Contribution to Proceedings of Computational Modelling Workshop,
Closing the Gap Between Neurophysiology and Behaviour: A Computational Modelling Approach
Editor: Dietmar Heinke
http://comp-psych.bham.ac.uk/workshop.htm
University of Birmingham, UK, May 31st-June 2nd 2007
This paper is at
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0801
Original longer version here
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0801a

Architectural and Representational Requirements for Seeing Processes and Affordances.

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

May 30, 2008

Abstract

This paper, combining the standpoints of philosophy and Artificial Intelligence with theoretical psychology, summarises several decades of investigation of the variety of functions of vision in humans and other animals, pointing out that biological evolution has solved many more problems than are normally noticed. Many of the phenomena discovered by psychologists and neuroscientists require sophisticated controlled laboratory settings and specialised measuring equipment, whereas the functions of vision reported here mostly require only careful attention to a wide range of everyday competences that easily go unnoticed. Currently available computer models and neural theories are very far from explaining those functions, so progress in explaining how vision works is more in need of new proposals for explanatory mechanisms than new laboratory data. Systematically formulating the requirements for such mechanisms is not easy. If we start by analysing familiar competences, that can suggest new experiments to clarify precise forms of these competences, how they develop within individuals, which other species have them, and how performance varies according to conditions. This will help to constrain requirements for models purporting to explain how the competences work. The paper ends with speculations regarding the need for new kinds of information-processing machinery to account for the phenomena.

Contents

1 Introduction: from Kant to Gibson and beyond

1.1 The role of vision in mathematical competences

The aim of the workshop was to discuss a computational approach to "Closing the gap between behaviour and neurophysiological level". My approach to this topic is to focus almost entirely on what needs to be explained rather than to present any neurophysiological model, though conjectures regarding some of the design features required in such a model are offered in Section 6.

I originally began studying vision to understand the role of visual processing in mathematical discovery and reasoning, for instance in proving theorems in elementary Euclidean geometry, but also in more abstract reasoning, for example about infinite structures, which can be visualised but cannot occur in the environment. Sloman (1962) was an attempt to defend the view of mathematical knowledge as both non-empirical and synthetic, proposed by Kant (1781), but rejected by many contemporary mathematicians and philosophers. This led me into topics linking many disciplines, including mathematics, psychology, neuroscience, philosophy, linguistics, education and Artificial Intelligence. A full understanding requires parallel investigations of many different kinds of vision: in insects and other invertebrates, in birds, in primates, and in different sorts of future robots.

This paper attempts to show how the role of vision in mathematical reasoning is connected with the ability in humans, some other animals and future human-like robots, to perceive and reason about structures and processes in the environment, including *possible* processes that are not actually occurring. This requires us to focus attention on aspects of vision that are ignored by most other researchers including: those who study vision as concerned with image structure (e.g. Kaneff, 1970); those who study vision as a source of geometrical and physical facts about the environment (e.g. Marr, 1982); those who regard vision as primarily a means of controlling behaviour (e.g. Berthoz, 2000 and many researchers in AI/Robotics and psychology who regard cognition as closely tied to embodiment); and those who regard vision as acquisition of information about affordances (e.g. J. J. Gibson (1979); E. J. Gibson and Pick (2000)).

1.2 Why complete architectures matter

The processes of seeing involve actual or potential interactions between sensory mechanisms, action-control mechanisms and more central systems. Those subsystems arise from different stages in our evolutionary history and grow during different stages in individual development. So the functions of vision differ both from one species to another and can change over time within an individual as the information-processing architecture grows. Some of those developments, are culture-specific. such as what language the individual learns to read, which gestures are understood and which building designs are encountered.

A full understanding of vision requires investigation of different multifunctional architectures in which visual systems with different collections of competences can exist. An architecture with more sophisticated 'central' mechanisms makes possible more sophisticated visual functions. For instance, a central mechanism able to use an ontology of causal and functional roles is required for a system that can see something causing, preventing, or

enabling something else to occur. The ability to make use of an ontology including mental states (a meta-semantic ontology) is required if a visual system is to be able to perceive facial expressions, such as happiness, sadness, surprise, etc. and make use of the information. The requirements are discussed further in Section 4.

Psychologists and ethologists, instead of merely asking which animals can do X, or at what age or under what conditions young children can do X, should adopt the design-based approach, and constantly ask "how could that work?". We can use that to generate collections of requirements for information-processing architectures and then investigate mechanisms that could support the observed variety of visual functions in robots. Very often, it is not obvious whether a particular theory suffices, so using a theory as a basis for designing, implementing and testing *working* artificial systems is a crucial part of the process of explaining how natural systems work. This paper uses that approach, basing some unobvious functions of vision on detailed analysis of requirements for human-like visual systems. It ends with some speculations about sorts of mechanisms (using a large collections of multi-stable, interlinked dynamical systems) that appear to be required for those functions, which do not yet exist in any known computational models, and which may be hard to identify in neural mechanisms.

Understanding how a complex system works includes knowing what would happen if various aspects of the design were different, or missing. So understanding how humans work requires us to relate the normal human case to other products of biological evolution, to cases of brain damage or genetic brain abnormality, and to possible future engineering products. We can summarise this as follows (building on Sloman, 1982, 1984, 1994, 1995, 2000): We need to study the space of possible sets of requirements or niches, *niche space*, and we need to study the space of possible designs for working systems that can meet different sets of requirements, *design space*. Finally, we need to understand the various relationships between regions of design space and regions of niche space and the complex ecosystem loops in which changes in one design or niche produce changes in other designs or niches.

By analysing the different sets of functions supported by different information-processing architectures we can come up with a theory-based survey of possibilities for a mind or a visual system. For each system design there is a specific "logical topography", a set of possible states, processes and causal interactions that can occur within the architecture, some involving also interactions with the environment. The set of possibilities generated by each architecture can be subdivided and categorised in different ways for different purposes – producing different "logical geographies" (Ryle, 1949). E.g. the purposes of common sense classification are different from the purposes of scientific explanation.

1.3 Varieties of representation: Generalised Languages (GLs)

A particularly important feature of any information-processing system is how it encodes the information it acquires and uses. Researchers designing computational models often have commitments to particular forms of representation, since those are the ones for which they have tools, e.g. tools for modelling neural nets or tools for symbolic computation using trees and graphs. Those commitments can severely restrict the kinds of questions they ask, and the answers they consider.

Pre-verbal children and many non-human animals can perceive and react to processes as

they occur. That requires mechanisms providing the ability to represent changes while they happen. Perhaps the same mechanisms, or closely related mechanisms, can be used to reason about processes that are not happening. If some other primates and very young children use internal "generalised languages" (GLs) as proposed in Sloman (1979) and elaborated in Sloman and Chappell (2007), that suggests that GLs supporting structural variability and compositional semantics evolved before external human languages used for communication, and that GLs also precede the learning of communicative language in individual humans.

We shall later give several examples of human visual competences, including geometric reasoning competences, that seem to require use of GLs, for example in Sections 3.1, 4.1, 4.4, 5, 5.8, and 6. The suggestion that GLs are used for all these purposes, including the representation of processes at different levels of abstraction, poses deep questions for brain science, as we'll see later.

2 Wholes and parts: beyond "scaling up"

The rest of this paper addresses a subset of the requirements for visual systems. Designs meeting those requirements must be able to work with designs for other components: they must "scale out".

2.1 Putting the pieces together: "scaling out"

To study vision we need to consider a larger set of problems, including understanding information-processing requirements for a human-like (or chimp-like, or crow-like) organism or robot to perceive, act and learn in the environment. Instances of designs for mechanisms providing particular competences (e.g. visual competences) must be capable of interacting with other components in a larger design that satisfies requirements for a *complete* animal or robot.

