much attention. The first difficulty arises because rival researchers argue about whose algorithm, architecture or form of representation is best, instead of studying the structure of the space of alternatives and the trade-offs involving those options. There usually is no 'best' alternative.

A related difficulty comes from studying only normal adult human capabilities in a specific community, ignoring not only the deep genetic heritage humans share with many other animals, but also the variations in human capabilities across ages, cultures, and personality types, and the varied consequences of brain damage or deterioration. Many of the facts that need to be explained are not even noticed.

Research planning is hard because identifying good ways to make progress toward very distant goals is hard when we do not yet have a clear vision of the intermediate research stages that can lead to those goals. A partial solution is suggested in Section 9.2.

Many further difficulties arise from limitations of our conceptual tools: After less than a century of intense investigation of mechanisable varieties of information-processing, what we still understand about forms of representation, about mechanisms for processing information, and about designs for architectures, is only a tiny fragment of what evolution produced over billions of years. Limitations of current materials science and mechatronic engineering capabilities may also have deep implications regarding still unknown requirements for varieties of information processing underlying animal intelligence, for example sharing of functions between hardware and software as described in Berthoz(2000).

3 Failing to see problems: ontological blindness

Funding and institutional problems, and inadequacies of our concepts and tools have been listed as brakes on progress. A deeper hindrance is the difficulty of identifying what needs to be explained or modelled, often arising from 'ontological blindness' discussed in (Sloman & Chrisley, 2005). Any information-user must use an ontology, which to a first approximation is the set of *types* of entities the user can refer to. Gaps or spurious types in the ontology can cause researchers to mis-identify what organisms are doing. So they fail to identify the various subfunctions that need to be modelled.

Section 4 illustrates this in connection with modelling human vision, although the points are also applicable to other animals and other aspects of cognition.

Another common mis-identification concerns varieties of learning. It is often assumed that humans, and therefore human-like robots, necessarily start with very limited innate knowledge about the environment and have to use very general, knowledge-free forms of learning. This assumption ignores the fact that most organisms start with almost all the knowledge and competence they will require, because of their need to be independent from birth or hatching (like deer that walk to their mother's nipple and run with the herd very soon after birth). If evolution can produce so much information in the genome, for some species, why does it apparently produce helpless and ignorant infants in altricial species, for example humans and hunting mammals. Chappell and Sloman (2007) propose that such animals start with deep meta-competences using generic information about the environment, including information about how to use the body to explore and learn about the details of a 3-D environment. This is an old idea as regards language learning (Chomsky, 1965), but language learning could be just a special case of the use of partly innately specified meta-competences that generate both new competences and new meta-competences through active investigation of the environment.

Other misrepresentations of the requirements for human-like systems include the assumption that intelligent agents must use a sense-think-act cycle. E.g. the web page of a leading AI department states: "An intelligent agent should also learn to improve its performance over time, as it repeatedly performs this sense-think-act cycle", ignoring the fact that in humans and most other animals many kinds of information processing proceed in parallel, e.g. simultaneously walking, planning where to walk, enjoying the view and listening to what a companion is saying, concurrently with many bodily control functions. Different misrepresentations of requirements come from paying close attention to implementation details and concluding that all intelligence is to be explained in terms of the dynamics of sensorimotor control loops, ignoring the fact that many humans can listen to and understand a story, or look at and appreciate a complex scene without producing any relevant motor responses, and some can even prove theorems in their heads about transfinite ordinals, or plan a musical composition without producing any sounds.

4 What are the functions of vision?

Marr (1982) suggested that the function of vision was to provide information about geometrical and physical properties of the environment, e.g. shape, size, location, motion, and colour of objects. Many readers thought this obviously correct. However, in 1979, Gibson had pointed out that there are far more subtle functions of perception, namely providing information about "affordances" that are abstract properties of the environment related to possible actions and goals of the perceiver. This was generalised in Sloman (1982, 1989), by drawing attention to human abilities to learn to see states of mind and to read writing, music, and various other formal notations.

On the first view, vision is the same for a lion and a lamb surveying the same terrain, whereas 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. Requirements for catching and eating meat are very different from requirements for grazing: vegetable matter does not attempt to escape, and grass does not require peeling or breaking open. Similar differences exist between birds that build nests from twigs and those that build nests using only mud, or between birds that fly and birds that do not, birds that dive for fish and birds that catch insects, and so on, suggesting that what is seen by two different animals looking at a scene can differ as much as what is seen by people who know different languages looking a page showing texts written in different languages.

4.1 The importance of mobile hands

More subtle differences in requirements for vision depend on having hands that can move and manipulate objects. Animals that can manipulate things only using a mouth or beak will have very strong correlations between grasping actions and patterns of optical flow, because of the rigid link between eyes and mouth or beak, whereas animals with mobile hands or claws must be able to represent actions that move things without changing the viewpoint. Independently movable hands can perform essentially similar grasping actions producing vastly different visual feedback, so using only sensorimotor relationships will make it difficult to represent what is common to many ways of grasping, namely:

> *Two (or more) surfaces facing each other move together in 3-D space until they are in contact with opposite sides of some other object that can then be lifted, rotated, or moved horizontally.*

An 'exosomatic' ontology, referring to changing relationships in the external environment, is required for this, rather than a 'somatic' ontology that refers only to changing sensory and motor signals. An exosomatic ontology makes it much simpler to transfer facts learned about grasping in one situation to grasping in another.

If grasping is represented amodally in terms of changing relations between 3-D surfaces, with changing causal consequences, then how the process is represented is independent of how it is sensed or what motor signals are required to produce it. Not representing the process in terms of "sensori-motor contingencies" has useful consequences. One is that information about the occurrence of such processes can be stored compactly if what needs to be remembered is that grasping defined as a process in 3-D space occurred, not how it was done or what sensor or motor signals were produced during the process. That allows generalisations to be learnt at a high level of abstraction, applicable to wide ranges of cases, e.g. grasping with the left hand, with the right hand, with both hands, or with teeth. Such amodal generalisations include such facts as:

(a) a tightly grasped object undergoes the same translation and rotation as the two grasping surfaces,

(b) the grasped object may fall if the pressure applied by the two surfaces is reduced,

(c) when a grasped object is released it accelerates downwards, and

(d) grasping changes the shape of non-rigid objects.

How such generalisations are encoded, how they are learnt, how they are retrieved when relevant, how they are used in making predictions, explaining facts, or forming plans, all need to be explained: and if the need is unnoticed then the models of the role of vision and cognition will be inadequate.

Another advantage of representing processes independently of the sensory and motor signals involved is that plans for future actions can be formed without considering details that can only be determined in the context of action. Information about previous actions or previously observed processes can also be stored in a more abstract form.

An amodal ontology allows what an individual learns by grasping to be transferred to grasping done by another individual, and vice versa. So actions done by others can be predicted, failures understood, and preventive measures taken (e.g. when a predator threatens, or a child is about to fail to reach a goal). Moreover, processes that are observed but produced by nobody (e.g. an apple falling) can suggest goals to be achieved. So-called 'mirror neurons' might best be construed in terms of use of an ontology that abstracts from sensorimotor details. Perhaps 'abstraction neurons' would have been a better label. Actor-independent representations of certain actions can allow parents to perceive 'vicarious affordances' for offspring or predators, enabling them to use causal reasoning to help or protect offspring and to obstruct or avoid predators. This makes possible the 'scaffolding' of learners by parents and teachers discussed by developmental psychologists. As noted in Chappell and Sloman (2007), there are hunting birds and mammals that give their young practice in dealing with prey, in a manner that suggests that they understand what their offspring need to do. Grasping is merely an example: use of landmarks, observed in ants and wasps, may be another. Use of amodal, exosomatic, representations of processes involved in actions is not a feature of all animal vision: but it is certainly a human competence and the extent to which other animals have it requires research.

