Evolvable Biologically Plausible Visual Architectures *

Aaron Sloman
School of Computer Science
The University of Birmingham
Birmingham, B15 2TT, UK
a.sloman@cs.bham.ac.uk
http://www.cs.bham.ac.uk/~axs/

Abstract

Much work in AI is fragmented, partly because the subject is so huge that it is difficult for anyone to think about all of it. Even within sub-fields, such as language, reasoning, and vision, there is fragmentation, as the sub-sub-fields are rich enough to keep people busy all their lives. However, there is a risk that results of isolated research will be unsuitable for future integration, e.g. in models of complete organisms, or human like robots. This paper offers a framework for thinking about the many components of visual systems and how they relate to the whole organism or machine. The viewpoint is biologically inspired, using conjectured evolutionary history as a guide to some of the features of the architecture. It may also be useful both for modelling animal vision and designing robots with similar capabilities.

1 Introduction

Seeing is believing — among many other things, and there's the rub. It is a biological fact that (in humans) vision feeds and corrects beliefs, but that's not all: it also participates in posture control, in tight feedback loops as we pick things up and in ballistic actions like throwing a ball into a bin; it can produce embarrassment, aesthetic pleasure or pain, sexual arousal, changes in how we hear speech [11], nausea (because what is seen disgusts us or because of perceived motion); and it can inform us that something is *impossible* (e.g. the chair going through the narrow doorway) and what *might* exist or is *likely* to exist [20], e.g., seeing possible courses of action (possible routes across a cluttered room) or how a mechanism works, or the danger that a construction is about to become unstable or the risk of the toddler falling into the fish-pond. These are all examples of Gibson's affordances [7, 16], generalised below.

It is not all one-way traffic: what we see can depend on what we already know (e.g. reading words, or seeing the difference between a pair of identical twins only after getting to know them), on what we want, what we are doing, or what we are afraid of as we walk through a forest in dim light. Seeing can also take many forms, including a clear and distinct percept, like the sight of your hand before your face, with all the details of skin texture, or a fleeting impression in the visual periphery, a subconsciously detected change in optical flow that makes you lose your balance [10], sensing the hostility as you enter a room, or experiencing the strange effect in Figure 1. Another biological fact is that much human visual processing, like much else in the mind, is not accessible to consciousness. E.g. we do not *experience* using optic flow to control our posture. So our experience of seeing is at best a partial guide to the functions even of our own vision.

Can we hope to explain and model all these aspects of biological vision, or build them into robots? Solutions found for isolated sub-problems may be constrained in ways that prevent integration. Integration will not occur if people working in different sub-fields do not communicate, painful though that may be when there is so much to do in one's own sub-field. But communication is not enough. We need a conceptual framework and a methodology to support attempts at integration. This paper offers one by sketching ideas about possible high

*Presented at British Machine Vision Conference (BMVC01) Manchester, September 2001

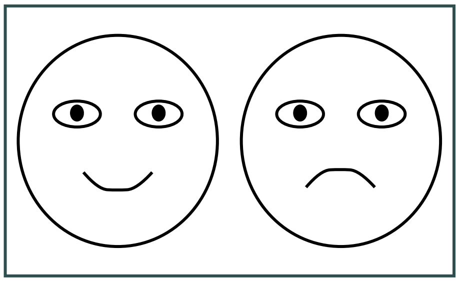

Figure 1: Some people, but not all, see a non-geometric difference between the eyes.

level architectures in which different processes can be combined in an organism, or robot. Though speculative, the framework is a result of many years of thinking about the problems[1] and inspiration from many researchers in AI, psychology, neuroscience, ethology, biological evolution, and philosophy. We use partly speculative evolutionary history as a coarse-grained guide to some architectural features. This biology-inspired framework may also be useful for the extraordinarily difficult engineering task of building complete robots with capabilities similar to humans.

None of this is intended as a critique of work on image analysis and interpretation that solves specific engineering problems without being integrated into a complete agent architecture. Such solutions can be judged in relation to their objectives.

