

# Some Requirements for Human-like Robots: Why The Recent Over-emphasis on Embodiment has Held up Progress\*

Aaron Sloman

School of Computer Science, University of Birmingham, B15 2TT, UK.

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

December 23, 2008

## Abstract

Some issues concerning requirements for architectures, mechanisms, ontologies and forms of representation in intelligent human-like or animal-like robots are discussed. The tautology that a robot that acts and perceives in the world must be embodied is often combined with false premises, such as the premiss that a particular type of body is a requirement for intelligence, or for human intelligence, or the premiss that all cognition is concerned with sensorimotor interactions, or the premiss that all cognition is implemented in dynamical systems closely coupled with sensors and effectors. It is time to step back and ask what robotic research in the past decade has been ignoring. I shall try to identify some major research gaps by a combination of assembling requirements that have been largely ignored and design ideas that have not been investigated – partly because at present it is too difficult to make significant progress on those problems with physical robots, as too many different problems need to be solved simultaneously. In particular, the importance of studying some abstract features of the environment about which the animal or robot has to learn (extending ideas of J.J.Gibson) has not been widely appreciated.

## 1 Introduction

I shall discuss a collection of competences of humans and other animals that appear to develop over time under the control of both the environment and successive layers of learning capabilities that build on previously learned capabilities. In the process I shall point out some important limitations in current ideas about embodiment and dynamical systems. To set the scene, I use a well known and influential paper by Rodney Brooks (1990), that expresses very clearly both views that are still widely held that I think are seriously mistaken, and also some important truths. However, some of those influenced by Brooks have been more mistaken than he was. I also believe he has changed his position since writing it.

There is certainly some truth in the claim made by Brooks and others that early work on symbolic AI ignored important issues relating to design of robots able to act in the world continuously in real time (also stressed in [32, 3, 34]). However it is usually forgotten that in early work on AI, including the SRI robot Shakey<sup>1</sup> and the Edinburgh robot Freddy [1], the available computing power was so limited (e.g. with CPU speeds measured in at most kilocycles per second and computer memories in at most kilobytes) as to rule out any real-time interaction. For example it could take 20 minutes for an AI visual system to find the upper

---

\*This is a preprint version of a paper to appear in B. Sendhoff et al. (Eds.): *Creating Brain-Like Intelligence*, LNAI 5436, pp. 248-277, 2009.

©Springer-Verlag Berlin Heidelberg 2009

<sup>1</sup><http://www.sri.com/about/timeline/shakey.html>

rim of a teacup. This meant that the only option for AI researchers with long term goals was to adopt an extremely slow and artificial, sense-think-act cycle for their experiments. Unfortunately despite its flaws, noted below, that cyclic 3-stage model has persisted with variations (e.g. sense-compute-act) even in some systems that reject symbolic AI methods.

Each new AI fashion criticises previous fashions for failing to achieve their predicted successes, alleging that they used the wrong mechanisms. This is a serious error. The main reason why predictions were wrong was that the *problems* were not understood, not that the wrong *mechanisms* were adopted. The problems of replicating human and animal competences, especially the problems of seeing and interacting with a structured 3-D environment, are very complex in ways that we are only now beginning to understand. Had the problems been understood in the early days of AI the over-optimistic predictions would not have been made. None of the tools and forms of representation so far proposed could have enabled the targets of AI to be met on the predicted time-scales. Moreover, none of them is adequate to the long term task, since, as various researchers have pointed out (e.g. Minsky in [20]), human-like robots will need information-processing architectures that combine different mechanisms running concurrently, doing different sorts of tasks, using different forms of information, processed in different ways.<sup>2</sup>

But before we can achieve that, we shall need to do a deeper and more complete *requirements analysis* in order to understand the problems that need to be solved.

## 2 The Seduction of Embodiment

One of the fashions that has immense plausibility for beginners (especially those not trained in philosophy) and has influenced large numbers of researchers over the last decade or two, has a number of aspects. It emphasises the fact that animals are embodied, claims that animal intelligence depends on the detailed morphology, including the sensors and effectors available, denies the need for symbolic representations and mechanisms for doing planning or logical reasoning, relies heavily on information being constantly available in the environment, emphasises the importance of the brain as a dynamical system linked to the environment through sensors and motors and in some cases makes heavy use of statistical learning and probabilistic control mechanisms. A good example is [18].

In working systems that allow machines to manipulate internal states with semantic contents it is common within this fashion to stress the importance of “symbol grounding” [11], i.e. the derivation of all semantic content from the contents of sensor signals and the patterns of association between sensory and motor signals. The idea of symbol grounding is closely related to the old philosophical theory of concept empiricism, discredited long ago by the work of Kant and finally buried by 20th century philosophers of science considering the role of theoretical terms in science (e.g. “electron”, “gene”, “valence”, etc.) that are primarily defined by their roles in explanatory theories.<sup>3</sup>

In his 1990 paper, Brooks presented the key ideas of so-called “nouvelle AI”, emphasising embodiment and sensory-motor interactions with the environment, including the proposal to

---

<sup>2</sup>I use the word “representation” to refer to whatever is used to encode information. It could be some physical structure or process, or a structure or process in a virtual machine. It could be transient or enduring. It may be used for a single function or for many functions. See also

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/whats-information.html>

<sup>3</sup>As explained in this tutorial: <http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#models>

dispense with symbolic representations, for instance in perceiving, planning, reasoning and learning, because

*“... the world is its own best model. It is always exactly up to date. It always contains every detail there is to be known. The trick is to sense it appropriately and often enough. ... The constructed system eventually has to express all its goals and desires as physical action, and must extract all its knowledge from physical sensors. ... We explore a research methodology which emphasises ongoing physical interaction with the environment as the primary source of constraint on the design of intelligent systems.”*

The claim was that once this point was understood and systems were built using a hierarchy of layers of sensorimotor control, everything would turn out to be “*rather simple once the essence of being and reacting are available*”.

Part of the argument was that it took far longer for evolution to reach the stage just before the first mammals evolved than to achieve all the subsequent developments. From this, it was inferred that once we had produced insect-like systems everything else would be “rather simple”, a view parodied in the title of Kirsch (1991).

Of course, Brooks was not the only defender of the emphasis on embodiment and the rejection of symbolic AI architectures and mechanisms.<sup>4</sup> It is not clear that he adopted the “symbol-grounding” approach and the closely related assumption that all a robot’s knowledge must be expressed in terms of sensorimotor contingencies, though other supporters of nouvelle AI did that explicitly or implicitly, e.g. by restricting the learning of robots to discovery of statistical patterns relating sensory and motor signals. Many of them now present robots performing tasks where much use is made of the morphology of the robot to reduce the software sophistication required. An extreme example of this approach that I have not yet seen presented but would illustrate the ideas well would be the claim that a tennis ball is an elegant design for a kind of robot for going down helter-skelters. The extensions required for climbing to the top of the helter-skelter unaided will take a little longer.

## 2.1 Morphology Can Be Important, For Some Tasks

I am not denying that choice of physical morphology can considerably simplify software design in some contexts. One of the oldest examples is the use of a “compliant wrist” to simplify problems of grasping and manipulation of 3-D objects e.g. in [7].

Some of the recent work on embodied robots has been very impressive. One of the most advanced is the BigDog robot designed by Marc Raibert and colleagues at Boston Dynamics<sup>5</sup> capable of being a load-bearer to help humans in difficult and varied terrain. The Honda Asimo, a humanoid robot is able to perform a number of non-trivial actions.<sup>6</sup> Brooks’

---

<sup>4</sup>The paper seems to me to be ambivalent between two positions: (a) wholesale rejection of symbolic AI as irrelevant, and (b) recommending that research on symbolic methods be *temporarily* shelved until after the techniques of nouvelle AI are well developed. Whatever his real views, many readers and followers in many countries interpreted Brooks’ work as demonstrating that symbolic AI is totally irrelevant, and one result of that is that many students who learn about AI are now taught nothing about AI techniques for planning, parsing, symbolic reasoning, logical inference, learning and reasoning by analogy, etc. though they may learn a lot of mathematics related to probabilistic representation and reasoning. Compare this comment by Cliff in [6]: “Most work in AI prior to the mid-1980’s is largely of historical interest, and will be discussed no further.”

<sup>5</sup>Shown in online videos here: <http://www.bostondynamics.com/>

<sup>6</sup>Illustrated at <http://world.honda.com/ASIMO/>

company iRobot makes and sells many robots.<sup>7</sup> The Shadow robot hand<sup>8</sup> has been used in demonstrations showing how the physical properties of the hand contribute to control functions.