All this imposes strong constraints on models of visual cognition that aim for completeness. Other constraints include the need to explain how visual servoing works, where precise, continuous motion is constantly monitored and modulated in sensorimotor feedback and feedforward loops.

This distinction between the use of vision in 'online' control of fine details of actions and its use in acquiring more abstract reusable information applicable to many different situations, using dorsal and ventral pathways respectively, has, in recent years, been wrongly described as a difference in perception of 'what' *vs.* 'where' or 'what' *vs.* 'how', and related to ventral and dorsal pathways. The mistake was corrected in Goodale and Milner (1992), albeit using a misleading contrast between perception and action, as opposed to two perceptual (visual) functions. Both kinds of perception can be used in actions, though their roles are different.

An 'objective' (exosomatic, amodal) ontology can also be used to represent a hand moving to grasp an object that is not in view because the eyes have moved to get a view of something else, or the grasped object is obscured by something. Absence of sensorimotor details also allows representations to be useful in planning or predicting future actions where what motor and sensory signals will be involved can be left unspecified – because that can vary according to detailed circumstances at the time of action. Without use of amodal abstractions, the combinatorial complexity of the process of searching for a plan or prediction or explanation would be far greater.

Grasping is, of course, just one example. All this still leaves unexplained how a visual system manages to derive amodal descriptions of 3-D structures and processes from sensory input, but research on that is in progress in AI, psychology and neuroscience e.g. (Hayworth & Biederman, 2006).

Ontological blindness to these possibilities leads some researchers to suppose that vision is merely (or mostly) object recognition, or that vision uses only image-based or more generally sensori-motor representations. There is now a vast amount of research on models of sensorimotor learning, using architectures that cannot do anything else, e.g. (Lungarella & Sporns, 2006). Such models cannot explain most of the uses of vision.

4.2 Seeing processes, affordances and empty spaces

Most vision researchers (including me for many years) have concentrated on perception of *static* scenes – blind to the fact that perception occurs in an environment that is in motion much of the time, with constantly changing viewpoints as animals move. So a major function of a vision is to provide information about which *processes* are occurring, not just about structures and relationships. In that case, viewing static scenes is a special case of perceiving processes in which nothing much happens. So perhaps the *primary* form of representation for visual information should be representations of processes extended in time and space, whereas many researchers (an exception being Grush, 2004), assume that vision produces information about objects, properties and relationships. Of course, that is *part* of what needs to be explained, but only part. If vision uses representations of processes including affordances concerned with what can and cannot happen or be done the representations must be rich in *conditional* information about processes.

One test for whether a vision system perceives affordances is how it sees empty 3-D or 2-D spaces. If all the functions of a visual system are concerned with perception of objects, then empty space cannot be seen, whereas humans can see an empty space as full of potential for various kinds of occurrences, 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 are. A bird holding a twig to be added to its partially built nest needs to see places where that twig would be useful, and a route by which it can be inserted. Someone like Picasso can see potential in a blank surface that most people cannot. Harold Cohen's AARON program, described in M. A. Boden (1990) and accessible at http://www.kurzweilcyberart.com, also has some grasp of 2-D affordances and how they change as a painting grows. 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 ever-increasing numbers of sides just inside and just outside the circle – one of many cases of mathematical use of the visual ability to represent processes. Other cases are discussed in Sloman (1978) and Anderson, Meyer, and Olivier (2001).)

So, vision researchers who focus only the task of extracting from the optic array information about things *that exist in the scene*, exhibit ontological blindness insofar as they ignore the role of vision in seeing *what does not yet exist but could exist*, that is, the positive and negative affordances. Many also ignore the importance of seeing *ongoing processes* in which both structures and affordances are changed over time.

4.3 Seeing without recognising objects

A vast amount of vision research is concerned with recognition. But that fails to address seeing without recognising objects, which involves acquiring information about spatial structure, relationships, affordances and processes. Perception of structure involves recognition, not of whole objects, but of image and scene *fragments*, such as occluding edges, bumps, dents in surfaces, partially visible edges, changing curvature, and specularities. Perception of spatial structures and relations can be the basis of recognition, and may sometimes be facilitated by recognition. But systems designed only for recognition of 'whole' objects must fail when confronted with new things! 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. Range-finders have been used to obtain 3-D structure, but usually 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. A useful survey is in Várady and Martin (1997).

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 tearing open and eating its prey). Likewise, when Jackie Chappell presented parakeets with cardboard 'polyflaps' (Sloman, 2006a), they played with, manipulated, and chewed them, despite never having seen them previously.

Infants spend much time developing competences related to various 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, and often does not require the whole objects to be classified or recognised.

A good theory might explain ways in which brain damage can differentially affect abilities to see various kinds of surface features and affordances, without removing the ability to see, as illustrated by *prosopagnosia*, an affliction in which the ability to recognize faces is lost.

In summary: there are deep and general forms of perception and learning that we need to understand in order to understand important aspects of vision on which many other competences build, in humans and other animals, including spatial and causal reasoning capabilities.

4.4 Many developmental routes to related cognitive competences

We should not assume that human visual competence (or any cognitive competence) depends on having a specific bodily form, e.g. having hands that can manipulate things. Babies born without arms, as occurred in 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 a kind of *vicarious* learning capability used when watching others do things we cannot do ourselves, using an exosomatic ontology, as discussed in Section 4.1. Perhaps a shared evolutionary heritage provides the ability to develop a core set of amodal forms of representation that enables severely disabled children to learn about structures, processes and affordances through watching others do things they cannot do. This ability to learn about, perceive and make use of vicarious affordances undermines some claims about cognition as intimately tied up with embodiment. It is arguable that having a human mind depends more on having had embodied ancestors than on being embodied.

4.5 The role of perception in ontology extension

At any particular time, an animal or child will have developed an ontology that is used in percepts, predictions and plans, all of which represent entities, relationships, processes and affordances in the environment. But things can go wrong: plans can fail and predictions can turn out false. The infant who takes a cut-out picture of an animal out of its recess and then later tries to replace it can fail, being surprised when it doesn't fit the recess. Such failures could trigger 'debugging' processes that sometimes lead the child to extend the high-level ontology, e.g. using low-level sensory features that were previously disregarded. For example the child may, somehow, extend its ontology to include the concept of the *boundary* of a flat object and the concept of two boundaries being *aligned*. After that change, the failure to get the puzzle piece into its recess may be overcome by performing additional actions to align the two boundaries. For this, the ontology will need to include processes like *sliding*, *rotating* and *coming into alignment*.

Some toys are cleverly designed to require a less complex ontology. For stacking cups that are symmetrical, boundaries need not be aligned during insertion. Making cups conical allows small bases to be inserted into larger openings, reducing the need for precision in placing. Careful observation of actions of infants and toddlers at various stages of development reveals subtle ways in which they encounter difficulties that seem to be based on not yet having a rich enough ontology that they later extend –

perhaps driven by detecting differences between actions previously seen as similar, or by modifying preconditions or consequences of actions to include relationships previously not representable. An 11 month old child is described in Sloman, Chappell, and CoSyTeam (2006) who was able to feed himself yogurt from a tub using a spoon to transfer the yogurt to his mouth, but failed to transfer yogurt to his leg because he merely placed the bowl of the spoon on his leg, apparently not realizing that it needed to be rotated. There are probably hundreds, or even thousands, of such processes of self-debugging leading to ontology extension in the first few years of life of a human child. Those extensions depend on types of objects (including types of food and clothing) in the environment, whose properties and behaviours can vary enormously from one part of the world to another and sometimes change as a result of cultural development. For example, many children born recently have acquired an ontology appropriate for interacting with a computer using a mouse, which none of their ancestors encountered. Some of the transitions in which new competences are acquired were studied by Piaget many years ago (1954), but the time may be ripe for renewed systematic study facilitated by the ability to use video recordings so that many different people can examine the same episode. Researchers with expertise in designing robots should have richer ontologies with which to perceive and think about what infants do, or fail to do.