This requirement to "scale out" contrasts with the frequently mentioned need to "scale up", namely coping successfully with larger and more complex inputs. Many human competences do not scale up, including parsing, planning, and problem-solving competences. It is possible to produce a highly efficient implementation of some competence that scales up very well but cannot be integrated with other human competences. Most current implementations of linguistic processing, vision, planning, or problem-solving do not scale out, because they are designed to work on their own in test situations, not to work with one another.

Being biologically inspired is not enough. Using observed human performance in a laboratory to specify a computing system that produces the same error rates, or learning rates or reaction times, need not lead to models that "scale out", i.e. can be extended to form part of a larger model meeting a much wider range of requirements. Not all mechanisms that perform like part of a system are useful parts of something that performs like the whole system. For example, we have machines that meet one requirement for a computer model of a good human chess player, namely that the machine is able to beat most human players. But if you add other requirements, such as that the player should be able to teach a weaker player, by playing in such a way as to help the weaker player learn both from mistakes and

from successes, then the obvious designs that do well as competent chess-players are not easily extendable to meet the further requirements. Similarly, a computer system can be comparable to humans at recognising certain classes of objects in images, without being extendable so that it can intelligently use the fact that same object looks different from different viewpoints. However, the ability of a design with functionality F1 to be expanded to include functionality F2 is partly a matter of degree: it depends on how much extra mechanism is needed to provide F2.

2.2 Biological *mechanisms* vs biological *wholes*

The relevance of biological *mechanisms* for AI (e.g. neural nets and evolutionary computations) has been acknowledged for several decades. What is relatively new in the computational modelling community is moving beyond studying what *biological mechanisms* can do, to finding out what *whole animals*, can do and how they do it, which is one way of studying *requirements* to be met by new designs. Brooks (1991) also recommended building complete systems, though his rejection of representations went too far, as we'll see.

Finding out how information flows around brain circuits connected to the eyes, has received a lot of attention prompted by new non-invasive brain-imaging machinery, but does not tell us what the information is, how it is represented, or what the information is used for. That can include things as diverse as fine-grained motor control, triggering saccades, aesthetic enjoyment, recognition of terrain, finding out how something works, or control of intermediate processes that perform abstract internal tasks.

It is possible to make progress in the case of simple organisms where neurophysiology corresponds closely to information-processing functions. Measuring brain processes may be informative in connection with evolutionarily older, simpler, neural circuits (e.g. some reflexes), but many newer functions are extremely abstract, and probably only very indirectly related to specific neural events. In those cases, results of brain imaging may show some correlations without telling us much about what the virtual machines implemented in brains are doing. For example, consider speakers of two very different languages with different phonetic structures, grammars, and vocabularies who hear and understand reports with the same semantic content. If similar parts of their brains show increased activity: that will not answer any of the deep questions about how understanding works. E.g. we shall be no nearer knowing how to produce a working model with the same functionality.

The journey towards full understanding has far to go, and may even be endless, since human minds, other animal minds and robot minds can vary indefinitely. Evaluating intermediate results on this journey is difficult. A Popperian emphasis on the need for falsifiable hypotheses can slow down scientific creativity. Instead, as proposed by Lakatos (1980), we need to allow that distinguishing degenerative and progressive research programmes can take years, or decades, and we need to understand explanations of possibilities, as well as laws (Sloman, 1978, ch.2).

2.3 Vision and mathematical reasoning

In his *Critique of Pure Reason*, Kant had claimed, in opposition to Hume, that there are ways of discovering new truths that extend our knowledge (i.e. they are "synthetic", not analytic) and which are not empirical. When trying to prove a theorem, mathematicians frequently use the ability to *see* both structural relationships, e.g. relations between overlapping figures, and also the possibility of changing such relationships, e.g. drawing a new construction line. Even without making the change it is often possible to visualise the consequences of such a change. If you look at a triangle with vertices labelled A, B and C, or simply imagine looking at one, you can *see* that it is possible to draw a straight line from any vertex, e.g. A, to the midpoint of the opposite side. You may also be able to work out that the two resulting triangles *must* have the same area.

Not all mathematical discoveries are based on visual reasoning. For example, very different discoveries, some of them documented in Sloman (1978, ch.8), occur as a child learns to count, and then discovers different uses for the counting process and different features of the counting process, such as the fact that the result of counting a collection of objects is not altered by rearranging the objects but can be altered by breaking one of the objects into two objects. Such mathematical discoveries depend on perceiving invariant structures and relationships in procedures that can be followed. Simultaneous perception of spatial and temporal relationships can lead to the discovery that any one-to-one mapping between elements of two finite sets can be converted into any other such mapping by successively swapping ends of mappings, This sort of discovery requires quite abstract visual capabilities. A self-monitoring architecture Sloman (2008) seems to be needed, allowing one process to observe that another process has consequences that do not depend on the particularities of the example, and which are therefore *necessary* consequences of the procedure.

2.4 Development of visual competences

Many people can see and think about geometrical possibilities and relationships. Very young children cannot see all of them, let alone think about them. Why not? And what has to change in their minds and brains to enable them to see such things? There are developments that would not normally be described as mathematical, yet are closely related to mathematical competences.

For example, a very young child who can easily insert one plastic cup into another may be able to lift a number of cut-out pictures of objects from recesses, and know which recess each picture belongs to, but be *unable* to get pictures back into their recesses: the picture is placed in roughly the right location and pressed hard, but that is not enough. The child apparently has not yet extended his or her ontology to include boundaries of objects and alignment of boundaries. Such learning may include at least three related aspects: (a) developing new forms of representation; (b) extending the learner's ontology to allow new kinds of things that exist; and (c) developing new ways of manipulating representations, in perception and reasoning. What the child playing with puzzle pieces has to learn, namely facts about boundaries and how they constrain possible movements, is a precursor to being able to think mathematically about bounded regions of a plane. Later mathematical education will build on general abilities to see structures and processes and to see how structures can constrain or facilitate processes,

as illustrated in Sauvy and Suavy (1974). We shall see that this is related to perception of proto-affordances.

3 Affordance-related visual competences: seeing processes and possibilities

3.1 Perceiving and reasoning about changes and proto-affordances

Visual, geometrical, reasoning capabilities depend on (a) the ability to attend to parts and relationships of a complex object, including "abstract" parts like the midpoint of a line and collinearity relationships, (b) the ability to discern the possibility of changing what is in the scene, e.g. adding a new line, moving something to a new location, (c) the ability to work out the consequences of making those changes, e.g. working out which new structures, relationships and further possibilities for change will come into existence.

Both the ability to see and make use of affordances and the ability to contemplate and reason about geometric constructions depend on a more primitive and general competence, namely the ability to see not only structures but also *processes*, and the closely related ability to see *the possibility* of processes that are not actually occurring, and also *constraints* that limit those possibilities. I call such possibilities and constraints "proto-affordances".

Gibson's affordances (J. J. Gibson, 1979) were concerned only with opportunities for *actions* that *the perceiver* could perform, whereas normal humans can see what is common between processes that they produce, processes that others produce and processes that are not parts of any intentional actions, e.g. two surfaces coming together. They can perceive and think about the *possibility* of processes occurring without regard to what they can do or what can affect them, e.g. noticing that a rock could start rolling down a hillside. This involves seeing proto-affordances, i.e. seeing that certain processes can and others cannot occur in a certain situation.

That ability underlies both the ability to perceive "vicarious" affordances, namely affordances for others, and also the ability to think about possible occurrences such as rain falling or the wind blowing tomorrow, without specifying a viewpoint. This uses an "exosomatic ontology", making it possible to refer to entities and processes that can exist outside the body, independently of any sensory or motor signals. A pre-verbal child or non-verbal animal that can see and reason about such proto-affordances and vicarious affordances is probably using a spatial GL to represent the possibilities, as suggested in Section 1.3.

Perceiving vicarious affordances for predators or for immature offspring can be biologically very important. It is an open research question which animals can reason about vicarious affordances, though both pre-verbal human children and some chimpanzees can perceive and react to affordances for others, as shown by Warneken and Tomasello (2006), illustrated in videos available at http://email.eva.mpg.de/~warneken/video.htm.