There is no evidence, however, that robots based *solely* on the principles of nouvelle AI will be able to perform all the demanding tasks for which human-like robots are often expected to be needed, including, for example, being a helpful companion to a disabled or blind person. Being able to select a suitable place to cross a busy road requires a vision system that can tell where the nearby pedestrian crossings are, select one, and plan a route to get from the current location to the required point on the other side of the road via the crossing. This functionality should cope with foreign towns where traffic lights and pedestrian crossings have different visual features. All of that could be avoided by having roadsides wired with hidden cables guiding robots to crossing points, but a robot that required such guidance would not be a human-like robot. Apart from the cameras and motors for moving them, parts of the body of robot (in contrast with information about parts of the body) are not needed for the process of using visual perception and prior knowledge to work out a sensible route, although the whole body is of course relevant to carrying out the plan.

To be fair to Brooks, since early 2002, I have heard him publicly acknowledge in conference presentations and discussions that the ideas of nouvelle AI are insufficient, since a mixture of old and new AI techniques is needed for progress in robotics. Some early strident proponents of embodiment and how much you can get “for free” by suitable physical design of robots have also recently adopted a more muted tone, while still failing to mention that intelligent robots may have to plan future actions, reason about unobserved events, or represent mental states of others, e.g. [24]. The issues are discussed in more detail in the special issue of *Cognitive Systems Research* on embodiment [54].

### 3 Fallacies in Nouvelle AI

Unfortunately there remain many believers in nouvelle AI, whether they use that label or not, writing books and papers, giving presentations and teaching young and impressionable students. So I shall try to explain the two main fallacies, and present an alternative synthesis. The first fallacy is in the use of the evolutionary argument, and the second amounts to ignoring many of the important cognitive competences of humans and some other animals, including primates and some nest-building birds such as corvids, competences that arise from the complex and diverse structure of the environment.

#### 3.1 The Argument from Evolutionary Timescales

The time required for evolution is not an indication of what would be hard for us. Evolution started from a very primitive state, and among other things, had to create mechanisms for creating mechanisms. It could turn out that the most difficult problem solved by evolution was using the physics and chemistry available on earth to produce any sort of organism that could grow itself, maintain and repair itself, including obtaining nutrients and disposing of waste materials, and in addition reproduce itself (including its behaviours) using mechanisms

---

<sup>7</sup>Illustrated here <http://www.irobot.com/sp.cfm?pageid=109>

<sup>8</sup><http://www.shadowrobot.com/hand/videos.shtml>

that support evolutionary adaptation. Mechanisms supporting autopoiesis are required for all biological organisms except perhaps the very simplest.

The robots produced by “nouvelle AI” do very little of that. They are pre-constructed and typically move around rather pointlessly (possibly doing something useful as a side-effect), or achieve arbitrary goals imposed from outside, as opposed to doing what is required to grow, maintain and reproduce themselves: a goal for robotics that is unlikely to be achieved in the foreseeable future except where quite complex components are provided ready made by human engineers. There are some impressive demonstrations of very narrow competences based on the sensorimotor approach, such as hands that will close round objects of a variety of shapes with which they come into contact, though nothing yet like a robot that can use such a hand to assemble toys from meccano parts, guided by a picture.<sup>9</sup>

### 3.2 The Failure to Identify Deliberative and Metacognitive Requirements

More importantly, the fact that some task is very hard, and that it provides the foundation for animal competences, does not imply that everything else is unimportant and easy to replicate.

It is one thing to build an elephant that can walk and maintain its balance on a hard flat surface. It is quite another to build one that can learn how to negotiate different sorts of terrain, including walking up a slippery muddy slope, can learn to distinguish edible and nutritious matter from other things and how to get to and consume such food (including pushing down a tree to get at the leaves at the top), and in addition can learn the layout of a large terrain and remember where to go at different times of the year, along with learning about the social hierarchy of its group and how to anticipate or interpret actions of other individuals. The immediate environment may be a good source of directly sensed information about the immediate environment, but it is not a source of immediately available information about remote locations, or about consequences of actions that have not yet been performed or about dispositional qualities of materials in the environment that need to be represented in a theory-based conceptual framework.

Elephants may not play chess, solve algebra problems, prove topological theorems, or design jetliners, but nothing follows about what they and other animals can do, or how what they do might be related to human capabilities they lack. They are still far more intelligent than current robots. In the course of evolution, changing physical environments and changing physical designs (and consequent physical needs and competences) in a species can both produce new opportunities and new design requirements in order that those opportunities are made use of. In particular, some niches require symbolic planning capabilities while others do not [41].

Some animals have to be able to work out how to achieve tasks such as nest building, picking berries, or dismembering a carcass after a kill to get at the edible parts, using complex independently movable manipulators. In general that can involve thinking two or more steps ahead in order to choose appropriate actions, sometimes comparing alternative branching futures – unless the required actions have been pre-compiled by evolution or learnt through trial-and-error searching, both potentially very slow processes compared with planning. For some species in some environments, neither evolution nor prior learning can provide a solution for every situation. There is strong evidence that at least one species of spider can plan a new route in advance of following it – the portia spider works out a route to its prey then

---

<sup>9</sup>Like the crane illustrated here

follows it even when it can no longer see the prey, making detours if necessary and avoiding branches that would not lead to the prey [50].<sup>10</sup>

For humans, tasks such as building shelters, tracking down prey, weaving cloth, attacking enemies, preparing defences against attack, designing and building weapons to help with hunting or warfare, and in recent times building many kinds of vehicles, buildings, and machines for making machines, all extend information processing requirements – specifically requirements concerned with representing and choosing between future possible actions and products. A robot with similar abilities would need, along with sensory and motor competences, an architecture supporting “fully deliberative” competences, as explained in [40].

More human-like robots would need to be able to develop an interest in various branches of mathematics, including geometry, topology, and transfinite set theory, and would explore ways of proving theorems in those domains, as discussed in [43]. Some would become interested in philosophical questions, or enjoy composing or reading novels. The notion that intelligence involves constant interaction with the environment ignores most of human nature. I’ll return to the implications of this in discussing varieties of dynamical systems required.

Simpler organisms may cope with “proto-deliberative” reactive mechanisms, in which a combination of sensed state and current goals can activate alternative responses, triggering a competitive mechanism to make a selection, and act accordingly. Some researchers confusingly use the label “deliberative” for this. So I use “fully deliberative” to label the ability to construct several representations of *branching* futures, represent and compare their relative strengths and weaknesses, and then select one as a plan and use it to generate behaviours, while being prepared to learn from mistaken decisions of that sort. Few other animals can do all this, but human-like robots will need to, as explained in [40], though how such competences are implemented is an open question.

## 4 Limitations of Symbolic AI

It was indeed unwise for early AI researchers to assume that after building their systems with planning and problem-solving competences everything else would be easy and that human-like robots could be achieved in a few decades. In contrast, Turing’s prediction in 1950 that a high proportion of the general population could be fooled by a text-based simulation program for about five minutes by the end of the century was far more modest (and I think turned out true).

Nevertheless, the fact that symbolic AI systems do not *suffice* does not mean that such systems are not *required* for human-like intelligence, as Brooks was obviously aware. He wrote (in section 6.1)

*“Thus the two approaches appear somewhat complementary. It is worth addressing the question of whether more power may be gotten by combining the two approaches. However, we will not pursue that question further here.”*

This paper offers reasons why the two approaches need to be combined, in contrast with many of the followers of Brooks’ approach who believe that symbolic AI has failed and can safely be ignored (e.g. Cliff, quoted in Note 4).

---

<sup>10</sup>Tarsitano states: “By visual inspection, they can select, before setting out, which detour routes do and do not lead to prey, and successfully perform a detour with no further visual contact with the prey”.

Of course, I am not claiming that *only* the specific symbolic algorithms developed in the 1960s should be included. The early planning and reasoning systems worked in simple contexts but were not sufficient for planning future actions in all cases since they did not take account of the amount of uncertainty and scope for error in perceptual information in many (though not all) contexts and they did not do justice to the cases where the consequences of performing specific actions could not always be relied on, e.g. because of slippage, unreliable components, effects of wear and tear on mechanisms, and the possibility of external (animate and inanimate) sources of interference with actions. However, the designers of Shakey and Freddy were not stupid and did allow for the possibility of actions not producing intended results. For example, the PLANEX module was available in Shakey to cope with such failures by re-planning, though that is not always a good solution.

## 5 Meta-semantic and Exosomatic Ontologies

Another point missed by some of the nouvelle AI community is that where there are other intelligent agents in the environment whose actions need to be understood, and sometimes obstructed or assisted, or if individuals need to monitor and improve their own thinking and planning procedures, as in Sussman's HACKER [49], then there is a requirement to represent things that themselves represent other things: i.e. meta-semantic competences are required. This is not the place to go into the details of that requirement, which includes coping with referential opacity. Some comments on meta-semantic competences and requirements for supporting them can be found in [44]. Warneken [53] demonstrated behaviours in pre-verbal children and in chimpanzees that indicate some sort of meta-semantic competence insofar as the experimental subjects are spontaneously able to identify goals of others and perform actions to achieve them.