2 Overview of the CogAff framework

We offer a framework for thinking about design options for information processing architectures for complete animal-like agents of various kinds, including insects, various sorts of vertebrates, primates and humans, and artificial software agents and robots. The information processing architecture need not map in any simple way onto brain physiology or computational hardware: it is a "virtual machine" (VM) architecture, in the standard sense in computer science, e.g. where the same Prolog or Java virtual machine may run on very different hardware architectures. Virtual machines are, of course, real machines, and can perform real tasks, such as controlling a chemical plant, solving equations, or re-formatting a document. (This point, and the implications for causation in VMs is discussed on my website.)

Our framework, called "CogAff" because it accommodates cognition and affect, is based on the observation that within an organism there can be different sorts of VM architectures and sub-architectures which evolved at different times, whose tasks are very different, and which can be sub-divided in different ways, as indicated crudely in Figure 2. For instance, in humans there are information processing mechanisms concerned with managing many aspects of bodily and mental function; mechanisms concerned with low level, fine grained control of walking, grasping and other actions; mechanisms concerned with thinking about possible futures, evaluating them, making plans; mechanisms concerned with self monitoring, self evaluation and self control; mechanisms concerned with being part of a social community requiring many forms of cooperative and non-cooperative, verbal and non-verbal interaction. Each of these mechanisms may have a complex internal architecture, and each may require specialised perceptual input and motor output capabilities, many of them served by vision.

We can distinguish *reactive*[2] mechanisms, where states or events detected by external or internal sensors immediately trigger external or internal responses, from *deliberative* mechanisms in which alternative possibilities for action can be considered, categorised, evaluated, and selected or rejected. More powerful deliberative mechanisms can do "what if" reasoning about the past or future, or even counterfactual reasoning

[1] See for instance [15, 16, 17, 18, 19, 20, 21, 24, 23, 22, 25]

[2] For some researchers the term "reactive" implies no change of internal state. For others, the term is not so restrictive: it includes finite-state automata, and various kinds of adaptive neural nets, but excludes *deliberative* capabilities of the sorts described below. I use the word in the more general sense, as does Nilsson in [14].

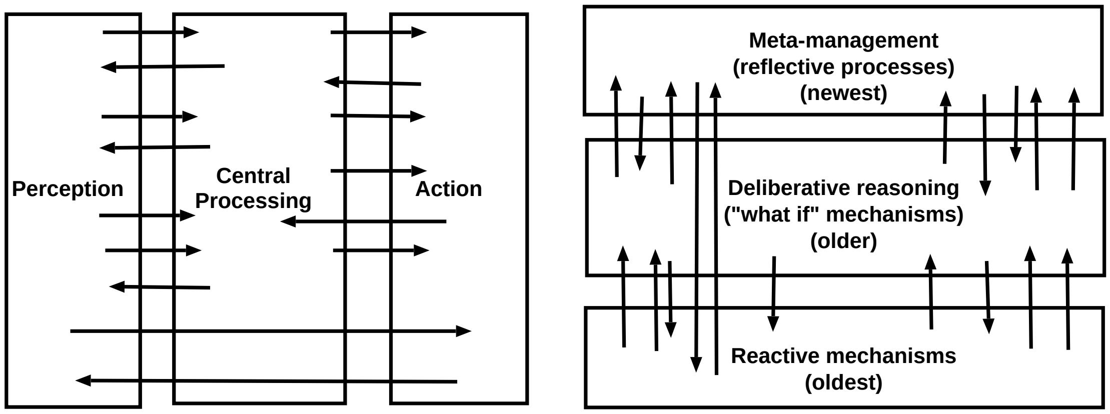

Figure 2: Two coarse divisions of information processing architectures
(a) Organisms and robots require perceptual mechanisms and action mechanisms of varying degrees of sophistication, along with some persistent internal state which may be modified over various time-scales: Nilsson's (1998) "triple tower" model. Arrows represent flow of information and control signals. Boundaries between "towers" need not be sharp. (b) Another architectural division concerns mechanisms that evolved at different times, providing reactive, deliberative and meta-management capabilities. Below we superimpose these divisions.

about how things might have been. The depth, precision and soundness of such reasoning can vary. Some organisms need a third *meta-management* layer [2, 25] to monitor, categorise, evaluate, and (partially) control processes occurring within the system. This requires explicit use of formalisms and concepts referring to internal virtual machine states.