5 Representational capabilities

In order to explain how a child extends an ontology we need to know what representations are used. What sort of representation does a child or chimp have of a three dimensional curved surface such as various parts of a spoon? How are the causal capabilities represented? There are many mathematical ways of representing shapes, for instance using differential equations, or using very large vectors of point features. But those representations may not be adequate for cognitive purposes if they are too difficult to derive from the available sensory information (e.g. because of noise, low resolution of the parts of the visual field, or lack of suitable algorithms). The mathematical representations may also be unsuited to the derivation of affordances, and hard to use in planning or in controlling actions. Explaining cognitive competences in dealing with a 3-D environment may require new forms of representation that capture spatial structure in a manner that abstracts from the precise details that would be represented in differential equations and collections of point features, and are better tailored to facilitating action selection and control. It is likely that evolution 'discovered' many more forms of representation and corresponding mechanisms than human mathematicians, scientists and engineers have so far thought of.

Besides the difficulty of specifying the forms of representation used, there is the problem of explaining how they are implemented in brain mechanisms. My impression is that despite vast advances in detailed tracing of neuronal connections, the study of chemical brain process, and the recent development of more and more fine-grained brain imaging devices, there is still very little understanding of how the mechanisms so far discovered are capable of supporting most of the cognitive functions we believe humans and other animals are capable of. For example, standard neural models assume that all structures and processes can be represented in the contents of large vectors of values of sensory and motor signals, possibly at different levels of abstraction. We seem to need different sorts of computations, involving different information structures, in order to make progress in modelling cognitive processes. Some of the requirements are identified in Trehub (1991) and some hypothetical neural mechanisms proposed. But it is not clear whether they can meet all the requirements.

Research on these topics is extremely difficult. Perhaps that explains why the tasks identified by Barrow and Tenenbaum (mentioned in 4.3) have largely been forgotten, while most vision researchers work on other tasks that do not involve detailed understanding of spatial structure and affordances. Great progress has been made in developing mechanisms with narrow competences, like object recognition or classification, object tracking, trajectory prediction, pushing or grasping simple objects (e.g. Saxena, Driemeyer, Kearns, Osondu, & Ng, 2006) and path traversal – all of which are worthy research topics, of course, but form only a relatively small subset of functions of vision. Other functions, not discussed here, include the role of vision in fine-grained control of actions (visual servoing), posture-control, perceiving varieties of motion, developing many kinds of athletic capabilities using vision, parking a car or other vehicle, 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 text first laboriously then later fluently, 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 seeing what the opponent is doing. This probably depends on the concurrent operation of mechanisms that perform fast and fluent well-learnt tasks reactively and mechanisms that have more abstract and flexible deliberative capabilities (Sloman, 2006b).

It might be fruitful to set up a multi-disciplinary project to expand our ontology for thinking about vision, including a comprehensive taxonomy of functions of vision, along with 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. Similar things could be done for other major cognitive functions. Such projects will benefit from the scenario-driven research described in Section 7.

5.1 Is language for communication?

Similar kinds of ontological blindness can afflict students of language. A common assumption is that the sole function of language is communication: meanings are assumed to exist and language is used to convey them. But many are blind to the deeper problem of how it is possible for a person, animal or machine to have meanings to communicate. Thoughts, percepts, memories, suppositions, desires, puzzles, and intentions all have semantic content, and can therefore exist only where there is something that encodes or expresses their content.

Many animals clearly perceive things, want things, try to do things, and learn things, despite not having human language capabilities. Similarly, very young children have intentions, desires, information gained from the environment, and things they want to communicate before they have learnt how to communicate in language (compare Halliday, 1975). They can be very creative, e.g. before having learnt to say "Look here", a child may move an adult's head to face something requiring attention.

Moreover, other animals can be attentive, afraid, puzzled, surprised, or repeatedly trying to do something, all of which involve states with semantic content. A dog that brings a stick to be thrown for it to catch need not have in its head a translation of the English sentence 'Please throw this so that I can catch it', for it may not use the same ontology as we do nor the same mode of composition of meanings, nor the same varieties of speech-act. All we can be sure of is that they must have *some* internal states, processes or structures that express or encode semantic content, and that allow the specific content to have consequences for internal and external behavior, even if the semantic content is not in a propositional form, or expressible in a language like English.

Many scientists now use 'language' in a general sense referring to anything that expresses semantic content, whether for oneself or another agent, especially if it allows both structural variability and compositional semantics, providing the ability to cope with novel information items of varying complexity (Sloman, 2006b). Sloman and Chappell (2007) use "g(generalised)-language" to refer to such forms of representation, including propositional and analogical representations (Sloman, 1971). So the previous paragraph implies that many animals and prelinguistic children use g-languages. G-languages capable of expressing meanings with complex structures must therefore have evolved before communicative languages (Sloman, 1979), for use *within* individual animals, rather than for communication *between* animals. From this viewpoint, the functions of language include perceiving, thinking, wanting, intending, reasoning, planning, learning, deciding, and not just communication. (Unlike Fodor, 1975 we are not claiming that individuals use a fixed innate language into which they translate everything else.)

This leaves open what those inner languages are like. Despite the claims of Brooks (1991), intelligent systems must use representations, at least in the widely-used sense of 'representation' that refers to something that provides, stores, or conveys usable information for some user. The requirements for g-languages are met by the forms of representation used in computational work on high-level vision, reasoning, planning, learning, problem solving, and are also met by external human languages; including both structural variability and compositional semantics, allowing fragments of information to be combined in different ways to form more complex information items that can then be combined with further information. Such representations can be used to express facts, hypotheses, conjectures, predictions, explanations, questions, problems, goals, and plans of varying form and complexity (Sloman, 2006b).

Not all mechanisms and architectures are capable of meeting the requirements:

structural variability, for example, rules out forms of meaning that are expressed only in fixed size vectors with atomic components, such as are often used as inputs and outputs of neural nets. Although no human can actually cope with unbounded complexity, we can argue, echoing Chomsky's (1965) distinction between competence and performance, that humans have virtual machines with unbounded complexity but their implementations in physical machines impose limits. This is also true of most programming formalisms (Scheutz, 2002; Sloman, 2002).

We need more investigation of both the variety of requirements for forms of representation and the variety of possible representations, instead of assuming that known forms will suffice. We also need to stop assuming that human languages and linguistic meanings are *sui generis* and ask whether they are outgrowths of rich forms of syntactic and semantic competence provided by internal g-languages in other animals and in prelinguistic children. This is not to deny that external languages (including pictorial and other forms of communication) allowed rapid acceleration of both learning in individuals and cultural evolution that are unique to humans. In particular, individuals who have learnt to use a human language for external communication are able to enrich the semantic contents expressed internally for their own purposes, e.g. in categorising their thoughts as confused, their desires as selfish, or their knowledge as incomplete. (Cultural learning is discussed further in Section 6.2.)