The ability to represent mental states and processes of information manipulation in other individuals is just one example among many of the requirement to acquire, manipulate and use information about things in the environment that cannot be sensed. For example, most of the properties of matter in the environment, including rigidity, impenetrability, compressibility, various kinds of flexibility, various kinds of elasticity, fragility, solubility, edibility, and many more, are not directly sensed but have to be inferred from other things, including how objects composed of such matter respond to various kinds of actions, e.g. lifting, pushing, prodding, twisting, attempts to break or tear, etc. It is commonly thought that colours are properties of light hitting the retina whereas many colour illusions refute that, and support the notion that what we see as colours are properties of surfaces indirectly inferred from sensory information [25].

Animals that make use of such properties of matter in decision-making must have the ability to acquire ontologies that go beyond patterns found in sensorimotor signals. I.e. somatic ontologies may suffice for insect-like robots [18] but will not suffice for more sophisticated animals (such as orangutans that make intelligent use of compliance in arboreal supports) and robots: exosomatic ontologies are needed [41].

Even within the domain of processes that can be sensed, such as visual and haptic sensing of movements of hands, it is arguable that amodal exosomatic ontologies (e.g. referring to structures and processes involving 3-D surfaces and their interactions, rather than the sensory signals produced by such things or the motor signals that can modify them) are far more useful and economical for acquiring and using knowledge about a wide variety of states

and processes, including ones that have not previously been encountered – e.g. perceiving a grasping process viewed from a new angle. For example, using an amodal exosomatic ontology allows generalisations from actions performed using one hand to similar actions performed using the other hand, or using two hands, or even actions performed by other agents (a process often “explained” by almost magical properties of so-called mirror neurones). But that kind of exosomatic ontology would be ruled out by those whose nouvelle AI includes a commitment to symbol-grounding theory.

Even the great mathematician Poincarè [26] apparently fell into the trap of assuming all concepts must be grounded in sensorimotor patterns:

*“But every one knows that this perception of the third dimension reduces to a sense of the effort of accommodation which must be made, and to a sense of the convergence of the two eyes, that must take place in order to perceive an object, distinctly. These are muscular sensations quite different from the visual sensations which have given us the concept of the two first dimensions.”*

I suspect he would have modified his views if he had been involved in designing robots that can perceive and reason about 3-D scenes.

The points just made are closely related to Gibson’s (1979) theory of perception as providing information about positive and negative affordances for action in the environment. However, Gibson’s work was just the beginning, as will be explained below in Section 11.

## 6 Developments Required in a Human-like Robot

The growth of competences in a human, from infancy onwards, includes many different sorts of development, such as the following, all of which appear to be required in a human-like robot – though what mechanisms could produce this combination is still not known.

- Early development includes learning to calibrate and control sensors and effectors. Recalibration, or adaptation, continues throughout life as the body grows, as new goals and new skills impose new requirements, as the environment changes, and as impairments caused by disease or injury produce new constraints.
- After learning to control some movements and acquiring related perceptual competences, infants can use that competence to perform experiments on both the physical environment and nearby older humans, at first using only genetically determined exploratory processes rather than conscious rational exploration [39, 5].
- After new concepts have been acquired they can be used to formulate new theories, in addition to new plans, goals, etc. This is most obvious in the adult developments of scientists and engineers, but I suggest much child development should be viewed in a similar way.
- New forms of representation can be used to formulate new conceptual building blocks, and to formulate and test new plans, goals, hypotheses, etc., sometimes producing new forms of reasoning, problem solving and learning, using those representations. This happens explicitly and overtly to adults (e.g. learning mathematics, physics, genetics, musical theory) and seems to be required to explain some qualitative changes of competence in infancy and childhood.
- Later, learning to observe and modulate learning processes starts, e.g. attempting actions that test and challenge current competences. This seems initially to use genetically determined meta-cognitive mechanisms and competences, later enhanced by cultural influences including games, sporting activities and academic learning.

- Higher level meta-cognitive mechanisms can later begin to monitor, debug and modulate the initial forms, in ways that depend on how the initial competences fare in coping with the environment in increasingly complex ways.
- Representing the individual's own information-processing or the information processing of others requires meta-semantic competence, which in turn requires architectural extensions that allow "encapsulated" representations to be used (e.g. representing the individual's possible future beliefs and goals, or the beliefs and goals of others). Encapsulation is required to prevent hypothetical descriptions being used by mechanisms that use the owner's beliefs, goals, etc.
- All of the above can help drive the development of linguistic and other communicative competences using external, shared, languages.
- Being able to communicate with others makes it possible to learn quickly what other individuals previously learnt more slowly and with greater difficulty, such as discovering a new way to make something.
- Layered learning processes start in infancy, and, in humans, can continue throughout adult life, extending the architecture as well as modifying and extending contents of parts of the architecture.

That list does not refer to results of esoteric laboratory experiments, but mostly summarises examples of human development that are fairly obvious in a reflective culture. Many people are already attempting to explain them.

However, the competences are usually studied separately, in different disciplines, or in different research groups that do not work together, for instance groups studying language learning and groups studying perceptual development or groups attempting to build computer models of such development. So, people who attempt to build working AI models or robots normally consider only a small subset of the competences shown by humans and other animals, and different researchers focus on different subsets. However, it is not obvious that models or explanations that work for a narrowly focused set of tasks can be extended to form part of a more general system: as explained in [38] systems that "scale up" (i.e. do not consume explosively increasing amounts of time or memory as problem complexity grows) do not always "scale out" (i.e. many of them cannot be combined with other competences in an integrated system).

## 7 Morphology and Development

It is sometimes thought that body morphology determines cognitive development. This may be true in the very early stages of human development but as new layers of competence are added, they depend more and more on the nature of the environment, including other intelligent agents (friendly and unfriendly conspecifics, prey, predators, and in future also robots), and less and less on the specific sensors and effectors of the learner. That is in part because many aspects of the environment cannot be directly sensed or acted on, no matter what the morphology, including: (a) past, future and spatially remote events, (b) the "hidden" properties of matter that are only discovered through scientific investigation and (c) hypothetical future processes considered when making plans or predictions.

Consequently, more abstract features of the environment can be learnt about and thought about by individuals with very different body morphologies and sensors. This is presumably why "thalidomide babies" who were born without limbs or with stunted limbs, and children born blind or deaf, or with motor control problems, manage to develop normal human minds

by the time they are adults. Another illustration is the way different robots with different sensors and motors that provide information about distance to nearby rigid surfaces (e.g. using sonar, laser range-finders, or vision) can use similar abstract competences to build maps of their environment and localise themselves in relation to the maps using SLAM (Simultaneous Localisation And Mapping) techniques.

People who emphasise strong relations between cognition and embodiment often forget that humans are capable of overcoming many physical obstacles produced by genetic deformity, illness, or accidents causing serious damage. There is no unique route from birth to the mind of an adult human. Instead, as clinical developmental psychologists know,<sup>11</sup> there are many different possible sequences of achievement of the same competences: they are only partially ordered.<sup>12</sup>

The fact that later stages of human cognitive development are reached by multiple routes, depending on the circumstances and constraints of individual learners, is contrary to the opinions of many politicians who attempt to define educational policies that assume a fixed developmental sequence. There is not a totally ordered sequence of transitions, only a partial ordering, with many options that depend on variations in both external circumstances and individual impairments. The fact that serious physical deformity does not prevent development of normal human vision and other aspects of normal adult human cognition, may depend on the fact that the individual's brain has many structures that evolved to meet the needs of ancestors with the full range of modes of interaction with the environment: Perhaps all the impaired individuals are using genetically specified forms of representation and theory-construction mechanisms that evolved to meet the requirements of humans with a full complement of normal limbs, digits, tongue, lips, and senses. The morphology of one's ancestors may be more important in determining one's overall cognitive potential than one's own morphology.

Depending on the impairment, more or less help may be required from gifted teachers who know what needs to be learnt. A gifted robot designer needs some of the characteristics of a gifted teacher: teaching is a form of programming – programming a very sophisticated largely autonomous machine.

## 8 Requirements for Models and Explanations

Analysis of combinations of different sorts of competence, including perceiving, reasoning, planning, controlling actions, developing new ontologies, playing, exploring, seeing explanations, and interacting socially, provides very demanding requirements to be met by human-like robots that develop through interacting with a rich and complex 3-D world. There are closely related and equally demanding requirements for an explanatory theory of how humans (and similar animals) do what they do.

There is some overlap between the themes of nouvelle AI and the viewpoint presented here. Both emphasise the importance of studying the environment when designing robots that learn

---

<sup>11</sup>I am grateful to Dr. Gill Harris, consultant at Birmingham Children's hospital for this observation.