3.2 Evolutionary significance of independently mobile graspers

There are commonalities between affordances related to doing things with left hand, with right hand, with both hands, with teeth and with tools such as tongs or pliers. In principle it is possible that all the means of grasping are represented in terms of features of the sensorimotor signals involved as in Lungarella and Sporns (2006), but the variety of such patterns is astronomical. If, however, grasping is represented more abstractly, in terms of 3-D relations between surfaces in space, using an amodal form of representation and an exosomatic ontology referring to things outside the body, the variety of cases can be considerably reduced: for instance very many types of grasping involve two surfaces moving together with an object between. Such abstract representation can be used for high level planning of actions involving grasping, including, the common requirement for the two grasping surfaces to be further apart during the approach than the diameter of the thing to be grasped. More detailed information about specific cases can then be used either when planning details, or during servo-controlled action execution. I suspect that biological evolution long ago "discovered" the enormous advantages of amodal, exosomatic, representations and ontologies as compared with representations of patterns in sensorimotor signals.

Although 2-D image projections are often helpful for controlling the fine details of an action during visual servoing (as noted in Sloman, 1982), using an exosomatic ontology including 3-D spatial structures and processes, rather sensorimotor signal patterns, makes it possible to learn about an affordance in one situation and transfer that learning to another where sensor inputs and motor signals are quite different: e.g. discovering the consequences of grasping an object with one hand then transferring what has been learned to two-hand grasping, or biting. This assumes that the individual can acquire generic, re-usable, mappings between 3-D processes in the environment and sensor and motor signal patterns.

4 Towards a more general visual ontology

4.1 Proto-affordances and generative process representations

Affordances for oneself and for others depend on the more fundamental "proto-affordances". A particular proto-affordance, such as the potential of one object to impede the motion of another, can be the basis for many action affordances. E.g. it could produce a negative affordance for an agent trying to push the moving object towards some remote location, or a positive affordance for a individual wishing to terminate the motion of a moving object.

An animal's or machine's ability to discover and represent proto-affordances, to combine and manipulate their representations, allows a given set of proto-affordances to generate a huge variety of affordances, including many never previously encountered, permitting creative problem-solving and and planning in new situations. The ability to combine proto-affordances to form new complex affordances apparently enabled the New Caledonian crow Betty to invent several ways of transforming a straight piece of wire into a hook in order to lift a bucket of food from a glass tube (Weir, Chappell, & Kacelnik, 2002).

Most AI vision researchers, and many psychologists assume that the sole function of visual perception is acquiring information about objects and processes that exist in the

environment, whereas a major function of vision (in humans and several other species) seems to include acquiring information from the environment about *what does not exist but could exist*. Mechanisms for seeing processes may also be involved in reasoning about consequences of possible processes. Such predictive and manipulative abilities are not innate, but develop over time. Examples in human infants and children are presented in E. J. Gibson and Pick (2000), though no mechanisms are specified.

4.2 Complex affordances: Combining process possibilities

Some animals learn, by playing in the environment, that affordances can be *combined* to form more complex affordances, because *processes can be combined to form more complex processes.* Reasoning about such complex processes and their consequences depends on the ability to combine simpler proto-affordances to form more complex ones.

Because processes occur in space and time, and can have spatially and temporally related parts, they can be combined in at least the following ways:

- processes occurring in sequence can form a more complex process;
- two processes can occur at the same time (e.g. two hands moving in opposite directions);
- processes can overlap in time, e.g. the second starting before the first has completed;
- processes can overlap in space, for example a chisel moving forwards into a rotating piece of wood;
- one process can modify another, e.g. squeezing a rotating wheel can slow down its rotation;
- one process can launch another, e.g. a foot kicking a ball.

The ability to represent a sequence of processes is part of the ability to form plans prior to executing them. It is also part of the ability to predict future events, and to explain past events, but this is just a special case of a more general ability to combine proto-affordances. How do animal brains represent a wide variety of structures and processes?

There have been various attempts to produce systematic ways of generating and representing spatial structures. However, the demands on a system for representing spatial *processes* (e.g. translating, rotating, stretching, compressing, shearing twisting, etc.) are greater than demands on generative specifications of spatial *structures*. The extra complexity is not expressible just by adding an extra dimension to a vector. The set of possible perceivable processes can be constrained by a context, but the remaining set can be very large, e.g. in a child's playroom, a kitchen, a group of people at a dinner table, a garden, a motorway, various situations in which birds build nests, and many more. A theory of how biological vision works must explain what kinds of information about spatial processes particular animals are capable of acquiring, how the information is represented, how it is used, and how the ability to acquire and use more kinds of information develops. It seems that in order to accommodate the variety of processes humans and other animals can perceive and understand, they will need forms of representation that have the properties we ascribed to spatial GLs in Section 1.3, with the benefits described in Sloman (1971).

4.3 Varieties of learning about processes

In the first few years of life, a typical human child must learn to perceive and to produce many hundreds of different sorts of spatial process, some involving its own body, some involving movements of other humans and pets, and some involving motion of inanimate objects, with various causes. Examples include topological processes where contact relationships, containment relationships, alignment relationships go into and out of existence and metrical processes where things change continuously. The very same physical process can include both metrical changes, as something is lowered into or lifted out of a container and discrete topological changes as contact, containment, or overlap relationships between objects and spatial regions or volumes change.

In each context there are different sets of 'primitive' processes and different ways in which processes can be combined to form more complex processes. Some simple examples in a child's environment might include an object simultaneously moving and rotating, where the rotation may be *closely coupled* to the translation, e.g. a ball rolling on a surface, or *independent of* the translation, e.g. a frisbee spinning as it flies. Other examples of closely coupled adjacent processes are: A pair of meshed gear wheels rotating; a string unwinding as an axle turns; a thread being pulled through cloth as a needle is lifted; a pair of laces being tied; a bolt simultaneously turning and moving further into a nut or threaded hole; a sleeve sliding on an arm as the arm is stretched; and sauce in a pan moving as a spoon moves round in it.

Many compound processes arise when a person or animal interacts with a physical object. Compound 3-D processes are the basis of an enormous variety of affordances. For example, an object may afford grasping, and lifting, and as a result of that it may afford the possibility of being moved to a new location. The combination of grasping, lifting and moving allows a goal to be achieved by performing a compound action using the three affordances in sequence. The grasping itself can be a complex process made of various successive sub-processes, and some concurrent processes. However, other things in the environment can sometimes obstruct the process: as in the chair example below (Figure 2).

In addition to producing physical changes in the environment, more abstract consequences of actions typically include the existence of new positive and negative affordances. The handle on a pan lid may afford lifting the lid, but once the lid is lifted not only is there a new physical situation, there are also new action affordances and epistemic affordances, e.g. because new *information* is available with the lid off. The action of moving closer to an open door also alters epistemic affordances (Figure 1).

4.4 Reasoning about interacting spatial processes

Processes occurring close together in space and time can interact in a wide variety of ways, depending on the precise spatial and temporal relationships. It is possible to learn *empirically* about the consequences of such interactions by observing them happen, and collecting statistics to support future predictions, or formulating and testing universal generalisations. However, humans and some other animals sometimes need to be able to *work out* consequences of novel combinations, for example approaching a door that is shut, while carrying something in both hands, for the first time. It does not take a genius to work out that an elbow can be used to depress the handle while pushing to open the door.