Reactive mechanisms and architectures evolved first and are most wide-spread in nature, in multitudinous forms. Deliberation evolved much later, and is much rarer. Sophisticated variants require a long term associative memory and symbolic reasoning capabilities using a short term re-usable memory in which structural descriptions can be built. This imposes demands on perceptual mechanisms to recognise more abstract categories, suitable for expressing generalisations, and on motor systems to accept more abstract "instructions". Meta-management evolved even later, and is rarer still.

Single-celled organisms, plants, insects and many other animals apparently lack any deliberative capability, though some mammals (and possibly some nest-building birds?) seem to be able to consider alternatives and then choose. Moreover, without meta-management, they will not be aware of what they are doing, just as insects perceive without knowing that they do (e.g. because they lack mechanisms and formalisms for self-description). Humans appear to have all three architectural layers though probably not at birth. Concepts used for self-categorisation may also be useful for describing mental states of others, and vice versa. For some social animals, for predators and for prey, being able to perceive mental state (e.g. intentions) can be very useful.

The three layers operate concurrently, and do not form a simple dominance hierarchy. For instance, the two top layers cannot directly change the contents of the reactive layer, though they may be able to change it indirectly through training, e.g., when a novice learns a new skill, such as driving a car, by following instructions and practising.

How finely to divide up the layers is partly a matter of taste: some authors e.g. [5] prefer to separate *reflexes* from the reactive layer, and some (e.g. Minsky) would prefer to split off some of the high level meta-management functionality into a separate layer. It is likely that a host of further subdivisions will later prove useful.

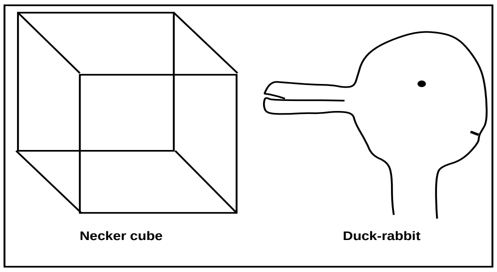

Figure 3: Does vision inform only about geometrical and physical properties, or does it include more abstract information, e.g. which way an animal is facing?

3 Implications for vision

If the visual system simultaneously serves the needs of all three layers, then perhaps it too has a complex architecture with layers that evolved at different times, achieving different sorts of functions, all of which build on the lowest level mechanisms shared by all sub-architectures. This generalises Gibson's notion of "affordance"[7]. He claimed that perceptual mechanisms give an organism information not only about physical features of the environment, but, more importantly, what the opportunities and obstacles to various kinds of actions are: a far more abstract type of information. Thus some objects are seen to offer positive affordances, such as support, passage or shelter, others negative affordances, such as obstruction, difficulty in grasping, danger, etc. Affordances are in part determined by the organism's own needs and capabilities. For instance, a type of organism that cannot grasp or never needs to grasp anything, or to recognise grasping in others, might never perceive graspability. This aspect of Gibson's theory is independent of other questionable aspects including his notion of perception as 'direct', and his rejection of the relevance of representations and computation to perception.

For Gibson, the positive and negative affordances perceivable by an animal concern the whole animal. However, the notion that there are many perceptual subsystems performing different tasks, even if they share a sensory organ, leads naturally to the notion that there are different affordances to be detected which are relevant to the needs and capabilities of different subsystems. So as central subsystems evolve with new needs and capabilities, this can drive evolution of the perceptual "tower" towards a *concurrent* stratified system with new specialised perceptual VM capabilities tailored to the needs of particular central subsystems. Examples might be the evolution of visual mechanisms for detecting 3-D shape categories, or for recognising individual faces, or mechanisms concerned with perception of mental state (see Figures 1 and 3) and social interactions.