<sup>12</sup>Famous examples of abilities to engage in rich and deep intellectual communication with other humans despite very serious physical disabilities include Helen Keller and Stephen Hawking.

to interact with it. This is also emphasised by McCarthy [19],<sup>13</sup> and Neisser [22].<sup>14</sup> This view is echoed in the work I have been doing with biologist Jackie Chappell [39, 5, 46]. However, the complexities of the environment are often grossly underestimated even by researchers who emphasise embodiment. The precise features of the environment that are relevant to different sorts of organisms in different sorts of contexts are often far from obvious, as J.J. Gibson’s work on affordances shows, for example.

There is also an old and subtle point, made by Strawson in [48], and discussed in [33], that in order to be able to refer to spatially remote, or past, or future particulars, a referring individual needs to be embedded in a web of causal links connecting the individual’s time and place with the referent or its context. That web of relationships may be needed to eliminate ambiguities of reference that cannot be removed simply by adding more descriptive content to the specification of some particular.

Despite the overlap between the position taken here and nouvelle AI, there is a very large difference of emphasis between the nouvelle AI claim that most of the information about the environment is easily available through use of sensors, provided it is sensed “appropriately and often enough”, or through sensorimotor interaction patterns, and my claim that learning about the environment involves developing deep explanatory theories using an ontology that extends beyond the reach of sensorimotor ontologies and beyond ontologies meeting the requirements of symbol-grounding theory (i.e. the requirements of concept empiricism). That is obvious for the ontologies of theoretical science, but it also applies to ontologies learnt much earlier, about different kinds of “stuff”, about internal states of complex systems, about remote places and times, and about social and mental phenomena.

Both the engineer’s robot-building tasks and the scientist’s explanation-building tasks need to take account of the distinctive features of 3-D environments in which objects of very varied structure, made of many different kinds of materials with different properties, can interact, including objects manipulated by humans, animals or robots, where the manipulation includes assembling and disassembling structures of varying complexity, and varying modes of composition, and some objects with their own externally unobservable mental states and processes. In particular the variety of *materials* of which objects can be composed, the properties of those materials, and the consequences of juxtapositions and interactions involving those materials, need to be expressed using theoretical, exosomatic, concepts, as remarked in Section 5, and discussed further in 12.2, below.

The evolutionary niches associated with such environments posed combinations of problems for our biological predecessors that need to be understood if we wish to understand the products of evolution.

## 9 Requirements for Human Visual Processing

In humans, the speed at which vision works when a person goes round a corner, or when coming out of railway stations or airports in new towns, or when watching TV documentaries

---

<sup>13</sup> “Instead of building babies as Cartesian philosophers taking nothing but their sensations for granted, evolution produced babies with innate prejudices that correspond to facts about the world and babies’ positions in it. Learning starts from these prejudices.” Footnote 2 adds: “There is a complication. Appropriate experience is often required for the genetically determined structures to develop properly, and much of this experience can be regarded as learning.”

<sup>14</sup> “We may have been lavishing too much effort on hypothetical models of the mind and not enough on analyzing the environment that the mind has been shaped to meet.”

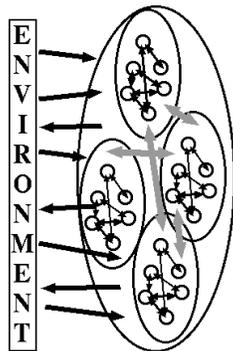


Figure 1: Dynamical systems as often conceived of: closely coupled through sensors and effectors with an environment.

about places never visited, etc. gives an indication of some of the processing requirements. An informal demonstration of the speed with which we can process a series of unrelated photographs presented at a rate of approximately one per second and extract quite abstract information about them unprompted is available online here

<http://www.cs.bham.ac.uk/research/projects/cogaff/misc/multipic-challenge.pdf>

No known mechanism comes anywhere near explaining how that is possible especially at the speed with which we do it. Reflection on a wide range of phenomena to be explained, including the speed and sophistication with which human vision copes with complex scenes presented with no warning, has led to a hypothesized architecture which at one level of description consists of a very complex dynamical system composed of a network of dynamical systems of different sorts, operating concurrently on different time scales, that grows itself, and which can be contrasted with the very much simpler kinds of dynamical system that have so far been investigated in biologically inspired robotics.

The next section illustrates crudely some of the differences between commonly considered dynamical systems and the kind of complexity that I suggest exists in virtual machines that run on human brains and the brains of some other animals, and will need to be replicated in human-like robots.

## 9.1 Two Extreme Kinds of Dynamical system

Figure 1 depicts the kind of dynamical system assumed by some of those who attempt to build “biologically inspired” robots. The key feature is very close coupling between internal states and the environment, resulting from close coupling between the internal dynamical systems and the sensory and motor signals of the animal or machine. Similar assumptions are found among some psychologists and neuroscientists, e.g. [2].

That can be contrasted with the kind of complexity depicted crudely in Fig 2 that I suggest is needed to explain visual and other capabilities in humans and some other species, and will be required for human-like robots: A very large collection of dynamical systems in virtual machines running concurrently on a brain is built up over time as a result of interaction with the environment (including toys and other artifacts, other animals, conspecifics and teachers).

The requirements for this sort of self-extending architecture are specified more detail in [42]. Many of the details were hypothesised to explain aspects of the speed and diversity of human vision revealed in the informal experiment mentioned above. Other aspects are related

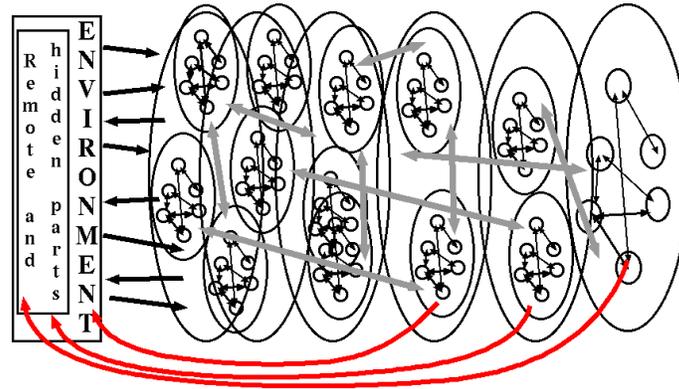


Figure 2: A more complex kind of dynamical system, composed of sub-systems of different sorts, concerned with different levels of abstraction, some continuous and some discrete, with sub-systems grown at different stages of development, some of them referring to relatively remote or inaccessible parts of the environment. At any time many of the systems are dormant, but can be rapidly reactivated.

to human abilities to think, plan, or converse while performing physical actions, and also to monitor and notice features of their own physical behaviour and their thought processes, while producing them. (Not everything can be monitored concurrently, however, on pain of infinite regress.) A human designer's ability to imagine construction of a type of object never previously encountered may use many different sorts of dynamical systems operating concurrently, even if the designer is lying flat on his back with his eyes shut, and cannot perform the assembly himself, but will need to use sophisticated machines that do not yet exist!

Such a system must depend on a large amount of previously acquired information about image and scene fragments and compatibility constraints between and within levels. The learnt information would need to be compiled into a large network of "specialist" multistable dynamical systems, most of which are dormant at any time, but which can very rapidly be triggered into activity through new sensory inputs activating fast global constraint-propagation mechanisms that "switch on" dormant sub-systems able to collaborate to build bridges between the sensory data and any available current high-level knowledge or expectations (e.g. that a busy street scene, or a beach scene, is around the corner), partly constrained by current goals, and in the case of vision, partly in registration with the optic array.

Once a global collection of subsystems with no detected inconsistencies<sup>15</sup> has been activated and settled down to compatible sub-states, further changes in sensory or motor processes can produce corresponding gradual changes in the "visual interpretation", for example when something in the perceived environment moves, or when the perceiver moves. Motion of the perceiver typically produces massive coordinated changes throughout the optic array which can help to disambiguate previous visual information in addition to providing new information about the changed location.

Topics for further investigation include: How to build dynamical systems that can very

<sup>15</sup>Pictures of impossible objects or scenes, e.g. Escher's drawings, show that in humans percepts do not need to be globally consistent.

rapidly configure themselves under control of static information both sensory and previously acquired, and then allow themselves to hand over control to changing sensory signals; how those dynamical systems are constructed over many years; how the constraint-propagation works; how the processes are influenced by current expectations, interests and goals; how the resulting percepts continue to be driven by changing sensory data; and how all of that is used in controlling and planning actions, and producing further learning. These are questions for future research.

The structural changes in percepts are both global and rapid, suggesting that they are changes in virtual machines that have only a loose relationship to underlying physical machines, whose global structures cannot change rapidly, though coordinated local changes (e.g. changes in synaptic weights) can happen rapidly.

However, that leaves the problem of how large numbers of local changes are coordinated so as to produce the required global changes at the virtual machine level. Available techniques for investigating and modelling neural structures and processes may have to wait until the theory is much further developed at the level of virtual machines, so as to guide the empirical neuroscience. Compare trying to work out what a modern computer is doing by studying electronic details, without being told anything about its operating system or the detailed functions that it performs.

