

# **How do animals gather useful information about their environment and act on it?**

Jackie Chappell\*

\*Centre for Ornithology, School of Biosciences  
University of Birmingham, Edgbaston, Birmingham B15 2TT, UK  
j.m.chappell@bham.ac.uk

## **Abstract**

Animals are much more successful than current robots in their ability to gather information from the environment, detect affordances, attribute causes to affects, and sometimes generate individually novel behaviour. What kinds of mechanisms might make this possible? I will discuss different mechanisms for acquiring information in animals, and their strengths and weaknesses given different life histories and niches. I will discuss experiments which have attempted to uncover the extent of animals' abilities to use information from their environment, and the mechanisms that might be used to accomplish this. The development of these kinds of competences (in evolutionary time and over the course of an individual's lifetime) is another interesting problem. Exploration and play seem to be very important for some kinds of behaviour, particularly flexible responses to novel problems, but there is also the possibility that animals come equipped with certain kinds of 'core knowledge', which might help to direct and structure the acquisition of more complex competences.

# How do animals gather useful information about their environment and act on it?



Jackie Chappell

Center for Ornithology  
School of Biosciences  
University of Birmingham  
[j.m.chappell@bham.ac.uk](mailto:j.m.chappell@bham.ac.uk)

1

# What is involved in gathering information and acting on it?

- How do you **perceive objects** in ways that allow manipulation?
- What do you pay **attention** to (filtering and selective attention)?
- How do you **detect affordances**?
- How do you **assign causality** to actions, events or agents?
- How can competences be **re-combined flexibly** to generate appropriate behaviour in novel contexts, or creativity?
- How does this all **develop**?

2

If you were trying to build a robot to behave spontaneously like the chimp in the following clip, how would you do it?

3

# Pal, 2.5 years old



video taken by Misato Hayashi, Primate Research Institute, Kyoto University, used with permission

Hayashi & Matsuzawa (2003) Animal Cognition

4

## Questions raised

- Why did she specifically pay attention to the blocks (**attention**)?
- What mechanism could have allowed Pal to learn that she could stack the blocks (detect the **affordances** of blocks)?
- Did she understand **causal relationships** (e.g. that hitting the blocks would make them fall)?
- Would she be able to stack other shapes or different objects (**re-combinable competences**)?
- How did this behaviour **develop**?

5

What kinds of mechanisms make it possible for animals to find out about affordances, attribute causes to effects and generate appropriate (sometimes novel) behaviour?

6

## What mechanisms do we know of?

- Developmentally-fixed behaviour - usually genetically determined
  - Fast and reliable, but inflexible
- Associative learning
  - Gradual process, but fairly flexible and surprisingly subtle
- Social learning
  - Can provide a short-cut to learning a novel behaviour
- Some extended learning mechanism—some 'core knowledge', new competences acquired, extended and re-combined through exploration and play?

7

## Developmentally-fixed behaviour



- Complex behaviour triggered by simple cues
- Useful when:
  - Limited opportunity for learning
  - Behaviour needs to be perfect on the first attempt (e.g. flight in cliff or tree-nesting birds)
  - There are time constraints (e.g. short life span)
- Common in precocial species where young are relatively independent from birth

8

## Associative learning

- Classical conditioning and operant conditioning
- Can lead to a complex chain of behaviour → novel responses to the environment
- Relatively slow and gradual process (though one-trial learning is possible)

9

## Social learning

- Learn from the behaviour of others:
  - Directly, by observation
  - Or via products of another's behaviour
- Can spread novel behaviour rapidly through a population → cultural transmission → cultural evolution

10

## Extended learning mechanism and exploration

- Animals can learn about the space of possible actions with an object, unusual properties etc.
- Time consuming, but possible for altricial species during development, when parent(s) care for offspring
- May also enable very rapid learning if 'chunks' of knowledge about the environment can be reused
- Exploration (not directly reinforced) may be very important

11

## What do you pay attention to?

- Some genetically-determined biases which limit the stimuli that form associations (e.g. taste conditioning in rats)
- Exploration → classification of some things as 'interesting'?

12

“Appropriateness” of the stimulus or response matters (Domjan & Wilson, 1972)

	Group taste	Group noise
Train	Sweet water → illness	Noisy water → illness
Test	Sweet water vs. Plain water	Noisy water vs. Silent water
<b>RESULT</b>	LEARNING	NO LEARNING
Train	Sweet water → shock	Noisy water → shock
Test	Sweet water vs. Plain water	Noisy water vs. Silent water
<b>RESULT</b>	NO LEARNING	LEARNING

So, natural selection constrains associations to those likely to be causally linked

13

## How to detect affordances?

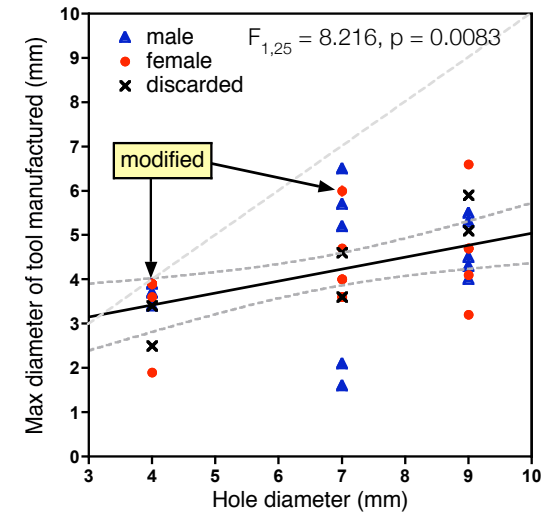
- Are affordances tied to specific stimuli, or can animals abstract more general properties?
- What is the role of experience?
- Is this an adaptation specific to the tool-using domain?

14

Making an appropriate tool for a novel task  
(New Caledonian crows)



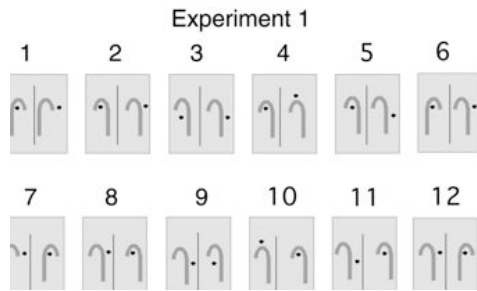
15



(Chappell & Kacelnik 2004)

16

## What do non-tool users understand about the function of tools?



(Santos et al. 2005)

17

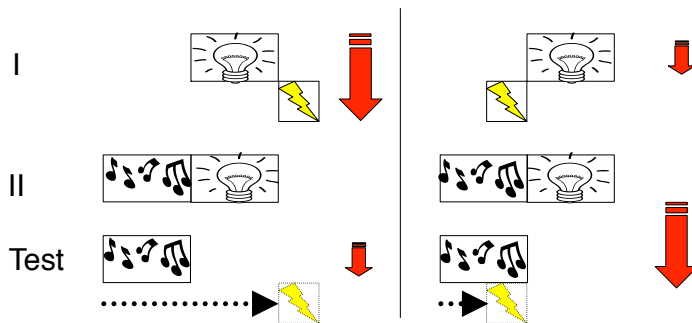
## How to assign causality?

- Probabilistically, through contingency and contiguity (Rescorla & Wagner 1972)
- Test hypotheses by performing interventions (Gopnik & Schultz 2004)
- Core knowledge about the structure of the world (acquired or developmentally fixed) → expectations about causal structure (not all causes are equally possible) (Carey & Spelke 1996)

18

## Animals can learn about the temporal relationship between events → causal attribution

(Barnet, Cole & Miller, 1997)



19

## What causes objects to fall?



Possibly gaining dynamic feedback from environment, and adjusting behaviour appropriately

20

## Re-combinable competences

- To what degree can animals re-combine existing competences to generate novel behaviour?
- How does this depend on previous experience?

21

## Pilfering in scrub jays: it helps to have been a thief to catch a thief

- Three groups:
  - **Observer + Pilferer**—had experience of both observing conspecifics caching, and of pilfering others caches
  - **Observer**—only experience with observing caching
  - **Pilferer**—listened to others caching, then allowed to pilfer caches

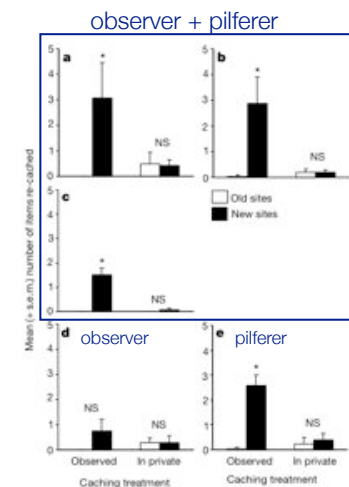
(Emery and Clayton 2001)

22

## Experimental protocol

- Birds allowed to cache food in a tray:
  - With an observer bird watching from an adjoining cage ('**observed**' trial)
  - With no bird watching them ('**in private**' trial)
- Then allowed to retrieve cache and also given opportunity to re-cache in old tray or a new one

23



- Pilferers re-cached food when observed caching (in new sites)
- Specific to the tray which was observed, not a general increase in re-caching
- Observation of caching not sufficient to prompt re-caching

(Emery and Clayton 2001)

24

## Novel manufacturing behaviour with a new material

- In an experiment on choice between a hooked wire and a straight one, Betty bent the hook spontaneously on the 5th trial
- In a subsequent experiment, she bent the hook and used it to remove the bucket on 9/10 trials

(Weir, Chappell & Kacelnik 2002)

25



[Weir, Chappell & Kacelnik 2002]

26

## What might the mechanism allowing re-combination of competences be?

- Built-in drive to explore (with no immediate reinforcement consequences)
- Cognitive structures (genetically determined) which might guide or constrain exploration ('bootstrapping' of behaviour)
- Construction of reusable 'chunks' which can themselves be recombined into more complex structures (e.g. language learning)

27

## How do these abilities develop?

- Exploration and play
  - Lack of neophobia—you can't discover properties of objects you never go near
- Altricial species often have a large amount brain development going on after birth/hatching
  - Is it important that the developing brain is exposed to the environment?
- To what degree are animals limited by their exploratory tendencies?

28



Are animals limited by species-specific representational capacities, or by their exploratory tendencies?