5.2 Varieties of complexity: 'Scaling up' and 'scaling out'

Another common kind of ontological blindness involves varieties of complexity. Early AI researchers discovered that combinatorial explosions threatened progress. If the solution to a problem involves n actions and for every action there are k options, then there are k^n possible action sequences, a number that grows exponentially with n. Because this quickly makes problems intractable, a common demand is that models should 'scale up', namely continue 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 call 'scaling out'. Vision and language illustrate this: Particular capabilities often depend on and contribute to other capabilities with which they can be combined. We have seen how impoverished theories of vision result from missing the role of vision in satisfying requirements for action and thought. Similarly, work on language that focuses entirely on linguistic phenomena, such as phonemics, morphology, syntax, and semantics, may fail to address such problems as:

- how language is used for non-communicative purposes (e.g., thinking, reasoning, having goals, desires, intentions, and puzzles);
- how it relates to and builds on capabilities that exist in young children or other animals that cannot use language;
- how it relates to forms of representations and mechanisms that evolved prior to human language; and

- how the process of learning a language relates to the evolutionary and developmental precursors of language.

A requirement for a model of how language or vision works, how plans are made and executed, how mathematical or other reasoning works, and how learning works, is that *the proposed mechanisms should be able to form a usefully functioning part of an integrated complete agent combining many other capabilities in different ways at different times.*

That 'scaling out' requirement looks obvious once stated, but its implications for the various components of the system are not obvious, and are often ignored. The kinds of combination required can vary. In simple models, sub-modules are given tasks or other input, and run for a while (as 'black boxes'), then produce results that can be used by other modules. (Like the modules in (Fodor, 1983).) Many proposed architectures assume that sort of structure: they are represented by diagrams with arrows showing unidirectional flow of information between modules. As mentioned in Section 3 some designers assume that there is a *sense-think-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 behavior is produced, and then the cycle repeats, as in the TOTE (Test-Operate-Test-Exit) units of Miller, Galanter, and Pribram (1960), and many more recent designs.

This is clearly wrong. A deeper integration is required: different competences can interact while they are running in parallel and before specific tasks are complete. For humans, many other animals, and robots with complex bodies and multiple sensors acting in a fast changing environment, the *sense-think-act* model fails to account for the variety of extended, concurrent, interacting, processes that are capable of mutual support and mutual modulation. (Compare chapter 6 in Sloman, 1978.)

For instance, while looking for an object, if you hear someone say "Further to the left", what you hear can interact with how you see and help you recognise what you were looking for. Someone looking at the well-known puzzle picture of a dappled dalmation may become able to see the animal on hearing "It's a dog". Likewise, while you are trying to work out what someone means by saying "Put the bigger box on the shelf with 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 the perceived size of the bigger box. Some of these interactions were demonstrated several decades ago (Winograd, 1972). Others are explored in the work of Grice (e.g. Grice, 1975). The interactions need not be produced by first 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 heard sentence even before it is complete.

There are well-documented examples of close interaction between vision and spoken language comprehension, including the 'McGurk effect' (McGurk & MacDonald, 1976) in which the same recorded utterance is heard to include different words when played with videos of speakers making different mouth movements. Interactions can also occur between active and currently suspended processes: 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.

That sort of interaction can even cause the current task to be dropped, with attention switching to a much more important, previously suspended task. 'Anytime' planners, which 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. For example, see this 2004 workshop on incremental parsing: http://homepages.inf.ed.ac.uk/keller/acl04_workshop/ .

Yet another example is combining expert chess competence with knowledge of capabilities of a young opponent to produce chess moves and verbal comments suited to helping the youngster learn. Much teaching requires that sort of mixing of competences: another example of the ability to scale out.

The ability to "scale up" has received far more attention from cognitive modellers, who often try to design mechanisms that are able to cope with increasingly complex inputs without being defeated by a combinatorial explosion. But that is not a requirement for modelling human competence: humans do not scale up!

5.3 Humans scale out, not up

There are many human capabilities that are nowhere near being matched by current machines, yet all of them seem to be complexity-limited, a point related to what Donald Michie (1991) called "the human window". Moreover, there are already many specialised forms of competence where machines far outperform most, or all, humans. Such models scale up, but not out: They have only very narrowly focused competence. Suitably programmed computers can do complex numerical calculations that would defeat all or most humans, but that does not enable those machines to explain what a number is or why it is useful to be able to do arithmetic. Chess programs, like Deep Blue, that use brute force mechanisms, can beat the vast majority of humans, but cannot 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 any artificial chess system 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. Some of the representational and architectural requirements for such states are discussed in Sloman, Chrisley, and Scheutz (2005).

Occurrences of different competences interacting are part of our everyday life, but we may be blind to them when planning our research. Solving the problems of deep integration of cognitive systems with multiple functions may turn out to be much more difficult than anyone anticipates. For example, it is at least conceivable that some powerful forms of information-processing were 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 at this web site http://www.cs.york.ac.uk/nature/gc7/.

6 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, their ability to share goals, 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. It seems that both kinds of passion come from an unscientific commitment, e.g. to religious (or 'romantic'?) reasons for *wanting* to think of humans as unique, or from 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.
The debate is misguided 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 behavior, their needs, their modes of reproduction and development, and how they process information, that are also facts about other animals.

This is a shallow response, however, because there is so much we do not yet know about how humans and other animals work, and what the similarities and differences actually are, and what the implications of those differences are. We still understand relatively little about how most animals work, partly because we do not 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.

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, 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 many 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. Consequently the correct descriptions of future machines will be equally problematic.

6.1 Altricial and precocial skills in animals and robots

Many people are unaware of the great differences between

> (a) the vast majority of species that seem to have their main competences determined genetically, e.g. grazing mammals that can run with the herd shortly after birth, and birds such as chickens that can peck for food soon after hatching, and
>
> (b) the small subset of species that are born helpless, physiologically under-developed, and apparently cognitively incompetent, yet end up as adults 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 former species are labelled 'precocial' species by biologists, and the latter 'altricial'. However, there is a spectrum of cases with different mixtures of precocial skills (genetically determined, preconfigured), and altricial skills ('meta-configured' competences generated by the individual and the environment through play, exploration and learning, using powerful meta-level bootstrapping mechanisms). The nature/nurture trade-offs between different design options are not well understood, although a preliminary analysis was offered in Sloman and Chappell (2005) and refined in Chappell and Sloman (2007) and (Sloman & Chappell, 2007). That analysis suggests that just as there is a spectrum of combinations of preconfigured (precocial) and meta-configured (altricial) skills in biological organisms, so will there also be such a spectrum in future robots, including robots developed as models of human cognition. Robots placed in environments where complex and unpredictable changes can occur over time will, like altricial animals, need to be able to bootstrap meta-configured competences their designers know nothing about, even though they start with a large collection of preconfigured skills, like precocial species. Where most of the environment is predictable in advance, a fully precocial design may function well, but it will not be a model of human, primate, or corvid cognition.

Some altricial species, especially humans, learn both very rapidly and in a wide range of environments, to cope with those environments. As a result, some young children have competences none of their ancestors had. In contrast, skills of precocial species (e.g. deer, chickens) are shaped only in minor ways by the environment in which they live, and altered mainly by slow, laborious training (e.g. circus training), unlike the spontaneous and rapid learning through play, in primates and some other species. At present the mechanisms supporting the latter learning are not well understood, and there are no learning mechanisms or self-constructing architectures under investigation that can account for this, although an idea suggested over 20 years ago by Oliver Selfridge is presented in Sloman and Chappell (2005). [1].