## 10 The Variety of Possible Architectures

There are different ways of formulating explanatory designs, at varying levels of abstraction (discussed in [37]). Alternative designs need to be analysed and compared against sets of possible requirements. Previous sections present requirements for a kind of self-extending, multi-functional, virtual machine information-processing architecture to explain a variety of human capabilities – a collection of requirements that no current AI systems or current neural theories (at least none known to the author) address.

### 10.1 Dynamical Systems for Symbolic AI

Those who reject symbolic AI often fail to realise that symbolic mechanisms may be needed for the more abstract, more loosely coupled, more discrete, dynamical systems, for example the ones further to the right in Figure 2, that enable you to learn ancient history, do algebra in your head, discuss philosophy, make plans for conference travel, read music, or read this paper. However, existing symbolic AI systems are clearly nowhere near human competence except in very narrow domains, e.g. playing certain board games (chess, but not Go), certain mathematical competences (e.g. in packages like Mathematica, Matlab and others), and some kinds of compiler optimisation.

Much of the technology now being used all over the world on the internet is more or less closely related to work done in symbolic AI, including rule-based expert systems in past decades (whose techniques have often been reinvented, because they are so useful). Inductive logic programming (ILP) techniques have contributed to molecular biology research on folding of protein structures [14]. Generalising and combining such specialised competences to match human flexibility is not trivial, but something like that will need to be done in combination with results of nouvelle AI.

It is not obvious which non-human animals have competences that require sub-systems that are only loosely coupled to current sensory and motor signals, or de-coupled from them,

for example the ability to think about past events, wonder about possible futures, form hypotheses about what is going on out of sight, plan journeys to remote locations, or think about the contents of other minds or their own minds.

Of course not all the errors lie in the nouvelle AI tradition: as Brooks points out, people who think symbolic AI will suffice for everything often fail to attend to the kinds of intelligence required for controlling continuous actions in a 3-D structured environment, including maintaining balance while pushing a broom, drawing a picture with a pencil, or playing a violin [32]. Many such activities require both sorts of mechanism operating concurrently, along with additional monitoring, evaluating, and learning (e.g. debugging) mechanisms. The fact that humans, other animals, and robots are *embodied* can lead, and has led, some researchers in the nouvelle AI tradition to assume, mistakenly, that only the first type of dynamical system (Fig 1) is required because they forget that some embodied systems can think about past, future, distant places, games of chess, what elephants can't do, transfinite ordinals and how to design intelligent systems. Reading and writing novels and history books that refer to physical environments does not require sensorimotor interaction with the environment referred to, but something far more abstract and symbolic.

The different roles of dynamical systems operating concurrently at different levels of abstraction need to be accommodated in any theory of motivation and affect, since motives, evaluations, preferences, and the like can exist at many levels and operate on different timescales.<sup>16</sup> It should be clear that the “traditional” *sense*→*think/decide*→*act* loop (presented in several AI textbooks) is much too restrictive to accommodate the requirements presented here.

## 10.2 Types of Architecture

We can crudely decompose the variety of sub-processes that occur in biological organisms in two dimensions shown in Figure 3: one dimension, related to what Nilsson called “Towers” in Chapter 25 of [23] is concerned with whether the processes are (1) perceptual/sensory, or (2) central or (3) concerned with effectors/actions; while the other dimension related to what Nilsson called “Layers”, depends on whether the processes are based on (1) reactive (often evolutionarily old) mechanisms, or (2) deliberative mechanisms (with the ability to consider and compare branching sets of possibilities) or (3) meta-management mechanisms (concerned with self-monitoring or control, or using meta-semantic competences in relation to other agents). Note that not everything reactive is closely coupled with the environment. Reactive processes can produce internal state-changes, such as activation of new motives, alteration of priorities, and triggering of learning processes.

Superimposing the tower and layer dimensions produces a 3 by 3 grid of types of sub-system as illustrated in Figure 3. The grid, as explained in [37], is only an approximation – more sub-divisions are to be found in nature than this suggests (in both dimensions, but especially the vertical dimension). That is in part because biological evolution can produce only discrete changes, not continuous changes. The discrete changes vary in size and significance: e.g. duplication is often more dramatic than modification, and can be the

---

<sup>16</sup> A critique of some shallow theories of emotion is presented in: “Do machines, natural or artificial, really need emotions?”

<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#cafe04> See also “What are emotion theories about?” (AAAI Spring Symposium 2004):

<http://www.cs.bham.ac.uk/research/projects/cogaff/04.html#200403>

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

Figure 3: The CogAff schema specifying, to a first approximation, types of (concurrently active) components that can occur in an architecture, with links between many of them.

start of a major new development. Further sub-divisions include different functions, different forms of representation, different types of learning, different mechanisms, different connections between modules, and different nature-nurture tradeoffs.

A type of reactive mechanism that might have evolved after others is an “alarm” mechanism as indicated in Figure 4, that gets inputs from various parts of the architecture has outputs that influence many parts of the architecture, and can detect patterns that require rapid global reorganisation of processing (e.g. freezing, fleeing, ducking out of sight, pouncing on prey, reconsideration of current plans, etc.). Copying and modification of early versions of such alarm mechanism might later have led to development of more sophisticated meta-management mechanisms for monitoring and control of processing. Such speculative theories can only be tested in the long term, to see whether they lead to progressive or degenerative research programmes, using the terminology of [17].

A special case of the CogAff schema that we have been developing for some years is the H-CogAff (Human-CogAff) architecture specification described in [36] and depicted crudely in Figure 4. The H-CogAff architecture overlaps with Minsky’s ideas in [21].<sup>17</sup> There is a lot more detail regarding these ideas in presentations and papers in the Birmingham CogAff and CoSy projects.<sup>18</sup>

### 10.3 Layered Perception and Action

A point that can cause puzzlement is the suggestion that the central parts of the architecture can be directly connected to higher levels in the perceptual and action subsystems. In this context ‘directly’ does not mean physically directly, but something more abstract. The CogAff schema, and the H-CogAff instance, are both concerned with virtual machine architectures rather than physical architectures. (Virtual machines were also emphasised in Section 9.1.) The different levels inside the perception and action sub-systems concurrently

<sup>17</sup>Minsky’s focus seems to be mainly on how to model or replicate certain aspects of human intelligence, whereas I am interested in attempting to understand the space of possible designs and niches explored by evolution and its relation to future possible artificial systems.

<sup>18</sup><http://www.cs.bham.ac.uk/research/projects/cogaff/>  
<http://www.cs.bham.ac.uk/research/projects/cogaff/talks/>  
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/>

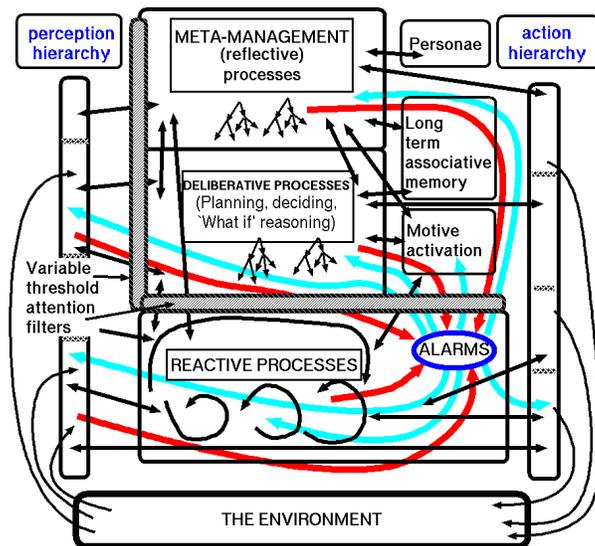


Figure 4: The H-CogAff architecture – an instance of the CogAff schema that seems to be required to account for many human competences. This diagram should be viewed as a special case of Fig 2 rotated 90 degrees counter-clockwise, though many details are still unknown. (The diagram is intended only to be suggestive, not self-explanatory.)

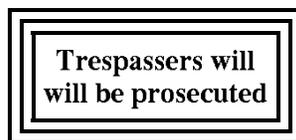


Figure 5: Anyone who sees this as a familiar phrase, without noticing anything wrong with it, is not aware of what is in the low level sensory data.

process information at different levels of abstraction closely related to what is going on in the environment, and partly in registration with representations of the environment (e.g. the optic array (Gibson) rather than what's on the retina (or V1), which is merely a device for sampling optic arrays.) This is related to gaps in current machine vision.

Most people find it obvious that the perceptual mechanisms involved in understanding speech, or hearing music, have to operate at different levels of abstraction in parallel, in a mixture of top-down and bottom-up processing, e.g. in the case of speech, processing low level acoustic features, phonemes, morphemes, phrases, intonation contours, etc.