- Representational view vs. Ecological view (Cummins-Sebree and Frigaszy, 2005)
- Capuchin monkeys spontaneously re-positioned canes to pull a food reward towards them, unlike tamarins
- Is this difference because of species differences in exploratory/manipulatory behaviour?

29

## Summary

- We need to combine the richness of animals' behaviour with the depth of knowledge of the mechanisms involved in artificial systems to explore this
- There is almost certainly more than one solution to the problem (*in vivo* and *in silico*)—the optimal solution depends on the 'habitat' of the agent
- Animals (and robots) need to be tested in ethologically valid ways to reveal their competences fully
- It's a very difficult (but interesting) problem!

30

# A Novel Computing Architecture for Cognitive Systems based on the Laminar Microcircuitry of the Neocortex – the COLAMN project

Michael Denham

Centre for Theoretical and Computational Neuroscience  
University of Plymouth, UK  
mdenham@plym.ac.uk

## Abstract

Understanding the neocortical neural architecture and circuitry in the brain that subserves our perceptual and cognitive abilities will be an important component of a “Grand Challenge” which aims at an understanding of the architecture of mind and brain. We have recently embarked on a new five-year collaborative research programme, the primary aim of which is to build a computational model of minimal complexity that captures the fundamental information processing properties of the laminar microcircuitry of the primary visual area of neocortex. Specifically the properties we aim to capture are those of self-organisation, adaptation, and plasticity, which would enable the model to: (i) develop feature selective neuronal properties and cortical preference maps in response to a combination of intrinsic, spontaneously-generated activity and complex naturalistic external stimuli; and (ii) display experience-dependent and adaptation-induced plasticity, which optimally modifies the feature selectivity properties and preference maps in response to naturalistic stimuli. The second aim of the research programme is to investigate the feasibility of designing VLSI circuitry which would be capable of realising the computational model, and thus demonstrate that the model can form the basis for a novel computational architecture with the same properties of self-organisation, adaptation, and plasticity as those displayed by the biological system. A basic premise of the research programme is that the neocortex is organised in a fairly stereotyped and modular form, and that in this form it subserves a wide range of perceptual and cognitive tasks. In principle, this will allow the novel computational architecture also to have wide application in the area of cognitive systems.

## 1 Introduction

The neocortex of the brain subserves sensory perception, attention, memory and a spectrum of other perceptual and cognitive functions, which combine to provide the biological system with its outstanding powers. It is clear that the brain carries out information processing in a fundamentally different way to today’s conventional computers. The computational architecture of the brain involves the use of highly parallel, asynchronous, nonlinear and adaptive dynamical systems, namely the laminar microcircuits of the neocortex. The neurons which make up a neocortical microcircuit (Silberberg et al, 2002; Mountcastle, 1997) are precisely connected to each other and to their afferent inputs through synapses in specific layers of the laminar cortical architecture, and on specific locations on their dendritic trees Thomson and Bannister 2003; Callaway, 1998). Each

synapse acts as a unique adaptive filter for the transmission of data into the circuit and between pairs of cells. Thus whilst a single neuron may connect to many hundreds of other neurons, a signal sent by one neuron will be interpreted by each target neuron in a unique way. Furthermore, these connections are not static but change their transmission characteristics dynamically and asynchronously, on a millisecond timescale, partly determined by their highly precise spatial location in the dendritic tree (Häusser et al, 2003) but also in relation to the function of the different neuronal types that they connect. In addition, both the synaptic connections and the transmission properties of the dendritic tree have the remarkable ability to continuously adapt and optimise themselves to meet the requirements of novel tasks and environments. This takes place both through unsupervised, self-organising modification of their dynamic parameters, and through optimisation of the synaptic and dendritic dynamics by spe-

cific adaptation-induced and experience-dependent plasticity mechanisms.

Capturing the fundamental information processing properties of the laminar microcircuitry of the neocortex in the form of a computer model could provide the foundation for a radical new generation of machines that have human-like performance in perceptual and cognitive tasks. Such machines would be capable of using self-organisation, adaptation, and plasticity mechanisms which are inherent in the neocortex, in order to deal with complex, uncertain and dynamically changing information. They would potentially be much more powerful, require minimal programming intervention, and be resilient to failures and errors. Creating the necessary understanding of these properties of the neocortex, expressing them as a computational model of minimal complexity, and translating this model into the design of a computer architecture capable of realisation in VLSI, will require the collaborative efforts of neuroscientists, computer scientists, mathematicians, and engineers.

## **2 The aims of the research programme**

The aim of this research programme is to create a new “brain-inspired” computational architecture which possesses the basic properties of self-organisation, adaptation and plasticity manifest in the laminar neural microcircuitry of the neocortex. The principal objective is a functional model of a “stereotypical” cortical microcircuit which captures these basic properties of the neocortex, and provides the basis for the design of a novel, modular computational architecture capable of realisation in a combination of analogue and digital VLSI circuits. The ultimate goal of this avenue of research is a “brain-inspired” architecture which will deliver human-like levels of performance for a wide range of perceptual and cognitive tasks, and deal with all sensory modalities.

This goal is well beyond the scope of the currently envisaged research programme; however, as a first step towards this goal, the programme will aim at capturing the fundamental properties of self-organisation, adaptation and plasticity of the neuronal circuitry in the primary visual area of the mammalian neocortex. This will allow us to build on the wealth of current neurobiological knowledge concerning the properties and interconnectivity of neurons and the behaviour of local and long-range neuronal circuitry in this area of neocortex in response to visual stimuli. It must be stressed that our aim is not to build a detailed, biologically-precise

model of neocortex, but rather it is to identify and capture in a minimally complex model these key fundamental properties that underlie its remarkable information processing capabilities.

The specific aims of the proposed research programme can therefore be summarised as follows:

1. To build a computational model of minimal complexity that captures the fundamental information processing properties of the laminar microcircuitry of the primary visual area of neocortex. Specifically the properties we aim to capture are those of self-organisation, adaptation, and plasticity, which would enable the model to:

- i. develop feature selective neuronal properties and cortical preference maps in response to a combination of intrinsic, spontaneously-generated activity and complex naturalistic external stimuli, and

- ii. display experience-dependent and adaptation-induced plasticity, which optimally modifies the feature selectivity properties and preference maps in response to naturalistic stimuli.

2. To investigate the feasibility of designing VLSI circuitry which would be capable of realising the computational model, and thus demonstrate that the model can form the basis for a novel computational architecture with the same properties of self-organisation, adaptation, and plasticity.

## **3 The research programme**

The research programme involves a high level of integration of activities in neurobiological modelling, experimental neurobiology and the VLSI circuit design. It is organised into a set of such activities, each of which addresses a well-defined aim of the research programme, as described below.

### **3.1 Novel neocortical neuron and circuit connectivity models**

A basic premise of the research programme is that the neocortex is organised in a fairly stereotyped and highly modular fashion. Although much is already known about the structure and functional connectivity of microcircuits in the neocortex, the current state of knowledge is only sufficient to inform the initial design and construction of the proposed computational model. Recent work eg Thomson and Bannister (2003), has contributed important and detailed insights into the synaptic connectivity and the dynamic and plastic aspects of information transmission along these synaptic connections within a cortical column. The research will draw on this and on further, on-going work in order to more fully

elucidate the neuronal and synaptic connectivity which it is necessary to capture within the computational model in order to endow it with the self-organisation, adaptation and plasticity properties of cortical microcircuits.

In particular, the behaviour of circuit models of spiking neurons strongly depends on the properties of their constituents, the individual neurons, as well as on the synaptic connectivity between them. For example, the phase diagrams describing the dynamics of sparsely connected networks of excitatory and inhibitory neurons, which can exhibit different synchronous and asynchronous states (Brunel, 2000) change fundamentally when current-based integrate-and-fire neurons are replaced by conductance-based integrate-and-fire neurons as the constituents of the network. In order to provide the cortical modelling and the VLSI designs with the best possible description of single neurons (in terms of both accuracy and computational efficiency), the plan is to construct new types of integrate-and-fire neuron models that represent the biophysical mechanisms operating in biological neurons in a more realistic way, including the role of neuronal dendrites in the transformation of synaptic input into spike output.

Models will be validated by direct comparison with experimental data from experiments in brain slices and in the intact animal in vivo which describe the input-output relation of different types of neurons both at a functional, eg Chadderton et al., 2004, and a biophysical level, eg Häusser et al., 2001. We will focus on those characteristics of real neurons that are currently not, or only with insufficient accuracy, captured by the integrate-and-fire or spike response models currently available. We expect that models including the subthreshold dynamics of voltage-dependent conductances, including oscillatory behaviour, as well as the shunt conductances associated with action potential firing, which provide only a partial reset of the membrane potential in the neuron, will lead to more realistic yet compact descriptions of the input-output relations of different types of cortical neurons.

Single-neuron models will be complemented by three-dimensional geometric models of synaptic connectivity based on anatomical and physiological data from a large dataset of anatomically and physiologically identified, synaptically connected neurons which are being generated in a number of laboratories. Together these will provide an intra- and interlaminar wiring diagram of the cortical microcircuit. The functional properties of the synaptic connections between different types of neurons will be described by statistical distributions of the amplitudes and time courses of the synaptic conductances,

including a representation of short- and long-term synaptic plasticity.

### **3.2 Functional analysis and modelling of the neocortical microcircuit**

It will be essential to provide constraints for the proposed computational model. This is a non-trivial but essential task if we are to ensure that the modelling work does not result in “parameter explosion”. In particular, it will be necessary to constrain the model on the basis of the functional properties of the neurons and their interconnectivity in the cortical microcircuit. *In vivo*, cortical cells receive input from several thousand synaptic connections simultaneously, and only a proportion of these are connections from other cells within the cortical microcircuit. Some aspects of the intra-columnar connectivity revealed by intracellular recordings will form an essential part of the function of the cortical column, while other aspects are unimportant details that are best ignored in the proposed computational model. Constraining the model therefore means deciding which aspects are important, and estimating the strength of their contribution relative to other, external inputs and influences. This will require combining new extracellular recording techniques and novel statistical analysis and modelling approaches. Silicon array electrode techniques make it possible to record spiking activity simultaneously from dozens of neurons throughout a cortical microcircuit, in the living brain while it is carrying out its natural information processing tasks. In the past, simple filter models have been used to predict responses of individual neurons in sensory cortex (Schnupp et al, 2001). The research programme will aim at a dramatic improvement in these simplistic models through the use of novel statistical modelling techniques.