Our conjecture is that in more complex animals there are essentially *several different* visual systems, all sharing a collection of physical resources, such as lenses, retina, eye muscles, optic nerves and parts of the brain dealing with some of the earliest stages of visual processing. Some of the sharing will involve mechanisms that process different information obtained from the eyes in parallel, using different parts of the brain e.g. when your posture control mechanism and your route-finding systems both use visual information simultaneously. However conflicts can arise, leading either to clashes or to sequential use, e.g. if different subsystems require different directions of gaze [17].

Different kinds of familiar visual ambiguity illustrate the variety of types of visual tasks. In the necker cube, shown in Figure 3, the visual flip is purely geometric, between interpretations of the image where there are different distances and orientations of surfaces and edges. This is consistent with Marr's view in [12] the 'quintessential fact of human vision – that it tells about shape and space and spatial arrangement'. In the duck-rabbit, however, there is no geometric flip: the change is much more abstract and involves both changes in how parts are identified (e.g. ears *vs* bill) and more abstract notions like "facing this way", "facing that way" which presupposes *perception of other organisms seen as perceivers*. There are many sorts of things humans can see besides geometrical properties: that one object is supported by another, that one object constrains motion of another (e.g. a window catch), that something is flexible or fragile, which parts of an organism are ears, eyes, mouth, bill, etc., which way something is facing, what action some person or animal is about to perform (throw, jump, run, etc.), whether an action is dangerous, whether someone is happy, sad, angry, etc., whether a painting is in the style of Picasso...

Investigating how such perception occurs will include investigating (among other things) (i) the precise nature of the information, (ii) architectures capable of using the information, (iii) formalisms that can be usefully employed to store such information and (iv) means by which such information can be produced and processed.

Our discussion of multiple functions for vision is consistent with the now familiar idea in neuropsychology that there are different dorsal and ventral visual pathways [8] performing different tasks. However our framework suggests that describing these as 'what' and 'where' pathways is misleading if the main difference is not so much a difference in *content*, as a difference in which sub-mechanisms, e.g. reactive or deliberative, online or ballistic, individual or social, *use* the information. In particular we conjecture that what we are *conscious* of seeing is what the meta-management system can access, probably a very small, specially processed subset of visual information used in various parts of the system. A machine so designed might re-discover what philosophers call 'sensory qualia.' One sort of blindsight occurs when those meta-management mechanisms are damaged while the reactive layer continues to function.