Philipona, O'Regan, and Nadal (2003) present a type of learning-by-doing through finding invariants in sensorimotor patterns. This may explain some ontological

[1] Illustrated in this PDF presentation http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0506

extensions, but does not account for the human-like exosomatic, amodal ontology discussed in Section 4. Another important process may be selection of actions and percepts as 'interesting' (Colton, Bundy, & Walsh, 2000). This requires architectural support for varieties of purely cognitive motivation as opposed to motivation based on physical and reproductive needs. We need to look closely at a variety of phenomena found in the animal world, including recent work on animal tool-making and use (e.g. Chappell & Kacelnik, 2002 and Chappell & Kacelnik, 2004). Related discussions and empirical data can be found in Cummins and Cummins (2005), Csibra and Gergely (2006) and Tomasello, Carpenter, Call, Behne, and Moll (2005). Perhaps future work on altricial robots will enable us to rewrite Piaget's (1954) theories.

6.2 Meta-semantic competence

Another feature important in humans and possibly some other animals is 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 (1994). (The word 'reflective' is sometimes used, but also often has other meanings – one of many examples of confused terminology in the study of architectures.) Closely related ideas have been developed in Minsky (2006), and Singh (2005), focusing mainly on attempts to model human competence. Sloman and Chrisley (2003) relate this to the concept of having qualia.

It seems that humans are not alone in having meta-semantic competence, but the richness of their meta-semantic competence, whether directed inwardly or outwardly is unmatched. 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, based, for example, on cultural learning. Late development does not rule out genetic determination, as should be clear from developments in puberty.

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, and to engage with them in various kinds of shared activities. There are deep problems concerned with referential opacity that need to be solved by such theories. For instance, normal modes of reasoning break down 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. Developmental psychologists study growth of understanding of these matters in children, but do not explain the mechanisms. Perhaps roboticists will one day. Multidisciplinary research is needed to investigate when meta-semantic capabilities evolved, why they evolved, how much they depend on learning as opposed to being preconfigured or meta-configured, how they are influenced by a culture, and what their behavioral consequences are. There are very few discussions of architectural and representational requirements for an organism or machine to represent, refer to, or reason about, semantic contents. Exceptions include McCarthy (1995) and Minsky (2006). Further work is needed for progress on integrated cognitive systems that scale out.

7 Using detailed scenarios to sharpen vision

One way to reduce ontological blindness to some of the functions of natural cognition, is to formulate design goals in terms of *very detailed* scenarios, an idea being taken up in the euCognition network's Research Roadmap project. If scenarios are described in minute detail, e.g. using imaginary 'film-scripts' for future demonstrations of human-like robots, then close attention to individual steps in the scenario can generate questions of the form: 'How could it do that?' that might not be noticed if a competence is described at too general a level. Moreover, we must not focus only on scenarios involving useful 'adult' robots. A three-year-old child who is well able to hold a pencil and make spirals and other things on a sheet of paper may be unable to copy a square drawn on the paper despite being able to trace a square, and to join up dots forming the corners of a square. This could inspire a scenario in which a robot learns to perceive and produce pictures of various sorts on a blank sheet. By trying to design a robot that starts with the abilities and limitations of the three-year-old, and later extends its abilities, we may hope to gain new insights into hidden complexities in the original copying task. (Incidentally, this is one of many examples where the core issues could be studied using a simulated robot: the cognitive development is not dependent on physical embodiment.)

7.1 Sample Competences to be Modelled

As mentioned in Section 4.5 a young child may be able to lift 'cut-out' pictures of various animals (a cat, a cow, an elephant) from a sheet of plywood, but be unable to replace them in their recesses until concepts like 'boundary' and 'alignment' have been added to his or her ontology. We can extend the example by analysing a sequence of intermediate competences, each of which can be achieved without going on to the next step:

- 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,

- being able to see that the picture is not yet in its recess,
- being able to randomly move and rotate the picture until the picture drops into its recess,
- seeing that the explanation of the picture's not going into its recess is that its boundary is not aligned with the boundary of the recess,
- being able to use the perceived mismatch between 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, and
- being able to help a younger child understand how to get the pictures back into their recesses.

This partially 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, algorithms, or architectures, can account for this competence? What needs to be added to the child's ontology at each stage 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? What sort of architecture can combine all these competences and the required forms of representation?

We should not assume that there is some *uniform* learning mechanism that is involved at all stages. Nor should we assume that all required forms of learning are present from the start: some kinds of learning may themselves be learnt. We need to distinguish kinds of meta-competence and ask which are learnt, and how they are learnt. The last example, the ability to help a younger child, has many precursor competences not in the list, that would need to be unpacked as part of a detailed analysis, including meta-semantic competences, such as being able to see and think about another individual as having goals, as perceiving objects, as performing intentional actions, as making mistakes, or as not knowing something.

7.2 Fine-grained Scenarios are Important

The need for 'fine grain' in scenario specifications is not always appreciated. 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 thing 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 behavior 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 behavior 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". If this attention to detail seems tedious, we need to remember that we are attempting to understand results of many millions of years of evolution.

7.3 Behavior specifications are not enough

Merely specifying a form of behavior to be demonstrated does not specify research goals, for, at one extreme, it may be that the behavior is largely pre-programmed by genetic mechanisms in an animal or by explicit programming in a robot (as in precocial species), or, at another extreme, 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). The scenario-based methodology avoids arguments over 'best' target scenarios or 'best' designs, allowing both extremes and also a variety of intermediate cases to be studied, so that we learn the detailed requirements for various combinations of competences, 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. A theory of ontological and representational development crying out for new research in computational models is presented in Karmiloff-Smith (1994). Compare the analysis of learning to count, in Chapter 8 of Sloman (1978). Cognitive robotics researchers should attend to discoveries of psychologists, students of animal behavior, neuroscientists, and clinicians who identify failures of competence arising out of various kinds of brain damage or deterioration. Examples of 'ritual behaviours' providing hints about the architecture are presented in Boyer and Lienard (2006).

8 Resolving fruitless disputes by methodological 'lifting'

Many choices have to be made when designing explanatory models, including 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 can or cannot, should or should not have emotions: See (H. A. Simon, 1967; Sloman & Croucher, 1981; Arbib & Fellous, 2005). Too often the disagreements become pointless squabbles about which design option is right or best. They are pointless if the terms used are ill defined, or if there is no *best* option, only a collection of *trade-offs*, as argued in the online presentation on whether intelligence requires emotions, at this web site: http://www.cs.bham.ac.uk/research/cogaff/talks/#cafe04.

8.1 Analyse before you choose

Instead of continuing these debates, we can shift the questions to a higher level, encouraging former opponents to become collaborators in a deeper project. Instead of debating whether neural or symbolic forms of representations should be used, we can instead explore the space of possible forms of representation, trying to understand the dimensions in which the formalisms differ, while trying to understand what the individual types are and are not good for, what mechanisms they require, and how they differ in relation to a range of meta-requirements, such as speed, accuracy, reliability, extendability and generality. Usually, the answers are not obvious, so if the options and trade-offs can be made clear by research addressing such 'meta-level' questions, then future researchers can choose options wisely on the basis of detailed task requirements, instead of following fashions or prejudice. When we understand the trade-offs fully we shall be in a much better position to do empirical and theoretic research to support various design choices.

An example is 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 influential meta-level paper (McCarthy & Hayes, 1969) produced a first draft list of criteria for adequacy of forms of representation, namely, metaphysical adequacy, epistemological adequacy, and heuristic adequacy (to which, e.g., learnability and evolvability in various environments could be added). That paper's emphasis on logic provoked a charge of narrowness in Sloman (1971), and a rebuttal in Hayes (1984). A recent development of this thread is a PhD thesis on proofs using continuous diagrams Winterstein (2005). Some steps toward a more general overview of the space of possible forms of representation are in Sloman (1993, 1996). However, the analysis of varieties of information processing in biological systems still has a long way to go.

