

Cross-Disciplinary Reflections: Philosophical Robotics

Aaron Sloman

University of Birmingham {axs@cs.bham.ac.uk}

12.1 Introduction

This chapter reports work done mostly by one member of the team – a philosopher with substantial AI programming experience, whose primary interests were in the very long term goals of the project, summarised in Chapter 1, including the goal of shedding light on problems solved by biological evolution, and who was not directly involved in the coding but who interacted closely with people who were, and with people outside the project, in several related disciplines. The majority of the work reported here is concerned with requirements, and gaps between those requirements and the current state-of-the-art in AI/Robotics, and related disciplines. A key feature of this work is its emphasis on study of aspects of the 3-D environment we and other animals inhabit, with which a Fido-like intelligent domestic robot (described in Chapter 1) would need to interact. This is an essential part of a strategy for developing a roadmap to bridge the gaps in the long term.

From the start, the CoSy project emphasised the need to study *requirements*, as a prerequisite for producing designs. This analysis supported the suspicions in the original proposal, namely that current state-of-the-art designs and implementations were nowhere near meeting the long term requirements. So our goal became to develop a methodology for identifying the gap more precisely, and a strategy for bridging the gap (or gaps). As an aid to these tasks, we organised several interdisciplinary events, including a tutorial at IJCAI'05, a Symposium at AISB'06 and the “Meeting of minds” workshop in Paris, in 2007. Most of this work did not fit into the formal work-plans and deliverables, but results were published in workshop and conference papers,

contributions to collections, the euCognition wiki, and online presentations and discussion papers,¹ along with web sites for the events we organised.²

Some of the results of that study are presented below, suggesting directions for future work and implications for other disciplines, including study of humans and other animals. The project's goal of producing and demonstrating a sequence of implemented, integrated systems starting in year 1 (see Section 11.1) required many detailed decisions to be taken before long term requirements could make much progress. As a result, the long-term requirements analysis proceeded in parallel with most of the design and implementation work, and only loose connections were possible. Had the project been funded for 10 or 15 years, without the need to produce publications and demonstrations every year from the start, things might have been different.

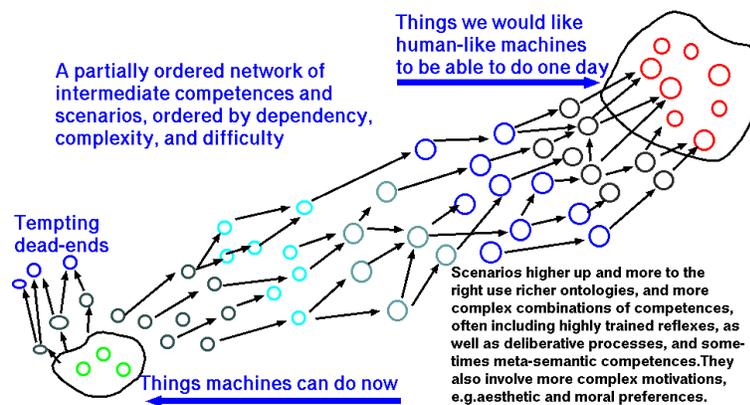


Fig. 12.1. How to develop a long term research roadmap based on a partially ordered network of scenarios developed by backward-chaining (presented at euCognition roadmap meeting Jan 2007).

As work on the two streams (requirements analysis and implementation) progressed, it became clear that the aforementioned gap was even greater than we had anticipated. This created a tension between the “safe” approach of taking existing techniques and attempting to combine them, where possible with additions, and the “risky” approach of trying to find ways to reduce some of the huge gaps between current techniques and animal/human competences.

¹ See [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] [32]

Length restrictions allow only a subset of the ideas to be presented here.

² <http://www.cs.bham.ac.uk/research/projects/cosy/conferences>
<http://www.cs.bham.ac.uk/research/projects/cogaff/gc/aisb06>
<http://www.cs.bham.ac.uk/research/projects/cosy/conferences/mofm-paris-07>

Attacking the gaps head-on would have required most of the four years to be spent on the study of long term requirements, and attempting to decompose those requirements in a backward chaining process of the sort described in [4] and depicted in Figure 9.1. Instead, a fairly small subset of the work in the whole project focused on long term requirements: most of the design and implementation work was an attempt to extend the current state of the art, as reported in preceding chapters, especially the state of the art in integrating different kinds of functionality.

The latter work did include some requirements analysis, e.g. for architectures and tools for integration of subsystems (Chapter 11). The specific short-term demonstrator goals were grown, modified and constrained through experience of trying to put pieces together by extending techniques originally developed for modules running in isolation. As a result of our growing appreciation of the gaps between the state of the art in AI and robotics and the functionality required for the futuristic ‘Fido’ scenario, most of the effort in the project went into feasible extensions of component capabilities, along with integration into a system that combined the components. Some of what that left out is presented in the rest of this chapter. (Choosing scenarios to work on raises many practical problems, some of which are discussed in Chapter 11.) Many of the detailed requirements became visible as a result of reflecting on what our implementations could not do, illustrating the importance of implementation-based requirements analysis! For this purpose rapid-prototyping tools without a prior commitment to any particular architecture are essential, as discussed in [33, 34] and Chapter 11.

12.2 Must an intelligent robot use language?

Since preverbal children and many animals that do not use language can interact with complex environments, including environments in which processes are occurring, e.g. during nest-building, fighting, manipulating sources of food or shelter, eating things that do not come ready carved up into bite-size chunks, etc., that shows that the ability to use a human language is not a prerequisite for such competences. However, it can be argued (as in [15]) that both pre-verbal humans and other intelligent animals must use forms of representation *internally* that support structural variability, context-sensitive compositional semantics, as well as the ability to introduce substantive extensions to the ontology. They must also be suitable for use in perception, planning, various kinds of learning, the expression of motives and preferences, and the generation and control of action.

Such features, especially structural variability and compositional semantics, are normally thought of as key features of languages used for *communication*, but they must have existed earlier in *internal* “languages”. What forms those representation used, and still use, is an open question. There are no obviously correct candidates, though they are unlikely to have the grammat-

ical or logical structures of languages that evolved for communication rather than for representing percepts and controlling actions. Neither do we claim that the internal languages are fixed at birth: they may be extended by bootstrapping and debugging processes, including substantive (non-definitional) ontology extension [17]. So our arguments are not endorsements of Fodor’s theory in [35].

Since some researchers object to the use of the word “language” to label something that is not used for communication, we call this notion of language that covers both forms of representation used for communication and forms used internally for perceiving, thinking, etc. ‘Generalised Language’ (GL), discussed in more detail in [36] [15] [27].

The argument that GLs used internally precede the use of verbal languages for communication both in evolution and in individual human development has many implications, including implications concerning requirements for future intelligent robots.

Looking at videos of pre-verbal infants and toddlers helped to draw attention both to gaps in their understanding of various aspects of the environment and to the depth and variety of their visual and manipulative competences, despite those gaps. Before a human child starts learning to talk there is already a deep understanding of, and interest in, many structures and processes in the environment [38], and those competences and interests are required for the language learning process. Learning a language is part of learning how to achieve collaborative goals in a shared, partially understood, 3-D environment, using an exosomatic ontology (defined below in Section 9.5.1), not just learning mappings between acoustic signals (spoken words) and other sensory signals. Because our robots were nowhere near human toddler competences in vision and manipulation, and did not have the rich internal information-processing formalisms (“internal languages”) postulated above (and in [36, 15]), their language learning processes described in previous chapters had to be totally different from human language learning, and much more artificial and restricted. It is to be hoped that this deficit can be remedied in future research.

12.3 The role of the environment

Thirty years ago, [39] suggested that work in AI could clarify or solve many philosophical problems, but our analysis of requirements in the CoSy project revealed further implications of the fact that important aspects of human and animal intelligence were evolutionary responses to the challenge of interacting with and manipulating movable, reconfigurable, 3-D objects of varying complexity in an extended environment, only part of which is perceivable at any time. A core feature of the challenge is perception of concurrently changing 3-D spatial (geometric and topological), causal and functional relationships between both whole objects and parts of objects, where some of the changes involve independently movable limbs and hands. Some of the requirements

are discussed in [30]. Most of the requirements deriving from that challenge appear not to have been noticed by roboticists, vision researchers or psychologists, and philosophers have not realised their significance for philosophy of science and philosophy of mathematics (see [21, 5]).

A very early robot with manipulative capabilities was the Edinburgh robot Freddy_II [40] developed around 1973, which could assemble two different objects (a toy car and a toy boat) from parts initially piled or scattered randomly on a table.³ Its speed and versatility were severely limited⁴ yet recent robots have not reached some of Freddy’s competences, even though many hardware and software components required have separately developed enormously. Why not?

Processes of learning and development in humans and some other animals depend on rich interactions with the environment in early months and years that lay a foundation on which many other aspects of human intelligence depend in later life. Many biologists study animal behaviours, and many developmental psychologists study spatial competences in infants and young children (e.g. the book by Gibson and Pick [41]). However, the methods of experiment accepted as producing significant results are so restrictive that most of the cognitive richness of processes required for interaction with manipulable objects, described in [20], and below, goes unnoticed. Likewise, limitations of current tools and techniques cause AI researchers to ignore most of the complexity in the environment, as do many researchers working on digital companions for the elderly or disabled [14]. Examples of what they ignore (e.g. differences between tracking moving 2-D image features and perceiving a 3-D process) are given later in Section 9.12.2.

Without a rich and deep, mostly culture-neutral, biologically rooted, understanding of space, time, and the 3-D physical environment in which animal activities are embedded, a robot is unlikely to be able to learn to talk, think, and perform tasks like a human adult – though it may cope with a very restricted subset in a very brittle way, as in many robot demos.

12.4 Analysing requirements is very hard

Analysis of requirements is a task whose complexity is largely unnoticed. Many researchers think it is sufficient to define some goal in terms a very high level description e.g. “recognition of everyday objects in everyday situations”, “coping flexibly with domestic tasks”, “engaging in natural conversation about some topic”. Sometimes, words like “reliability”, “flexibility”, “robustness”,

³ While this chapter was in preparation, a remarkable video of Freddy_II was made available here http://en.wikipedia.org/wiki/Freddy_II

⁴ In 1973, 384KBytes of memory was a luxury, computer speeds were measured in kilocycles, and it could take several minutes just to find the bounding contours in an image, ruling out concurrent perception and action, or visual servoing.

“intelligence”, “autonomy”, “versatility”, “extendability”, and “maintainability” are used to indicate design features, even though such words refer to very different characteristics in different contexts. For example, the requirements for *robustness* are very different in an operating system, a word processor, a theorem prover, a medical expert system, and a lawnmower. (Discussed in more detail in [42].)

Sometimes benchmark tasks (e.g. fixed sets of images for training and testing) are used to replace imprecise requirements. But the benchmarks often lack “ecological validity”, directing research down narrow paths and diverting attention from the problems the benchmarks were intended to characterise.

Close examination of problems with which various environments confront humans and other animals reveals richness and diversity of problems and solutions that usually escape notice. The diversity of possible solutions indicates a need to study trade-offs between alternative competences and alternative designs, rather than performance metrics and bench-marks (see Section 9.11).

By collecting partially ordered sets of scenarios (ordered by both difficulty and dependency), we can identify short, medium and long term challenges to be met in specifying designs, as indicated in Figure 9.1. Some of the scenarios should include multiple interacting competences with the interactions described in film script detail. Systematically varying features of the scenarios, can demonstrate the inadequacy of designs tailored to *limited* sets of examples.

This task of generating scenarios is very close to the process of conceptual analysis in philosophy, where theories (e.g. about the nature of desire, intention, attention, perception, belief, understanding, etc.) need to be tested by production of examples. In principle, this is an area where philosophers and AI researchers should be able to interact, partly because good philosophers have already developed the ability to think up examples to challenge theories.

We developed templates for scenarios,⁵ and a scenario-generation methodology based on a 2-D grid of types of competence against types of object, with complexity as a third dimension, summarised in [2]. Unfortunately the task of generating and analysing demanding scenarios proved difficult for researchers who had never previously done anything like it, especially while facing great challenges in their own sub-fields. Eventually, scenarios limited by short-term feasibility were developed, described in Chapters 9, 10 and 11.