### **3.3 Learning rules for the development of stable self-organised feature selectivity**

The role of self-organisation in the stimulus-dependent development of orientation selectivity was first suggested by von der Malsburg and recently reviewed by Miller et al (1999) and Sur and Leamey, 2001. The latter suggest that spontaneous patterns of neural activity in the absence of visual stimuli may be sufficient in the early periods of development, after the initial cortical circuitry has been established, for the early development of orientation selectivity, but that the formation of orientation selectivity is strongly influenced by input activity to the developing cortex (Sur and Leamey, 2001; Sur et al, 1988). Experiments show that input activity has an influence on synaptic connections in the

cortical circuitry which gives rise to orientation map development and long-range horizontal intracortical connections in layers 2/3. A cortical microcircuit model would thus need to embody development of the dynamical interactions provided by intracortical connections in an activity-instructed self-organising process of map development. As yet, it would appear that no biological models exist which implement this activity-dependent self-organising process of development.

It has been demonstrated experimentally that the self-organised modification of synapses depends on the precise timing of spikes, causing the neuron to evolve in such a way as to be driven by its fastest and most reliable inputs. Therefore it seems reasonable to hypothesise that the learning rules which govern the self-organised emergence of cortical orientation selectivity should yield populations of selective cells, large enough to perform fast and reliable computation, yet small enough to be efficient. The investigation of these issues will lead to an understanding of how learning rules can self-organise the synaptic interconnectivity in the cortical microcircuit to produce a stable, sparse coded orientation selective network. An important component of this work will be to investigate how the stability of feature selectivity might be helped by recurrent interactions between neurons. These connections could break the symmetry and stabilise the synaptic weights, improving the stability of feature selectivity. The precise details of the learning rules are expected to be of crucial importance for the final selectivity patterns learned. This holds for both rate based as for spike timing dependent learning rules (van Rossum et al, 2000).

### **3.4 Neural coding of feature selectivity properties of cortical circuits**

Intimately related to the investigation of developmental self-organisation learning rules is the question of neural coding, i.e. of how neuronal populations represent sensory information. This is even more evident in the case of spike timing dependent learning rules. For instance, if stimulus features are coded across the cortical microcircuit by either precise spike times of individual neurons or by synchronous neuronal activity across neurons, it is of importance to know how such coding affects learning. Likewise, the developmental learning rules which result in specific patterns of synaptic connectivity have to support the selective neural coding of the stimulus feature set. A major objective of this part of the research programme will be to understand what advantages the laminar architecture of

the neocortex offers in terms of efficiency of information representation.

By using mathematical analysis techniques based on the principles of information theory, the role of columnar organization in cortical information representation has recently been investigated (Panzeri et al, 2003), but it is clear that the laminar organisation can provide both advantages and constraints that are as important. It has been shown that real cortical neurons encode information by timing of individual spikes with millisecond precision (Panzeri et al, 2001a) and investigated what mechanisms are need to read out this information, eg dendritic processing must be important to decode information if most information is encoded by the “label” of which neuron fired each spike, and not very important if instead neurons can sum up all spikes at the soma and still conserve all information (Panzeri et al, 2003).

Research in this part of the programme will extend these ideas by investigating in detail the information processing capabilities of laminar cortical circuits. In particular, we will determine (i) the “neuronal code” used in different laminae, i.e. which of the features (e.g. spike count, precise spike times, synchronization) characterizing the responses of neuronal populations in different laminae convey the most sensory information (ii) whether the precise synaptic connectivity within the laminar neocortical architecture is to some extent “optimal” for fast information transmission from one neuron/layer to another neuron/layer. By “optimal” we mean that the observed wiring comes close enough in terms of transmitted information to the best possible one. By fast we mean that all of this information must be transmitted by the model synaptic system in time scales as fast as the cortical ones (Panzeri et al, 2001b).

### **3.5 Learning rules for experience-dependent and adaptation-induced plasticity in the developed cortical microcircuit**

It is well known that the ability to detect small orientational differences can be significantly improved through training on a visual discrimination task over an extended period of time. This perceptual learning process is also seen to have a long-lasting effect, indicating that it must be the result of some form of long-term synaptic plasticity in the brain. Other characteristics of the learning, which can be psychophysically observed, such as the lack of transfer of the learning from one orientation to the orthogonal orientation or from one learned retinal location to a nearby nonoverlapping location, indicate that

the plasticity must involve the primary visual cortex, where the neurons have localised orientation selectivity and small receptive fields. Orientation plasticity has also been demonstrated in response to continuous visual stimulation for a period of seconds to minutes, a process known as adaptation. The results suggest that adaptation-induced orientation plasticity involves changes in circuit connectivity which then define a new preferred orientation. As proposed in the review by Dragoi and Sur (2003), the changes in orientation selectivity following adaptation imply a circuit mechanism that reorganizes responses across a broad range of orientations, and suggest that adaptation-induced orientation plasticity in primary visual cortex is a self-organised emergent property of a local cortical circuitry acting within a non-uniform orientation map. Research in this part of the programme will investigate the learning rules necessary to support the proposed emergence of adaptation-induced modification of orientation selectivity, and whether such learning rules can also support long-term experience-dependent plasticity of orientation selectivity.

### **3.6 Novel neocortical neuron and circuit connectivity models**

Spatiotemporal response properties of neurons in fully developed primary sensory areas are not static but can change on various timescales. Dynamic changes of response properties on long timescales have been assigned to adaptation and plasticity mechanisms. But responses also change on fast time-scales of a few to a few hundred milliseconds revealing rich dynamic features that result from the neural and synaptic activation dynamics and ongoing interactions between neurons within and across cortical microcircuits eg Bringuir et al (1999). Recent experiments eg Ringach et al (2002) indicate even more complex responses of cortical neurons and circuits to naturalistic stimuli. Spatiotemporal responses look similar to those for simple bar or grating stimuli, but there are also significant differences (Ringach et al, 2002). In part these differences seem to be related to influences from outside the classical receptive field: These experiments provide evidence that spatiotemporal response properties of cortical neurons are dynamically shaped in quite intricate ways by intrinsic neuronal and synaptic activation dynamics, interactions between neurons within the microcircuit, and longer ranging synaptic recurrent, feedforward and feedback circuits. These dynamical properties may underlie the surprisingly fast and adaptable information processing within the cortical microcircuit.

A general analytical approach has recently been described (Wennekers, 2002) that relates differently tuned enhanced and suppressed phases in a spatiotemporal response function to feedforward or recurrent pathways between participating cell classes. Although useful for some spatiotemporal phenomena, much of the complexity in real neural responses remains unexplained by such models. Models of complex spatiotemporal phenomena which incorporate the influence of ongoing and spreading activity, or responses to real-world stimuli, are still scarce.

### **3.7 Feasibility analysis for VLSI circuit design**

A major aim of the research programme is to use the computational model of the neocortical laminar microcircuit to define an efficient and implementable VLSI “building block” for a novel computational architecture. The mapping of the model of the cortical microcircuit into the VLSI circuit design for a novel computational architecture will require the investigation of detailed issues with respect to the numerical accuracy, performance, power consumption and area cost of novel analogue and digital circuit alternatives. These investigations will form the activity of this workpackage. We envisage a structure for the VLSI design based upon an analogue VLSI spiking neural substrate, interconnected via a digital VLSI address-event communication network, all controlled by a software configuration and control system. However, the integration of low-level neural models implemented by analogue VLSI circuits, with digital VLSI for signal routing and communication will need to go far beyond the simple protocols currently used by the neuromorphic engineers, and presents a major challenge. The research will also aim at understanding the implications, in both directions, for including or omitting certain components in the computational model, and assessing the relationship between the levels of description chosen for the computational model and the constraints of VLSI circuit design. In addition, the issues of optimisation (area, power) are present both in the neocortical microcircuit, and in silicon, so some direct analogies on a physical level will be investigated, eg the possible arrangement of the physical layout of devices in a way which is inspired by the 3-dimensional laminar architecture of the cortical microcircuit, in which connections between the cortical microcircuit “building blocks” are predominantly within or between certain layers.

## **4 Summary**

The neocortex of the brain subserves sensory perception, attention, memory and a spectrum of other

perceptual and cognitive functions, which combine to provide the biological system with its outstanding powers. It is clear that the brain carries out information processing in a fundamentally different way to today's conventional computers. The computational architecture of the brain clearly involves the use of highly parallel, asynchronous, nonlinear and adaptive dynamical systems, namely the laminar neural microcircuits of the neocortex. The fundamental aim of this research programme is to create a new brain-inspired computing architecture which possesses the basic properties of self-organisation, adaptation and plasticity manifest in the neural circuitry of the neocortex. The objective is a modular architecture based on a representation of a "stereotypical" cortical microcircuit. The research will focus on the laminar microcircuits of the primary visual cortex in order to build on the wealth of neurobiological knowledge concerning the behaviour and interconnectivity of neurons in this area of neocortex. However the wider objective would be to use the laminar microcircuitry of primary visual cortex as an exemplar for a stereotypical neocortically-inspired architecture. This will allow the architecture to be deployed in a wide range of perceptual tasks, and potentially also in cognitive tasks such as decision making, with minimal changes to the basic circuitry. The aim is not simply to build a detailed, biologically-precise model of primary visual cortex, but rather the challenge is to identify and capture the key fundamental principles and mechanisms that underlie the remarkable and ubiquitous information processing power of the neocortex.

## Acknowledgements

The collaborators in this research programme are: Alex Thomson (School of Pharmacy, London); Michael Häusser and Arnd Roth (UCL); Jan Schnupp (Oxford); Mark van Rossum and David Willshaw (Edinburgh); Stefano Panzeri, Piotr Dudek and Steve Furber (Manchester); Thomas Wennekers and Susan Denham (Plymouth). I would like to fully and gratefully acknowledge their individual contributions to the description of the aims, objectives and research activities which I have set out on their behalf in this paper. Any mistakes herein are however my own. The COLAMN research programme is supported by a grant from the UK Engineering and Physical Research Council under the Novel Computation Initiative.