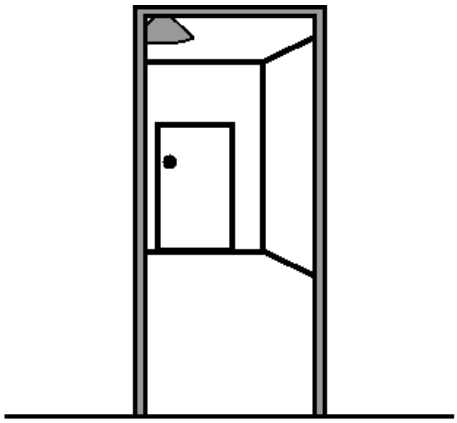

Figure 1: *As you move nearer the door you will have access to more information about the contents of the room, and as you move further away you will have less. Moving left or right will change the information available in a different way.*

E. J. Gibson and Pick (2000) state on page 180 that the affordance of a tool can be discovered in only two ways, by *exploratory activities* and by *imitation*. They apparently failed to notice a third way of discovering affordances, namely *working out* what processes are possible when objects are manipulated, and what their consequences will be.

A requirement for human-like visual mechanisms is that they should produce representations that can be used for *reasoning* about novel spatial configurations and novel combinations of processes, i.e. the kind of reasoning that led to the study of Euclidean geometry. (An example of the need to "scale out".) Likewise, young children can reason about spatial processes and their implications long before they can do logic and algebra and to some extent even before they can talk. As remarked in Sloman and Chappell (2007), this has implications for the evolution and development of language.

By "exploratory activities" Gibson and Pick referred to physical exploration and play. The missing alternative, working things out, can also involve exploratory activities, but the explorations can be done with *representations* of the objects and processes instead of using the actual physical objects. The representations used can either be entirely mental, e.g. visualising what happens when some geometrical configuration is transformed, or physical, for instance 2-D pictures representing 3-D structures, with processes represented by marks on the pictures (Sloman, 1971; Sauvy & Suavy, 1974).

Although reasoning with representations in place of the objects is sometimes fallacious, nevertheless, when done rigorously, it is mathematical inference rather than empirical inference. As Lakatos (1976) showed, the methods of mathematics are far from infallible. But that does not make them empirical in the same way as the methods of the physical sciences are. The issues are subtle and complex: see (Sloman, 2008).

The ability to think about and reason about novel combinations of familiar types of process is often required for solving new problems, for example realising for the first time that instead of going from A to B and then to C it is possible to take a short cut from A to C, or realising that a rigid circular disc can serve as well as something long and thin (like a screwdriver) to lever something up.

4.5 Creative reasoning about processes and affordances

Gibson and Pick describe various kinds of "prospectivity" that develop in children but they focus only on empirically learnt kinds of predictive rules, and ignore the child's growing ability to design and represent novel complex multi-stage processes and work out that they can achieve some goal.

The ability to work things out is facilitated and enhanced by the ability to form verbal descriptions, as in inventing stories, but the ability to use a public human language is not a prerequisite for the ability, as can be seen in the creative problem-solving of pre-verbal children and some other animals. Both seem to be able to work out the consequences of some actions using intuitive geometric and topological reasoning.

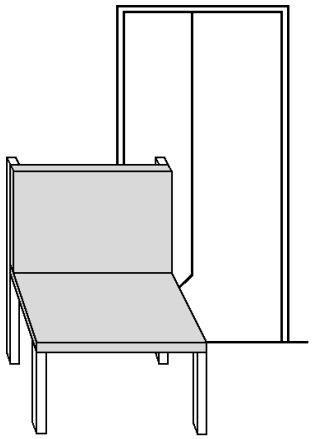

Figure 2: *A person trying to move a chair that is too wide to fit through a door can work out how to move it through the door by combining a collection of translations and 3-D rotations about different axes, some done in parallel, some in sequence. Traditional AI planners cannot construct plans involving continuous interacting actions.*

Figure 2 illustrates affordances that interact in complex ways when combined, because of the changing spatial relationships of objects when processes occur. A large chair may afford lifting and carrying from one place to another, and a doorway may afford passage from one room to another. But the attempt to combine the two affordances may fail when the plan is tried, e.g. if it is found during execution that the chair is too wide to fit through the doorway. After failing, a learner may discover that a combination of small rotations about different axes combined with small translations can form a compound process that results in the chair getting through the doorway. An older child may be able to see the possibility of the whole sequence of actions by visualising the whole process in advance, working out by visual reasoning how to move the chair into the next room, and then doing it. It is not at all clear what sort of brain mechanism or computer mechanism can perform that reasoning function or achieve that learning.

A child can also learn that complex affordances with positive and negative aspects can be disassembled so as to retain only the positive aspects. Children and adults often have to perform such "process de-bugging". A colleague's child learnt that open drawers could be closed by pushing them shut. The easiest way was to curl his fingers over the upper edge

of the projecting drawer and push, as with other objects. The resulting pain led him to discover that the pushing could be done, slightly less conveniently, by flattening his hand when pushing, achieving the goal without the pain. Being able to do geometrical reasoning enables a child who is old enough to work out *why* pushing with a flat hand prevents fingers being caught between the two surfaces.

4.6 The need for explanatory theories

Many humans can perform such reasoning by visualising processes in advance of producing them, but it is not at all clear what representations are used to manipulate information about the shapes and affordances of the objects involved.

There has been much work on giving machines with video cameras or laser scanners the ability to construct representations of 3-D structures and processes that can be projected onto a screen to show pictures or videos from different viewpoints, but, except for simple cases, robots still lack versatile spatial representation capabilities that they can use for multiple purposes such as manipulating objects, planning manipulations, and reasoning about them.

AI planning systems developed in the 1960s, such as the STRIPS planner (Fikes & Nilsson, 1971) and more complex recent planners (surveyed in Ghallab, Nau, & Traverso, 2004), all make use of the fact that knowledge about affordances can be abstracted into reusable information about the preconditions and effects of actions. Doing that provides a new kind of *cognitive* affordance: concerned with acting on information structures, demonstrating the possibility of combining knowledge about simple actions to provide information about complex actions composed of simple actions. However that work assumed that the information about actions and affordances can be expressed in terms of implications between propositions expressed in a logical formalism. Planning processes search for a sequence (or partially ordered network) of discrete actions that will transform the initial problem state into the desired goal state. But we need a richer mechanism to handle actions that involve interactions between continuously changing structural relations, like the changes that occur while an arm-chair is being rotated and translated simultaneously, or a sink is being wiped clean with a cloth.

Sloman (1971) challenged the then AI orthodoxy by arguing that intelligent machines would need to be able to reason geometrically as well as logically, and that some reasoning with diagrams should be regarded as being valid and rigorous, and in some cases more efficient than reasoning using logic, because logical representations are topic-neutral and sometimes lose some of the domain structure that can be used in searching for proofs. But it became clear that, although many people had independently concluded that AI techniques needed to be extended using spatial reasoning techniques, neither I nor anyone else knew how to design machines with the right kinds of abilities, even though there were many people working on giving machines the ability to recognise, analyse and manipulate images, or parts of images, often represented as 2-D rectangular arrays, though sometimes in other forms, e.g. using log-polar coordinates, e.g. Funt (1977). Other examples were presented in Glasgow, Narayanan, and Chandrasekaran (1995). More recent examples are Jamnik, Bundy, and Green (1999); Winterstein (2005).

4.7 An objection: blind mathematicians

It could be argued that the description of mathematical reasoning as "visual" must be wrong because people who have been blind from birth can reason about shapes and do logic and mathematics even though they cannot see (Jackson, 2002). That argument ignores the fact that some of the visual apparatus produced by evolution to support seeing and reasoning about structures and processes in the environment is in brain mechanisms that perform some of their functions without optical input: like the normal ability to see what *can* change in a situation when those changes are not occurring, or the ability to visualise a future sequence of actions that are not now being performed, and therefore cannot produce retinal input.

People who have been blind from birth may still be using the bulk of the visual system that evolved in their ancestors, just as sighted people may be using it when they dream about seeing things, and when they visualise diagrams with their eyes shut. The fact that individuals with different disabilities acquire a common humanity via different routes is an indication of how much of human mentality is independent of our specific form of embodiment. (Though not independent of the forms of our ancestors!)