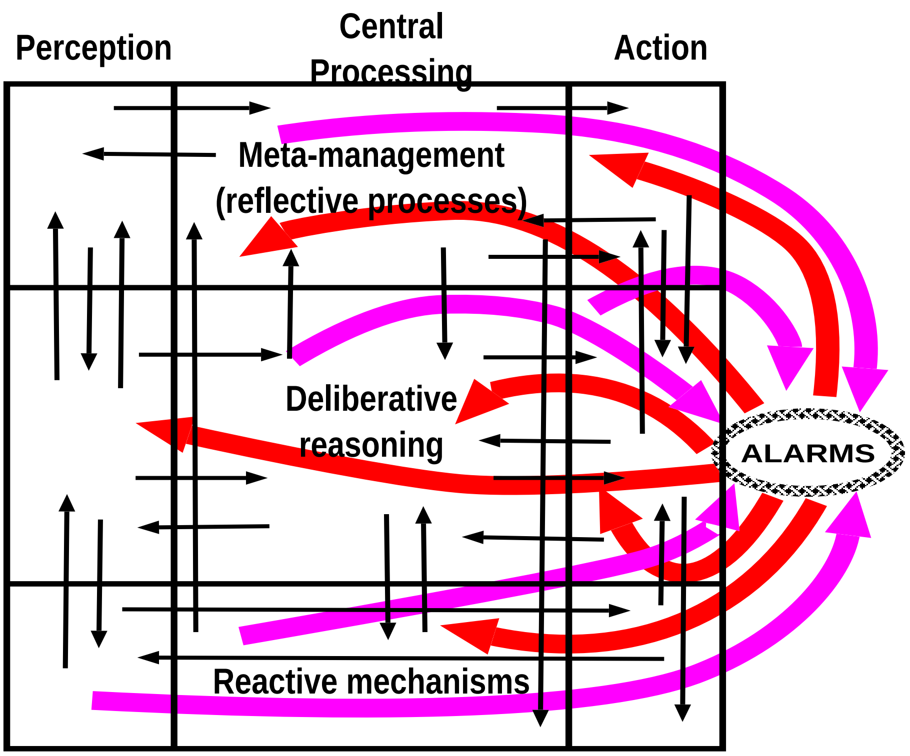

Figure 4: Superimposing the previous divisions

The divisions of Fig 2 can be superimposed, producing new sub-types of components, with functions defined by relationships to other parts of the system. A fast reactive alarm system receives inputs from, and can send control signals to, many components. Shaded arrows represent information flowing to and from it. Being purely reactive and pattern driven it will typically be stupid and capable of mistakes, but may be trainable. An insect's architecture might include only the bottom layer, with alarm mechanisms. Some animals appear to have reactive and deliberative layers. Humans have all three.

4 The variety of biological architectures

The CogAff schema in Figure 4 is not so much a specific architecture as an indication of varieties of roles that components of an architecture can have. Many organisms and artificial systems will have only a subset of the components, and some artificial systems may not fit the schema, e.g. distributed software agents. We conjecture that information processing architectures of individual biological organisms are adequately covered, though not whole colonies. An insect might have only the bottom, reactive, layer, possibly including alarm mechanisms — unlike purely deliberative AI systems. Other animals and some robots may have a hybrid reactive and deliberative mechanism, with varying sophistication in the deliberative mechanism.

Visual mechanisms for primitive reactive agents might detect edges, optical flow, image statistics, surface orientation, etc. More sophisticated reactive agents may recognize, and react to, more global structures, e.g. objects to eat, or mate with.

Visual systems feeding deliberative capabilities need to characterise objects, states of affairs, or action patterns at a level of abstraction supporting predictive generalisations (e.g. "if it sees me it will run").

Co-evolution of meta-management and social perception could lead to the ability to perceive mental states of other agents (Figure 3 (b)), solving the "mind-body" problem. But there is much we don't yet understand about possible evolutionary trajectories [22].

The architectural layers described here should not be confused with Marr's three methodological levels. Several multi-layer architectures occur in the AI literature, though superficially similar diagrams may be used for very different designs. E.g., the 'triune brain' architecture in [1] looks partly like our three layered system, but closer inspection reveals a pipelined architecture: Information flows in through low level sensors, then up the central hierarchy, then, after high level decision making, down through the central pillar and out through low level transducers. We call this an 'Omega' architecture because the information flow pattern within the CogAff diagram is roughly Ω-shaped. Such "peephole" perception and action mechanisms contrast with"multi-window" perception and action permitted by CogAff. Another alternative is Brooks' *subsumption* architecture [3, 4]: it allows several layers, but they are all reactive (i.e. non-deliberative), entirely within the bottom level of the CogAff schema.

Detailed requirements for various architectural components need to be analysed further. Because of the nature of access to a large content addressable associative memory store and also the requirement for a re-usable temporary storage space for 'what if' descriptions, a deliberative system is likely to be slow, discrete and serial, compared with fast, parallel, and largely analog reactive mechanisms. For this reason, in a hybrid reactive and deliberative system, it may be necessary to have an "attention filter" with dynamically varying filter threshold to protect the resource-limited deliberative mechanism from being interrupted too often during urgent and intricate tasks (as shown schematically in Figure 5) – explaining why soldiers in battle don't notice some injuries. Alarm states or intense perceptual inputs may be capable of exceeding the filter threshold, sometimes producing emotions [25].