12.5 Robotics and Philosophy of Science

12.5.1 Ontologies and Laws

A project like CoSy illustrates limitations of conventional philosophy of science. Chapter 2 of [39]⁶ explained how science is an attempt to understand

⁵ E.g. in <http://www.cs.bham.ac.uk/research/projects/cosy/scenarios>

⁶ Online at <http://www.cs.bham.ac.uk/research/projects/cogaff/crp/chap2.html>

both the form and the content of the world, where the form has two aspects: (a) what sorts of things are possible (an ontology), and (b) how those possibilities are limited, e.g. in *laws* such as “All *As* are *Bs*”, which rules out the possibility of something being an *A* and not a *B*. But that presupposes an ontology that includes the possibility of *A* things and *B* things.

Conventional philosophy of science emphasises (b), not (a), whereas the deepest scientific advances are of type (a), substantially extending our ontologies, and thereby allowing new questions and theories to be formulated, e.g. adding atomic theory, or evolution by natural selection to our ontology. Such advances require new concepts, extending the pre-existing ontology substantively, i.e. adding new concepts that cannot be defined in terms of old ones, though they may be connected via what Carnap called “meaning postulates” in [43]. Substantive kinds of ontology extension must occur in children [15].

Deep research in cognition and robotics requires researchers to extend their scientific ontologies if they are to produce more intelligent machines. Following [30] these can be labelled “designer ontologies”, in contrast with the ontologies required by robots (or animal being modelled!) “application ontologies”. Intelligent robots will also need mechanisms capable of substantive (non-definitional) ontology extension. Mechanisms for extension of sensorimotor ontologies by dimensionality reduction, presented in Chapter 3, may be useful, but cannot add new dimensions, e.g. required for interpreting 2-D motions as projections of 3-D rotations. Similarly, mechanisms using sensorimotor statistics to induce new concepts useful for prediction, will not be enough: contrary to “Symbol-grounding” theory, which is a serious impediment to progress, as explained in [17]. In particular, what a child develops and what a robot will need includes “exosomatic” concepts referring not to sensorimotor patterns, but to objects and processes in the environment that could exist independently of the observer.

12.5.2 No “right” or “best” designs

Philosophers often try to specify *necessary* conditions for something to be a mind. This leads to shallow and unsatisfactory theories. Attempting to identify *one* best design for intelligent systems would be like physicists attempting to study only the substances that exist in some particular spatio-temporal region, e.g. Rome in 1630. Instead, there is a broad class of *possible* active information-processing systems, a class that includes myriad varieties of organism produced on this planet by biological evolution, and probably even more produced in other parts of the universe, or in the future on earth, and some that are possible, but never will be produced.

Different designs can be evaluated in relation to different sets of requirements. Sets of requirements for organisms are referred to as “niches”. So the task is to understand the space of sets of designs, the space of sets of requirements and the relationships between them (Figure 9.2). Since individuals can

develop and since species can evolve, there are also trajectories within those spaces and we need to understand how those trajectories work.

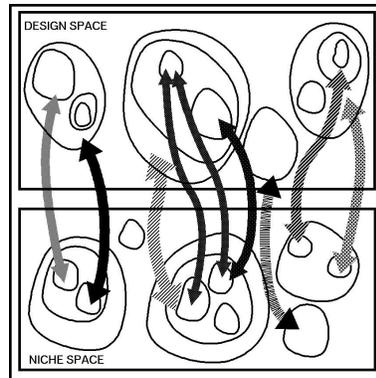


Fig. 12.2. *The space of possible designs, the space of possible niches (sets of requirements) and the varied relationships and trade-offs between them. Both spaces have many discontinuities. There are trajectories of various sorts through both spaces, including development and learning done by individuals, evolution of species, social/cultural evolution, and in the case of artificial systems design changes. In an ecosystem there will be complex feedback loops involving trajectories in both design space and niche space.*

This broad-minded approach to both philosophy and AI, presented in [44, 45, 46], is hard for most researchers, especially as progress is inevitably slow. Studying the full spaces in depth is impossible, but we can explore limited regions (“neighbourhoods”) in design space and niche space ([47]).

Detailed specification of such regions cannot be done in a research proposal: it is the *result* of research. As the work on integration in CoSy progressed, decisions had to be taken about what was and what was not being addressed, such as whether the robot should be able to perceive *processes* (like its hand moving), whether to include recognition of objects or perception of 3-D structure, since they required very different mechanisms, whether it should interact sensibly with more than one human at a time, which kinds of failure in performing tasks it should be able to detect and remedy, which aspects of verbal interaction should be capable of influencing visual processing, or vice versa. Some choices between project sub-goals, and the design problems they led to, were not visible to participants before the project started. Such “invisibility” is reduced as more researchers gain experience in integrated projects.

12.5.3 A science of explosive diversity

The problem of description arises partly because individual designs for “complete” working systems can vary enormously: in their architectures, e.g. in the

variety of components they contain; the forms of representation they use; the kinds of information they acquire, manipulate and use; the variety of connections between sub-systems; and whether the architecture is static or grows itself, as certainly happens in humans, but not yet in CoSy. This diversity makes it very hard to compare designs, especially when represented in complex diagrams using arbitrary diagrammatic notations. We need a better way to talk about such complexity.

Much of the variation between designs is closely related to the different challenges posed by the kinds of environment that different systems need to interact with, and a detailed study of interesting designs must be linked to a detailed study of the relevant environments, a point emphasised also by Neisser [48] and Gibson [49] (discussed further below). [12] illustrates some aspects of the complex feedback between evolution of designs and evolution of niches, from microbes in chemical soups to articulated animals surrounded by diverse rigid and non-rigid 3-D structures, and also other intelligent systems.

12.5.4 Individual variability

Not only the diversity of designs, but also the diversity of states and processes possible for *instances* of a design needs to be studied. Instances of more complex designs are capable of more diverse states, processes, and forms of development over time, illustrated by the vast diversity of human minds. We mention some high level concepts that may help, in Section 9.7.

Some organisms, often labelled “precocial”, remain largely unchanged throughout their life (apart from parameter adjustments within a fixed framework), or follow patterns of change common to all members of their species (e.g. microbes, insects, and probably most other invertebrates) whereas others, the so-called “altricial” species, start highly incompetent, and develop under the influence of complex feedback from the environment.

A paper (with J.Chappell) was presented at IJCAI 2005 [6] arguing that the precocial-altricial spectrum is just as relevant to robots as to animals. A sequel was an invited journal paper [1], presenting ideas about multi-layered bootstrapping processes based on a combination of features of the genome and specific features of the environment revealed by exploratory play, as in Figure 9.3. These ideas are still being developed, but have already had some influence. We propose to apply these ideas in research on primate competences.

12.5.5 The “designer stance” in biology

Robotic research exposes questions not normally asked by researchers in animal behaviour. Animal behaviour researchers should adopt what McCarthy [50] calls “the designer stance” and ask “What mechanisms, forms of representations, and architectural features, would I need to put into a robot to enable it to do *that*”. Robotics-inspired questions can draw attention to previously unnoticed fine details of competences generating behaviours, and may also

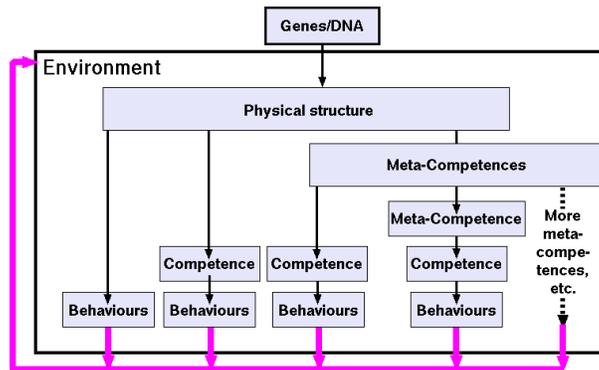


Fig. 12.3. Varieties of control routes from the genome to competences and behaviours: on the left most details are determined by the genome, whereas towards the right there are more complex staggered or layered processes of learning to learn, producing sequences of more sophisticated competences and meta-competences. (Based on [1].)

transform some nature-nurture debates, e.g. by showing that the demands of interacting with a complex, structured, manipulable 3-D environment may be at least as important in driving evolution of cognitive powers as the need for social interaction, which has gained far more attention.

E.g. the work on altruism in young humans and chimpanzees in [38] investigates motivation, but takes for granted the cognitive competences mentioned above in Section 9.2, required for all of: perceiving what is going on, inferring somebody’s intention (e.g. “trying to get books into a cupboard”), planning actions to achieve that intention, deciding to execute actions to unblock a precondition (walking forward and opening the door), and performing the actions. If all that, including the use of meta-semantic capabilities, can occur in pre-verbal children, and in simpler forms in chimpanzees and other animals, that raises deep questions about the pre-verbal forms of representation available to animals. There are also deep questions about where those representations come from – are they innate, or, as seems more likely, since evolution is unlikely to have provided an innate concept of a cupboard door, are they a result of developmental processes including processes driven by meta-competences that cause exploratory behaviours that produce learning as indicated in Figure 9.3, explained in [1].

12.5.6 Should requirements refer to laws of behaviour?

For animals with very large collections of capabilities, changing motives, plans and beliefs, and frequently changing environments, there may not be any *laws* that characterise their behaviour – only *possibilities* that need to be explained. So although some members of the cosy team studied psychological

research literature, as reported in Chapter 8, it is important not to assume that laboratory results provide evidence of laws to which human-like robots had to conform – they merely show what sometimes happens, especially when the subjects are from a single culture, restricted age ranges, and numbers of subjects are often well below 100.

Even for a particular design there may not be well defined *laws* predicting characteristics or behaviour, only a rich space of possibilities. There may, however, be laws concerned with “low level” features of transducers and implementation mechanisms, and there will be some limitations on what is possible for any particular implementation of a design, for instance limitations of processing speeds, and capacity limits. Nevertheless, how any particular individual instance of an “altricial design”, behaves within the limits permitted by the architecture and mechanisms used, far from conforming to exceptionless laws will typically depend in complex ways on its goals, preferences, interests, what the individual has learnt, what it was doing previously, etc. – conditions that can vary enormously across individuals and from time to time for each individual. In addition for deterministic dynamical systems with non-linear feedback prediction can be impossible if measurements of initial states have bounded precision.⁷ So the kind of research that is appropriate to a design-based science of cognitive systems, natural and artificial, will not fit the popular conception of science as mainly a search for exceptionless laws (or even high probability generalisations) but will require us primarily to attempt to understand varieties of possibilities that particular designs support in combination with particular contexts and personal histories [39, Ch 2].

12.6 Environment-neutral requirements and limitations

Some requirements and limitations that arise when they are not satisfied are not concerned with specific types of environment: they are “topic-neutral”. Other requirements will be discussed later. The main topic-neutral requirements and limitations are concerned with architectures and with forms of representation available or missing.

12.6.1 Forms of representation

Limitations due to forms of representation available include: if full predicate logic is not available, it will be difficult to express or reason about non-singular propositions such as “There is no tiger in the room” or “Every person in the room is taller than at least one other person in the room” or the corresponding questions, or goals (e.g. “Find chairs for everyone”). However “compiled” versions of some of these can be expressed as procedures (e.g. a procedure for

⁷ See <http://www.ecmwf.int/research/predictability/background>.

fetching chairs, with appropriate stopping condition). This will enable a robot to do things, but not to describe or think about what it is doing.

Systems lacking modal operators (e.g. “possible”, “necessary”, “impossible”, “contingent”) may be unable to represent the difference between an *empirical* generalisation such as that pools of water sometimes merge while being counted and a *necessary* truth such as that counting a fixed set of objects in different orders must give the same result. (See Section 9.13.) Without modal operators it will also be impossible to represent what is *possible* or *impossible*, and therefore affordances will be inexpressible. (See Section 9.12.5. Words of natural language related to affordances, such as “graspable” refer to what is possible, and therefore involve an implicit modal operator.)