## References

Silberberg G, Gupta A, Markram H (2002) *Trends Neurosci* 25, 227-230

- Mountcastle V (1997) *Brain* 120:701-722
- Thomson AM, Bannister AP (2003) *Cereb Cortex* 13, 5-14
- Callaway EM.(1998) *Annu Rev Neurosci* 21:47-74
- Häusser M, Spruston N, Stuart GJ (2000) *Science*, 290:739-744;
- Brunel N (2000) *Network* 11:261-280
- Chadderton P, Margrie TW, Häusser M (2004) *Nature* 428:856-860
- Häusser M, Major G, Stuart GJ (2001) *Science*. 291:138-41.
- Schnupp JW, Mrsic-Flogel TD, King AJ (2001) *Nature* 414:200-204
- Miller KD, Erwin E, Kayser A. (1999) *J Neurobiol.* 41:44-57
- Sur M, Leamey CA (2001) *Nature Reviews Neurosci.* 2:251-262
- Sur M, Garraghty PE, Roe AW (1988) *Science* 242:1437-41
- van Rossum M, Bi GQ, Turrigiano GG (2000), *J. Neurosci.* 20, 8812-8821
- Panzeri S, Petroni F, Petersen RS, Diamond ME (2003) *Cerebral Cortex* 13: 45-52
- Panzeri S, Petersen R, Schultz SR, Lebedev M, Diamond ME (2001a) *Neuron* 29: 769-777
- Panzeri S, Rolls ET, Battaglia F, Lavis R (2001b) *Network* 12: 423- 440
- Dragoi V, Sur M (2003) In Eds. Chalupa LM and Werner JS, *The Visual Neurosciences*, MIT Press.
- Bringuir V, Chavane F, Glaeser L, Fregnac Y (1999) *Science* 283, 695-699
- Ringach DL, Hawken M, Shapley R (2002) *J Vision* 2, 12-14
- Wennekers T. (2002) *Neural Computation* 14: 1801-25

# High-Performance Computing for Systems of Spiking Neurons

Steve Furber and Steve Temple\*

\*The University of Manchester  
School of Computer Science

Oxford Road, Manchester M13 9PL, UK

steve.furber, steven.temple@manchester.ac.uk

Andrew Brown†

†The University of Southampton

Department of Electronics and Computer Science  
Southampton, Hampshire, SO17 1BJ, UK

adb@ecs.soton.ac.uk

## Abstract

We propose a bottom-up computer engineering approach to the Grand Challenge of understanding the Architecture of Brain and Mind as a viable complement to top-down modelling and alternative approaches informed by the skills and philosophies of other disciplines. Our approach starts from the observation that brains are built from spiking neurons and then progresses by looking for a systematic way to deploy spiking neurons as components from which useful information processing functions can be constructed, at all stages being informed (but not constrained) by the neural structures and microarchitectures observed by neuroscientists as playing a role in biological systems. In order to explore the behaviours of large-scale complex systems of spiking neuron components we require high-performance computing equipment, and we propose the construction of a machine specifically for this task – a massively parallel computer designed to be a universal spiking neural network simulation engine.

## 1 Introduction

### 1.1 Neurons

The basic biological control component is the neuron. A full understanding of the ‘Architecture of Brain and Mind’ (Sloman, 2004) must, ultimately, involve finding an explanation of the phenomenological observations that can be expressed in terms of the interactions between the neurons that comprise the brain (together with their sensory inputs, actuator outputs, and related biological processes).

Neurons appear to be very flexible components whose utility scales over systems covering a vast range of complexities. Very simple creatures find a small number of neurons useful, honey bees find it economic to support brains comprising around 850,000 neurons, and humans have evolved to carry brains comprising  $10^{11}$  neurons or so. The component neuron used this range of complexities is basically the same in its principles of operation, so in some sense it has a universality similar to that enjoyed by the basic logic gate in digital engineering.

There is a further similarity between neurons and logic gates: both are multiple-input single-output components. However, while the typical fan-in (the number of inputs to a component) and fan-out (the

number of other components the output of a particular component connects to) of a logic gate is in the range 2 to 4, neurons typically have a fan-in and fan-out in the range 1,000 to 10,000. (It is easy to show that that mean fan-in and fan-out in a system are the same – they are just different ways of counting the number of connections between components.)

A more subtle difference between a logic gate and a neuron is in the dynamics of their internal processes. Whereas a logic gate implements a process that is essentially static and defined by Boolean logic, so that at any time from a short time after the last input change the output is a well-defined stable function of the inputs, a neuron has complex dynamics that includes several time constants, and its output is a time series of action potentials or ‘spikes’. The information conveyed by the neuron’s output is encoded in the timing of the spikes in a way that is not yet fully understood, although rate codes, population codes and firing-order codes all seem offer valid interpretations of certain observations of spiking activity.

Accurate computer models of biological neurons exist, but these are very complex. Various simpler models have been proposed that capture some of the features of the biology but omit others. The difficulty lies in determining which of the features are



essential to the information processing functions of the neuron and which are artefacts resulting from the way the cell developed, its need to sustain itself, and the complex evolutionary processes that led to its current form.

## 1.2 Neural microarchitecture

The universality of the neuron as a component is also reflected in certain higher-level structures of the brain. For example, the cortex displays a 6-layer structure and a regularity of interconnect between the neurons in the various layers that can reasonably deserve the application of the term ‘microarchitecture’. The same regular laminar cortical microarchitecture is in evidence across the cortex in regions implementing low-level vision processes such as edge-detection and in regions involved in high-level functions such as speech and language processing. This apparent ‘universality’ (used here to describe one structure that can perform any function) of the cortical microarchitecture suggests there are principles being applied here the understanding of which could offer a breakthrough in our understanding of brain function.

In contrast to the regularity and uniformity of the microarchitecture, the particular connectivity patterns that underpin these structures appear to be random, guided by statistical principles rather than specific connectivity plans. The connectivity is also locally adaptive, so the system can be refined through tuning to improve its performance.

## 1.3 Engineering with neurons

As computer engineers we find the neuron’s universality across wide ranges of biological complexity to be intriguing, and there is a real challenge in understanding how this component can be used to build useful information processing systems. There is an existence proof that this is indeed possible, but few pointers to how the resulting systems might work.

There are other ‘engineering’ aspects of biological neurons that are interesting, too. We have already mentioned the regularity of neural microarchitecture. The power-efficiency of neurons (measured as the energy required to perform a given computation) exceeds that of computer technology, possibly because the neuron itself is a very low performance component. While computer engineers measure gate speeds in picoseconds, neurons have time constants of a millisecond or longer. While computer engineers worry about speed-of-light limitations and the number of clock cycles it takes to get a signal across a chip, neurons communicate at a few metres per second. This very relaxed performance at the technology level is, of course, compensated by the very high levels of parallelism and connectivity of the

biological system. Finally, neural systems display levels of fault-tolerance and adaptive learning that artificial systems have yet to approach.

We have therefore decided to take up the challenge to find ways to build useful systems based upon spiking neuron components (for example, Furber, Bainbridge, Cumpstey and Temple, 2004), and we hope that this will lead to mutually-stimulating interactions with people from many other disciplines whose approach to the same Grand Challenge, of understanding the Architecture of Brain and Mind, will be quite different from our own.

## 2 Relevance to GC5

What has any of this engineering really got to do with the Grand Challenge of understanding the Architecture of Brain and Mind?

As this is aimed at a broad audience, not many of whom are computer engineers, we will digress briefly to consider what computer engineers may bring to this Grand Challenge. To begin with, it is useful to appreciate the skills and mindset that a computer engineer, for better or for worse, possesses. What can a person whose stock-in-trade consists of logic gates, microchips and printed circuit boards contribute to the bio-psycho-philosophical quest to understand the workings of the mind?

### 2.1 A Computer Engineer’s manifesto

To a computer engineer ‘understand’ has a specific meaning that is different from what a scientist means by the same word, which is in turn probably different from the meanings used by other disciplines. To the scientist, understanding is to have a repeatably-verifiable explanation of a phenomenon. To the engineer, understanding means to be able to go away and build another artefact that works in the same way. The scientist’s analysis reduces a complex phenomenon into its basic components; this is complemented by the engineer’s ability to take those components, or components that encapsulate the same essential behaviour, and synthesize them back into a functioning system.

Thus, when a computer engineer claims to ‘understands’ how a mobile phone works, the statement can be interpreted as meaning that they can (at least in principle) explain when every one of the 100 million or so transistors switches, why it switches, what will happen if it fails to switch, and so on. OK, we might get on less secure ground when describing the chemistry of the lithium-ion battery and the details of the radio and antenna design or the higher levels of the software. And when it comes to explaining why the plastic case is pink and the buttons are arranged in swirling patterns with no obvious ergonomic objective we are completely lost! But back in

the familiar territory of the digital transistor circuits we have a vocabulary comprising baseband processors, DSPs, maximum likelihood error correctors, RAMs, buses, interrupts, and so on, that together provide a language of description at multiple levels of abstraction from an individual transistor to the lower levels of the system software. This enables us to describe in very fine detail how the phone works and, more particularly, how you might make another working phone at lower cost and with better battery life.

This is the approach we bring to understanding the Architecture of Brain and Mind. In neuroscience we see that there are pretty accurate models of the basic component from which brains are built – the neuron. There are some rather sketchy and limited descriptions of how these components are interconnected and how they behave in natural networks, and there is rather better information about their macro-level modular organisation and gross activity. The weakest part of the neuroscientists’ analysis (for very good reason – it is hard to apply reductionist principles to systems whose interesting characteristics depend on their organizational complexity) is at the intermediate levels between the component neurons (where analysis is applicable) and the macro-organisation (where mean field statistics work).

This intermediate level is precisely the level at which the computer engineer may have something to offer. Assembling basic components into functional units, implementing useful computational processes based on networks of dynamical systems, these are all second nature to the computer engineer once we have come to grips with the spiking neuron as a component. As we observed earlier, it even looks a bit like a logic gate – several inputs but only one output.