As explained below,<sup>19</sup> the same is true of vision, with different intermediate results of visual processing using ontologies at different levels of abstraction (not merely part-whole hierarchies), loosely in registration with the optic array. A working demonstration program was developed in the mid-70s, but funding was not continued, partly because referees were convinced (following Marr) that vision systems work purely bottom-up and that the rich information in natural scenes makes sophisticated visual architectures unnecessary.

One way to convince yourself that vision operates at different levels of abstraction is to look at well-known spontaneously changing ambiguous images, such as the necker cube, the

<sup>19</sup>And in <http://www.cs.bham.ac.uk/research/projects/cogaff/crp/chap9.html> Chapter 9 of [31]

old/young woman, the duck-rabbit, and others,<sup>20</sup> and try to describe what changes. When they ‘flip’ so that you are seeing something different, nothing changes in the image or in the 2-D experience: so what is seen as different is something that requires an ontology distinct from image contents. Likewise if you see a happy and a sad face, there may be geometric differences (e.g. direction of curvature of mouth) but happiness and sadness are not geometric properties. Figure 5 is a variant of a familiar example where some people fail to see anything wrong with the phrase in the box despite staring at it for some time, indicating that what they are conscious of is a high level percept, even though it is easy to establish that lower level information was taken in (as some of them discover if they cover the figure and try to make a copy).

Many actions also require multi-level concurrent control, at different levels of abstraction. This is obvious for speech, nearly as obvious for performance of music, and perhaps less obvious for walking. But the way you walk can simultaneously involve low level control of motion and balance, higher level control of direction towards your target (e.g. a door) and perhaps more abstract expressions of mood or attitude (e.g. walking with a swagger) – a form of social communication with people known to be watching.

Those different influences on the behaviours will come from different layers of the internal architecture performing different tasks concurrently, using very different forms of representation, and the mechanisms in the action “tower” will need to combine all those influences in producing the final low level output.

Arnold Trehub’s 1991 book [51] (now online) presents closely related ideas in his “retinoid” system. Trehub proposed that the main perceptual data-structures representing what is in the environment at different levels of abstraction and different distances from the viewer are not in registration with the contents of the primary visual cortex (or the retina) because that mapping is constantly changing under the influence of saccades, and vestibular and other information about how the eyes and head are moving in 3-D space, whereas the information about contents of the immediate environment should not change so fast.

The architecture diagrams presented here cannot express all that complexity, and are merely intended to challenge some “obvious” but mistaken views of perception and action (often expressed in architecture diagrams showing perception and action as small boxes), which may be correct only for a subset of organisms, e.g. most invertebrates, though the portia spider, mentioned in Section 3.2 seems to have the ability to perceive potential routes.

An ecosystem, or even the whole biosphere, can be thought of as a complex *virtual* machine with multiple concurrent threads, producing multiple feedback loops of many kinds (including not just scalar feedback but structured feedback – e.g more like a sentence than a force or voltage). Individual organisms in the ecosystem, insofar as they have designs described here, will also include multiple virtual machines. All these virtual machines are ultimately implemented in physics and chemistry, but their states and processes and causal interactions are not describable in the language of physics and chemistry. We don’t yet know enough about what variety of virtual machines can be implemented on the basis of physical principles.

---

<sup>20</sup>E.g available here <http://lookmind.com/illusions.php>, and here [http://www.planetperplex.com/en/ambiguous\\_images.html](http://www.planetperplex.com/en/ambiguous_images.html)

## 11 Perception of Actual and Possible Processes

### 11.1 Beyond Gibson

Work on an EU-funded cognitive robotics project, CoSy ([www.cognitivesystems.org](http://www.cognitivesystems.org)) begun in 2004, included analysis of requirements for a robot capable of manipulating 3-D objects, e.g. grasping them, moving them, and constructing assemblages, possibly while other things were happening.

That analysis, reported in [47], revealed (a) the need for representing scene objects with parts and relationships (as everyone already knew), (b) the need for several ontological layers in scene structures (as in [31, Ch 9]), (c) the need to represent “multi-strand relationships” because not only whole objects but also parts of different objects are related in various ways (e.g. a corner of one object touching an edge of another), (d) the need to represent “multi-strand processes”, because when things move the multi-strand relationships change, e.g. with metrical, topological, causal, functional, continuous, and discrete changes occurring concurrently, and (e) the need to be able to represent *possible* processes, and constraints on possible processes [35, 45].

### 11.2 Proto-affordances

I call such possibilities and constraints restricting them “proto-affordances”, because they are the substratum of affordances, and more general: unlike J.J. Gibson’s affordances, the proto-affordances are not restricted to processes the perceiver can produce, or processes relevant to the perceiver’s possible goals.

Moreover, proto-affordances can be combined in various ways because processes in the environment can be combined, including serial composition, parallel composition of non-interacting processes, and most importantly parallel composition of *juxtaposed* processes involving interacting objects, for instance one rotating gear wheel causing another to rotate, a hand catching a ball or two moving objects colliding. So complex proto-affordances, involving possibilities of and constraints on possible processes can be formed by synthesising simpler proto-affordances.

It is not clear how many animal species can use this possibility. It requires the use of an exosomatic ontology, not a sensorimotor ontology, since moving interacting objects (e.g. boulders rolling down a hill bumping into trees and other boulders) can exist without being perceived or acted on.

### 11.3 Epistemic Affordances

Gibson came very close to discussing what could be called *epistemic* affordances, concerned not with opportunities and obstacles for possible processes in the environment or actions of the perceiver, but with availability and limitations of *information* potentially usable by the perceiver. Intelligence often involves discovering which actions or other processes can usefully change epistemic affordances, e.g. removing uncertainty about a part of the environment by moving to a new viewpoint, or moving an occluding object – another example of composition of processes: a physical process involving motion and a more abstract process of changing available information.

#### 11.4 Affordances in the Environment

Not all perceived changes are produced or could be produced by the perceiving agent. Moreover, seeing that a process is possible, e.g. an apple dropping off a tree, or that there are constraints on possibilities, e.g. because a table is below the apple, does not require the process to be one that the perceiver desires to produce. So perception of proto-affordances and perception of processes in the environment makes it possible to take account of far more than one's own actions, their consequences and their constraints. One important consequence is perception of *vicarious affordances*, affordances for others, e.g. collaborators, prey, predators and offspring learning how to interact with the environment, who may sometimes need to be helped, rescued, or warned about risks. As discussed in Section 5, some of those competences require an *exosomatic* form of representation of processes; one that is not tied to the agent's sensorimotor contingencies. That possibly is ignored by researchers who focus only on sensorimotor learning and representation, and base all semantics on "symbol-grounding".

#### 11.5 Representing Actual, Possible and Impossible Processes

The ability to perceive a multi-strand process requires the ability to have internal representations of the various concurrently changing relationships. Some will be continuous changes, including those needed for servo-control of actions, while others may be discrete changes that occur as topological relations change or goals or preferences become satisfied or violated. Such process-perception could include concurrent changes occurring in different concurrently active dynamical systems depicted in Figure 2. How processes can be represented in perceivers, and how the representation differs when the process is actually being perceived and when it is merely being thought about or reasoned about remain open questions. The fact that we can perceive and think about processes that we cannot produce ourselves (e.g. the sky darkening, or a flock of birds swarming) rules out theories that assume perceiving processes uses motor subsystems that could produce those processes.

Perceiving *possibilities of* and *constraints on* processes in the environment requires additional forms of representation and mechanisms for manipulating them without being driven by sensory input, since there is no sensory input from a process that does not exist yet. An important research topic is how an animal or robot can represent proto-affordances and their composition, using the results in planning, generating and controlling behaviour. This ability develops both in the life of an individual and across generations in a culture.

It seems that mechanisms closely related to those used for perceiving multi-strand processes can also be used to predict outcomes of possible processes that are not currently occurring (e.g. when planning), and to explain how perceived situations came about. Both may use a partial simulation of the processes (compare [10, 29]), though simulation is not enough in general: it may be necessary to save and compare "snapshots" of portions of the processes, for example in order to think about pros and cons of alternative future actions. The mechanisms for reasoning about processes and affordances seem also be closely related to the development of mathematical competences in humans discussed below in Section 12.2.

21

---

<sup>21</sup>Examples are given in the discussion of fully deliberative competences in [40] and in this discussion paper on predicting changes in action affordances and epistemic affordances: <http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0702>

## 12 How Evolution Produced Mathematicians?

### 12.1 The Importance of Kinds of Matter

Some predictions need to go beyond geometry and topology. A child, or robot, learning about kinds of process that can occur in the environment needs to be able to extend indefinitely the ontologies she uses, and not merely by defining new concepts in terms of old ones: there are also substantive ontology extensions, refuting symbol-grounding theory, as explained above.