The forms of representation available and the architecture can also constrain what *questions* can be formulated (internally or externally), a topic discussed in more detail in an early CoSy requirements deliverable [30]. Being able to formulate questions that can generate information seeking goals is an important aspect of being an autonomous learner.

Being able to think about or communicate with other intelligent individuals requires the use of forms of representation that support meta-semantic competences: the ability to refer to things that refer. This is also required for certain kinds of introspection and self-understanding. Meta-semantic competences require the ability to make a distinction between representations whose function is to refer to the world and those whose function is to represent what someone or something else is referring to, possibly erroneously. This is generally referred to by philosophers as “referential opacity”. It is not clear when human children have that competence or whether other animals have it, though the ability to understand or tell stories, or engage in “make believe” play requires it. We did not attempt to provide this level of sophistication in our robots so in a sense they could not think about other individuals as having information processing capabilities, including beliefs, intentions, desires, etc. This necessarily limited the forms of interaction that were possible. For example, without such meta-semantic competence a robot cannot consider whether a human lacks information, or whether been misperceived, or a communication misunderstood.

There is no general agreement on how referentially opaque forms of representation should be dealt with. Many researchers hope that an extension to logic, e.g. using new logical operators, will suffice. I suspect that the best solution is to extend architectures, to allow the same form of representation to have different roles in different parts of the system, so that a form of representation may be taken as believed to be true if it occurs in one part of the system, and as a specification of someone else’s belief if it occurs in another part of the system. Similar remarks can be made about differences between propositions, questions, goals, conjectures, memories, fantasies, etc.

Another important feature of a form of representation may be the way it can be used to control search in certain classes of problem. Being restricted to Fregean (logical, applicative) forms of representation, without any subsystems

able to reason with analogical or other forms of representation may be a serious handicap for some classes of problem, as explained in [51] and [39, Chapter 7]. Moreover, in some cases probabilistic representations are more useful than categorical representations. In other cases hybrids are useful. Often it is useful to combine many detailed information items with histograms that can reveal otherwise unnoticed global patterns – e.g. most edge-features in one part of an image are horizontal, and vertical in another. We did not attempt to produce a catalogue of forms of representation that could be used to select optimal candidates for various parts of the system. Instead, in most parts of *CoSy*, forms of representation were used that were traditional for work of that sort in AI. It may be useful at some future date to investigate whether this has restricted progress (either in *CoSy* or in the whole field of robotics).

12.6.2 Architectures

Limitations directly related to architectural features include: whether a system can do certain tasks in parallel or not; whether certain behaviours are interruptable; whether there are some low level “cognitively impenetrable” components (Pylyshyn in [52]) whose functioning is not affected by other sub-systems; whether some competences can be modified (e.g. debugged or speeded up) as a result of self-monitoring during performance; whether information about the environment is represented on different scales (e.g. local and global); whether information is stored about what has recently happened (episodic memories); whether there are mechanisms that check for consistency within and between percepts, previously acquired generalisations, episodic memories; whether information changes can trigger side effects via constraint propagation (as suggested in connection with vision, below in Section 9.12.6 and Figure 9.8); whether there is an “alarm” system (or several) that trigger rapid reorganisation in response to a detected threat or opportunity (Section 9.10). This is not a complete list.

Since the environment endures over time with some changing and some static features, while sensory contents change on much faster time-scales, it can be useful to have an architecture in which the main source of information about the environment for most of the system is *not* the contents of sensory signals but an enduring, incrementally updated representation of the environment. In some cases this should be a-modal so that different sensory sources can be used in parallel adding or modifying different features, and acting as cross-checks.

Both motion of perceived objects and the perceiver’s own motion (including saccades) entail such requirements for a visual architecture, discussed by Trehub in [53]. In particular the mapping between an enduring scene representation and image pixels will be constantly changing, though not the mapping between the scene representation and the *optic array*, if the viewpoint is fixed. However if information at different levels of abstraction is represented, the very notion of registration becomes blurred, though a sort of registration

is indicated in figure 6 of [39, Chapter 9]. Further complexities are required if manipulation or motion of either objects viewed, or the viewer, causes different parts of an object to be visible. The requirement to represent hidden parts (e.g. the far side of a cube after rotation, or the contents of a box after shutting its lid) goes beyond representing what is available in the optic array. Unfortunately, it was not possible in the time available to meet these architectural requirements, except in restricted ad hoc fashion, though the dialogue system (Chapter 8) presupposed some of them, and the Explorer system (Chapter 5) used SLAM techniques to represent far more information about currently unperceived entities than PlayMate.

12.7 De-fusing diversity: Towers and layers

Designs for whole systems can vary in uncountable ways, if all possible alternatives for every design decision are considered. This raises the urgent question: is there any way a science of intelligent systems can impose some intelligible and useful structure on the space of possibilities, or is it just a morass composed of an enormous collection of special cases?

12.7.1 Generative frameworks

One way of trying to impose structure is to specify a generative framework by specifying a collection of basic building blocks for behaviours and ways of combining them into arbitrarily complex systems. That was the approach adopted by Turing, which produced a class of machines that was later shown could be generated in alternative ways, e.g. Turing machines, production systems, lambda calculus, Curry combinators, logical inference mechanisms, etc.

Another approach is to take some supposedly general architectural framework such as SOAR or ACT-R (both summarised in [54]) and then show how all other architectures of interest can be subsumed by the chosen one.

There are two problems with this generative approach, despite its great power and usefulness in computer science. First the set of machine-types generated is too restricted, being composed entirely of systems whose behaviour consists only of components in the selected framework, e.g. discrete serial steps in a Turing machine. However that limitation can sometimes be overcome by allowing the components to be combined in different ways, to generate types of machine with possibly asynchronous concurrently active components, e.g. as in digital circuit design formalisms, or Milner's pi calculus [55], or Hewitt's actor formalism [56]. The remaining restriction to discrete processes can be removed by allowing additional analog mechanisms that support continuous variation, e.g. conductors, oscillators, capacitors, etc, or mechanical devices, such as gears, pulleys, springs, strings, etc. and analog to digital and digital to analog converters. The use of generative representations has the problem

of being “bottom up”, making it hard to get a high-level, top-down view of the space of designs.

A closely related problem is that for the kinds of behaving systems we are interested in there is always an external environment, and we need to find ways of characterising systems not merely in terms of how they are built, from processing components, or what goes on inside them, but in terms of what sorts of environments they can interact with, and how they interact with them – i.e. they are characterised in terms of combinations of types of *functionality* rather than combinations of types of *mechanism*. A framework for presenting those ideas was loosely inspired by Nilsson’s discussion in [57, Ch 25] of architectural “towers” and architectural “layers”.

12.7.2 Subdivision into towers of functionality

Systems that act in an environment can be described as having three major (possibly overlapping) sub-systems with different functions:

- a *perceptual* sub-system that gains information from the environment, processing information derived from physical transducers that produce internal signals from incoming energy in various ways;
- an *action* sub-system that emits energy in various forms (especially applied forces) using transducers controlled by internal signals;
- and between those sub-systems an arbitrarily complex collection of “central” mechanisms that interact with the perceptual and action sub-systems but may also do many other things.

This gives us Nilsson’s three towers, which he called the perception tower, the action tower and the model tower. For our purposes, the label “model” is too narrow, though our label “central” is also inadequate.

We can further subdivide types of tower according to which kinds of sensory information they use (visual, auditory, haptic, magnetic, etc.), which kinds of outputs their effectors produce (luminescence, acoustic, pressure-applying, squirting, throwing, blowing, changing shape, etc.), what kinds of information they can derive from the sensors, and what changes they can produce in the environment – including cases where acting and sensing are tightly linked (Gibson, [58]). e.g. altering gaze in order to obtain different visual information or squeezing something in order to gain haptic information. (See Chapter 3.)

12.7.3 Subdivision into layers of functionality

A different way of subdividing systems or sub-systems “top-down”, i.e. in terms of their functionality, is by distinguishing different ways of mediating sensing and acting. This is orthogonal to the previous divisions, and can be thought of as providing three layers, though our partly biologically-inspired

division into layers is not exactly the same as Nilsson's (or most others, e.g. [59], Gat [60], and Minsky [61]).

Reactive layer: The evolutionarily oldest, and easiest to implement mechanisms are systems that respond to sensory input by immediately generating a short term response (externally or internally) without representing any consequences of that response, and without reasoning about multiple possibilities before selecting that response. We can call those “reactive” mechanisms, while acknowledging that they can vary enormously in complexity, including the number of intermediate processing stages between sensing and acting, and also whether some of the things sensed or altered are internal, e.g. sensing internal energy levels and damping down internal levels of activity, or switching sub-systems on or off. Such reactive systems may be either discrete or continuous, and may or may not include feedback control, adaptive learning, or other kinds of sophistication. In more complex cases (often implemented in neural net mechanisms) the inputs and corresponding outputs are *input and output patterns*, where a collection of sensors acting concurrently trigger a collection of coordinated outputs. There are also intermediate cases. Other possibilities include temporally extended triggering inputs (e.g. using thresholds) and temporally extended outputs (e.g. running away from something). Moreover, reactive systems can include “proto-deliberative” mechanisms, described in [10] where competing output actions are stimulated at the same time, but only one wins on the basis of some mechanism for evaluating alternatives. (Unfortunately, some researchers confusingly label this “deliberative”.)

Deliberative layer: A second type of layer can provide various kinds of deliberative sub-system, which vary in the sophistication of the predictive, or control functions they support, as discussed in [10]. The common feature involves the ability to respond to input, or the formation of a goal, by considering not only alternative responses but also the consequences of those responses before selecting an action. More sophisticated versions can consider alternative action sequences before deciding – as many of the earliest AI systems did. The ability to explore branching futures depends crucially on sensory information being “chunked” into discrete categories, so that associations can be learnt between discrete cases, avoiding the need to handle infinitely branching futures (see [10]). (Nilsson's top two layers are included in our second layer.)

Meta-management layer (with meta-semantic competences): The third, biologically most recent, type of layer, not discussed in Nilsson's chapter, is able to do two important kinds of things, namely monitor some of its own (semantically rich) internal states and processes, including characterising them in some explicit, structured formalism (unlike hierarchical analog control systems), and representing other individuals as also having such internal states, describable using meta-semantic competences (explained above in Section 9.6.1).

Although three layers have been distinguished, there are intermediate cases not yet mentioned, as well as other useful ways of dividing up functionality (e.g. Minsky [61] has six layers). A possible source of confusion is that all mechanisms must ultimately be implemented in reactive mechanisms of some kind.

12.8 The CogAff architecture schema – one small step

By the time the CoSy project started, we had attempted in previous work (mentioned in Chapter 1) to make the task of exploring design- and niche-space more tractable by using a generic schema, in which the tower and layer classifications presented above were superimposed, thus forming the nine cell CogAff architecture schema, depicted in Figure 9.4 and mentioned in Chapter 1 of this book. This uses a modified version of Nilsson’s distinction between layers and towers, where differences between towers arise from different relationships to the environment, and differences between layers are defined in terms of differences in evolutionary age, kind of computational and representational sophistication, and types of mechanism used – though I now prefer to emphasise types of functionality rather than types of mechanism. For instance, the bottom left box could include several types of low level sensory processing, and the top left perception of communicative actions, intentions, moods, etc. Likewise the bottom right box could contain low level motor outputs of various kinds and the top right box gestures, linguistic utterances, and expressive behaviours. The schema is clearly an oversimplification, but provides a crude initial framework for comparing a wide range of architectures, according to which boxes are used, what they contain, which forms of representation are used in the boxes, and how the components are connected. Some of the problems of combining different sorts of functionality, using different forms of representation are discussed in Chapter 2. In contrast with the “generative” approach to diversity, this framework is neutral as to what the smallest processing components are. Examples of systems that do not fit this schema are systems composed of large numbers of autonomous individuals, so-called “multi-agent systems”.