The intrinsic dynamics of a neuron may confound the computer engineer who is used to working only with digital circuits that are controlled by the extrinsic straitjacket of a clock signal, but a small minority of us are proficient in building circuits whose sequential behaviour is intrinsic – members of the class of digital circuit generally described as asynchronous or self-timed. The knowledge we hold on how to build reliable, highly complex asynchronous digital systems *may just* provide us with new insights into the highly complex asynchronous neural systems that provide the hardware platform upon which the brain and mind are built.

## 2.2 GC5 methodology

Our approach to this Grand Challenge is essentially bottom-up, which will complement the top-down and middle-out approaches that are better-suited to those who bring different skills and mindsets from other disciplines.

The bottom-up approach starts from the concept of a neuron as a basic component, and then seeks useful compositions of neurons to create (and implement) increasingly higher levels of functional abstraction. These compositions may be inspired by neuroscience; for example, we have an involvement in the EPSRC-funded COLAMN project which has as its goal the creation of novel computational architectures based on the laminar microarchitecture of the neocortex, with considerable input from the ‘wet’ neuroscientists in the project. Or they may be designed in the abstract; for example our earlier work on  $N$ -of- $M$  coded sparse distributed memories (Furber, Bainbridge, Cumpstey and Temple, 2004) – with at best tenuous relevance to biology.

A feature of this research is that it can yield a positive outcome in two distinct ways. It may contribute to the scientific objective of understanding the architecture of brain and mind, and/or it may contribute to the engineering objective of delivering better/different/novel models of computation. Either of these outcomes would justify our engagement, and with a following wind we might just achieve both...

In order to pursue this research agenda we need a sandpit in which we can experiment with neuron components on a large scale, hence the massively parallel high-performance computer theme that we will turn to shortly. This large-scale engineering project brings with it additional research aspects relating to fault-tolerance, autonomic computing, self-healing, networks-on-chip, and so forth, all of which add to the engineering challenge but probably contribute little to the GC5 agenda.

## 3 Objectives

We have set ourselves the objective of simulating a billion spiking neurons in real time while making as few assumptions as possible about what a neuron is and how the neurons are connected. We approach this by viewing a neural system as an event-driven dynamical system – a hybrid system where a (large) set of components, each of which operates in continuous time (and is characteristically described by a set of differential equations), interact through discrete events.

In order to retain complete flexibility in the internal neural dynamics we implement the real-time differential equation solvers (which will typically use discrete-time fixed-point approximations) in software, and then exploit the high speeds of electronic signalling to communicate the discrete inter-neuron communication events around the system in a time which is close to instantaneous on the time-scales of the neuron dynamics. This allows us to use a virtual mapping from the physical structure of the

biological system we are modelling to the physical structure of the electronic system we are running the model on.

## 4 Neural computation

Any computation system must achieve a balance between its processing, storage and communication functions. It is useful to consider how these three functions are achieved in neural systems.

### 4.1 Processing

The neuron itself performs the processing function. It produces output events in response to input events through a non-linear transfer function, which we will model using suitable differential equations whose complexity is limited only by the available computing power.

The simplest neuron models process inputs by taking a linear sum of the inputs, each weighted by the strength of its respective synapse. When the inputs are spike events the multiplication implied by the weighting process reduces to repeated addition. Multiplication by repeated addition is usually inefficient, but here many inputs are likely to be inactive at any time and multiplication by zero by repeated addition is supremely efficient!

The weighted input sum is then used to drive the neural dynamics. A leaky-integrate-and-fire (LIF) model applies an exponential decay to the effect of each input, but if enough inputs fire close together in time to push the total activation past a threshold, the neuron fires its output and resets. More sophisticated models have more complex dynamics. For example, the models by Izhikevich (2004) are based on mathematical bifurcation and display a more diverse range of biologically-relevant behaviours than the LIF model.

### 4.2 Communication

Communication in neural systems is predominantly through the propagation of spike ‘events’ from one neuron to the next. The output from the neuron’s body – its soma – passes along its axon which conveys the spike to its many target synapses. Each synapse uses chemical processes to couple the spike to the input network – the dendritic tree – of another neuron.

Since the spike carries no information in its shape or size, the only information is which neuron fired and when it fired. In a real-time simulation the timing is implicit (and the communication, being effectively instantaneous, preserves the timing), so all we need to communicate is the identity of the

neuron that fired, and we must send that to every neuron to which the firing neuron connects.

In the biological system the identity of a firing neuron is spatially encoded – each neuron has its own physical axon. In our system we cannot implement an equivalent level of physical connectivity so instead we use logical encoding by sending a packet identifying the firing neuron around a network that connects all of the components together.

### 4.3 Storage

It is in the storage of information that the neuron’s story becomes most complex. There are many processes that can be seen as storing information, some operating over short time scales and some very long-term. For example:

- the neural dynamics include multiple time constants, each of which serves to preserve input information for some period of time;
- the dynamical state of the network may preserve information for some time;
- the axons carry spikes at low speeds and therefore act as delay lines, storing information as it propagates for up to 20ms;
- the coupling strength of a synapse is, in many cases, adaptive, with different time constants applying to different synapses.

The primary long-term storage mechanism is synaptic modification (within which we include the growth of new synapses).

In a real-time modelling system we expect the modelling to capture the neural and networks dynamics, and hence the contributions these mechanisms make to information storage. The axon delay-line storage does not come so easily as we have deliberately exploited the high speeds of electronic signalling to make spike communication effectively instantaneous in order to support a virtual mapping of the physical structures. It is likely that the axon delay is functionally important, so we must put these delays back in, either by delaying the issue of the spike or by delaying its effect at the destination. Either solution can be achieved in software, but both have drawbacks, and this remains one of the trickier aspects of the design to resolve to our complete satisfaction.

The final storage process is the most fundamental: synaptic weight adaptivity. Here we require long-term stability and support for a range of learning algorithms. We will exploit the fact that digital semiconductor memory is a mass-produced low-cost commodity, and the proposed machine is built around the use of commodity memory for storing synaptic connectivity information.

Indeed, as we shall see in the next section, the major resources in a neural computation system revolve around the synapses, not around the neural dynamics.

## 5 Computing requirements

Various estimates have been offered for the computational power required to run a real-time simulation of the human brain based on reasonably realistic neuron models. The answer generally comes out in the region of  $10^{16}$  instructions per second, which is some way beyond the performance of a desktop PC or workstation, but not far beyond the performance of the petaFLOP supercomputers currently in design.

The route to this performance estimate can be summarized as follows: the brain comprises around  $10^{11}$  neurons, each with of the order of 1,000 inputs. Each input fires at an average rate of 10 Hz, giving  $10^{15}$  connections per second, and each connection requires perhaps 10 instructions.

Note that this estimate is based on the computing power required to handle the synaptic connections. Modelling the neuron dynamics is a smaller part of the problem:  $10^{11}$  neurons each requiring a few 10s of instructions to update their dynamics perhaps  $10^3$  times a second, requiring in total an order of magnitude less computing power than the connections.

A similar calculation yields the memory requirements of such a simulation:  $10^{14}$  synapses each require of the order of a few bytes, so around  $10^{14}$  bytes of synaptic connection data are required.

At present the only way a machine of such capacity can be conceived is to employ a massively parallel architecture. This is likely to remain true even with future developments in CMOS technology as further increases in clock speed and individual processor throughput are unlikely to be great, as evidenced by the recent trend towards multi-core processors from all of the leading microprocessor vendors. The future of the microprocessor is in chip multiprocessors, and the future of high-performance computing is in massively parallel systems.

Fortunately, the problem of simulating very large numbers of neurons in real time falls into the class of ‘embarrassingly’ parallel applications, where the available concurrency allows the trade-off of processor performance against the number of processors to be totally flexible. The issue, then, is to determine how such a system might be optimised. What are the relevant metrics against which to make decisions on the systems architecture?

We propose that the primary metrics should be *performance density* (measured in MIPS/mm<sup>2</sup> of silicon) and *power-efficiency* (measured in

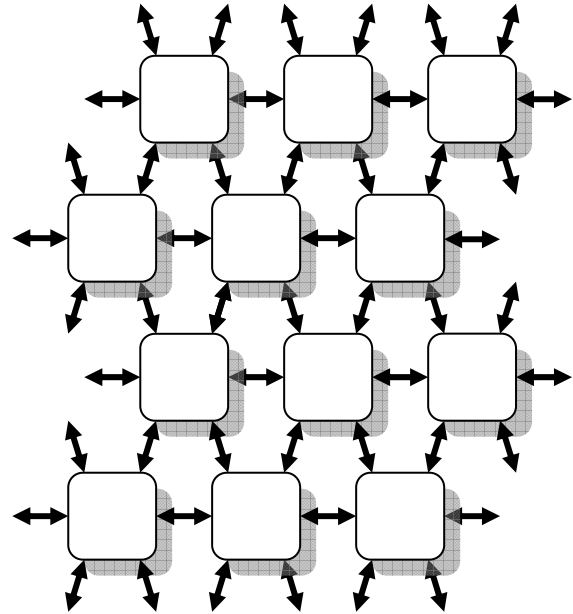


Figure 1: The system architecture.

MIPS/watt). The former is the primary determinant of the capital cost of the machine, while the latter influences both the capital cost – in terms of the cooling plant – and the running cost – a machine such as this demands a significant electrical power budget.

A choice then has to be made between using a large number of high-performance processors or an even larger number of more power-efficient embedded processors. Here the metrics can be our guide – embedded processors win handsomely on power-efficiency, and to a lesser extent on performance density, over their much more complex high-end counterparts.

That, then sets the course for this work. The objective is to build a machine, based on large numbers of small processors, that has the potential to scale up to the levels of parallelism and performance necessary to model a brain in real time. Admittedly, modelling a complete human brain is some way beyond our current goals, but we should be able to model substantial parts of the human brain and complete brains of less complex species with what we propose here, which is a machine capable of modelling a billion spiking neurons in real time.