Concepts of “kinds of stuff” are examples. Whereas many perceived processes involve objects that preserve all their metrical relationships, there are also many deviations from such rigidity, and concepts of different kinds of matter are required to explain those deviations. For example, string and wire are flexible, but wire retains its shape after being deformed. An elastic band returns to its original length after being stretched, but does not restore its shape after bending. Some kinds of stuff easily separate into chunks if pulled, e.g. mud, porridge, plasticine and paper. A subset of those allow restoration to a single object if separated parts are pressed back together. A subset of those will retain a weakness at the joint. There are also objects that are marks on other objects, like lines on paper, and there are some objects that can be used to produce such marks, like pencils and crayons. Marks produced in different ways and on different materials can have similar structures. Lines drawn in sand with a finger or a stick are partly similar to lines drawn with a surface marker, and partly different: they can serve similar diagrammatic or symbolic functions, without depending in the same way on colour-contrasts to be perceived.

As demonstrated by the Sauvys, in [28], children, and presumably future robots, can learn to play with and explore strings, elastic bands, pieces of wire, marks on paper and movable objects in such a way as to learn about many different sorts of process patterns. Some of those are concerned with rigid motions some not. Many of their examples use patterns in non-rigid motions that can lead to development of topological concepts.

### 12.2 From Empirical to Mathematical Truths

Consideration of a space of niches and a space of designs for different sorts of animal and different sorts of machine reveals nature/nurture tradeoffs, and indicates hard problems that AI researchers, psychologists and neuroscientists need to address. It was suggested in [43] that a robot or animal that learns through play and exploration in a complex, changing 3-D environment needs certain competences (e.g. the ability to perceive and reason about both action affordances and epistemic affordances and to discover invariant non-empirical features of spatial processes) that also underlie human abilities to do mathematics (including geometry, topology, number theory and set theory).

A child may learn that going round a house in one direction enables doors and windows to be seen in a certain temporal order, whereas going round in the opposite direction reverses that order. (This requires perceptual memory capabilities that need further discussion.) That empirical discovery may later be recognized to allow no exceptions (if the building does not change) so that it can be totally relied on in future: it is seen not to be empirical.

This is not a case of using observed correlations to raise a probability to a maximal value but something deeper: discovering an invariant intrinsic to a class of processes. Another example is discovering at first empirically that counting a set of objects in different orders always produces the same result, and then later understanding that there cannot be exceptions to that (if no objects are added, removed, merged or split). Impossibilities may also be

discovered at first empirically then found to be necessary. A child may use a rubber band and some pins to make various shapes, e.g. a square, a triangle, a star, etc., and then discover the minimum number of pins required to make each particular shape, e.g. a six pointed star, or an outline capital **H**. How can the child be *certain* that the number found is the smallest? At first it may just be an empirical discovery, which remains open to refutation. But later it is seen to be a necessary truth. How?<sup>22</sup>

### 12.3 It's Not a Matter of Probabilities

The ability to reason about what is and is not possible is required for reasoning about affordances, and is also at the basis of mathematical competence. This 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 much 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 *enduring 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 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. even the excellent [8].

I suspect that further work will show that such learning processes vindicate the claim of Immanuel Kant (1781) (against David Hume) that mathematical knowledge is both non-empirical and non-analytic [30]. However, the representational and architectural requirements for a robot that can make such discoveries is still an open research question. (Piaget's theory of "formal operations" was an attempt to address similar problems, but he lacked the conceptual tools required to design working systems.)

A caveat: The claim that some discoveries turn out to be non-empirical implies neither that the knowledge is innate, nor that the discovery was somehow pre-programmed genetically (compare [27]). Further, the processes are not infallible: mistakes can occur and such discoveries may have "bugs" as Lakatos [16] demonstrated using the history of Euler's theorem about polyhedra – though that is sometimes wrongly taken to imply that mathematical knowledge is empirical in the same way as knowledge about the physical properties of matter. Lakatos used the label "quasi-empirical". The discovery of bugs in proofs and good ways to deal with them is an important feature of mathematical learning. This can be triggered by perceiving counter-examples in the environment, but often it is sufficient merely to *think* of a possible counter-example. That is not enough to refute an empirical theory, since what

---

<sup>22</sup>In <http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#mkm08> and [43] I suggest that there are many "toddler theorems", and have begun to collect examples.

is thinkable need not be possible in practice. Noticing real or apparent exceptions to a mathematical generalisation, whether perceived or imagined, can lead a learner to investigate properties that distinguish different sub-cases.

Viewing human mathematical competence as a side-effect of evolutionary and developmental processes meeting biological needs can shed new light both on old philosophical problems about the nature of mathematical knowledge and on problems in developmental psychology and education, especially mathematical education. It also helps us understand requirements for human-like robots.

## 13 Conclusion

AI and Cognitive Science still have much to do before the visual and other cognitive mechanisms of robots can match the achievements of a typical human child or even a nest-building bird, or young ape or hunting mammal. There have been important advances in the design of physically embodied behaviour-based robots (some of it inspired by the work of Brooks and his colleagues) and there have been important advances in mechanisms and applications of symbolic AI. Achieving the longer term scientific and engineering goals of AI will require the two approaches to be brought together, along with a more general synthesis of research in different sub-fields of intelligence.

What is important about embodiment (e.g. what drove the most significant evolutionary developments in primate and bird cognition) is not the specific morphology of the learner, but 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. For many of the points made by J.J. Gibson all that is important is that we can see and can change our viewpoint and viewing direction, which is also true of many animals with very different shapes, e.g. birds and lions. In many cases there is no interaction or manipulation (e.g. we cannot interact with long dead ancestors, but we can think about them), and when interaction or manipulation does occur, it does not seem to be essential for human development whether it uses mouth, hands, feet, or some tool or prosthetic device, for what we learn about is not primarily about our own sensorimotor processes but about what can happen in our environment. This is also why, as mentioned in Section 7, children with various physical deformities and disabilities (e.g. thalidomide babies) can grow up able to communicate about the same environment as those without the disabilities, though the process of learning is different in its details.

In contrast, much of the work on embodied cognition in robots has focused on the terribly narrow problem of learning about sensorimotor relationships in robots with very specific morphologies. (There are some exceptions.)

If we are to make significant progress in developing robots that understand our world, it will be necessary for AI researchers to abandon factional disputes, and stop claiming that there is one key mechanism that is required, and instead study in far more detail than hitherto both the features of the environment that impose demands on animal and robot competences and the variety of ways in which biological evolution, including human evolution, has responded to those demands. In particular, it is important to understand that it is not easy to determine empirically what biological mechanisms are and how they work (e.g. brain mechanisms); instead, by working from carefully observed competences towards mechanisms, we may come

up with important ideas generating questions to guide the researchers who attempt to study the mechanisms directly. After all, symbolic computation was a human competence long before AI began, and was the original inspiration for symbolic AI, so there must be biological mechanisms that make it possible (even if nobody knows what they are), and Brooks was right to suggest that a synthesis of complementary approaches might be essential to progress in the long run.

### Acknowledgements

I would like to thank Bernhard Sendhoff and Olaf Sporns for useful critical comments on an earlier draft. Many of the ideas reported in this paper were developed as part of the requirements analysis activities in the EU-funded CoSy robotic project: <http://www.cognitivesystems.org>, especially in discussions with Jeremy Wyatt. Jackie Chappell helped with biological examples, and nature-nurture tradeoffs.

### References

- [1] A. P. Ambler, H. G. Barrow, C. M. Brown, R. M. Burstall, and R. J. Popplestone. A Versatile Computer-Controlled Assembly System. In *Proc. Third Int. Joint Conf. on AI*, pages 298–307, Stanford, California, 1973.
- [2] Alain Berthoz. *The Brain's sense of movement*. Perspectives in Cognitive Science. Harvard University Press, London, UK, 2000.
- [3] V. Braitenberg. *Vehicles: Experiments in Synthetic Psychology*. The MIT Press, Cambridge, MA, 1984.
- [4] R. A. Brooks. Elephants Don't Play Chess. *Robotics and Autonomous Systems*, 6:3–15, 1990. <http://people.csail.mit.edu/brooks/papers/elephants.pdf>.
- [5] Jackie Chappell and Aaron Sloman. Natural and artificial meta-configured altricial information-processing systems. *International Journal of Unconventional Computing*, 3(3):211–239, 2007. <http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609>.
- [6] Dave Cliff. Biologically-Inspired Computing Approaches to Cognitive Systems: a partial tour of the literature. Technical Report HPL-2003-11, Hewlett-Packard Labs, Bristol, UK, 2003. <http://www.hpl.hp.com/techreports/2003/HPL-2003-11.html>.
- [7] M.R. Cutkosky, J.M. Jourdain, and P.K. Wright. Testing and Control of a Compliant Wrist. Technical Report CMU-RI-TR-84-04, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, March 1984. [http://www.ri.cmu.edu/pubs/pub\\_73.html](http://www.ri.cmu.edu/pubs/pub_73.html).
- [8] Eleanor J. Gibson and Anne D. Pick. *An Ecological Approach to Perceptual Learning and Development*. Oxford University Press, New York, 2000.
- [9] J. J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, MA, 1979.
- [10] R. Grush. The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27:377–442, 2004.
- [11] S. Harnad. The Symbol Grounding Problem. *Physica D*, 42:335–346, 1990.
- [12] Eva Jablonka and Marion J. Lamb. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. MIT Press, Cambridge MA, 2005.