It is not claimed that this is the only way of dividing architectural components or that the divisions are sharp. Different AI researchers who use layered architectures describe their layers differently (e.g. Brooks [59], Gat [60], and Minsky [61]) though some of those differences can be subsumed within the CogAff Schema. Sun [63], describes an architecture (CLARION) made of several components each with two layers corresponding roughly to the two bottom layers of CogAff. This can be re-described as a two-layer architecture, where the two layers are subdivided into components, each closely linked to a component in the other layer. Exploring all these ways of characterising architectures, to see whether there is some useful over-arching form of representation is a topic for further research.

Perception	Central Processing	Action
	Meta-management (reflective processes) (newest)	
	Deliberative reasoning ("what if" mechanisms) (older)	
	Reactive mechanisms (oldest)	

Fig. 12.4. The CogAff schema [62] loosely based on a combination Nilsson's ideas of towers and layers, provides a framework for comparing a wide range of architectures. However, as explained in the text, it is only crude beginning.

To a first approximation the CogAff schema provides a sort of “grammar” for architectures. There are many special cases of that schema, including special cases that use only the reactive layer e.g. purely reactive subsumptive architectures [59], and reactive insect-like architectures that include a reactive “global alarm” mechanism receiving inputs from all parts of the system and capable of modulating or redirecting all parts very quickly.

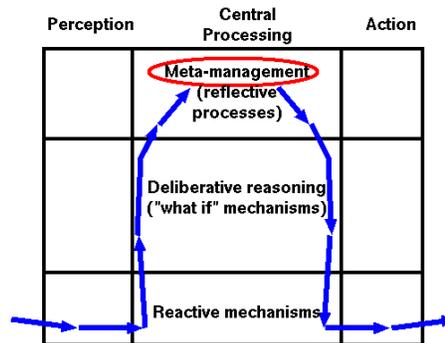


Fig. 12.5. A popular type of architecture, which could be called the “Omega” architecture, depicted here, is a special case of the CogAff schema.

12.8.1 Omega architectures

A popular type of architecture incorporates a sequential multi-layer pipeline: sensory information comes in via low level sensors (‘bottom left’), gets abstracted as it goes up through higher central layers, until action options are

proposed near the top, where decisions are taken, after which control information flows down through the layers and out to the motors ('bottom right'). This can be called an 'Omega' architecture because the pattern of information flow is shaped like a the Greek letter Ω , as shown in Figure 9.5. Many models in AI and psychology have this style e.g. Albus in [64]. The influential 'contention scheduling' model of Shallice and collaborators [65] is a variant in which the upward information flow activates a collection of competing units where winners are selected by a high level mechanism, on the basis of possibly learnt preferences. Their later work added a further layer the "Supervisory Attentional System" (SAS), for dealing with novel situations. Such systems are examples of a general three step cyclic pipeline model for processing information:

REPEAT: (1) Sense. (2) Think and decide. (3) Act.

The CogAff schema accommodates such sequential pipeline architectures, as a special case, but also permits alternatives where the mechanisms in different layers are *concurrently* active, and various kinds of information constantly flow within and between the sub-systems in both directions. An important feature of embodied animals and robots with multiple physical sub-systems concerned with locomotion, perception, manipulation, communication, and internal functions, is that they illustrate the need for such concurrency. For example, the CoSy Explorer needs to control its movements at the same time as it uses its visual and other perceptual systems to check whether it has reached its target and simultaneously processes some speech input. In general the sorts of concurrency required in a robot with multiple interacting sub-systems refute philosophical functionalist theories that use a finite state machine model of mind, e.g. Block in [66]. This functionality is also inconsistent with sequential "sense think act" models.

Many researchers do not understand the need for anything but Omega architectures: they think of perception and action as essentially low level processes of transduction, so they use small boxes for them in architecture diagrams, possibly connected to low level internal processes. That is roughly how perception and action are currently implemented in the CoSy robots, (except perhaps for the linguistic components). Section 9.12 below explains why towers are needed for vision and action in more advanced robots, though only primitive forms have so far been implemented in CoSy, described in previous chapters.

12.9 Beyond the CogAff schema

The subdivisions in the CogAff schema are not offered as the only subdivisions to be used in describing designs. A study of the products of evolution will reveal many intermediate cases, requiring a finer-grained subdivision of

types of component. The task of designing working systems, as in CoSy, can also drive further development of such a conceptual framework, identifying discontinuities in requirements and designs as illustrated in [12] and below, challenging philosophical and other theories that propose a small number of major steps in evolution (e.g. expanding the useful overview in [67]). Some finer-grained distinctions between layers are described in [10].

One of high level features of a design not represented in the CogAff schema is the extent to which it learns or develops. It is clear that the human architecture is not fixed at birth but develops over time, including acquiring new layers of competence, new forms of representation (including use of new languages, mathematical and scientific notations, musical notations, technical diagrams, maps, circuit diagrams, etc.), new ontologies and new reactive skills related to the other developments. There may be still unknown *internal* forms of representation that humans and other animals develop after birth.

In a more detailed survey we would need to divide up systems in terms of different patterns of change and development, with at one extreme a completely fixed system, followed by a system that is fixed except for parameters that can be adjusted, followed by more and more complex forms of learning and development, as illustrated in Figure 9.3. It is not clear that researchers in AI and Cognitive Science have so far produced a comprehensive taxonomy of the sort that would be required. The CogAff schema may turn out to provide a useful way of indicating that different sub-systems develop in different ways – for instance many reactive systems learning by parameter adaptation, the high level systems growing by developing new languages, building explanatory theories, extending their ontologies, and absorbing new values from the surrounding culture, and all of them developing by acquiring new links between subsystems (e.g. compiling new reactive versions of pre-existing deliberative capabilities). Fido, the domestic robot in Chapter 1 may need many of these forms of development.

Another set of divisions between boxes is concerned with forms of representation. AI researchers have frequently noted the importance of choosing forms of representation that are suited to particular problems, e.g. Minsky in 1961 [68] and McCarthy and Hayes in 1969 [69]. The latter extended Chomsky's ideas about adequacy of grammatical formalisms, by distinguishing metaphysical adequacy, epistemological adequacy and heuristic adequacy. Following ideas of Marr, it is now commonplace to distinguish viewer-centred, object-centred and room-centred forms of representation, among others. The functional differences between the layers and the towers in the CogAff schema will be only loosely connected with differences in form of representation. Further work is needed to work out which sorts of representation fit where. A special case is the use of linguistic forms of representation.

12.9.1 Where are the linguistic mechanisms?

Analysis of requirements for our robots raised the question whether linguistic competences fit naturally into a small subset of the boxes in the CogAff schema, or whether they need additional boxes, or whether they are to be distributed in all the main portions of the architecture.

The answer seems to be that linguistic competence is distributed through many parts of the architecture. It is obvious that linguistic perception of speech requires multi-level, processing dealing concurrently with acoustic, phonetic, morphological, syntactic, semantic and pragmatic information. So for speech the requirement for a *tower* of perception is clear. Similar comments apply to reading printed or hand-written text. Likewise speech production requires multiple levels of processing as ideas, sentences, phrases, word-selection, and production details such as tempo, intensity and intonation are determined, along with self-monitoring that can lead to self-correction. Similar remarks can be made about production of written communications. So producing linguistic utterances requires an action tower, rather than just a simple system for feeding signals to a transducer. These two points rule out an Omega architecture for a system with human language capabilities.

Use of language internally or externally can also function as an enhancer for other sub-systems, e.g. concerned with planning, reasoning, hypothesis formation, prediction, motivation, and conflict resolution. Moreover, while explicitly learning a new foreign language seems at first to exercise mostly the two upper levels of the schema, as expertise develops that seems to make increasing use of automatic reactive sub-systems, in the lowest layer.

Further development of these ideas and their implications for imposing structure on the space of possible designs, remains a topic for further research, especially if a future project can develop an intelligent pre-verbal toddler-robot as a basis for acquiring linguistic competence, so that we can more clearly understand what difference language makes, an issue that was not in CoSy because linguistic competence was integrated from the start. Although that appeared a reasonable strategy at the time, and may be an appropriate strategy for particular engineering applications of AI, it is arguable that the failure to produce a robot that behaved intelligently before adding linguistic abilities has seriously distorted our research because proper human-like linguistic competence should be a later addition to more basic general animal competences, extending sophisticated functionality, including communicative capabilities, that existed without language.

CoSy does have a number of such capabilities, including perceptual and planning capabilities, so in principle they could have been used to generate a kind of animal intelligence (e.g. producing play, exploration and learning). However that would have required a different way of putting things together so that motives need not be derived from a human linguistic communication. We did discuss mechanisms for “architecture-based” motivation, as opposed to the kinds of “reward-based” motivation often assumed to be necessary.

Architecture-based motivation, which, it is arguable, is characteristic of most biological organisms, involves having one or more portions of the architecture where possible future states or processes, or constraints on states and processes can be described, and which, if present will tend to generate and modulate planning, decision-making, and the prioritising of actions. In that case other modules in the architecture can be triggered by various occurrences to create these motivational representations and insert them where they can become effective: for example an auditory mechanism detecting a strange noise, and automatically generating a motive to investigate the source of the noise and adding it to the store of current motives (along with other relevant information [70]). If several mechanisms cause conflicting motives to be generated, that could be detected and might trigger a conflict resolution mechanism to deal with the conflict. All such mechanisms for architecture-based motivation are probably innate in simplest animals (and machines), while products of learning and development play a role in more sophisticated types.

Pressure of time, and our commitment to integrate the subsystems developed by all the partners, prevented us investigating this kind of possibility further. It might in future demonstrate how a robot might learn by play and exploration without being reward-driven.

12.9.2 Varieties of compositional semantics

One of the major differences between the reactive layer and other layers in the CogAff schema is that the upper two layers can make use of formalisms supporting structural variation (e.g. predictions, plans and percepts of varying complexity) whereas the reactive level mainly uses atomic symbols (e.g. measures) or fixed dimensional vectors of symbols.

Where there is structural variation, and old structures can be combined to form new more complex ones (like sentences with sub-clauses), the semantic content of complex structures is usually assumed to be based on compositional semantics: the meaning of any complex expression is a function of the meanings of the parts and the structure of the expression. A problem addressed in CoSy led to an extension of this idea, as follows.

A question arose as to how certain words and phrases such as ‘to the left of’ used in an instruction to place something should be interpreted when there is a large area to the left. The solution at first proposed was to use a probabilistic semantics to select the “best” target location in the region under consideration, on the assumption that the probabilities would be derived from previous experiences. A probabilistic mechanism was therefore implemented.

Despite the empirical evidence, I found the arguments for the probabilistic interpretation unconvincing because they did not take account of the importance of context. I thought the empirical, probabilistic, data were likely to be a side-effect of deeper, more powerful and general mechanisms.

So an alternative theory was proposed, according to which Gricean principles of communication can be combined with a form of compositional se-

mantics that allows context to play a role anywhere in a semantic structure. So on this view, the proper way to obey the request “Put the pen at the left of the book” is to use an understanding of what the pen is needed for, and what it can interact with, along with the perceived set of spatial relationships, to select a target location, rather than using a probability calculation based on previously observed placings. Where the context does not determine a selection, any location within reach of the person who asked for the pen will do.

This turned out to be a special case of an important general idea, that transforms many supposedly vague expressions (e.g. “heap”, “big”, “long”, “efficient”), into expressions with a gap to be filled by the context of use, based on general knowledge and intelligence. For example, the number of stones required for a heap, depends on why a heap is required: to hold down a tarpaulin in a strong wind, to provide a base for one end of a bridge, to provide a platform on which to stand to see over a wall, and so on. These ideas led to a long and complex discussion paper [8] still under development.

An implication of this, is that the process of sentence comprehension can intrinsically include deep integration with reasoning, memory and visual perception. For this to work, not only the language sub-systems, but also the reasoning, memory and perceptual sub-systems must be designed so as to support the integration. That is a powerful challenge to system designers working on those sub-systems in isolation.

It is a powerful challenge anyway, and I expect it will be many years before machines can handle such context sensitivity in a useful way: it will first require development of a great deal of knowledge about the world and what people can want or intend or fear or prefer to happen in it.