## 6 SpiNNaker

A spinnaker is a large foresail that enables a yacht to make rapid progress in a following wind (see reference to ‘following wind’ in Section 2.2 above!). We have adopted SpiNNaker as a name for our project because it comes close to a contraction of ‘a (uni-

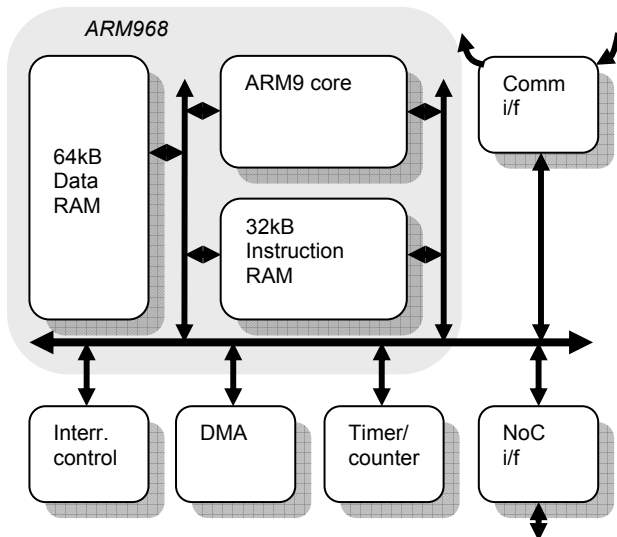


Figure 2: Processor subsystem organization.

versal) Spiking Neural Network architecture’, provided you say it quickly enough. Again, this is our goal: to build a computer system that is as universal as we can make it in its ability to simulate large systems of spiking neurons, preferably in real time.

The following description of the system is largely extracted from Furber, Temple and Brown (2006).

## 6.1 System architecture

The system is implemented as a regular 2D array of nodes interconnected through bi-directional links in a triangular formation as illustrated in Fig. 1. The 2D mesh is very straightforward to implement on a circuit board and also provides many alternative routes between any pair of nodes which is useful for reconfiguration to isolate faults. (Nothing in the communications architecture precludes the use of a more complex topology if this proves advantageous.)

Each node in the network comprises two chips: a chip multiprocessor (CMP) and an SDRAM, with the integer processing power of a typical PC but at much lower power and in a compact physical form. The six bidirectional links support a total of 6 Gbit/s of bandwidth into and out of the node. A system of 100 x 100 nodes will deliver a total of 40 teraIPS, sufficient to simulate perhaps 200 million spiking neurons in real time, and will have a bisection bandwidth of 200 Gbit/s.

## 6.2 ARM968 processor subsystem

For the reasons already outlined, we choose to base the system around a massively-parallel array of

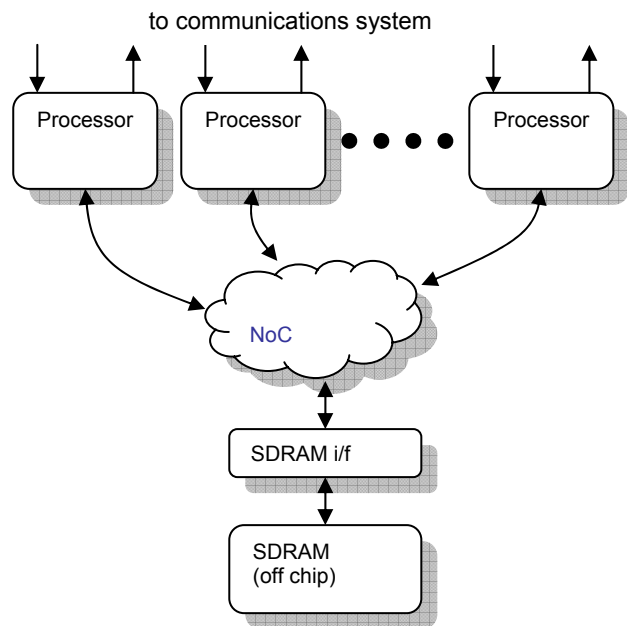


Figure 3: The CMP system NoC.

power-efficient embedded processors, and have chosen the ARM968 as offering the best balance of performance, area, power-efficiency and ease of use for our purposes. The ARM968 is a synthesizable ARM9 processor core with tightly-coupled instruction and data memories, and an integral on-chip bus (ARM Ltd, 2004). Each processor subsystem comprises a processor, instruction and data memory, timers, interrupt and DMA controllers and a communications NoC interface (Fig. 2).

We estimate that a 200 MIPS integer embedded ARM9 processor should be able to model 1,000 leaky-integrate-and-fire (or Izhikevich) neurons, each with 1,000 inputs firing on average at 10 Hz, in real time. The synaptic connectivity information for these neurons requires around 4 Mbytes of memory and the neuron state requires around 50 Kbytes of memory. These estimates have led us to adopt a hybrid architecture where the synaptic data is held in an off-chip SDRAM while the neural state data is held in on-chip memory local to each embedded processor. A processing node in our system therefore comprises two ICs: a chip multiprocessor (CMP) with about twenty 200 MIPS embedded ARM9 processors, and an SDRAM chip. The synaptic data is accessed in large blocks and this enables an SDRAM bandwidth of around 1 GByte/s to provide this data at the required rate.

The processors on a CMP share access to the SDRAM using a self-timed packet-switched Network-on-Chip (NoC). This fabric will use the CHAIN technology (Bainbridge and Furber, 2002), developed at the University of Manchester and commercialized by Silistix Ltd, which gives a through-

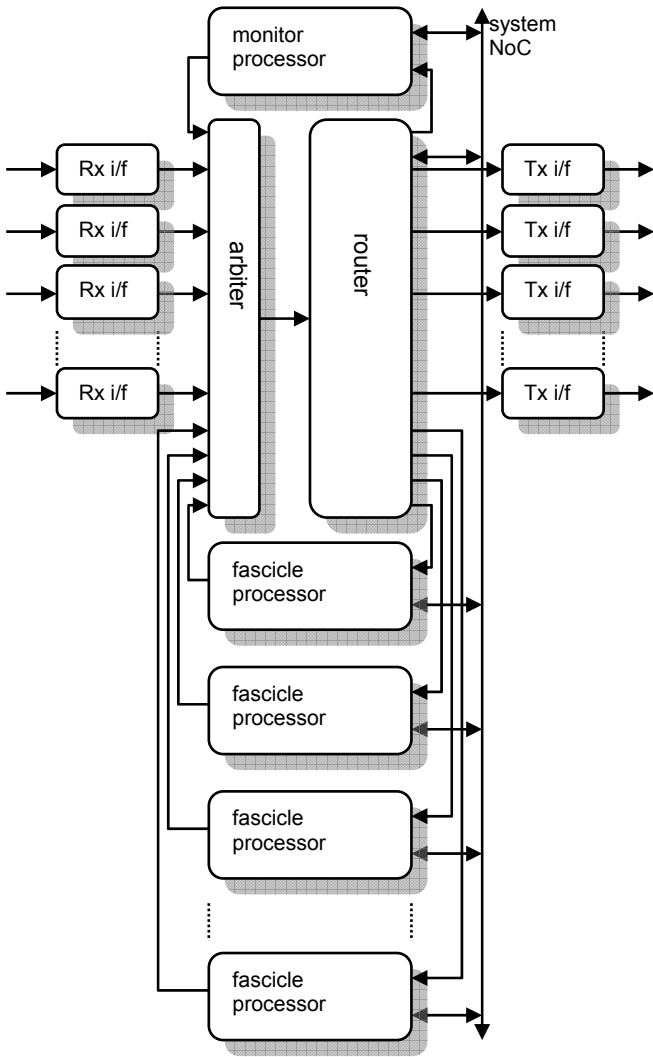


Figure 4: The communications NoC.

put of around 1 Gbit/s per 6-wire link (Bainbridge, Plana and Furber, 2004). The organization of the system NoC that connects the processor subsystems to the SDRAM is shown in Fig. 3.

### 6.3 The communications system

The major challenge in designing a scalable multi-chip neural modeling system is to emulate the very high connectivity of the biological system. The high fan-in and fan-out of neurons suggests that an efficient multicast communication system is required. We propose a communication NoC fabric based upon address-event signaling, but carried over a second self-timed packet-switched fabric rather than the usual bus-based fabric. The self-timed fabric decouples the many different clock domains within and across the CMPs.

The inter-chip communication uses a self-timed signalling system on an 8-wire inter-chip link that employs a self-timed 2-of-7 non-return-to-zero (NRZ) code (Bainbridge, Toms, Edwards and Furber, 2003) with an NRZ acknowledge. 16 of the 21 possible 2-of-7 codes are used to carry four bits of data, and a 17<sup>th</sup> code carries end-of-packet (EOP). Each 8-wire link has a capacity of around 1 Gbit/s when connecting two CMPs on the same circuit board, matching the on-chip bandwidth of a CHAIN link, and the self-timed protocol guarantees correct operation (albeit at a lower data rate) when the CMPs are on different circuit boards, automatically adapting to the addition delays incurred by any signal buffering that may be required.

The complete communications subsystem on a CMP is illustrated in Fig. 4. The inter-chip links are accessed via input protocol converters ('Rx i/f' in Fig. 4) that translate the off-chip 2-of-7 NRZ codes to the on-chip CHAIN codes, and output protocol converters ('Tx i/f') that perform the inverse translation. Each of the on-chip processing subsystems ('fascicle processor') is also a source of network traffic and a potential destination. All of the on- and off-chip sources are merged through an asynchronous arbiter into a single stream of packets that passes through the multicast router which will, in turn, propagate the packet to a subset of its on- and off-chip outputs. The monitor processor is identical to a fascicle processor but is dedicated to system management functions rather than neural modeling. It is chosen from among the fascicle processors at boot time; the flexibility in its selection removes another possible single point of failure on the CMP, improving fault tolerance.

The heart of the communication subsystem is the associative multicast router which directs every incoming packet to one or more of the local processors and output links using a routing key based on the source ID and a route look-up table.

### 6.4 Fault-tolerance

The scale of the proposed machine demands that it be designed with a high degree of fault-tolerance. Since the neural system we are modelling has intrinsic fault-tolerant properties (healthy humans lose about one neuron a second throughout their adult life; neurodegenerative diseases incur much higher loss rates) this capacity will be transferred to the simulated system to some degree. However, many of the techniques we employ to map the natural system onto the electronic model concentrate distributed biological processes into single points of failure in the model: a single microprocessor models a thousand neurons; a single inter-chip link carries the spikes on perhaps a million axons. Thus we must

engineer some additional resilience into the electronic system.