Omega and subsumption models have a rigid control hierarchy, but that is not the only possibility. Our framework allows systems where all the layers and the alarm system(s) operate concurrently, each (partially) capable of interrupting and redirecting the others.

The meta-management layer does not need to be a permanently fixed, rigid system. Instead, a collection of high level culturally determined "personae" may be available in some sort of database within the architecture, turned on and off by different contexts and causing global features of the behaviour to change, e.g. switching between bullying and servile behaviour. One of the sub-functions of vision may be to facilitate learning behaviour-patterns in a social context (compare so-called "mirror neurons"). Different global states may trigger different (previously learnt) visual sub-mechanisms, for instance when reading music, driving a car or gazing into a lover's eyes.

Besides the sorts of components already reviewed, a human-like organism would need components such as: long term associative memories, mood controllers (altering global processing states), motive generators (Frijda's concerns [6]), standards & values, attitudes, skill-compilers, motive comparators, formalisms, inference mechanisms, etc., many of them linked to concurrently active perceptual mechanisms, e.g. parents

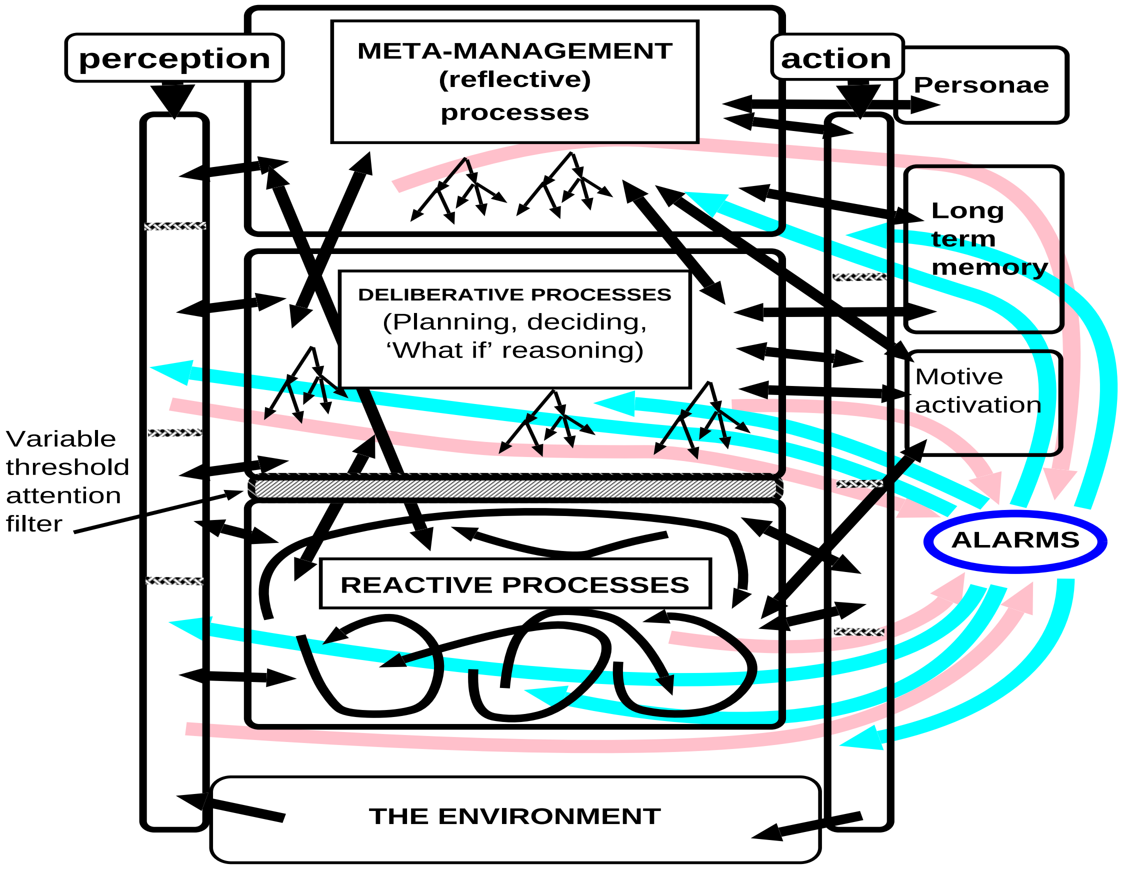

Figure 5: The "Human-like" sub-schema H-Cogaff

The reactive, deliberative and meta-management layers evolved at different times, requiring discontinuous changes in the design, and providing significantly new capabilities. An attention filter with dynamically varying threshold may be used to protect resource-limited higher level functions. Some aspects of the alarm system apparently correspond to the brain's limbic system, and frontal lobes implement some meta-management functions.

reacting to perceived threats to offspring.

Most of this has been or will be discussed in more detail elsewhere. For now all I am trying to do is draw attention to some of the surprising diversity of functions of vision, or more generally perception, in the three-layered, human-like, H-Cogaff architecture sketched in Figure 5.

5 Future work

For vision researchers, the main conclusion is that concurrently active internal processes have diverse needs, requiring distinct forms of visual processing, performed either in parallel by different mechanisms, or sequentially as resources are switched between tasks. As we have seen, this includes such things as posture control, route planning and aesthetic processes, which can all be served concurrently by different (multi-window) visual mechanisms sharing some lower level mechanisms. The perceptual needs of concurrently active subsystems are defined not (only) by the physical/geometrical nature of the environment, but by the functions and causal relationships within the larger architecture, of the subsystem and its *capabilities*, including processing and representational capabilities. Therefore "affordances" available to an animal are a function of the sub-system that uses them, not just features of the environment. Different sub-systems use different affordances, and possibly different formalisms and ontologies. Thus simply studying physical aspects of objects and the physical processes of image formation may divert attention from the most important perceptual processes.