12.10 The H-CogAff Special Case

One special case of the CogAff schema, labelled H-CogAff (Human-Cogaff), developed at Birmingham in the decade before CoSy [62], and also mentioned in Chapter 1, assumes all the boxes are occupied and that there are many connections, including connections to a reactive alarm system, and connections linking various layers in multi-level perception and action sub-systems to different parts of the central “tower”, as explained in [62].

This special case and the diversity allowed by the CogAff schema were mentioned in our proposal. It was hoped that some of the ideas could be tested and the specifications refined and extended as a result of work on the robots. We did develop some new ideas about requirements for such an architecture as a result of the requirements analysis done during the project and work on some of the sub-systems. In particular [10] showed the need for several important subdivisions in the deliberative layer, and there were changes to our ideas about vision, language and the types of compositional semantics required in an intelligent system, as mentioned above.

Implications for meta-management were related to requirements for motive generation and selection in the robot, but so far only very simplified versions of those mechanisms have been implemented, as they sufficed for the scenarios chosen. Another important function of meta-management is mentioned below in connection with mathematical learning. These ideas are still under development, and have not yet fed into working systems (so they are not mentioned in Chapter 2). In some cases, this will require major advances. In general, insofar as meta-management involves the ability to monitor, represent, modulate or extend internal processes, it will require meta-semantic competence: namely the ability to represent things that have semantic content.

This will require dealing with referential opacity, and many other things that have so far not arisen in the CoSy project because the scenarios addressed have been relatively simple. However, as some of the discussions among the team working on integration revealed, the problem is already just below the surface in the current system because we are close to problems where the robot needs to think about and reason about what a human knows, wants, intends, can see, etc. Future work will need to address these problems.

Some researchers, e.g. John McCarthy, favour dealing with such cases by introducing new modal logics with special modal operators for Knows, Wants, etc. I think a better solution uses architectural mechanisms providing a kind of ‘encapsulation’ e.g. where supposed/possible beliefs (one’s own or someone else’s) are treated using the usual forms of representation but not allowed to have the normal causal powers of beliefs. For instance, imagining a situation where you believe a hungry lion is running towards you should not make you decide to run away. On the other hand if you know that John thinks the bush in the shadows is a hungry lion you should be able to work out that he may wish to run away, and therefore decide to tell him that what he thinks is a lion is a bush. It is well known that young children take some time to develop the ability to handle referential opacity. It is not clear whether a future robot will also have to grow the ability, as opposed to having it pre-programmed. This may be related to whether pre-verbal competences need to be well developed before language learning starts.

12.11 Study trade-offs not special cases

The diversity of designs and sets of requirements mentioned above is very daunting. One way of trying to make sense of this diversity is to consider divisions at a fairly high level of abstraction to start with, as proposed in previous sections.

Another strategy for imposing structure on the problem is to think in terms of *trade-offs*. Instead of arguing about how things must be, or trying find ways of categorising whole systems using evaluation functions, we can look at costs and benefits of different options within a single design problem. These costs and benefits can be thought of as niche-relative disadvantages and advantages,

without assuming that these must all have numerical values. Often the trade-offs need to be expressed descriptively rather than numerically: e.g. design X makes certain kinds of learning impossible, but enables predators of type Y to be avoided – not-unlike the style of consumer reports on multi-functional objects such as cars, cookers, computers, etc. (These issues are addressed in this old paper on “better” [71], which suggests a semantic structure with a component for context, not unlike the analysis of vague words, above.)

12.11.1 Nature-nurture trade-offs

A particularly interesting case is the trade-off between pre-programming all the behaviour required in a system (as evolution in effect does for the majority of species, a feature for which the label “precocial” is often used) and instead producing an “altricial” organism or machine that starts off superficially highly incompetent but has a sophisticated meta-competence for developing itself in a certain class of environments, including acquiring many new competences through interactions with the environment [41, 72, 6, 1, 27] This seems to be the option evolution has developed for many mammals, especially hunting mammals and primates, as well as many birds).

In between the two extremes are different combinations of “precocial” (or preconfigured) competences and “altricial” (or meta-configured in the terminology of [1]) competences. Jackie Chappell and I argued at IJCAI’05 [6] that both extremes and intermediate cases will be required in robots.

The CoSy robots have a particular combination of pre-programmed and learnt competences dictated largely by limitations of our time and available techniques: future work should explore more possibilities in a more principled way, after a deep analysis of trade-offs. There is work in progress in building a version of the PlayMate that discovers things about the world and effects of its actions in Birmingham⁸ and Freiburg.⁹ We are still a long way from the kind of bootstrapping mentioned in Section 9.5.4.

12.11.2 Image-scene tradeoffs in visual processing

One trade-off that provoked much discussion, especially in the context of the PlayMate scenario was the trade-off between trying to perform visual tasks by

⁸ Work by Marek Kopicki, partially reported in <http://www.cs.bham.ac.uk/~msk/report8/report8.final.pdf> <http://www.cs.bham.ac.uk/~msk/report9/report9.pdf>

⁹ See Jürgen Sturm, Christian Plagemann, Wolfram Burgard. Adaptive Body Scheme Models for Robust Robotic Manipulation. In *Proceedings of Robotics: Science and Systems (RSS)*, Zürich, Switzerland, 2008.
Jürgen Sturm, Christian Plagemann, Wolfram Burgard. Unsupervised Body Scheme Learning through Self-Perception. In *IEEE International Conference on Robotics and Automation (ICRA)*, Pasadena, CA, USA, 2008.
<http://www.informatik.uni-freiburg.de/~sturm/media/resources/public/zora-7dof-demo.avi>

using and learning about 2-D image features, relationships, models and processes, as opposed to using and learning about 3-D structures and processes, which could be either inferred from image information or used to project to image information, in a mixture of bottom up and top-down processing.

An illustration of the need for a 3-D ontology is shown by the ambiguity of the Necker cube, in which the experience of a pattern of straight lines flips between seeing a cube with one of two spatial orientations, and where the relative distances change. Since *nothing* changes in the image when the cube flips it is clear that something not in the image must be represented in each of the two states.

It is a small step from there to argue that even in unambiguous images where a cube is seen the percept must represent 3-D structure, not just identify a portion of the image and its features and relationships. This kind of representation using an exosomatic ontology (i.e. referring to something outside the perceiver, that can exist independently of whether it is sensed or not) will be crucial in a robot with the competences we aimed for in CoSy. However at present the vision systems deployed are very limited in their understanding of 3-D structures.

Another tradeoff whose potential importance emerged in the analysis of requirements was the tradeoff between representing only static structures and then attempting to represent processes, including actions in terms of collections of representations of static structures, *vs* constructing representations of processes that could be used during the perception of processes and also when reasoning about possible past or future or unperceived processes. This topic is discussed in [23] and some of the online presentations.

12.11.3 Trade-offs related to noise and uncertainty

Another trade-off discussed at various times and especially during the final year is the trade-off between (a) dealing with noise and uncertainty by making use of probability distributions and performing probabilistic inferences, as opposed to (b) finding a form of representation that avoids noise and uncertainty (in a particular situation) by using a high level of abstraction. For example if it is impossible to decide for certain whether a curved line has constant curvature or not (i.e. is a circular arc), then instead of working with a probability distribution over a range of possible curvatures, simply describe it as curved. In some cases where there is uncertainty the robot could avoid reasoning about probabilities by getting new information. E.g. if the size of an object is uncertain because it is partly occluded, instead of dealing with probable lengths choose a different viewpoint, or temporarily move the occluding object. Another important possibility seems to be to notice that there may be regions of definiteness where an answer is “yes” and regions of definiteness where an answer is “no”, and a phase boundary where the answer is uncertain. In that case, it may be possible to move away from the phase boundary when it is encountered. E.g. if you can’t tell whether your trajectory will or

will not cause you to bump into the edge of a wall, it will typically be possible to ensure that you will definitely not bump into it if you aim more to one side. These points are discussed in this draft document [13] mentioned below in connection with changing affordances.

There are other trade-offs that surfaced during the project, most of which still need further work.

12.12 Requirements for visual systems

12.12.1 Why do perception and action need towers?

There are many different sorts of requirements that could be considered for visual systems. For example, work on systems that can be trained to recognise objects in images of cluttered scenes is challenging and useful for many application domains, but it does not necessarily provide the kind of visual competence required for a robot with a movable hand to be able to work out a good way to pick up an object with a complex shape, subject to varying constraints – e.g. it may be fragile, or full of liquid, or very hot, or too large to be grasped by one hand, or partially obstructed by another object. In some sense it is obvious that the ability to perceive 3-D shape, including distinguishing parts with different features and relationships and seeing their relationships to one another and to other objects must, for humans and other animals, and presumably also future intelligent robots, have priority over recognition of objects, since it is possible to see and interact with (e.g. picking up, climbing over, disassembling) something you do not recognise. This is related to the theories of Gibson mentioned later.

Humans can perceive and interact with objects in poor light, can see spatial structures and potential for actions of various sorts even in low resolution noisy images, as illustrated in Figure 9.6. A more detailed analysis of some of the requirements can be found in [73].

12.12.2 Multi-strand process perception

Analysis of requirements for PlayMate in year 1 [30] revealed the importance, for a robot doing 3-D manipulations, of being able to perceive and think about “multi-strand relationships” (relationships holding between different parts of two or more objects, in addition to relations between the whole objects). Some of the relationships will be metrical relationships of size, distance, angle, volume, curvature, etc. Others will be qualitative relationships, such as containment, contact, being above, overlapping, being nearer, being between, etc. The latter may change discontinuously while the former change continuously. Other relationships may be causal or functional, e.g. pushing, supporting, stretching, compressing, etc. All of these need to be perceived by a robot with the capabilities we were aiming for in the PlayMate, and to some



Fig. 12.6. *Despite low resolution, poor lighting, and noise in this image, people easily perceive a collection of objects with definite spatial relationships, even though what is perceived (including shapes, orientations, curvature, relative thickness, etc.) is not perceived with very great precision. A challenge is to devise forms of representation that (a) are derivable from images despite poor image quality, and (b) have sufficient definiteness to allow actions to be planned and executed reliably. (From [73]).*

extent also in the Explorer. However, the work on perception in CoSy was not able to address most of this.

When actions are performed the different sub-relations can change in parallel, producing “multi-strand processes”. Some of the changes are continuous and others discrete (e.g. topological). This is related to the need for perception of static structures to use multiple ontologies in parallel (not just part-whole layers). Examples of the use of different ontologies in seeing a static scene include seeing lines on a page, seeing the lines forming a face with parts seen as eyes, nose, mouth, cheeks, etc., and seeing the face as happy, or sad. When the well known ambiguous duck-rabbit figure (Figure 9.7) flips the perceived 2-D image features do not change, but different ontologies are involved in the two percepts, e.g. ears vs bill. Furthermore there is a meta-semantic ontology involved insofar as the duck is seen as looking one way and the rabbit as looking the other way. Around 30 years ago, the Popeye program described in [39, Ch 9] used a mixture of top-down and bottom-up processing, and stored knowledge, to interpret messy pictures in terms of different ontology layers, with distinct part-whole relationships in each layer. Although it was noticed many years ago that seeing static *structures* could involve perception of structures at different levels of abstraction, using different ontologies, the need for perception of *processes* at different levels of abstraction went unnoticed, though something like this was noted by Grush [74] in 2004.

This generates requirements for human, animal and robot vision that appear not to have been widely appreciated. This is because when things change there can also be changes going on simultaneously at different levels of abstraction. For instance looking at a video of a rotating wire frame cube, involves seeing changing light and dark portions of the image and simultaneously seeing edges, corners and faces of the cube moving around in 3D space, changing positions, orientations and relative distances. That is a relatively simple case compared with what is required for the PlayMate robot to see what it is doing

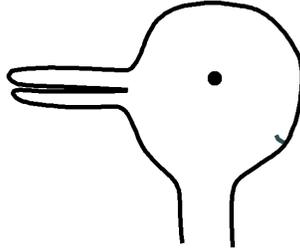


Fig. 12.7. *Many ambiguous figures flip between two percepts without anything in the image changing. This often gives important clues to the multi-layered ontologies that can be involved in visual perception. What changes when this flips between the two views?*

when it manipulates objects and far more complex shapes than wire frame cubes are involved, in addition to changing causal and functional relationships. At present there is nothing in CoSy that represents multi-strand processes, although some fragments are there. However the visual challenge of seeing our robot's arm and hand seem to be well beyond the state of the art in vision, especially seeing it moving. These ideas were presented in an invited talk at a multidisciplinary workshop in May-June 2007 and will be published in the proceedings: [23].