4.8 Use of abstraction is not metaphor [omitted in published version]

The claim that visual mechanisms using abstract patterns can support reasoning of the sort done in mathematics should not be confused with the common claim that spatial concepts are used as metaphors for non-spatial topics, for instance the claim that we must use spatial metaphors in thinking about numbers or about time. Such claims are based on a failure to understand that there are high level domain-neutral concepts (e.g. "order", "between", "more than", "is a subset of") which are *equally applicable* to many different domains for example because all those domains have some common topological features. There are many totally ordered sets, including points on a line and times in a temporal interval. Seeing that there is an abstract, generally applicable, pattern (i.e. total ordering of a set), is different from seeing structural mappings between partly similar instances.

Using manipulable structures in one domain to represent patterns in another domain for the purpose of reasoning about them, because both domains share some features (without necessarily being isomorphic), is a different matter from using the first domain as a metaphor for the second domain: metaphors do not provide valid inferences, although they are often usefully suggestive.

These ideas are not new: Several famous examples of visual proofs are presented in Nelsen (1993). Many theorists, including great logicians such as Frege (see Merrick, 2006) and mathematicians such as Poincaré, 1905, have pointed at the use of visualisation and spatial reasoning capabilities in mathematics and logic. It will be clear from earlier comments about exosomatic ontologies and representations in Section 3 that I do not agree with Poincaré's claim "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.

5 Studying mechanisms *vs.* studying requirements.

5.1 The importance of requirements

Researchers often launch into seeking designs, on the assumption that the requirements are clear, e.g. because they think everyone knows what visual systems do, or because they merely try to model behaviours observed in particular experiments, or give too much importance to particular benchmark tests. This focus can lead researchers (and their students) to ignore the question of what else needs to be explained. As already remarked, successful models or explanations of a limited set of behaviours may not scale out. A deeper problem, is that there is as yet not even a generally agreed ontology for discussing requirements and designs: we do not have an agreed set of concepts for describing cognitive functions with sufficient detail to be used in specifying requirements for testable working systems.

5.2 Mistaken requirements

Sloman (1989) lists nine common mistaken assumptions about vision systems. This paper extends that list. For example it is tempting to suppose that a requirement for 3-D vision mechanisms is that they must construct a 3-D model of the perceived scene, with the components arranged within the model isomorphically with the relationships in the scene. Such a model can be used to generate graphical displays showing the appearance of the environment from different viewpoints. However, impossible figures like Escher's drawings and the Penrose triangle show that we are able to see a complex structure without building such a model, for there cannot be a model of an impossible scene. Perhaps, instead, a visual system constructs a large collection of fragments of information of various sorts about surfaces, objects, relationships, possible changes and constraints on changes in the scene, with most of the information fragments represented in registration with the optic array, though in an amodal form. Unlike an integrated model, this could support many uses of the information.

This idea generalises the notion of an "aspect graph", in which distinct 2-D views of a 3-D object are linked to form a graph whose edges represent actions that a viewer can perform, such as moving left, or right or up or down, to alter the view. This idea can be generalised so that more actions are included, such as touching or pushing, or grasping an object and more changes are produced such as two objects coming together or moving apart, or an object rotating, or sliding or tilting, or becoming unstable, etc.

Much current research in vision and robotics focuses on mechanisms that manipulate only representations of sensorimotor phenomena, e.g. statistical patterns relating multi-modal sensor and motor signals, and making no use of amodal exosomatic ontologies and forms of representation. Exceptions include SLAM (self localisation and mapping) mechanisms that create an exosomatic representation of building or terrain layout, and use that to plan routes. The great advantage of exosomatic representations is sometimes that a single representation of a process in the environment, such as two fingers grasping a berry need not specify variations in sensor and motor signals that depend on precisely how the grasping is done and the viewpoint from which the process is observed, and which other objects may partially occlude relevant surfaces.

If a visual system's representation of a 3-D scene is made up of many piecemeal

representations of fragments of the scene and the possible effects of processes involving those fragments, then in principle those fragments could form an inconsistent totality. So it would seem that an intelligent robot or animal must constantly check whether it has consistent percepts. However, since no portion of the 3-D environment is capable of containing impossible objects, there is normally no need for such a visual system to check that all the derived information is consistent, except in order to eliminate ambiguities, which can often be done more easily by a change of viewpoint. This is just as well since in general consistency checking is an intractable process.

5.3 Obvious and unobvious requirements: ontological blindness

Many people, e.g. Neisser (1967); Fidler and Leonardis (2007), have noticed the need for hierarchical decomposition of complex perceived objects and the usefulness of a mixture of top down, bottom up and middle out processing in perception of such objects. In addition to such *part-whole* hierarchies there are also *ontological* layers, as illustrated in the Popeye program described in Chapter 9 of Sloman (1978). The program was presented with images made of dots in a rectangular grid, such as Figure 3, which it analysed and interpreted in terms of:

- a layer of dot configurations (which could, for example, contain collections of collinear adjacent dots);
- a layer of line configurations, where lines are interpretations of "noisy" sets of collinear dots, and can form configurations such parallel pairs, and junctions of various sorts;
- a layer of 2-D overlapping, opaque 'plates' with straight sides and rectangular corners, which in Popeye were restricted to the shapes of cut-out capital letters, such as "A", "E", "F", "H" etc. represented in a noisy fashion by collections of straight line segments;
- a layer of sequences of capital letters represented by the plates, also in a "noisy" fashion because the plates could be jumbled together with overlaps;
- a layer of words, represented by the letter sequences.

The program illustrated the need to use ontologies at different levels of abstraction processed concurrently, using a mixture of top-down, bottom-up and middle-out processing, where lower levels are not *parts* of the higher levels, though each ontological layer has part-whole hierarchies. Going from one ontological layer to another is not a matter of grouping parts into a whole, but *mapping* from one sort of structure to another, for instance, interpreting configurations of dots as representing configurations lines, and interpreting configurations of lines as representing overlapping 2-D plates.

Text represented using several ontological layers may be regarded as a very contrived example, but similar comments about ontological layers can be made when a working machine is perceived, such as the internals of an old fashioned clock. There will be sensory layers concerned with changing patterns of varying complexity in the optic array. A perceiver will have to interpret those changing sensory patterns as representing 3-D surfaces and their relationships, some of which change over time. At a higher level of abstraction there are functional categories of objects, e.g. levers, gears, pulleys, axles, strings, and various more or less complex clusters of such objects, such as escapement mechanisms.

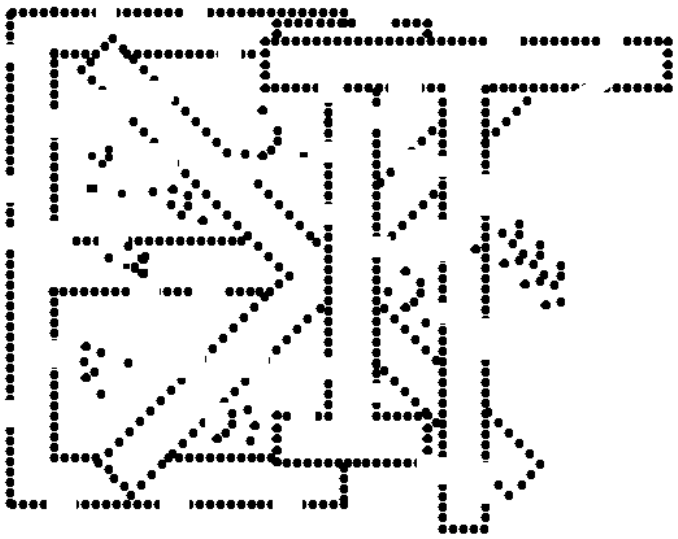

Figure 3: *This illustrates configurations of dots presented to the Popeye program in Chapter 9 of Sloman (1978), which attempted to find a known word by concurrently looking for structures in several ontological layers, with a mixture of top-down, bottom-up and middle-out influences. If the noise and clutter were not too bad, the program, like humans could detect the word before identifying all the letters and their parts. It also degraded gracefully.*

Many vision researchers appreciate the need for a vision system to move between a 2-D ontology and a 3-D ontology. For a recent survey see Breckon and Fisher (2005). The need for such layers will be evident to anyone who works on vision-based text-understanding. However it is rare to include as many ontological categories, in different layers as I claim are needed by an intelligent human-like agent interacting with a 3-D environment, or to relate those layers to different processing layers in the central architecture as explained in Sloman (2001).