One way to investigate some of this in more detail is to use evidence from brain damage: as illustrated in [8, 9] differentially disabled sub-systems provide clues concerning the architecture. Such results can be combined with studies of visual development, comparisons between different species, evolutionary studies and meticulous task analysis for design of robots of various kinds.

There is a huge amount of work still to be done, defining the tasks of all the various components of the architecture, and designing mechanisms that can perform those tasks, including mechanisms for preventing, or detecting and resolving, conflicts between processes running concurrently.

An important early task might be to combine information from a variety of disciplines, including ethology, developmental psychology, robotics, and brain science, to produce a first draft *taxonomy of types of affordances* that might be useful at various stages of evolution, or development. On that basis we should try to analyse ways in which sub-mechanisms using those affordances can use compatible representations, develop shared mechanisms, and collaborate on various sub-problems. Differences between precocial and altricial species (where the latter are born or hatched immature and helpless) may derive partly from how sophisticated the affordances are which young animals need to learn to detect and react to. There are trade-offs between evolving innate mechanisms for all the tasks and evolving a generic mechanism for learning by acting while the brain is growing. (E.g. the former requires longer and more varied evolutionary histories and larger DNA structures and brains — possibly explosively large).

Our framework supports Minsky's [13] use of the "society" metaphor for minds containing a collection of more or less distinct, concurrently interacting, collaborating and competing sub-systems. An even deeper understanding of their relationships emerges if we regard them as forming a co-evolved "ecosystem of mind" where each component has a niche partly determined by the others, as well as the external environment.

Concern about the limitations of current theories of vision is not new. Ullman writes "the recognition of common objects is still way beyond the capabilities of artificial systems, or any recognition model proposed so far" ([26] p.1). I have tried to show that far more than object recognition is at stake. Understanding this, and linking it with the ecosystem of mind view is a useful step towards achieving de-fragmentation in vision, and in AI generally.

Acknowledgements

This research is partly funded by the Leverhulme foundation. Many of the ideas are inspired by colleagues in the Cognition and Affect project http://www.cs.bham.ac.uk/research/cogaff/, especially Brian Logan and Matthias Scheutz, who helped me think about architectures, and Jane Riddoch and Glyn Humphreys whose empirical work on vision and brain damage helped to open my eyes.