12.12.3 How to acquire useful ontologies

If the above is correct, then that requires a visual system not merely to be able to learn to see part-whole hierarchies on different levels but also more abstract interpretation layers. This is related to the need for an intelligent system to use amodal exosomatic ontologies. How the ontologies develop and how the semantics of newly defined symbols are generated if they are not definable in terms of old symbols or sensorimotor patterns is a complex issue, discussed in more detail in [17].

12.12.4 Varieties of complexity reduction

Chapter 3 discusses varieties of learning/development that involve reducing complexity by reducing dimensionality of sensory motor information acquired by active exploration and experiment.

The complexity reduction involves moving to a new vector space with fewer dimensions, but preserving the “space-occupancy” form of representation. (An analogical form of representation in the sense of [51].)

We also need to understand other processes that are important in humans and some other animals and which could be essential for some future robots. A very different kind of transformation is the move from representing some

portion of the world in a pixel/voxel based form (i.e. occupancy of portions of some vector space) to representing discrete, enduring, re-identifiable objects with features and relationships – possibly changing features and relationships, where the same object can occupy different spatial regions/volumes at different times.

Consider this example:

A 1000x1000 video with 1000 frames, containing changing blue red and white pixels might be described either in terms of a huge three-valued array, or like this:

- A big blue roughly circular pulsating blob moves right with increasing speed against a white background.
- A smaller red triangular blob rotates clockwise about its centroid while moving left with decreasing speed.
- The second blob starts to the right of the first blob and is temporarily obscured by the first blob as it moves.

These descriptions summarise the contents of a 1000x1000x1000 array. The example illustrates a trade-off between simplicity of conceptual apparatus and simplicity of descriptions:

The conceptual apparatus required to describe all the array cells is very simple: just a uniform formalism based on three coordinates and a colour label, whereas the object-based summary description above achieves massive complexity reduction by using a much more sophisticated ontology, including a distinction between space and time dimensions, the notion of a spatially extended, temporally enduring, but changing object, (with complex criteria for identity of the object), etc.

Further, the three sentence, object-based, summary above describes not only the one scenario with 10^9 cells, but a huge number of different scenarios with slightly different contents (different blob sizes, slightly different shapes, different speeds and accelerations, etc.) though the boundaries between what is and what is not included are somewhat vague because of the vagueness of ‘big’, ‘roughly circular’, ‘pulsating’, etc. and the fact that speeds and accelerations are not specified. This kind of vagueness is related to the point about context-sensitive compositional semantics in Section 9.9.2.

I think biological evolution somehow discovered the need for something like the object-based form of representation in a subset of organisms. This seems to be the basis of human abilities to use logic and to use verbal descriptions. I don’t know how many organisms can use the object-based type of representation. It seems very likely that cats, monkeys, nest-building birds, primates, need it. I don’t know about frogs, flies, paramecia, etc. These questions could simulate new kinds of research in animal cognition, and new questions to be asked about cognitive development in humans. Issues like this are discussed in Brian Cantwell Smith’s book [75], though without specific implementation recommendations.

Much work in robotics (including CoSy) assumes the need to use an ontology of enduring but changeable objects – though often the ontology is assembled in a piecemeal unprincipled way. There is much work still to be done defining the long term ontological requirements first of all for a pre-verbal child-like robot and then for a robot learning to use a human language for communication. It is sometimes assumed that there must be some innate ontology on the basis of which everything else is constructed. However, that assumption ignores the possibility of an ongoing process of testing and debugging of the current ontology which could lead both to rejection of some of the innate components and to construction of extensions that are not definable in terms of the starting ontology. Something like that happens in the history of science, so it is possible in principle. How to get a robot to do that is a topic for future research – one of many long term requirements for human-like robots.

12.12.5 Beyond J.J. Gibson’s affordances

J.J. Gibson [49] drew attention to the requirement for a perceiver’s ontology to include positive and negative affordances: namely features of the environment that are relevant to enabling or obstructing actions that the perceiver can perform that might be relevant to achieving goals. Analysis of requirements for a robot has shown that Gibson’s ideas need to be extended:

- An agent can perceive the possibility of processes that are not produced by the agent, and also perceive things that will enable or prevent such processes. (For more details see [76].) I call this perception of *proto-affordances*. Affordances will then map onto a small subset of proto-affordances.
- There is a kind of affordance that Gibson did not explicitly distinguish from an action affordance, namely an *epistemic affordance*: concerned with aspects of the environment that support or obstruct the perceiver’s acquisition of new information. E.g. turning a face of a cube towards you creates the epistemic affordance that is the possibility of getting information about features of the face. Likewise the rotation will have moved at least one other face out of sight, creating a negative epistemic affordance for that face. One of the important things a child has to learn is what sorts of epistemic affordances the environment provides, and also what actions it can perform to increase the epistemic affordances (gain access to more information).

The ability to perceive and reason about both proto-affordances, including affordances for other individuals, and epistemic affordances is relevant to many of things a future domestic robot may be required to do. In particular the epistemic affordances are very relevant to visual servoing. A partial analysis is in [13], which was written in response to difficulties in getting the PlayMate to grasp things reliably at the CoSy review in 2007.

The ability to detect that one lacks some information requires self-monitoring and a meta-semantic competence and therefore belongs in the meta-management architectural layer. The ability to perceive epistemic affordances and action affordances, and to work out which action affordances will change which epistemic affordances (e.g. moving nearer the door to a room will enable you to see more of the contents of the room), seems to be a feature of development in very young children and some animals that has not been studied except for very restricted contexts using verbal interactions, e.g. asking a child whether looking at or feeling a hidden object will provide information about its shape or its colour. In some animals the actions required to alter epistemic affordances may be genetically compiled into reactive subsystems, but animals that have to learn to manipulate epistemic affordances in environments that evolution could not anticipate will need mechanisms for acquiring such competences. For example, as you walk through a car park there are vast amounts of visual information available in the changing optic array, about changing colours, angles subtended, altering occlusions, various kinds of optical flow, and moving highlights and reflections on curved and planar surfaces.

Using that information to compute, in parallel, shapes, orientations, curvatures, distances, spatial relationships, surface properties, etc. is unlikely to be based entirely on genetically determined competences since nothing like car-parks existed to influence the selection of our distant ancestors. So there must be processes of learning, still to be studied. I suspect it will require massively parallel constraint propagation mechanisms operating at different levels of abstraction, on different time scales. It will also involve perception operating in parallel at different levels of abstraction, with different levels in the perceptual systems connecting to different levels in central systems, as indicated in the CogAff perception tower.

12.12.6 Implications of speed of human and animal visual perception

Some experiments help to demonstrate familiar but not often noticed features of human visual competence: the speed at which very high level percepts seem to be constructed even when there are no expectations about what the next scene will be. For example, see the demonstration in [18].

Considerations of the sort presented there led to the conjecture that the architecture of a human-like visual system includes a multilayer collection of dynamical systems linked by a constraint propagation network, with different layers operating in parallel at different speeds, performing different sorts of tasks, some of them representing only sensorimotor ontologies whereas others, more remote from the sensorimotor interface can also represent ontologies referring to unperceivable, relatively inaccessible parts of the universe, including future actions and events. These ideas are discussed further in [23] See Figure 9.8.

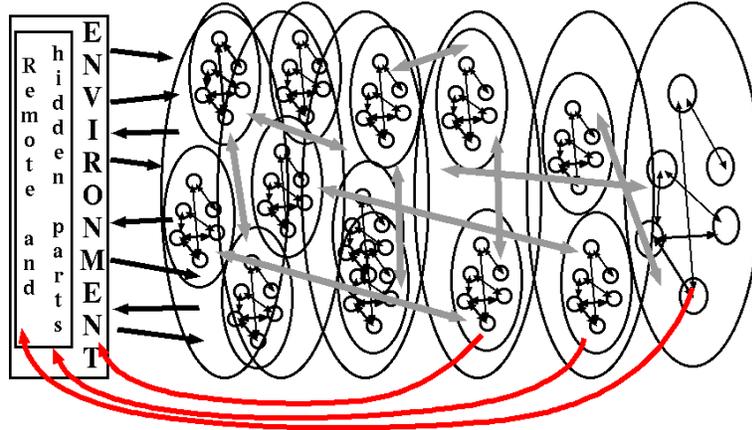


Fig. 12.8. A perceptual system composed of multiple dynamical systems linked in a constraint network, some of them dormant while others operate concurrently more and less remote from the sensorimotor interface, with more remote sub-systems able to refer to un-sensed aspects of reality. Some involve continuous dynamics, others discrete changes. The various perceptual (e.g. visual) sub-systems would also be connected with more central processing sub-systems and in some cases also with action sub-systems, e.g. for reflexes to work. Dynamical systems further from the sensorimotor interface can use semantic contents referring to things in the environment that are more remote from the sensorimotor interface, as indicated by the long arrows.

12.13 Learning to be a mathematician

The ability to see proto-affordances and also epistemic affordances has very deep implications. By studying requirements for a robot to be able to cope with novel configurations, we see that some things that at first are learnt as empirical generalisations, can later be regarded as mathematical (i.e. necessary, not empirical) truths, as appears to happen in young children, though this has not been noticed by developmental psychologists as far as I know. An example was mentioned in Section 9.6, namely a child coming to realise that counting a fixed set of objects in different orders must necessarily give the same result, even though initially this was learnt as an empirical generalisation.

There are many unsolved problems about how this transition happens, but it seems to be closely connected with learning about affordances and how they are related to structures of objects in a principled way that is not just empirical. When these matters are fully understood by the child, animal or robot they enable novel problems to be solved by creative reasoning about action and epistemic affordances. This depends on coming to realise that some true generalisations are not just empirical generalisations that might one day be tested. When fully understood they can

be seen to be, in effect, mathematical theorems, even if the learner does not explicitly notice this fact. This topic was discussed, with more examples, in a recent paper [21] and further expanded in this slide presentation: <http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#mkm08>

If we can make progress with modelling this kind of learning, not only will it be a contribution to robotics, and to developmental psychology, it will also favour Kant's philosophy of mathematics [77] over Hume's (and Russell's [78]).

12.13.1 Two kinds of causation

The kind of learning that uses a transition from empirical to non-empirical understanding also supports a notion of causation that is more like Kant's than Hume's conception of causation (which is essentially statistical and is the precursor of modern Bayesian notions of causation). In work done with Chappell a start has been made in using these ideas to analyse kinds of causal competence in other animals as well as in humans competences.¹⁰ It is very likely that future robots of many kinds will need that kind of causal understanding. However, our robots were not confronted with problems requiring the ability to think about causation.

This discussion points to a need for a form of learning that is very different from the heavily Bayesian (probabilistic/statistics-based) forms of learning that currently attract the most attention. A possible initial mechanism for this would be to allow some features of what has been learnt empirically to trigger a change in the way structures or processes in the environment are represented – e.g. a change from lots of sensorimotor conditional probabilities to representing 3-D objects moving around in a locally euclidean space. That form of representation of processes will have strong implications for what is and is not possible. If the distinctions between kinds of material are included in the representations (e.g. some things are impenetrable others not, some are rigid, others not) then properties of matter can play a role in some of the reasoning. For example if one end of a rigid rod is rotated in a plane then the far end *must* move in a circular arc. If one of two meshed gear wheels made of rigid impenetrable material is rotated, the other *must* rotate in the opposite direction. It is often thought that there are only two ways a young child or animal can discover useful affordances, namely either by empirical trial and error, or by learning from what someone else does (through imitation or instruction). However, our discussion shows that there is a third way, namely by *working out* the consequences of combining spatial processes in advance of their occurrence. This point seems to be missed by many developmental psychologists, e.g. [41].