The highly regular structure of the machine comes to our aid here. If a processor fails we can migrate its workload to another, on the same or on a different chip. This will almost certainly lead to a glitch in the system's real-time performance, but our goal is to minimise the size of this glitch and to build a system that is continuously monitoring its own performance and migrating its workload to minimise congestion, so a major failure just puts a higher transient demand on the workload management processes.

An inter-chip link failure (whether permanent or transient, perhaps due to local congestion) will be handled in the first instance at the hardware level, redirecting traffic automatically via an adjacent link, before invoking the performance management software to carry out a more permanent solution.

At all stages in the design we are exploring opportunities to identify mechanisms that support real-time fault-tolerance, some of which exploit the intrinsic fault-tolerance of neural systems but many of which will contribute to a separate research agenda in the area of autonomic, self-healing systems.

## 7 Conclusions

The Grand Challenge of understanding the Architecture of Brain and Mind is a multidisciplinary quest that will require many complementary approaches to run concurrently, each feeding off the others as sources of inspiration, ideas and sanity checks. The system synthesis approach of computer engineers such as ourselves may have something to contribute as a component of the overall process. An understanding of complex asynchronous interactions within digital systems seems highly relevant to the task of understanding the complex asynchronous interactions between neurons.

In our quest to understand the dynamics of systems of asynchronous spiking neurons we hope to contribute both to providing tools that help understand biological brains and also to the creation of novel computational systems that are inspired by biology, but whose link to biology may ultimately become tenuous.

To this end we propose to construct a massively-parallel computer that implements a universal spiking neural network architecture, SpiNNaker. Based on a chip multiprocessor incorporating around twenty 200 MIPS embedded ARM968 processors, and employing a communications infrastructure specifically designed to support the multicast routing required for neural simulation, this system will scale to hundreds of thousands of processors modelling up to a billion neurons in real time. It will form

a 'sandpit' in which we, and others with similar interests, can experiment with large-scale systems of spiking neurons to test our network topologies and neural models in order to validate (or disprove) our theories of how neurons interact to generate the hardware platform that underpins the Architecture of Brain and Mind.

## Acknowledgements

This work is supported by the EPSRC Advanced Processor Technologies Portfolio Partnership Award. Steve Furber holds a Royal Society-Wolfson Research Merit Award. The support of ARM Ltd and Silistix Ltd for the proposed work is gratefully acknowledged.

## References

- ARM Ltd. ARM968E-S Technical Reference Manual. DDI 0311C, 2004.  
<http://www.arm.com/products/CPUs/ARM968E-S.html>
- W. J. Bainbridge and S. B. Furber. CHAIN: A Delay-Insensitive Chip Area Interconnect. *IEEE Micro*, 22(5):16-23, 2002.
- W. J. Bainbridge, L. A. Plana and S. B. Furber. The Design and Test of a Smartcard Chip Using a CHAIN Self-timed Network-on-Chip. *Proc. DATE'04*, 3:274, Paris, Feb 2004.
- W. J. Bainbridge, W. B. Toms, D. A. Edwards and S. B. Furber. Delay-Insensitive, Point-to-Point Interconnect using m-of-n codes. *Proc. Async* :132-140, Vancouver, May 2003.
- S. B. Furber, W. J. Bainbridge, J. M. Cumpstey and S. Temple. A Sparse Distributed Memory based upon N-of-M Codes. *Neural Networks* 17(10):1437-1451, 2004.
- S. B. Furber, S. Temple and A. D. Brown. On-chip and Inter-Chip Networks for Modelling Large-Scale Neural Systems. *Proc. ISCAS'06*, Kos, May 2006 (to appear).
- E. M. Izhikevich. Which Model to Use for Cortical Spiking Neurons? *IEEE Trans. Neural Networks*, 15:1063-1070, 2004.
- A. Sloman (ed.). The Architecture of Brain and Mind. UKCRC Grand Challenge 5 report, 2004.

# Principles Underlying the Construction of Brain-Based Devices

Jeffrey L. Krichmar  
The Neurosciences Institute  
10640 John J. Hopkins Drive  
San Diego, CA 92121 USA  
krichmar@nsi.edu

Gerald M. Edelman  
The Neurosciences Institute  
10640 John J. Hopkins Drive  
San Diego, CA 92121 USA  
edelman@nsi.edu

## Abstract

Without a doubt the most sophisticated behaviour seen in biological agents is demonstrated by organisms whose behaviour is guided by a nervous system. Thus, the construction of behaving devices based on principles of nervous systems may have much to offer. Our group has built series of brain-based devices (BBDs) over the last 14 years to provide a heuristic for studying brain function by embedding neurobiological principles on a physical platform capable of interacting with the real world. These BBDs have been used to study perception, operant conditioning, episodic and spatial memory, and motor control through the simulation of brain regions such as the visual cortex, the dopaminergic reward system, the hippocampus, and the cerebellum. Following the brain-based model, we argue that an intelligent machine should be constrained by the following design principles: (i) it should incorporate a simulated brain with detailed neuroanatomy and neural dynamics that controls behaviour and shapes memory, (ii) it should organize the unlabeled signals it receives from the environment into categories without a priori knowledge or instruction, (iii) it should have a physical instantiation, which allows for active sensing and autonomous movement in the environment, (iv) it should engage in a task that is initially constrained by minimal set of innate behaviours or reflexes, (v) it should have a means to adapt the device's behaviour, called value systems, when an important environmental event occurs, and (vi) it should allow comparisons with experimental data acquired from animal nervous systems. Like the brain, these devices operate according to selectional principles through which they form categorical memory, associate categories with innate value, and adapt to the environment. Moreover, this approach may provide the groundwork for the development of intelligent machines that follow neurobiological rather than computational principles in their construction.

## 1 Introduction

Although much progress has been made in the neurosciences over the last several decades, the study of the nervous system is still a wide open area of research. This is not due to a lack of first-rate research by the neuroscience community, but instead it reflects the complexity of the problem. Therefore, novel approaches to the problem, such as computational modelling and robotics, may be necessary to come to a better understanding of brain function. Moreover, as our models and devices become more sophisticated and more biologically realistic, the devices themselves may approach the complexity and adaptive behaviour that we associate with biological organisms and may find their way in practi-

cal applications. In this review, we will outline what we believe are the design principles necessary to achieve these goals (Krichmar and Edelman, 2005; Krichmar and Reeke, 2005). We will illustrate how these principles have been put into practice by describing two recent brain-based devices (BBDs) from our group.

## 2 Brain-Based Modelling Design Principles

### 2.1 Incorporate A Simulated Brain With Detailed Neuroanatomy And Neural Dynamics



Models of brain function should take into consideration the dynamics of the neuronal elements that make up different brain regions, the structure of these different brain regions, and the connectivity within and between these brain regions. The dynamics of the elements of the nervous system (e.g. neuronal activity and synaptic transmission) are important to brain function and have been modelled at the single neuron level (Borg-Graham, 1987; Bower and Beeman, 1994; Hines and Carnevale, 1997), network level (Izhikevich et al., 2004; Pinsky and Rinzel, 1994), and synapse level in models of plasticity (Bienenstock et al., 1982; Song et al., 2000; Worgotter and Porr, 2005). However, structure at the gross anatomical level is critical for function, and it has often been ignored in models of the nervous system. Brain function is more than the activity of disparate regions; it is the interaction between these areas that is crucial as we have shown in Darwins IV through X (Edelman et al., 1992; Krichmar and Edelman, 2005; Krichmar et al., 2005b; Seth et al., 2004). Brains are defined by a distinct neuroanatomy in which there are areas of special function, which are defined by their connectivity to sensory input, motor output, and to each other.

## 2.2 Organize the Signals from the Environment into Categories Without a *a priori* Knowledge or Instruction

One essential property of BBDs, is that, like living organisms, they must organize the unlabeled signals they receive from the environment into categories. This organization of signals, which in general depends on a combination of sensory modalities (e.g. vision, sound, taste, or touch), is called *perceptual categorization*. Perceptual categorization in models (Edelman and Reeke, 1982) as well as living organisms makes object recognition possible based on experience, but without *a priori* knowledge or instruction. A BBD selects and generalizes the signals it receives with its sensors, puts these signals into categories without instruction, and learns the appropriate actions when confronted with objects under conditions that produce responses in value systems.

## 2.3 Active Sensing and Autonomous Movement in the Environment

Brains do not function in isolation; they are tightly coupled with the organism's morphology and environment. In order to function properly, an agent, artificial or biological, needs to be situated in the real world (Chiel and Beer, 1997; Clark, 1997). Therefore, models of brain function should be em-

bodied in a physical device and explore a real as opposed to a simulated environment. For our purposes, the real environment is required for two reasons. First, simulating an environment can introduce unwanted and unintentional biases to the model. For example, a computer generated object presented to a vision model has its shape and segmentation defined by the modeller and directly presented to the model, whereas a device that views an object hanging on a wall has to discern the shape and figure from ground segmentation based on its own active vision. Second, real environments are rich, multimodal, and noisy; an artificial design of such an environment would be computationally intensive and difficult to simulate. However, all these interesting features of the environment come for "free" when we place the BBD in the real world. The modeller is freed from simulating a world and need only concentrate on the development of a device that can actively explore the real world.

## 2.4 Engage in a Behavioural Task

It follows from the above principle that a situated agent needs to engage in some behavioural task. Similar to a biological organism, an agent or BBD needs a minimal set of innate behaviours or reflexes in order to explore and initially survive in its environmental niche. From this minimal set, the BBD can learn and adapt such that it optimizes its behaviour. How these devices adapt is the subject of the next principle, which describes value systems (see section 2.5). This approach is very different from the classic artificial intelligence or robotic control algorithms, where either rules or feedback controllers with pre-defined error signals need to be specified *a priori*. In the BBD approach, the agent selects what it needs to optimize its behaviour and thus adapts to its environment.

A second and important point with regard to behavioural tasks is that it gives the researcher a metric by which to score the BBD's performance. Moreover, these tasks should be made similar to experimental biology paradigms so that the behavioural performance of the BBD can be compared with that of real organisms (see section 2.6 below).