References

[1] J.S. Albus. *Brains, Behaviour and Robotics*. Byte Books, McGraw Hill, Peterborough, N.H., 1981.

[2] L.P. Beaudoin. *Goal processing in autonomous agents*. PhD thesis, School of Computer Science, The University of Birmingham, 1994. (Available at http://www.cs.bham.ac.uk/ research/ cogaff/).

[3] R. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2:14–23, 1986. 1.

[4] R. A. Brooks. Intelligence without representation. *Artificial Intelligence*, 47:139–159, 1991.

[5] Darryl N Davis. Reactive and motivational agents: Towards a collective minder. In J.P. Mueller, M.J. Wooldridge, and N.R. Jennings, editors, *Intelligent Agents III — Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages*. Springer-Verlag, 1996.

[6] Nico H. Frijda. *The emotions*. Cambridge University Press, Cambridge, 1986.

[7] J.J Gibson. *The Ecological Approach to Visual Perception*. Lawrence Earlbaum Associates, 1986. (originally published in 1979).

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

[9] G.W. Humphreys and M.J. Riddoch. *To See But Not To See: A Case Study of Visual Agnosia*. Erlbaum, Bloomington, 1987.

[10] D.N. Lee and J.R. Lishman. Visual proprioceptive control of stance. *Journal of Human Movement Studies*, 1:87–95, 1975.

[11] H. McGurkand J. MacDonald. Hearing lips and seeing voices. *Nature*, 264:746–748, 1976.

[12] D. Marr. *Vision*. Freeman, 1982.

[13] M. L. Minsky. *The Society of Mind*. William Heinemann Ltd., London, 1987.

[14] N.J. Nilsson. Teleo-reactive programs for agent control. *Journal of Artificial Intelligence Research*, 1:139–158, 1994.

[15] A. Sloman. *The Computer Revolution in Philosophy*. Harvester Press (and Humanities Press), Hassocks, Sussex, 1978.

[16] 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.

[17] A. Sloman. The mind as a control system. In C. Hookway and D. Peterson, editors, *Philosophy and the Cognitive Sciences*, pages 69–110. Cambridge University Press, Cambridge, UK, 1993.

[18] A. Sloman. Varieties of formalisms for knowledge representation. *Computational Intelligence*, 9(4):413–423, 1993. (Special issue on Computational Imagery).

[19] A. Sloman. How to design a visual system – gibson remembered. In D.Vernon, editor, *Computer Vision: Craft, Engineering and Science*. Springer Verlag, Berlin, 1994.

[20] 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. Morgan Kaufmann Publishers, 1996.

[21] A. Sloman. Towards a general theory of representations, in Peterson (1996). In D.M.Peterson, editor, *Forms of representation: an interdisciplinary theme for cognitive science*, pages 118–140. Intellect Books, Exeter, U.K., 1996.

[22] A. Sloman. Interacting trajectories in design space and niche space: A philosopher speculates about evolution. In M.Schoenauer et al., editor, *Parallel Problem Solving from Nature – PPSN VI*, Lecture Notes in Computer Science, No 1917, pages 3–16, Berlin, 2000. Springer-Verlag.

[23] A. Sloman. Diagrams in the mind. In M. Anderson, B. Meyer, and P. Olivier, editors, *Diagrammatic Representation and Reasoning*. Springer-Verlag, Berlin, 2001.

[24] A. Sloman. How many separately evolved emotional beasties live within us? In R. Trappl and P. Petta, editors, *Emotions in Humans and Artifacts*. MIT Press, Cambridge MA, (to appear).

[25] A. Sloman and B.S. Logan. Evolvable architectures for human-like minds. In G. Hatano, N. Okada, and H. Tanabe, editors, *Affective Minds*, pages 169–181. Elsevier, Amsterdam, 2000.

[26] S.Ullman. *High-level vision: Object recognition and visual cognition*. MIT Press, Cambridge Mass, 1996.