Some of these implicit theories with strong implications may have been pre-programmed genetically in some animals, as a precocial or pre-configured

¹⁰ See the slide presentations

<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#wonac>

competence. In others, a genetically pre-programmed tendency to perform experiments, discover generalisations, and then build a new layer of theory to make sense of the results may reach similar competences more indirectly and more slowly, but with more scope for subsequent modification. (An example of a nature-nurture trade-off.) It seems that a new born human is already predisposed to try to interpret perceived structures and processes as inhabiting a 3-D space, as might an “altricial” robot be. However, various observations could trigger a process of abduction leading to an enriched explanatory theory about the nature of the environment, e.g. allowing that not only are there movable objects, but some are rigid and some are impenetrable. These notions would be understood in such a way as to disallow representations of certain things bending, being dented, breaking, etc. and also disallowing the representation of part of one object passing through another. (Compare McCarthy on “The well designed child” [50].)

More subtle theories would need to be developed by a child to allow it to learn facts about the nature of mappings between two structures or two processes that allow the discovery that it is not just an *empirical* fact that counting a row of objects left to right produces the same result as counting the same row right to left. Contrast the theories of Rips et al. [79]

There is still much work to be done on the architectural and representational underpinnings for these layered processes of learning and theory construction. Originally it was hoped that some such forms of discovery (e.g. in relation to counting) could be made by a version of PlayMate that was able to point at a set of objects in sequence and ask questions about what it had done.¹¹ But so far we do not have vision systems, episodic memory or action sub-systems capable of being used for this sort of task.

12.14 Confusions about the role of embodiment

What’s important about embodiment (e.g. what drove the most significant evolutionary developments in primate and bird cognition) seems to have been the need to be able to perceive and interact with 3-D structures and processes (including manipulating, assembling and disassembling 3-D structures) and the need to be able to think about spatially located events, processes and entities in the past, remote spatial regions, and the future.

In contrast, much of the work on embodied cognition in robots, and much of the philosophical concern with the importance of embodiment has focused on the terribly narrow problem of learning about sensorimotor relationships. (There are some exceptions.) For a detailed critique of these ideas see [25].

The single most important reason why embodiment influences cognition in humans is that we are part of a very complex 4-D world that extends way

¹¹ As proposed in

<http://www.cs.bham.ac.uk/research/projects/cosy/PlayMate-start.html>

beyond what we can experience or interact with at any time, but which we can think about, plan about, learn about, ask questions about, find or construct routes to, use to explain what we perceive, build and test theories about, etc. People born blind, or without limbs (like Alison Lapper, the artist), or with four legs instead of two, or born as conjoined twins (two heads sharing a torso and legs) can develop those human cognitive competences.

That contrasts with disembodied AI systems that interact only with and think only about, some abstract information structure, such as a financial database, the internet, mathematical proofs, or a board game whose physical implementation is irrelevant. They don't need to be embodied or even to know anything about the 4-D environment and where they are in it or which of its occupants could affect them.

12.15 Developing the revolution in philosophy

Work in progress, partly inspired by the above work related to CoSy shows that old ideas about conceptual analysis in philosophy (the study of “logical geography” in Ryle’s terminology [80]) needs to be reconsidered as a special subset of a larger task: investigation of complex aspects of reality that generate a rich “logical topography” that can be divided up in different ways, supporting different “logical geographies”. A long paper explaining these ideas is under development [81].

It is not common for an AI project to be promoted as a contribution to philosophy, although there have been philosophers involved in AI projects (e.g. Daniel Dennett in COG, Bruce Buchanan in Dendral, Selmer Bringsjord, John Pollock, etc.) and at least one philosopher, Margaret Boden, has made major contributions to the history and philosophy of AI.

My concern was not merely to contribute to philosophy, but also to clarify problems in biology, psychology and eventually brain science.

There have been many AI projects involving philosophers, though usually as collaborators helping with engineering goals. Some philosophers interested in technical philosophical issues, for example, issues concerned with how scientific theories relate to evidence, or issues concerned with how a modal logic could be used in reasoning about permissions and obligations have tried using AI languages, tools and techniques to develop and test their philosophical theories (Herbert Simon, Paul Thagard).

A recent Edinburgh PhD thesis by Alison Pease took the philosophical analysis of the history of Euler’s theorem by Lakatos as the basis for a model of mathematical exploration.¹²

In some cases a particular philosophical viewpoint, for example a philosophical theory about meaning (e.g. symbol-grounding theory), or a theory concerning the nature of emotions, or of consciousness, or of the importance

¹² Available at <http://homepages.inf.ed.ac.uk/apease/research/phd.html>

of embodiment, has influenced the design of a working AI system, especially, in recent years a flood of work influenced by the bad philosophy of symbol-grounding, (criticised in Section 9.5).

If the claims in Section 9.13 about the processes of transforming empirical discoveries to something like mathematical theorems can be substantiated and modelled, this will have deep significance for several aspects of philosophy, including philosophy of mind, philosophy of mathematics, philosophy of causation, and philosophical questions about evolution.

12.16 Further documentation on these ideas

For anyone interested in finding out the extent of the impact of CoSy on thinking about the themes presented here there are three sources of further information (still growing):

- The online repository of papers, discussion notes and presentations at the Birmingham CoSy site:
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/>
- The collection of online presentations at seminars, workshops and conferences, available here:
<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/>
- The disorganised collection of html discussion notes in this web site
<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/>

12.17 Why other disciplines need AI

Very often philosophers and psychologists who attempt to think about possible cases in order to express their theories lack the experience of designing, implementing, testing and debugging working systems, so that they use levels of description that no engineer could take as a specification for a working system: the verbal descriptions used (sometimes with accompanying diagrams) are so non-specific as either to determine no possible working implementation, or as to determine very many different implementations with very different properties that the original proposers would never have considered.

One of the ways of changing this is for AI researchers to involve more people from other disciplines both in the detailed analysis of sets of requirements, and also in the processes of designing, implementing and testing, so that the depth and precision of future theories in the other disciplines can be improved over time.

12.18 Conclusion: The future

Work on the follow on EU-funded CogX project (2008-2012)¹³ will provide opportunities to develop a subset of the ideas presented here, though limitations of AI technology will remain a constraining factor for some time to come.

In parallel with that a proposal is under development for a collaborative project, with two biologists, Jackie Chappell and Susannah Thorpe, to investigate some of the cognitive competences displayed by orangutans moving through trees. Unlike most other animals, including other apes, Sumatran orangutans are able to use the compliance of branches intelligently to overcome difficulties caused by large gaps and the inability of some of the branches to support their weight. These problems and the achievements of the animals will be analysed from the standpoint of a robot designer and related to controversies about the evolutionary origins of ape intelligence, in which too much weight is sometimes given to social requirements.

Attempts will also be made to use some of the lessons learnt during the CoSy project, mentioned in Section 9.13, in collaborative research with developmental psychologists (yet to be identified) investigating the role in young children of the ability to transform empirical generalisations into proto-mathematical knowledge that can be used with confidence in solving novel problems. A start is being made by developing a collection of “toddler theorems”.¹⁴

References

1. J. Chappell, A. Sloman, [Natural and artificial meta-configured altricial information-processing systems](#), *International Journal of Unconventional Computing* 3 (3) (2007) 211–239,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609>.
URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/ijuc.pdf>
2. A. Sloman, [How to Put the Pieces of AI Together Again](#), Tech. Rep. COSY-TR-0608, University of Birmingham, School of Computer Science, poster summary for AAAI’06 Members Poster Session, Boston July 2006. 2-Page abstract at
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0608> Poster at
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0603> (2006).
URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/aaai06-member.pdf>

¹³ Described in <http://www.cs.bham.ac.uk/research/projects/cogx/>

¹⁴ See this recent workshop paper [21] and two related online slide presentations:
<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#math-robot>
<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#toddlers>

3. A. Sloman, [Introduction to Symposium GC5: Architecture of Brain and Mind Integrating high level cognitive processes with brain mechanisms and functions in a working robot](#), in: Proceedings of the AISB '06 Adaptation in Artificial and Biological Systems, Bristol, 2006,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0602>.
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/sloman-aisb06-gc5-intro.pdf>
4. A. Sloman, [Putting the Pieces Together Again](#), in: R. Sun (Ed.), Cambridge Handbook on Computational Psychology, Cambridge University Press, New York, 2008, Ch. 26, pp. 684–709,
<http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#710>.
 URL <http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-sunbook.pdf>
5. A. Sloman, [The Well-Designed Young Mathematician](#), Artificial Intelligence 172 (18) (2008) 2015–2034,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0807>.
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/sloman-aij-08.pdf>
6. A. Sloman, J. Chappell, [The Altricial-Precocial Spectrum for Robots](#), in: Proceedings IJCAI'05, IJCAI, Edinburgh, 2005, pp. 1187–1192,
<http://www.cs.bham.ac.uk/research/cogaff/05.html#200502>.
 URL <http://www.cs.bham.ac.uk/research/projects/cogaff/alt-prec-ijcai05.pdf>
7. A. Sloman, B. Schiele (Eds.), [Tutorial on Learning and Representation in Animals and Robots](#), IJCAI'05, Edinburgh, 2005,
<http://www.cs.bham.ac.uk/research/projects/cosy/conferences>.
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/conferences/>
8. A. Sloman, [Spatial prepositions as higher order functions: And implications of Grice's theory for evolution of language.](#), Research Note COSY-DP-0605, School of Computer Science, University of Birmingham, Birmingham, UK (2005).
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/spatial-prepositions.html>
9. A. Sloman, [Polyflaps as a domain for perceiving, acting and learning in a 3-D world](#), in: Position Papers for 2006 AAAI Fellows Symposium, AAAI, Menlo Park, CA, 2006, <http://www.aaai.org/Fellows/fellows.php> and <http://www.aaai.org/Fellows/Papers/Fellows16.pdf>.
 URL <http://www.aaai.org/Fellows/Papers/Fellows16.pdf>
10. A. Sloman, [Requirements for a Fully Deliberative Architecture \(Or component of an architecture\)](#), Research Note COSY-DP-0604, School of Computer Science, University of Birmingham, Birmingham, UK,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0604> (May 2006).
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/fully-deliberative.html>
11. A. Sloman, [Sensorimotor vs objective contingencies](#), Research Note COSY-DP-0603, School of Computer Science, University of Birmingham, Birmingham, UK,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0603> (May

- 2006).
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/sensorimotor.html>
12. A. Sloman, *Diversity of Developmental Trajectories in Natural and Artificial Intelligence*, in: C. T. Morrison, T. T. Oates (Eds.), *Computational Approaches to Representation Change during Learning and Development*. AAAI Fall Symposium 2007, Technical Report FS-07-03, AAAI Press, Menlo Park, CA, 2007, pp. 70–79,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0704>.
 URL <http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-aaai-representation.pdf>
 13. A. Sloman, *Predicting Affordance Changes (Alternatives ways to deal with uncertainty)*, Tech. Rep. COSY-DP-0702, School of Computer Science, University of Birmingham, unpublished discussion paper
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0702> (HTML) (Nov 2007).
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/changing-affordances.html>
 14. A. Sloman, *Requirements for Digital Companions: It's harder than you think*, position Paper for Workshop on Artificial Companions in Society: Perspectives on the Present and Future Organised by the Companions project. Oxford Internet Institute (25th–26th October, 2007)
<http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#711> (October 2007).
 URL <http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-oi-2007.pdf>
 15. A. Sloman, *What evolved first and develops first in children: Languages for communicating? or Languages for thinking? (Generalised Languages: GLs)*, presentation given to Birmingham Psychology department.
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0702> (2007).
 URL <http://www.cs.bham.ac.uk/research/projects/cogaff/talks/glang-evo-ai1.pdf>
 16. A. Sloman, *Why Some Machines May Need Qualia and How They Can Have Them: Including a Demanding New Turing Test for Robot Philosophers*, in: A. Chella, R. Manzotti (Eds.), *AI and Consciousness: Theoretical Foundations and Current Approaches AAAI Fall Symposium 2007*, Technical Report FS-07-01, AAAI Press, Menlo Park, CA, 2007, pp. 9–16,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0705>.
 URL <http://www.cs.bham.ac.uk/research/projects/cogaff/sloman-aaai-consciousness.pdf>
 17. A. Sloman, *Why symbol-grounding is both impossible and unnecessary, and why theory-tethering is more powerful anyway.*,
<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#models> (2007).
 URL <http://www.cs.bham.ac.uk/research/projects/cogaff/talks/models.pdf>
 18. A. Sloman, *A Multi-picture Challenge for Theories of Vision* (2008).
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/presentations/multipic-challenge.pdf>