- [13] I. Kant. *Critique of Pure Reason*. Macmillan, London, 1781. Translated (1929) by Norman Kemp Smith.
- [14] R.D. King, D.A. Clark, Shirazi J., and Sternberg M.J. Inductive logic programming used to discover topological constraints in protein structures. pages 219–26, 1994.
- [15] D. Kirsch. Today the earwig, tomorrow man? *Artificial Intelligence*, 47(1):161–184, 1991. <http://adrenaline.ucsd.edu/kirsh/articles/earwig/earwig-cleaned.html>.
- [16] I. Lakatos. *Proofs and Refutations*. Cambridge University Press, Cambridge, UK, 1976.
- [17] I. Lakatos. The methodology of scientific research programmes. In J. Worrall and G. Currie, editors, *Philosophical papers, Vol I*. Cambridge University Press, Cambridge, 1980.
- [18] M. Lungarella and O. Sporns. Mapping information flow in sensorimotor networks. *PLoS Computational Biology*, 2(10:e144), 2006. DOI: 10.1371/journal.pcbi.0020144.
- [19] J. McCarthy. The Well-Designed Child. *Artificial Intelligence*, 172(18):2003–2014, 2008. <http://www-formal.stanford.edu/jmc/child.html>.
- [20] M. L. Minsky. *The Society of Mind*. William Heinemann Ltd., London, 1987.
- [21] M. L. Minsky. *The Emotion Machine*. Pantheon, New York, 2006.
- [22] U. Neisser. *Cognition and Reality*. W. H. Freeman., San Francisco, 1976.
- [23] N.J. Nilsson. *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann, San Francisco, 1998.
- [24] R. Pfeifer, F. Iida, and G. Gomez. Designing intelligent robots - on the implications of embodiment. *Journal of Robotics Society of Japan*, 24(07):9–16, 2006. [http://www.robotcub.org/misc/review3/07\\_Pfeifer\\_Iida\\_Gomez\\_RSJ.pdf](http://www.robotcub.org/misc/review3/07_Pfeifer_Iida_Gomez_RSJ.pdf).
- [25] D.L. Philipona and J.K. O’Regan. Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties. *Visual Neuroscience*, 23(3-4):331–339, 2006.
- [26] Henri Poincaré. *Science and hypothesis*. W. Scott, London, 1905. <http://www.archive.org/details/scienceandhypoth00poinuoft>.
- [27] L. J. Rips, A. Bloomfield, and J. Asmuth. From Numerical Concepts to Concepts of Number. *The Behavioral and Brain Sciences*, In Press.
- [28] J. Sauvy and S. Suavy. *The Child’s Discovery of Space: From hopscotch to mazes – an introduction to intuitive topology*. Penguin Education, Harmondsworth, 1974. Translated from the French by Pam Wells.
- [29] M.P. Shanahan. A cognitive architecture that combines internal simulation with a global workspace. *Consciousness and Cognition*, 15:157–176, 2006.
- [30] A. Sloman. ‘Necessary’, ‘A Priori’ and ‘Analytic’. *Analysis*, 26(1):12–16, 1965. Now online <http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#701>.
- [31] A. Sloman. *The Computer Revolution in Philosophy*. Harvester Press (and Humanities Press), Hassocks, Sussex, 1978. <http://www.cs.bham.ac.uk/research/cogaff/crp>.
- [32] A. Sloman. Image interpretation: The way ahead? In O.J. Braddick and A.C. Sleight., editors, *Physical and Biological Processing of Images (Proceedings of an international symposium organised by The Rank Prize Funds, London, 1982.)*, pages 380–401. Springer-Verlag, Berlin, 1982. <http://www.cs.bham.ac.uk/research/projects/cogaff/06.html#0604>.
- [33] A. Sloman. What enables a machine to understand? In *Proc 9th IJCAI*, pages 995–1001, Los Angeles, 1985. <http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#4>.

- [34] A. Sloman. On designing a visual system (towards a gibsonian computational model of vision). *Journal of Experimental and Theoretical AI*, 1(4):289–337, 1989.  
<http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#7>.
- [35] A. Sloman. Actual possibilities. In L.C. Aiello and S.C. Shapiro, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifth International Conference (KR '96)*, pages 627–638, Boston, MA, 1996. Morgan Kaufmann Publishers.  
<http://www.cs.bham.ac.uk/research/cogaff/96-99.html#15>.
- [36] A. Sloman. The Cognition and Affect Project: Architectures, Architecture-Schemas, And The New Science of Mind. Technical report, School of Computer Science, University of Birmingham, 2003. <http://www.cs.bham.ac.uk/research/projects/cogaff/03.html#200307> (Revised August 2008).
- [37] A. Sloman. Cross-Disciplinary Reflections: Philosophical Robotics. Research Note: Draft chapter for a book on the CoSy project COSY-TR-0806, School of Computer Science, University of Birmingham, 2008. <http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0806>.
- [38] A. Sloman. Putting the Pieces Together Again. In Ron Sun, editor, *Cambridge Handbook on Computational Psychology*, chapter 26, pages 684–709. Cambridge University Press, New York, 2008. <http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#710>.
- [39] A. Sloman and J. Chappell. The Altricial-Precocial Spectrum for Robots. In *Proceedings IJCAI'05*, pages 1187–1192, Edinburgh, 2005. IJCAI.  
<http://www.cs.bham.ac.uk/research/cogaff/05.html#200502>.
- [40] Aaron 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, May 2006.  
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0604>.
- [41] Aaron Sloman. Diversity of Developmental Trajectories in Natural and Artificial Intelligence. In C. T. Morrison and T. Tim Oates, editors, *Computational Approaches to Representation Change during Learning and Development. AAAI Fall Symposium 2007, Technical Report FS-07-03*, pages 70–79, Menlo Park, CA, 2007. AAAI Press.  
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0704>.
- [42] Aaron Sloman. Architectural and representational requirements for seeing processes, proto-affordances and affordances. In Anthony G. Cohn, David C. Hogg, Ralf Möller, and Bernd Neumann, editors, *Logic and Probability for Scene Interpretation*, number 08091 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2008. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.
- [43] Aaron Sloman. Kantian Philosophy of Mathematics and Young Robots. In S. Autexier, J. Campbell, J. Rubio, V. Sorge, M. Suzuki, and F. Wiedijk, editors, *Intelligent Computer Mathematics*, LLNCS no 5144, pages 558–573, Berlin/Heidelberg, July 2008. Springer.  
<http://www.cs.bham.ac.uk/research/projects/cosy/papers#tr0802>.
- [44] Aaron Sloman. Varieties of Meta-cognition in Natural and Artificial Systems. In M. T. Cox and A. Raja, editors, *Workshop on Metareasoning, AAAI'08 Conference*, pages 12–20, Menlo Park, CA, 2008. AAAI Press. <http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0803>.
- [45] Aaron 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.  
<http://www.cs.bham.ac.uk/research/projects/cosy/papers#tr0801>.
- [46] Aaron Sloman and Jackie Chappell. Computational Cognitive Epigenetics (Commentary on [12]). *Behavioral and Brain Sciences*, 30(4):375–6, 2007.  
<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0703>.

- [47] Aaron Sloman and Cosy-partners. CoSy deliverable DR.2.1 Requirements study for representations. Technical Report COSY-TR-0507, The University of Birmingham, UK, 2005. <http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0507>.
- [48] P. F. Strawson. *Individuals: An essay in descriptive metaphysics*. Methuen, London, 1959.
- [49] G.J. Sussman. *A computational model of skill acquisition*. American Elsevier, 1975.
- [50] M. Tarsitano. Route selection by a jumping spider (*Portia labiata*) during the locomotory phase of a detour. *Animal Behaviour*, 72, Issue 6,:1437–1442, December 2006. <http://dx.doi.org/10.1016/j.anbehav.2006.05.007>.
- [51] A. Trehub. *The Cognitive Brain*. MIT Press, Cambridge, MA, 1991. <http://www.people.umass.edu/trehub/>.
- [52] A.M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950. (reprinted in E.A. Feigenbaum and J. Feldman (eds) *Computers and Thought* McGraw-Hill, New York, 1963, 11–35).
- [53] F. Warneken and M. Tomasello. Altruistic helping in human infants and young chimpanzees. *Science*, pages 1301–1303, 3 March 2006. DOI:10.1126/science.1121448.
- [54] T. Ziemke. Situated and Embodied Cognition. *Cognitive Systems Research*, 3(3), 2002. (Editor’s introduction to special issue.).