Many discussions of representations and mechanisms fail to take account of requirements for an integrated agent with a complex body embedded in a partially unknown and continuously changing richly structured 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, and planning a response, as well as aesthetic and emotional responses. Work on requirements for complete architectures in systems interacting with a rich dynamic environment has begun to address this complexity, but such work is still in its infancy. Gaps in our knowledge are easily revealed by analysis of requirements for detailed scenarios. For example, requirements for a robot to see its hand grasping and moving a complex object in the proximity of other complex objects include representing 'multi-strand processes', in which different relationships between parts of different objects change concurrently, some continuously (e.g. getting closer) and some discretely (e.g. coming into contact, and changing affordances).

8.2 The need to survey spaces of possibilities

'Meta-level' analysis of a space of possibilities (for forms of representation, for mechanisms, for architectures, etc.) should 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 debates are 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 trade-offs, then how we *label* the options is of lesser consequence. Agreeing on labels may sometimes arise from agreement 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 current state of teaching regarding whether to use symbolic forms of representation, or artificial neural nets and numerical/statistical formalisms and methods causes harm. Learners often simply pick up the prejudices of their teachers and, in some cases, 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: several applicants with MSc or PhD degrees in AI/Robotics had never encountered a symbolic parser, problem solver, or planning system and had apparently never heard of STRIPS or any other planning system. (An excellent introduction is Ghallab, Nau, and Traverso (2004).)

Similarly, although there have been many proposed architectures, some of them surveyed in Langley and Laird (2006), students who learn about a particular sort of architecture may never learn about very different alternatives. A generation of researchers trained with blinkered vision will not achieve the major advances in such a difficult field, even if different subgroups have different blinkers. To summarise:

- Before choosing the best X, try to understand the space of possible Xs
- Often there is no best X, but a collection of trade-offs
- Instead of trying to determine precise boundaries between Xs and non-Xs, it is often more fruitful to investigate varieties of X-like things, the dimensions in which they vary, and the trade-offs: often the X/non-X distinction evaporates and is replaced by a rich taxonomy of cases.

8.3 Towards an ontology for types of architectures

Over the last two decades, there has been a shift of emphasis in research on computational models from investigations of *algorithms* and *representations* for specific tasks, to the study of *architectures* combining many components performing different tasks. Various specific architectures have been proposed, some of them surveyed in Langley and Laird (2006).[2] That survey illustrates how unfortunate definitions can blinker vision, for it *defines* an architecture as something that cannot change, thereby excluding research into whether infants start with a limited architecture extended under the influence of the environment (Chappell & Sloman, 2007; Petters, 2006).

The research community has so far not developed an agreed analysis of requirements for different sorts of architectures nor an adequate ontology for describing and comparing alternatives. 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 will label as *deliberative* any system in which sensory inputs can activate rival responses, one of which is selected by a competitive process; whereas Sloman (2006b) calls that *proto-deliberative*, following AI tradition in reserving the label *deliberative* for mechanisms that search for and manipulate representations of variable structure and complexity, using compositional semantics. A richer meta-level ontology for types of architectures would allow a variety of intermediate cases. Some researchers use the label 'reactive' to exclude internal state change, whereas others allow reactive systems to learn and have changing goals, as long as they lack deliberative mechanisms for constructing and comparing hypothetical alternatives. As indicated in Section 6.2, the word "reflective" is also used with different meanings when describing architectures or components of architectures. Papers in the Cognition and Affect project http://www.cs.bham.ac.uk/research/cogaff/ present the *CogAff schema* as a first draft attempt to provide a more principled ontology for possible architectures, which will need to be related to *niche space*, the space of possible sets of requirements.

Researchers wishing to move beyond the present terminological mess can assume that biological evolution produced many intermediate cases not yet understood, some occurring during early stages of human infant and child development, though observing processes in virtual machines that bootstrap themselves is a task fraught with difficulties (Sloman & Chappell, 2005). We need to understand intermediate cases that occurred in nature if we are to match designs for working models to the variety produced by evolution, whether for scientific or for practical purposes. A better ontology for architectures may also help us develop better tools to support cognitive modelling (cf. Ritter, 2002; Kramer & Scheutz, 2007).

[2]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.

9 Assessing scientific progress

A psychologist once commented that whenever he heard researchers giving seminars on computational models, they talked about what they were going to do, and occasionally what they had done, but rarely about what they had *discovered*. Can the cognitive modelling research community map out intended *advances in knowledge* – as opposed to merely forming plans for *doing things*, however worthwhile? A partial answer was given in Sections 7 and 8: There is scientific work to be done producing systematic meta-level theories about varieties of forms of representation, mechanisms, architectures, functions, and requirements, that define the spaces from which we can choose components of designs and explanatory theories. That can provide a framework for further work on 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. For scientific purposes, merely building systems that work is of limited value, if we do not understand how they work and why they are better or worse than other possible designs, etc., or better in some contexts and worse in others.

Much funded applied research is defined in terms of specific practical goals, for example 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), helping with domestic chores (www.ai.rug.nl/robocupathome) or identifying potential terrorists at airports. In addition to identifying specific, somewhat arbitrary, target systems, however interesting and important, we should attempt to identify a structured set of scientific goals that advance our *knowledge and understanding*, as opposed to merely advancing our *practical 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 will probably never have the clarity and rigour of mathematics at the start of the 20th Century, because cognitive science encompasses the study of all forms of cognition, including future products of evolution and human-machine integration. But we can attempt to identify important questions that need to be answered.

9.1 Organising questions

Just as mathematicians showed that answering some questions will enable others to be answered, or at least simplified, so should cognitive modellers try to identify relations between unsolved problems. For example, if we can describe in detail 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 doing only research with a narrow focus, such as language learning, 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 may help us overcome the difficulty of identifying what needs to be explained, referred to as "ontological blindness" in Section 3. 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 learned.

So instead of trying only to propose specific scientific goals, over which there is likely to be strong disagreement regarding priorities, perhaps researchers can 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 described in Section 7.

9.2 Scenario-based backward chaining research

Suppose we describe *in great detail* a variety of scenarios involving various kinds of human-like or animal-like behavior whose achievement is far beyond the current state of the art. The dishwasher-loading, and picture-puzzle scenarios in Section 7 are examples, but we could produce hundreds more, relating to everyday competences of humans of different ages and sorts as well as other animals. If we then analyse requirements for producing the detailed behaviors, 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.

9.3 Assessing (measuring?) progress

By careful analysis of long-term and intermediate goals, and working backward 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. That will define milestones for measuring 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 should lead to tests that can show which hypothesis is better, or whether each is better only for a subset of cases. Sometimes one hypothesis will eventually turn out to be better at defining a long-term "progressive" research program in the sense of Lakatos (1980).

We can also work forward 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 short-term choices after we have sketched at least some of the terrain a long way ahead: otherwise more attractive short term goals will be selected.[3]

9.4 Replacing rivalry with collaboration

We can 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, and the possible consequences of solving them; and
- the *prioritising, justification,* or *selection* of research problems: deciding what is important and should be funded.

People can collaborate and reach agreement on the former while disagreeing about the latter. The process of collaborating on the first 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 is rumoured to have silenced an objector by saying 'Truth is more important than importance'.)

Understanding the science better will enable us to discuss the benefits of different ways of allocating scarce research resources. Work on clarifying and analysing a problem can contribute to a decision to postpone research on the problem, by revealing a hard prior problem, or by clarifying the relative costs and benefits of different options. Meta-level theoretical work revealing good routes to intermediate goals can be a significant contribution to knowledge about a hard problem, especially analysis of which mechanisms, formalisms, architectures, or knowledge systems, will or will not be sufficient to support particular types of scenarios (compare the role of complexity theory in software engineering.)