19. A. Sloman, *A New Approach to Philosophy of Mathematics: Design a young explorer, able to discover “toddler theorems”*, online presentation <http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#toddler> (2008).
URL <http://www.cs.bham.ac.uk/research/projects/cogaff/talks/sloman-toddler1.pdf>
20. A. Sloman, *Architectural and representational requirements for seeing processes, proto-affordances and affordances*, in: A. G. Cohn, D. C. Hogg, R. Möller, B. Neumann (Eds.), *Logic and Probability for Scene Interpretation*, no. 08091 in *Dagstuhl Seminar Proceedings*, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, Dagstuhl, Germany, 2008.
URL <http://drops.dagstuhl.de/opus/volltexte/2008/1656/pdf/08091.SlomanAron.Paper.1656.pdf>
21. A. Sloman, *Kantian Philosophy of Mathematics and Young Robots*, in: S. Autexier, J. Campbell, J. Rubio, V. Sorge, M. Suzuki, F. Wiedijk (Eds.), *Intelligent Computer Mathematics*, LNCS no 5144, Springer, Berlin/Heidelberg, 2008, pp. 558–573,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers#tr0802>.
URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/math-ai-sloman.pdf>
22. A. Sloman, *Varieties of Meta-cognition in Natural and Artificial Systems*, in: M. T. Cox, A. Raja (Eds.), *Workshop on Metareasoning*, AAAI’08 Conference, AAAI Press, Menlo Park, CA, 2008, pp. 12–20,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0803>.
URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/sloman-meta-aaai08.pdf>
23. A. Sloman, *Architectural and representational requirements for seeing processes and affordances*, in: *Computational Modelling in Behavioural Neuroscience: Closing the gap between neurophysiology and behaviour.*, Psychology Press, London, 2009.
URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/sloman-newmod.pdf>
24. A. Sloman, *Machines in the Ghost*, in: D. Dietrich, G. Fodor, G. Zucker, D. Bruckner (Eds.), *Simulating the Mind: A Technical Neuropsychanalytical Approach*, Springer, Vienna & New York, 2009, pp. 124–148,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0702>.
URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/sloman-enf07.pdf>
25. A. Sloman, *Some Requirements for Human-like Robots: Why the recent over-emphasis on embodiment has held up progress*, in: B. Sendhoff, E. Koerner, O. Sporns, H. Ritter, K. Doya (Eds.), *Creating Brain-like Intelligence*, Springer-Verlag, Berlin, 2009, pp. 248–277,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0804>.
URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/sloman-honda.pdf>
26. A. Sloman, J. Chappell, *Altricial self-organising information-processing systems*, *AISB Quarterly* (121) (2005) 5–7,
<http://www.cs.bham.ac.uk/research/cogaff/05.html#200503>.
URL <http://www.cs.bham.ac.uk/research/projects/cogaff/summary-gc7.pdf>

27. A. Sloman, J. Chappell, [Computational Cognitive Epigenetics \(Commentary on Jablonka and Lamb: Evolution in Four Dimensions \(2005\)\)](#), Behavioral and Brain Sciences 30 (4) (2007) 375–6,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0703>.
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/jablonka-sloman-chappell.pdf>
28. A. Sloman, J. Chappell, T. C. Team, [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, Internet, 2006, <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/>.
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/presentations/assc10-poster.pdf>
29. A. Sloman, J. Chappell, the CoSy PlayMate team, [Orthogonal Recombinable Competences Acquired by Altricial Species \(Blankets, string, and plywood\)](#), Research Note COSY-DP-0601, School of Computer Science, University of Birmingham, Birmingham, UK,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers#dp0601> (January 2006).
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/orthogonal-competences.html>
30. A. Sloman, Cosy-partners, [CoSy deliverable DR.2.1 Requirements study for representations](#), Tech. Rep. COSY-TR-0507, The University of Birmingham, UK, <http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0507> (2005).
 URL <http://cognitivesystems.org/files/dr-02-01-rev1.pdf>
31. A. Sloman, B. C. P. Team, J. Chappell, [Poster: Acquiring Orthogonal Recombinable Competences](#), in: H. Bekkering (Ed.), Proceedings CogSys-II, Radboud University Nijmegen, NL, 2006,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0601>,
 Conference url: <http://www.socsci.ru.nl/CogSys2>.
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/presentations/cogsys2-poster.pdf>
32. A. Sloman, J. Wyatt, N. Hawes, J. Chappell, G.-J. M. Kruijff, [Long Term Requirements for Cognitive Robotics](#), in: Cognitive Robotics: Papers from the 2006 AAAI Workshop: Technical Report WS-06-03,
<http://www.aaai.org/Library/Workshops/ws06-03.php>, AAAI Press, Menlo Park, CA, 2006, pp. 143–150,
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0604>.
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/cogrob06.pdf>
33. A. Sloman, [The SimAgent TOOLKIT – for Philosophers and Engineers \(And Some Biologists, Psychologists and Social Scientists\)](#) (1996).
 URL <http://www.cs.bham.ac.uk/research/projects/poplog/packages/simagent.html>
34. A. Sloman, B. Logan, [Building cognitively rich agents using the Sim_agent toolkit](#), Communications of the Association for Computing Machinery 42 (3) (1999) 71–77.

- URL
<http://www.cs.bham.ac.uk/research/projects/cogaff/96-9%9.html#49>
35. J. Fodor, *The Language of Thought*, Harvard University Press, Cambridge, 1975.
 36. A. Sloman, *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, Aslib, London, 1979, pp. 1–15.
 URL
<http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#43>
 37. E. Jablonka, M. J. Lamb, *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*, MIT Press, Cambridge MA, 2005.
 38. F. Warneken, M. Tomasello, *Altruistic helping in human infants and young chimpanzees*, *Science* (2006) 1301–1303 DOI:10.1126/science.1121448.
 39. A. Sloman, *The Computer Revolution in Philosophy*, Harvester Press (and Humanities Press), Hassocks, Sussex, 1978.
 URL <http://www.cs.bham.ac.uk/research/cogaff/crp>
 40. A. P. Ambler, H. G. Barrow, C. M. Brown, R. M. Burstall, R. J. Popplestone, *A Versatile Computer-Controlled Assembly System*, in: *Proc. Third Int. Joint Conf. on AI*, Stanford, California, 1973, pp. 298–307.
 41. E. J. Gibson, A. D. Pick, *An Ecological Approach to Perceptual Learning and Development*, Oxford University Press, New York, 2000.
 42. A. Sloman, *A First Draft Analysis of some Meta-Requirements for Cognitive Systems in Robots*, contribution to euCognition wiki, also available as, <http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0701> (2007).
 URL <http://www.cs.bham.ac.uk/research/projects/cosy/papers/meta-requirements.html>
 43. R. Carnap, *Meaning and necessity: a study in semantics and modal logic*, Chicago University Press, Chicago, 1947.
 44. A. Sloman, *The structure of the space of possible minds*, in: S. Torrance (Ed.), *The Mind and the Machine: philosophical aspects of Artificial Intelligence*, Ellis Horwood, Chichester, 1984.
 URL <http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#704>
 45. A. Sloman, *Explorations in design space*, in: A. Cohn (Ed.), *Proceedings 11th European Conference on AI*, Amsterdam, August 1994, John Wiley, Chichester, 1994, pp. 578–582.
 46. A. Sloman, *Interacting trajectories in design space and niche space: A philosopher speculates about evolution*, in: *et al.* M.Schoenauer (Ed.), *Parallel Problem Solving from Nature – PPSN VI, Lecture Notes in Computer Science*, No 1917, Springer-Verlag, Berlin, 2000, pp. 3–16.
 URL
<http://www.cs.bham.ac.uk/research/projects/cogaff/00-02.html#62>
 47. M. Scheutz, *The evolution of simple affective states in multi-agent environments*, in: D. Cañamero (Ed.), *Proceedings AAAI Fall Symposium 01*, AAAI Press, Falmouth, MA, 2001, pp. 123–128.
 48. U. Neisser, *Cognitive Psychology*, Appleton-Century-Crofts, New York, 1967.
 49. J. J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, MA, 1979.

50. J. McCarthy, [The Well-Designed Child](#), *Artificial Intelligence* 172 (18) (2008) 2003–2014.
URL <http://www-formal.stanford.edu/jmc/child.html>
51. A. Sloman, [Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence](#), in: *Proc 2nd IJCAI*, William Kaufmann, London, 1971, pp. 209–226.
URL <http://www.cs.bham.ac.uk/research/cogaff/04.html#200407>
52. Z. Pylyshyn, [Is vision continuous with cognition? The case for Cognitive impenetrability of visual perception.](#), *Behavioral and Brain Sciences* 22 (3) (1999) 341–423.
URL <http://rucss.rutgers.edu/ftp/pub/papers/ZPbbs98.pdf>
53. A. Trehub, [The Cognitive Brain](#), MIT Press, Cambridge, MA, 1991.
URL <http://www.people.umass.edu/trehub/>
54. P. Langley, J. Laird, [Cognitive architectures: Research issues and challenges](#), Tech. rep., Institute for the Study of Learning and Expertise, Palo Alto, CA., <http://csl.stanford.edu/~langley/papers/final.arch.pdf> (2006).
55. R. Milner, [The Polyadic pi-Calculus: A Tutorial](#), Tech. Rep. LFCS report ECS-LFCS-91-180, The University of Edinburgh.
URL <http://www.lfcs.inf.ed.ac.uk/reports/91/ECS-LFCS-91-180/>
56. C. Hewitt, P. Bishop, R. Steiger, [A Universal Modular ACTOR Formalism for Artificial Intelligence](#), in: *Proc. Third Int. Joint Conf. on AI*, 1973.
57. N. Nilsson, [Artificial Intelligence: A New Synthesis](#), Morgan Kaufmann, San Francisco, 1998.
58. J. Gibson, [The Senses Considered as Perceptual Systems](#), Houghton Mifflin, Boston, 1966.
59. R. Brooks, [A robust layered control system for a mobile robot](#), *IEEE Journal of Robotics and Automation* RA-2 (1986) 14–23, 1.
60. E. Gat, [On three layer architectures](#), in: D. Kortenkamp, R. P. Bonasso, R. Murphey (Eds.), *Artificial Intelligence and Mobile Robots*, AAAI Press, 1998.
61. M. L. Minsky, [The Emotion Machine](#), Pantheon, New York, 2006.
62. A. Sloman, [The Cognition and Affect Project: Architectures, Architecture-Schemas, And The New Science of Mind.](#), Tech. rep., School of Computer Science, University of Birmingham (2003).
URL <http://www.cs.bham.ac.uk/research/projects/cogaff/03.html#200307>
63. R. Sun, [The CLARION cognitive architecture: Extending cognitive modeling to social simulation](#), in: R. Sun (Ed.), *Cognition and Multi-Agent Interaction*, Cambridge University Press, New York, 2006, pp. 79–99,
<http://www.cogsci.rpi.edu/~rsun/sun.clarion2005.pdf>.
64. J. Albus, [Brains, Behaviour and Robotics](#), Byte Books, McGraw Hill, Peterborough, N.H., 1981.
65. R. Cooper, T. Shallice, [Contention scheduling and the control of routine activities](#), *Cognitive Neuropsychology* 17 (4) (2000) 297–338.
66. N. Block, [On a confusion about the function of consciousness](#), *Behavioral and Brain Sciences* 18 (1995) 227–47.
67. D. Dennett, [Kinds of minds: towards an understanding of consciousness](#), Weidenfeld and Nicholson, London, 1996.