5.4 Seeing mental states

Perception of intelligent agents in the environment involves yet another level of abstraction, insofar as some perceived movements are interpreted as actions with purposes, for instance a hand moving towards a cup. If eyes and face are visible, humans will often see not just actions but also mental states, such as focus of attention in a certain direction, puzzlement, worry, relief, happiness, sadness, and so on. Insofar as these are all *seen* rather than inferred in some non-visual formalism, the percepts will be at least approximately in registration with the optic array. Happiness is seen in one face and not in another. The requirement for perceptual mechanisms to use an ontological layer that includes mental states raises many problems that will not be discussed here, for example the need to be able to cope with referential opacity. Representing something that is itself an information user requires meta-semantic competences. These subtleties are ignored by researchers who train computer programs to label pictures of faces using words such as "happy", "angry", and claim that their programs can recognise emotional states.

5.5 Perceiving 2-D processes in the optic array

2-D processes involving changes in the optic array are also important, as J.J. Gibson pointed out. As noted in Sloman (1982), apart from perception of static scenes, vision is also required for online control of continuous actions (visual servoing) which requires different forms of representation from those required for perception of structures to be described, remembered, used in planning actions, etc.

Sometimes a 2-D projection is more useful than a 3-D description for the control problem, as it may be simpler and quicker to compute, and can suffice for particular task, such as steering a vehicle through a gap. But it is a mistake to think that only continuously varying percepts are involved in online visual control of actions: there is also checking whether goals or sub-goals have been achieved, whether the conditions for future processes have not been violated, whether new obstacles or new opportunities have turned up, and so on. Those can involve checking discrete conditions.

Unfortunately research on ventral and dorsal streams of neural processing has led some researchers (e.g. Goodale & Milner, 1992) to assume that control of action is separate from cognition, or worse, that spatial perception ("where things are") is a completely separate function from categorisation ("what things are"), apparently ignoring the fact that what an object is may depend on where its parts are in relation to one another, or where it is located in a larger whole.

5.6 Layered ontologies

We have seen, in Section 5.3, that in addition to part-whole decomposition, perception can use layered ontologies. For example, one sub-ontology might consist entirely of 2-D image structures and processes, whereas another includes 3-D spatial structures and processes, and another kinds of 'stuff' of which objects are made and their properties (e.g. rigidity, elasticity, solubility, thermal conductivity, etc.), to which can be added mental states and processes, e.g. seeing a person as happy or sad, or as intently watching a crawling insect. The use of multiple ontologies is even more obvious when what is seen is text, or sheet music, perceived using different geometric, syntactic, and semantic ontologies.

From the combination of the two themes (a) the content of what is seen is often processes and process-related affordances, and (b) the content of what is seen involves both hierarchical structure and multiple ontologies we can derive a set of requirements for a visual system that makes current working models seem very inadequate.

Many people are now working on how to cope with pervasive problems of noise and ambiguity in machine vision systems. This has led to a lot of research on mechanisms for representing and manipulating uncertainty. What is not always noticed is that humans have ways of seeing high level structures and processes whose descriptions are impervious to the low level uncertainties: you can see that there definitely is a person walking away from you on your side of the road and another walking in the opposite direction on the other side of the road, even though you cannot tell the precise locations, velocities, accelerations, sizes, orientations and other features of the people and their hands, feet, arms, legs, etc. The latter information may be totally irrelevant for your current purposes.

Some notes on this can be found in this discussion paper on predicting affordance changes, including both action affordances and epistemic affordances: http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0702

5.7 Seeing is prior to recognising

Much research on visual perception considers only one of the functions of perception, namely recognition. It is often forgotten that there are many things we can see, and can act in relation to, that we cannot recognise or label, and indeed that is a precondition for learning to categorise things.

When you see a complex new object that you do not recognise you may see a great deal of 3-D structure, which includes recognising many types of *surface fragment*, including flat parts, curved parts, changes of curvature, bumps, ridges, grooves, holes, discontinuous curves where two curved parts meet, and many more. In addition to many surface fragments, many of their relationships are also seen, along with relationships between objects, and also between different parts of objects. When those relations change we get different *processes* occurring concurrently, some at the same level of abstractions, others at different levels. They can be seen without necessarily recognising the objects involved.

Biological considerations suggest that, for most animals, perception of processes must be the most important function, since perception is crucial to the control of action, in a dynamic, sometimes rapidly changing environment that can include mobile predators and mobile prey, and where different parts of the environment provide different nutrients, shelter, etc. So from this viewpoint perception of structures is just a special case of perception of processes – processes in which not much happens.

Unfortunately, not only has very little (as far as I know) been achieved in designing visual systems that can perceive a wide range of 3-D spatial structures, there is even less AI work on perception of *processes*, apart from things like online control of simple movements which involves sensing one or two changing values and sending out simple control signals, for instance "pole balancing" control systems. There seems also to be very little research in psychology and neuroscience on the forms of representations and mechanisms required for perception of processes involving moving or changing structures, apart from research that merely finds out who can do what under what conditions. Examples of the latter include Heider and Simmel (1944.), Michotte (1962) and Johansson (1973). Finding out which bits of the brain are active does not answer the design questions.

Addressing those deficiencies, including, for instance, explaining how GLs for process representation could work, should be a major goal for future vision research, both in computational modelling but also in neuroscience. Some speculations about mechanisms are presented in Section 6.

5.8 Seeing possible processes: proto-affordances

We have already noted that an important feature of process perception is the ability to consider different ways a process may continue, some of them conditional on other processes intervening, such as an obstacle being moved onto or off the path of a moving object.

Many cases of predictive control include some element of uncertainty based on imprecise measurements of position, velocity or acceleration. This sort of uncertainty can be handled using fuzzy or probabilistic control devices which handle intervals instead of point values.

However there are cases where the issue is not uncertainty or inaccuracy of measurement but the existence of very different opportunities, such as getting past an obstacle by climbing over it, or going round it on the left or on the right. It may be very clear what the alternatives are, and what their advantages and disadvantages are. E.g one alternative may involve a climb that requires finding something to stand on, while another requires a heavy object to be pushed out of the way, and the third requires squeezing through a narrow gap.

The ability to notice and evaluate distinct possible futures is required not only when an animal is controlling its own actions but also when it perceives something else whose motion could continue in different ways. How the ability to detect such vicarious affordances is used may depend on whether the perceived mover is someone (or something) the perceiver is trying to help, or to eat, or to escape from.

In simple cases, prediction and evaluation of alternative futures can make use of a simulation mechanism. But the requirement to deal explicitly with alternative possibilities requires a more sophisticated simulation than is needed for prediction: a predictive simulation can simply be run to derive a result, whereas evaluation of alternatives requires the ability to start the simulation with different initial conditions so that it produces different results. It also requires some way of recording the different results so that they can be used later for evaluation or further processing.

The ability to cope with branching futures in a continuous spatial environment poses problems that do not arise in "toy" discrete grid-based environments. The agent has to be able to chunk continuous ranges of options into relatively small sets of alternatives in order to avoid dealing with explosively branching paths into the future. How to do this may be something learnt by exploring good ways to group options by representing situations and possible motions at a high level of abstraction.

Learning to see good ways of subdividing continuous spatial regions and continuous ranges of future actions involves developing a good descriptive ontology at a higher level of abstraction than sensor and motor signals inherently provide. The structure of the environment, not some feature of sensorimotor signals makes it sensible to distinguish the three cases: moving forward to one side of an obstacle, moving so as to make contact with the obstacle and moving so as to go to the other side.