By making construction, analysis and ordering of possible scenarios, along with analysis of corresponding design options and trade-offs, an explicit community-wide task (like the Human Genome project), we separate the task of identifying 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

[3]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, and was later refined in the context of the EU-funded CoSy robotic project. There is now a Research Roadmap project in the EU. See
http://www.cs.bham.ac.uk/research/cogaff/gc/targets.html,
http://www.eucognition.org/wiki/index.php?title=Research_Roadmap

done competitively. This can also reduce the tendency for research groups or sub-communities to specify their own evaluation criteria independently of what others are doing, a symptom of an immature and fragmented science. This can also provide a means of evaluating research *proposals*. Computational modelling researchers often propose to do what previous researchers had proposed to do, but failed to do, provoking the question: Why should the new proposals be taken seriously? New proposals are too often 'forward chaining' proposals regarding how known techniques, formalisms, and architectures will be used to solve hard problems: a well-tried recipe for failure. Perhaps, if more research is selected on the basis of detailed 'backward-chaining' analysis of long-term task requirements for integrated systems, a major change in the fortunes of research projects will follow.

10 Conclusion

Previous chapters have mainly focused on achievements. This one has reviewed some gaps that still need to be filled, outlining some ways of accelerating progress toward the development of models that are more human-like, using deeper and more comprehensive theories of human and animal cognitive competences and their development. There are many gaps and much work still to be done. For instance, most of what can be done by one- to two-year old toddlers is far beyond anything we can now model. We also cannot yet model finding something *funny* or *aesthetically pleasing*, neither of which is a matter of producing any behaviour.

Perhaps this partial overview will help provoke researchers to address new problems, such as how 'scaling out' happens, and new ways of thinking about the long-term challenge of integrating multiple competences. Perhaps 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. This will require overcoming serious institutional impediments to such developments. It may also require the invention of new forms of computation.[4]

Acknowledgements

Some of this work was inspired by participation in the EU-Funded project *CoSy: Cognitive Systems for Cognitive Assistants*, described at http://www.cognitivesystems.org, and owes much to discussions with members of the CoSy project and also interactions with, or publications of, Margaret Boden, Jackie Chappell, Ron Chrisley, John McCarthy, Marvin Minsky, Matthias Scheutz, and the late Push Singh. Ron Sun

[4]The references that follow should not be treated as a comprehensive bibliography. They are merely samples of what is available. Increasingly, the important up-to-date literature is not in printed documents but in web pages, including draft books and papers, online reviews, and the like, so several web sites have been listed as sources of further information. There are many more. The author accepts responsibility for refusal to provide retrieval dates for urls.

and anonymous referees made useful editorial suggestions. Peggy Rote very kindly took great trouble to provide help at the copy-editing stage. The copy-editor cannot be blamed for the author's intransigence in rejecting most proposed changes of wording or punctuation.[5]

References

Anderson, M., Meyer, B., & Olivier, P. (Eds.). (2001). *Diagrammatic representation and reasoning.* Berlin: Springer-Verlag.

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

Barrow, H., & Tenenbaum, J. (1978). Recovering intrinsic scene characteristics from images. In A. Hanson & E. Riseman (Eds.), *Computer vision systems.* New York: Academic Press.

Beaudoin, L. (1994). *Goal processing in autonomous agents.* Unpublished doctoral dissertation, School of Computer Science, The University of Birmingham, Birmingham, UK. (http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#38)

Berthoz, A. (2000). *The brain's sense of movement.* London, UK: Harvard University Press.

Boden, M. (2006). *Mind As Machine: A history of Cognitive Science (Vols 1–2).* Oxford: Oxford University Pres.

Boden, M. A. (1990). *The creative mind: Myths and mechanisms.* London: Weidenfeld & Nicolson.

Boyer, P., & Lienard, P. (2006). Why ritualized behavior? Precaution Systems and action parsing in developmental, pathological and cultural rituals. *Behavioral and Brain Sciences*, *29*(6), 595–650. (http://www.bbsonline.org/Preprints/Boyer-04042005/)

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.

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

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

Chappell, J., & Sloman, A. (2007). Natural and artificial meta-configured altricial information-processing systems. *International Journal of Unconventional Computing*, *3*(3), 211–239. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609)

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press,.

Colton, S., Bundy, A., & Walsh, T. (2000). On the notion of interestingness in

[5] http://www.cs.bham.ac.uk/~axs/publishing.html expounds my views on most copy-editors I have encountered.

automated mathematical discovery. *International Journal of Human-Computer Studies*, *53*(3), 351–375.
Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In M. H. Johnson & Y. Munakata (Eds.), *Processes of Change in Brain and Cognitive Development. Attention and Performance XXI* (pp. 249–274). Oxford: Oxford University Press. (http://www.cbcd.bbk.ac.uk/people/gergo/a&p_pedagogy.pdf)
Cummins, D., & Cummins, R. (2005). Innate modules vs innate learning biases. *Cognitive Processing: International Quarterly Journal of Cognitive Science*, *3*(3-4), 19–30.
Fodor, J. (1975). *The Language of Thought*. Cambridge: Harvard University Press.
Fodor, J. (1983). *The modularity of mind*. Cambridge MA: MIT Press.
Ghallab, M., Nau, D., & Traverso, P. (2004). *Automated Planning, Theory and Practice*. San Francisco, CA: Elsevier, Morgan Kaufmann Publishers.
Gibson, J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates. ((originally published in 1979))
Goodale, M., & Milner, A. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25.
Grice, H. (1975). Logic and Conversation. In D. Davidson & G. Harman (Eds.), *The Logic of Grammar* (pp. 64–75). Encino, CA:: Dickenson.
Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, *27*, 377–442.
Halliday, M. (1975). *Learning How to Mean: Explorations in the development of language*. London: Edward Arnold.
Hayes, P. (1984). Some problems and non-problems in representation theory. In R. Brachman & H. Levesque (Eds.), *Readings in knowledge representation* (pp. 3–22). Los Altos, California: Morgan Kaufmann.
Hayworth, K. J., & Biederman, I. (2006). Neural evidence for intermediate representations in object recognition. *Vision Research (In Press)*, *46*(23), 4024–4031. (doi:10.1016/j.visres.2006.07.015)
Hendler, J. (2005, March/April). A Letter from the Editor: Fathoming Funding. *IEEE Intelligent Systems*, *20*(2), 2–3. (http://ieeexplore.ieee.org/iel5/9670/30619/01413163.pdf?arnumber=1413163)
Jablonka, E., & Lamb, M. J. (2005). *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge MA: MIT Press.
Karmiloff-Smith, A. (1994). Precis of: Beyond Modularity: a Developmental Perspective on Cognitive Science. *Behavioral and Brain Sciences*, *17*(4), 693-706. (http://www.bbsonline.org/documents/a/00/00/05/33/index.html)
Kramer, J., & Scheutz, M. (2007). Development environments for autonomous mobile robots: A survey. *Autonomous Robots*, *22*(2), 101–132. (DOI 10.1007/s10514-006-9013-8)
Lakatos, I. (1980). The methodology of scientific research programmes. In J. Worrall & G. Currie (Eds.), *Philosophical papers, Vol I*. Cambridge University Press.