In addition to "chunking" of possibilities on the basis of differences between opportunities for an animal or robot to move as a whole, there are ways of chunking them on the basis of articulation of the agent's body into independently movable parts. For example, if there are two hands available, and some task requires both hands to be used, one to pick an object up and the other to perform some action on it (e.g. removing its lid) then each hand can be considered for each task, producing four possible combinations. However if it is difficult or impossible for either hand to do both tasks, then detecting that difficulty in advance may make it clear that the set of futures should be pruned by requiring each hand to do only one task, leaving only two options.

In humans, and some other species, during the first few years of life a major function of play and exploration in an infant is providing opportunities for the discovery of many

hundreds of concepts that are useful for chunking sets of possible states of affairs and possible process, and learning good ways to represent them so as to facilitate predicting high level consequences, which can then be used in rapid decision-making strategies.

The ability to perceive not just what is happening at any time but what the possible branching futures are – including, good futures, neutral futures, and bad futures from the point of view of the perceiver's goals and actions, is an aspect of J.J. Gibson's theory of perception as being primarily about *affordances for the perceiver* rather than acquisition of information about some objective and neutral environment. However, I don't think Gibson considered the need to be able to represent, compare and evaluate multi-step branching futures: that would have been incompatible with his adamant denial of any role for representations and computation.

6 Speculation about mechanisms required: new kinds of dynamical systems

Preceding sections have assembled many facts about animal and human vision that help to constrain both theories of how brains, or the virtual machines implemented on brains work, and computer-based models that are intended to replicate or explain human competences. One thing that has not been mentioned so far is the extraordinary speed with which animal vision operates. This is a requirement for fast moving animals whose environment can change rapidly (including animals that fly through tree-tops). An informal demonstration of the speed with which we can process a series of unrelated photographs and extract quite abstract information about them 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.

6.1 Sketch of a possible mechanism

Perhaps we need a new kind of dynamical system. Some current researchers (e.g., Beer, 2000) investigate cognition based on dynamical systems composed of simple "brains" closely coupled with the environment through sensors and effectors. We need to extend those ideas to allow a multitude of interacting dynamical systems, some of which can run decoupled from the environment, for instance during planning and reasoning, as indicated crudely in Figure 4. During process perception, changing sensory information will drive a collection of linked processes at different levels of abstraction. Some of the same processes may occur when possible but non-existent processes are imagined in order to reason about their consequences.

Many dynamical systems are defined in terms of continuously changing variables and interactions defined by differential equations, whereas our previous discussion, e.g. in Section 4.3, implies that we need mechanisms that can represent discontinuous as well as continuous changes, for example to cope with topological changes that occur as objects are moved, or goals become satisfied. Another piece of evidence for such a requirement is the sort of discrete 'flip' that can occur when viewing well known ambiguous figures such as the Necker cube, the duck-rabbit, and the old-woman/young-woman picture. It is significant that such internal flips can occur without any change in sensory input.

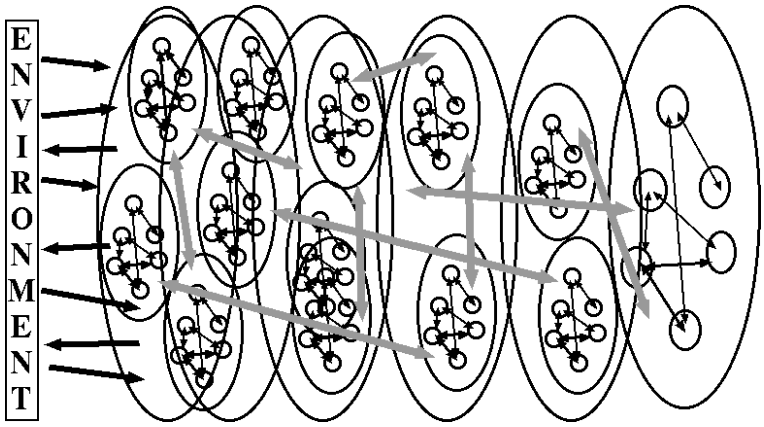

Figure 4: *A crude impressionistic sketch indicating a collection of dynamical systems some closely coupled with the environment through sensors and effectors others more remote, with many causal linkages between different subsystems, many of which will be dormant at any time. Some of the larger dynamical systems are composed of smaller ones. The system does not all exist at birth but is grown, through a lengthy process of learning and development partly driven by the environment, as sketched in Chappell and Sloman 2007*

It is possible that adult human perception depends on the prior construction of a very large number of multi-stable dynamical systems each made of many components that are themselves made of "lower level" multistable dynamical systems. Many of the subsystems will be dormant at any time, but the mechanisms must support rapidly activating an organised, layered, collection of them partly under the control of current sensory input, partly under control of current goals, needs, or expectations, and partly under the control of a large collection of constraints and preferences linking the different dynamical systems.

On this model, each new perceived scene triggers the activation of a collection of dynamical systems driven by the low level contents of the optic array and these in turn trigger the activation of successively higher level dynamical systems corresponding to more and more complex ontologies, where the construction process is constrained simultaneously by general knowledge, the current data, and, in some cases, immediate contextual knowledge. Subsystems that are activated can also influence and help to constrain the activating subsystems, influencing grouping, thresholding, and removing ambiguities, as happened in the Popeye program described in Section 5.3.

As processes occur in the scene or the perceiver moves, that will drive changes in some of the lower level subsystems which in turn will cause changes elsewhere, causing the perceived processes to be represented by internal processes at different levels of abstraction. Some the same mechanisms may be used when when possible but non-existent processes are imagined in order to reason about their consequences.

On this view, a human-like visual system is a very complex multi-stable dynamical system:

- composed of multiple smaller multi-stable dynamical systems
- that are grown over many years of learning,
- that may be (recursively?) composed of smaller collections of multi-stable dynamical

systems that can be turned on and off as needed,

- some with only discrete attractors, others capable of changing continuously,
- many of them inert or disabled most of the time, but capable of being activated rapidly,
- each capable of being influenced by other sub-systems or sensory input or changing current goals, i.e. turned on, then kicked into new (more stable) states bottom up, top down or sideways,
- constrained in parallel by many other multi-stable sub-systems,
- with mechanisms for interpreting configurations of subsystem-states as representing scene structures and affordances, and interpreting changing configurations as representing processes,
- using different such representations at different levels of abstraction changing on different time scales,
- where the whole system is capable of growing new sub-systems, permanent or temporary, some short-term (for the current environment) and some long term (when learning to perceive new things), e.g.
  - learning to read text
  - learning to sight read music
  - learning to play tennis expertly,

  etc.

That specification contrasts with "atomic-state dynamical systems", described in Sloman (1993) as dynamical systems (a) with a fixed number of variables that change continuously, (b)with one global state, (c)that can only be in one attractor at a time (d) with a fixed structure (e.g. a fixed size state vector).

The difficulties of implementing a dynamical system with the desired properties (including components in which spatial GLs are manipulated) should not be underestimated. The mechanisms used by brains for this purpose may turn out to be very different from mechanisms already discovered.

7 Concluding comments

In Sloman (1989) it was proposed that we need to replace 'modular' architectures with 'labyrinthine' architectures, reflecting both the variety of components required within a visual system and the varieties of interconnectivity between visual subsystems and other subsystems (e.g. action control subsystems, auditory subsystems, and various kinds of central systems).

One way to make progress may be to start by relating human vision to the many evolutionary precursors, including vision in other animals. If newer systems did not replace older ones, but built on them, that suggests that many research questions need to be rephrased to assume that many different kinds of visual processing are going on concurrently, especially when a process is perceived that involves different levels of abstraction perceived concurrently, e.g. continuous physical and geometric changes relating parts of visible surfaces

and spaces at the lowest level, discrete changes, including topological and causal changes at a higher level, and in some cases intentional actions, successes, failures, near misses, etc. at a still more abstract level. The different levels use different ontologies, different forms of representation, and probably different mechanisms, yet they are all interconnected, and all in partial registration with the optic array (not with retinal images, since perceived processes survive saccades).