Langley, P., & Laird, J. (2006). *Cognitive architectures: Research issues and challenges* (Tech. Rep.). Palo Alto, CA.: Institute for the Study of Learning and Expertise. (http://cll.stanford.edu/ langley/papers/final.arch.pdf)

Lungarella, M., & Sporns, O. (2006). Mapping information flow in sensorimotor networks. *PLoS Computational Biolology*, *2*(10:e144). (DOI: 10.1371/journal.pcbi.0020144)

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

McCarthy, J. (1995). Making robots conscious of their mental states. In *Aaai spring symposium on representing mental states and mechanisms*. Palo Alto, CA: AAAI. (http://www-formal.stanford.edu/jmc/consciousness.html)

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

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

McGurk, H., & 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.

Miller, G., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behaviour*. New York: Holt.

Minsky, M. L. (1963). Steps towards artificial intelligence. In E. Feigenbaum & J. Feldman (Eds.), *Computers and thought* (pp. 406–450). New York: McGraw-Hill.

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

Minsky, M. L. (2006). *The Emotion Machine*. New York: Pantheon.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Petters, D. (2006). *Designing Agents to Understand Infants*. Unpublished doctoral dissertation, School of Computer Science, University of Birmingham, UK, Birmingham, UK. (http://www.cs.bham.ac.uk/research/projects/cogaff/06.html#0605)

Philipona, D., O'Regan, J. K., & 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*. New York: Ballantine Books. (Last chapter online http://www.marxists.org/reference/subject/philosophy/works/fr/piaget2.htm)

Ritter, F. (Ed.). (2002). *Techniques for Modeling Human Performance in Synthetic Environments: A Supplementary Review*. Ft. Belvoir, VA: Defense Technical Information Center. (http://ritter.ist.psu.edu/papers/SOAR-Jun03.pdf)

Saxena, A., Driemeyer, J., Kearns, J., Osondu, C., & Ng, A. Y. (2006). Learning to

Grasp Novel Objects using Vision. In *10th International Symposium of Experimental Robotics, ISER 2006.*

Scheutz, M. (Ed.). (2002). *Computationalism: New Directions.* Cambridge, MA: MIT Press.

Simmons, R. G., Goldberg, D., Goode, A., Montemerlo, M., Roy, N., et al. (2003). Grace: An autonomous robot for the aaai robot challenge. *AI Magazine*, *24*(2), 51–72.

Simon, H. (1997). *Allen Newell (1927-1992) (Biographical Memoir).* National Academy of Sciences. (http://stills.nap.edu/readingroom/books/biomems/anewell.html)

Simon, H. A. (1967). Motivational and emotional controls of cognition. In H. A. Simon (Ed.), *Models of thought* (pp. 29–38). Newhaven, CT: Yale University Press.

Singh, P. (2005). *EM-ONE: An Architecture for Reflective Commonsense Thinking.* Unpublished doctoral dissertation, MIT, Cambridge, MA. (http://web.media.mit.edu/˜push/push-thesis.html)

Sloman, A. (1971). Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence. In *Proc 2nd ijcai* (pp. 209–226). London: William Kaufmann. (http://www.cs.bham.ac.uk/research/cogaff/04.html#200407)

Sloman, A. (1978). *The computer revolution in philosophy.* Hassocks, Sussex: Harvester Press (and Humanities Press). (http://www.cs.bham.ac.uk/research/cogaff/crp)

Sloman, A. (1979). The primacy of non-communicative language. In M. MacCafferty & K. Gray (Eds.), *The analysis of Meaning: Informatics 5 Proceedings ASLIB/BCS Conference, Oxford, March 1979* (pp. 1–15). London: Aslib. (http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#43)

Sloman, A. (1982). Image interpretation: The way ahead? In O. Braddick & A. Sleigh. (Eds.), *Physical and Biological Processing of Images (Proceedings of an international symposium organised by The Rank Prize Funds, London, 1982.)* (pp. 380–401). Berlin: Springer-Verlag. (http://www.cs.bham.ac.uk/research/projects/cogaff/06.html#0604)

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 (Ed.), *Forms of representation: an interdisciplinary theme for cognitive science* (pp. 118–140). Exeter, U.K.: Intellect Books.

Sloman, A. (2002). The irrelevance of Turing machines to AI. In M. Scheutz (Ed.), *Computationalism: New Directions* (pp. 87–127). Cambridge, MA: MIT Press. (http://www.cs.bham.ac.uk/research/cogaff/00-02.html#77)

Sloman, A. (2006a). Polyflaps as a domain for perceiving, acting and learning in a 3-D world. In *Online Position Papers for 2006 AAAI Fellows Symposium.* Menlo Park, CA: AAAI. (http://www.aaai.org/Fellows/fellows.php and

http://www.aaai.org/Fellows/Papers/Fellows16.pdf)
Sloman, A. (2006b, May). *Requirements for a Fully Deliberative Architecture (Or component of an architecture)* (Research Note No. COSY-DP-0604). School of Computer Science, University of Birmingham. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0604)
Sloman, A., & Chappell, J. (2005). The Altricial-Precocial Spectrum for Robots. In *Proceedings IJCAI'05* (pp. 1187–1192). Edinburgh: IJCAI. (http://www.cs.bham.ac.uk/research/cogaff/05.html#200502)
Sloman, A., & Chappell, J. (2007). Computational Cognitive Epigenetics (Commentary on (Jablonka & Lamb, 2005)). *Behavioral and Brain Sciences*, *30*(4). (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0703)
Sloman, A., Chappell, J., & CoSyTeam, T. (2006, June). How an animal or robot with 3-D manipulation skills experiences the world. In *The tenth annual meeting of the Association for the Scientific Study of Consciousness, Oxford.* ASSC. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0602, Poster for ASSC10, Oxford June 2006. Also at ASSC10 Eprints Archive: http://eprints.assc.caltech.edu/112/)
Sloman, A., & Chrisley, R. (2003). Virtual machines and consciousness. *Journal of Consciousness Studies*, *10*(4-5), 113–172.
Sloman, A., Chrisley, R., & Scheutz, M. (2005). The architectural basis of affective states and processes. In M. Arbib & J.-M. Fellous (Eds.), *Who Needs Emotions?: The Brain Meets the Robot* (pp. 203–244). New York: Oxford University Press. (http://www.cs.bham.ac.uk/research/cogaff/03.html#200305)
Sloman, A., & Chrisley, R. L. (2005, June). More things than are dreamt of in your biology: Information-processing in biologically-inspired robots. *Cognitive Systems Research*, *6*(2), 145–174.
Sloman, A., & Croucher, M. (1981). Why robots will have emotions. In *Proc 7th int. joint conference on AI* (pp. 197–202). Vancouver: IJCAI.
Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–735.
Trehub, A. (1991). *The Cognitive Brain.* Cambridge, MA: MIT Press. (http://www.people.umass.edu/trehub/)
Turing, A. (1950). Computing machinery and intelligence. *Mind*, *59*, 433–460. ((reprinted in E.A. Feigenbaum and J. Feldman (eds) *Computers and Thought* McGraw-Hill, New York, 1963, 11–35))
Várady, T., & Martin, J., R.R. andCox. (1997). Reverse Engineering of Geometric Models - An Introduction. *Computer Aided Design*, *29*(4), 255–268. (http://ralph.cs.cf.ac.uk/papers/Geometry/RE.pdf)
Winograd, T. (1972). Procedures as a Representation for Data in a Computer Program for Understanding Natural Language. *Cognitive Psychology*, *3*(1).
Winterstein, D. (2005). *Using Diagrammatic Reasoning for Theorem Proving in a Continuous Domain.* Unpublished doctoral dissertation, University of Edinburgh, School of Informatics.

(http://www.era.lib.ed.ac.uk/handle/1842/642)