It is very important to take account of the fact that those ontologies are not to be defined only in terms of what is going on inside the organism (i.e. in the nervous system and the body) since a great deal of the information an organism needs is not about what is happening in it, but what is happening in the environment, though the environment is not some unique given (as implicitly assumed in Marr's theory of vision (1982), for example) but is different for different organisms, even when located in the same place. They have different niches.

As Ulric Neisser pointed out in his (1976) it is folly to study only minds and brains without studying the environments those minds and brains evolved to function in.

One of the major points emphasised here is that coping with our environment requires humans to be able to perceive, predict, plan, explain, reason about, and control processes of many kinds, and some of that ability is closely related to our ability to do mathematical reasoning about geometric and topological structures and processes. So perhaps trying to model the development of a mathematician able to do spatial reasoning will turn out to provide a major stepping stone to explaining how human vision works and producing convincing working models. Perhaps it will show that Immanuel Kant got something right about the nature of mathematical knowledge, all those years ago.

8 Acknowledgements

Many of the ideas reported in this paper were developed as part of the requirements analysis activities in the EU funded CoSy project www.cognitivesystems.org. I am especially indebted to Jeremy Wyatt and other members of the Birmingham CoSy team. Mohan Sridharan kindly read and commented on a draft. Jackie Chappell has helped with biological evidence, the work on nature-nurture tradeoffs and the difference between Humean and Kantian causal understanding. Dima Damen helped me improve Figure 4. Shimon Edelman made several very useful comments. I thank Dietmar Heinke, the editor, for organising the workshop, driving the production of this volume and his patience with me.

References

Beer, R. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, *4*(3), 91–99. ((http://vorlon.case.edu/ beer/Papers/TICS.pdf))

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

Breckon, T. P., & Fisher, R. B. (2005). Amodal volume completion: 3D visual completion. *Computer Vision and Image Understanding*(99), 499–526. (doi:10.1016/j.cviu.2005.05.002)

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

Fidler, S., & Leonardis, A. (2007). Towards Scalable Representations of Object Categories: Learning a Hierarchy of Parts. In *Proceedings Conference on Computer Vision and Pattern Recognition,* (pp. 1–8). Minneapolis: IEEE Computer Society. (http://vicos.fri.uni-lj.si/data/alesl/cvpr07fidler.pdf)

Fikes, R., & Nilsson, N. (1971). STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, *2*, 189–208.

Funt, B. V. (1977). Whisper: A problem-solving system utilizing diagrams and a parallel processing retina. In *Ijcai* (p. 459-464). Cambridge, MA: IJCAI'77.

Ghallab, M., Nau, D., & Traverso, P. (2004). *Automated Planning, Theory and Practice*. San Francisco, CA: Elsevier, Morgan Kaufmann Publishers.

Gibson, E. J., & Pick, A. D. (2000). *An Ecological Approach to Perceptual Learning and Development*. New York: Oxford University Press.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Glasgow, J., Narayanan, H., & Chandrasekaran, B. (Eds.). (1995). *Diagrammatic reasoning: Computational and cognitive perspectives*. Cambridge, MA: MIT Press.

Goodale, M., & Milner, A. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25.

Heider, F., & Simmel, M. (1944.). An experimental study of apparent behaviour. *American Journal of Psychology*, *57*, 243-259.

Jablonka, E., & Lamb, M. J. (2005). *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge MA: MIT Press.

Jackson, A. (2002). The World of Blind Mathematicians. *Notices of the American Mathematical Society*, *49*(10). (http://www.ams.org/notices/200210/comm-morin.pdf)

Jamnik, M., Bundy, A., & Green, I. (1999). On automating diagrammatic proofs of arithmetic arguments. *Journal of Logic, Language and Information*, *8*(3), 297–321.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, *14*, 201–211.

Kaneff, S. (Ed.). (1970). *Picture language machines*. New York: Academic Press.

Kant, I. (1781). *Critique of pure reason*. London: Macmillan. (Translated (1929) by Norman Kemp Smith)

Lakatos, I. (1976). *Proofs and Refutations*. Cambridge, UK: Cambridge University Press.

Lakatos, I. (1980). The methodology of scientific research programmes. In J. Worrall & G. Currie (Eds.), *Philosophical papers, Vol I.* Cambridge: Cambridge University Press.

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

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

Merrick, T. (2006). What Frege Meant When He Said: Kant is Right about Geometry. *Philosophia Mathematica*, *14*(1), 44–75. (doi:10.1093/philmat/nkj013)

Michotte, A. (1962). *The perception of causality*. Andover, MA: Methuen.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Neisser, U. (1976). *Cognition and Reality*. San Francisco: W. H. Freeman.

Nelsen, R. B. (1993). *Proofs without words: Exercises in visual thinking*. Washingon DC: Mathematical Association of America.

Poincaré, H. (1905). *Science and hypothesis.* London: W. Scott. (http://www.archive.org/details/scienceandhypoth00poinuoft)

Ryle, G. (1949). *The concept of mind.* London: Hutchinson.

Sauvy, J., & Suavy, S. (1974). *The Child's Discovery of Space: From hopscotch to mazes – an introduction to intuitive topology.* Harmondsworth: Penguin Education. (Translated from the French by Pam Wells)

Sloman, A. (1962). *Knowing and Understanding: Relations between meaning and truth, meaning and necessary truth, meaning and synthetic necessary truth.* Unpublished doctoral dissertation, Oxford University. (http://www.cs.bham.ac.uk/research/projects/cogaff/07.html#706)

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

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

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

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

Sloman, A. (1984). The structure of the space of possible minds. In S. Torrance (Ed.), *The mind and the machine: philosophical aspects of artificial intelligence.* Chichester: Ellis Horwood.

Sloman, A. (1989). On designing a visual system (towards a gibsonian computational model of vision). *Journal of Experimental and Theoretical AI*, *1*(4), 289–337. (http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#7)

Sloman, A. (1993). The mind as a control system. In C. Hookway & D. Peterson (Eds.), *Philosophy and the cognitive sciences* (pp. 69–110). Cambridge, UK: Cambridge University Press. (http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#18)

Sloman, A. (1994). Explorations in design space. In A. Cohn (Ed.), *Proceedings 11th european conference on AI, amsterdam, august 1994* (pp. 578–582). Chichester: John Wiley.

Sloman, A. (1995). Exploring design space and niche space. In *Proceedings 5th scandinavian conference on AI, trondheim.* Amsterdam: IOS Press.

Sloman, A. (2000). 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* (pp. 3–16). Berlin: Springer-Verlag.

Sloman, A. (2001). Evolvable biologically plausible visual architectures. In T. Cootes & C. Taylor (Eds.), *Proceedings of British Machine Vision Conference* (pp. 313–322). Manchester: BMVA.

Sloman, A. (2008, March). *Kantian Philosophy of Mathematics and Young Robots* (To appear in proceedings MKM08 No. COSY-TR-0802). UK: School of Computer

Science, University of Birmingham.
(http://www.cs.bham.ac.uk/research/projects/cosy/papers#tr0802)

Sloman, A., & Chappell, J. (2007). Computational Cognitive Epigenetics (Commentary on (Jablonka & Lamb, 2005)). *Behavioral and Brain Sciences*, *30*(4), 375–6. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0703)

Warneken, F., & Tomasello, M. (2006, 3 March). Altruistic helping in human infants and young chimpanzees. *Science*, 1301-1303. (DOI:10.1126/science.1121448)

Weir, A. A. S., Chappell, J., & Kacelnik, A. (2002). Shaping of hooks in New Caledonian crows. *Science*, *297*, 981.

Winterstein, D. (2005). *Using Diagrammatic Reasoning for Theorem Proving in a Continuous Domain*. Unpublished doctoral dissertation, University of Edinburgh, School of Informatics. (http://www.era.lib.ed.ac.uk/handle/1842/642)