## 2.5 Adapt Behaviour when an Important Environmental Event Occurs

Biological organisms adapt their behaviour through value systems, which provide non-specific, modulatory signals to the rest of the brain that bias the outcome of local changes in synaptic efficacy in the direction needed to satisfy global needs. Stated in

the simplest possible terms, behaviour that evokes positive responses in value systems biases synaptic change to make production of the same behaviour more likely when the situation in the environment (and thus the local synaptic inputs) is similar; behaviour that evokes negative value biases synaptic change in the opposite direction. Examples of value systems in the brain include the dopaminergic, cholinergic, and noradrenergic systems (Aston-Jones and Bloom, 1981; Hasselmo et al., 2002; Schultz et al., 1997) which respond to environmental cues signalling reward prediction, uncertainty, and novelty. Theoretical models based on these systems and their effect on brain function have been developed (Doya, 2002; Friston et al., 1994; Montague et al., 1996; Yu and Dayan, 2005) and embedded in real world behaving devices (Arleo et al., 2004; Krichmar and Edelman, 2002; Sporns and Alexander, 2002).

## 2.6 Comparisons with Experimental Data Acquired from Animal Models

The behaviour of BBDs and the activity of their simulated nervous systems must be recorded to allow comparisons with experimental data acquired from animals. The comparison should be made at the behavioural level, the systems level, and the neuronal element level. These comparisons serve two purposes: First, BBDs are powerful tools to test theories of brain function. The construction of a complete behaving model forces the designer to specify theoretical and implementation details that are easy to overlook in a purely verbal description and it forces those details to be consistent among them. The level of analysis permitted by having a recording of the activity of every neuron and synapse in the simulated nervous system during its behaviour is just not possible with animal experiments. The results of such situated models have been compared with rodent hippocampal activity during navigation, basal ganglia activity during action selection, and attentional systems in primates (Burgess et al., 1997; Guazzelli et al., 2001; Itti, 2004; Prescott et al., 2006). Second, by using the animal nervous system as a metric, designers can continually make their simulated nervous systems closer to that of the model animal. This, in turn, allows the eventual creation of practical devices that may approach the sophistication of living organisms.

## 3 Illustrative Examples of Brain-Based Devices

In this section, we will use our group's two most

recent BBDs as illustrative examples of the above principles. The first example, Darwin X (Krichmar et al., 2005a; Krichmar et al., 2005b), is a BBD that develops spatial and episodic memory by incorporating a detailed model of the hippocampus and its surrounding regions. The second example is a BBD capable of predictive motor control based on a model of cerebellar learning (McKinstry et al., 2006).

### 3.1 An Embodied Model of Spatial and Episodic Memory

Darwin X was used to investigate the functional anatomy specific to the hippocampal region during a memory task. Darwin X incorporates aspects of the anatomy and physiology of the hippocampus and its surrounding regions, which are known to be necessary for the acquisition and recall of spatial and episodic memories. The simulated nervous system contained 50 neural areas, 90,000 neuronal units, and 1.4 million synaptic connections. It included a visual system, a head direction system, a hippocampal formation, a basal forebrain, a value or reward system, and an action selection system. Darwin X used camera input to recognize the category and position of distal visual objects and used odometry to develop head direction sensitivity.

Darwin X successfully demonstrated the acquisition and recall of spatial and episodic memories in a maze task similar to the Morris water maze (Morris, 1984) by associating places with actions. The association was facilitated by a dopaminergic value system based on the known connectivity between CA1 and nucleus accumbens and frontal areas (Thierry et al., 2000). The responses of simulated neuronal units in the hippocampal areas during its exploratory behaviour were comparable to neuronal responses in the rodent hippocampus; i.e., neuronal units responded to a particular location within Darwin X's environment (O'Keefe and Dostrovsky, 1971).

Darwin X took into consideration the macro- and micro-anatomy between the hippocampus and cortex, as well as the within the hippocampus. In order to identify different functional hippocampal pathways and their influence on behaviour, we developed two novel methods for analyzing large scale neuronal networks: 1) Backtrace - tracing functional pathways by choosing a unit at a specific time and recursively examining all neuronal units that led to the observed activity in this reference unit (Krichmar et al., 2005a), and 2) Causality - a time series analysis that distinguishes causal interactions within and between neural regions (Seth, 2005). These analyses allowed us to examine the information flow through the network and highlighted the importance of the perforant pathway from the en-

torhinal cortex to the hippocampal subfields in producing associations between the position of the agent in space and the appropriate action it needs to reach a goal. This functional pathway has recently been identified in the rodent (Brun et al., 2002).

As with other BBDs in the Darwin series, Darwin X follows the brain-based modelling principles. It is a physical device in a real world that carries out a task similar to that conducted with animal models. It adapts its behaviour based on its value system, and the dynamics of its nervous system were analyzed during its behaviour and compared with the responses of real nervous systems.

### 3.2 A Model of Predictive Motor Control Based On Cerebellar Learning and Visual Motion

Recently, our group constructed a BBD which included a detailed model of the cerebellum and cortical areas that respond to visual motion (McKinstry et al., 2006). One theory of cerebellar function proposes that the cerebellum learns to replace reflexes with a predictive controller (Wolpert et al., 1998). Synaptic eligibility traces in the cerebellum have recently been proposed as a specific mechanism for such motor learning (Medina et al., 2005). We tested whether a learning mechanism, called the delayed eligibility trace learning rule, could account for the predictive nature of the cerebellum in a real-world, robotic visuomotor task.

The BBD's visuomotor task was to navigate a path designated by orange traffic cones. The platform for this task was a Segway Robotic Mobility Platform modified to have a camera, a laser range finder, and infrared proximity detectors as inputs. The BBD's nervous system contained components simulating the cerebellar cortex, the deep cerebellar nuclei, the inferior olive, and a cortical area MT. The simulated cortical area MT, which responds to visual motion, was constructed based on the suggestion that the visual system makes use of visual blur for determining motion direction (Geisler, 1999; Kregelberg et al., 2003). The simulated nervous system contained 28 neural areas, 27,688 neuronal units, and 1.6 million synaptic connections. Using an embedded Beowulf computer cluster of six compact personal computers, it took roughly 40 ms to update all the neuronal units and plastic connections in the model each simulation cycle. Initially, path traversal relied on a reflexive movement away from obstacles that was triggered by infrared proximity sensors when the BBD was within 12 inches of a cone. This resulted in clumsy, crooked movement down the path. The infrared sensor input was also the motor error signal to the cerebellum via simulated climbing fibre input. Over time, the cerebellar circuit predicted the correct motor response based

on visual motion cues preventing the activation of the reflex and resulting in smooth movement down the centre of the path. The system learned to slow down prior to a curve and to turn in the correct direction based on the flow of visual information. The system adapted to and generalized over different courses with both gentle and sharp angle bends.

The experiments, which depend both on the dynamics of the delayed trace eligibility learning and on the architecture of the cerebellum, demonstrated how the cerebellum can predict impending errors and adapt its movements. Moreover, by analyzing the responses of the cerebellum and the inputs from the simulated area MT during its behaviour, we were able to predict the types of signals the nervous system might select to adapt to such a motor task. The BBD's nervous system categorized the motion cues that were predictive of different collisions and associated those categories with the appropriate movements. The neurobiologically inspired model described here prompts several hypotheses about the relationship between perception and motor control and may be useful in the development of general-purpose motor learning systems for machines.

## 4 Conclusions

Higher brain functions depend on the cooperative activity of an entire nervous system, reflecting its morphology, its dynamics, and its interaction with its phenotype and the environment. BBDs are designed to incorporate these attributes such that they can test theories of brain function. Like the brain, they operate according to selectional principles through which they form categorical memory, associate categories with innate value, and adapt to the environment. These BBDs also provide the groundwork for the development of intelligent machines that follow neurobiological rather than computational principles in their construction.

## Acknowledgements

This work was supported by grants from the Office of Naval Research, DARPA, and the Neurosciences Research Foundation.

## References

- Arleo, A., Smeraldi, F., and Gerstner, W. (2004). Cognitive navigation based on nonuniform Gabor space sampling, unsupervised growing networks, and reinforcement learning. *IEEE Trans Neural Netw* 15, 639-652.

- Aston-Jones, G., and Bloom, F. E. (1981). Nonrepinephrine-containing locus coeruleus neurons in behaving rats exhibit pronounced responses to non-noxious environmental stimuli. *J Neurosci* *1*, 887-900.
- Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J Neurosci* *2*, 32-48.
- Borg-Graham, L. (1987). Modelling the electrical behavior of cortical neurons - simulations of hippocampal pyramidal cells., In *Computer Simulation in Brain Science*, R. M. J. Cotterill, ed. (Cambridge: Cambridge University Press).
- Bower, J. M., and Beeman, D. (1994). *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural Simulation System.: TELOS/Springer-Verlag*.
- Brun, V. H., Otnass, M. K., Molden, S., Steffenach, H. A., Witter, M. P., Moser, M. B., and Moser, E. I. (2002). Place cells and place recognition maintained by direct entorhinal-hippocampal circuitry. *Science* *296*, 2243-2246.
- Burgess, N., Donnett, J. G., Jeffery, K. J., and O'Keefe, J. (1997). Robotic and neuronal simulation of the hippocampus and rat navigation. *Philos Trans R Soc Lond B Biol Sci* *352*, 1535-1543.
- Chiel, H. J., and Beer, R. D. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends Neurosci* *20*, 553-557.
- Clark, A. (1997). *Being there. Putting brain, body, and world together again.* (Cambridge, MA: MIT Press).
- Doya, K. (2002). Metalearning and neuromodulation. *Neural Netw* *15*, 495-506.
- Edelman, G. M., Reeke, G. N., Gall, W. E., Tononi, G., Williams, D., and Sporns, O. (1992). Synthetic neural modeling applied to a real-world artifact. *Proc Natl Acad Sci U S A* *89*, 7267-7271.
- Edelman, G. M., and Reeke, G. N., Jr. (1982). Selective networks capable of representative transformations, limited generalizations, and associative memory. *Proc Natl Acad Sci U S A* *79*, 2091-2095.
- Friston, K. J., Tononi, G., Reeke, G. N., Sporns, O., and Edelman, G. M. (1994). Value-dependent selection in the brain: simulation in a synthetic neural model. *Neuroscience* *59*, 229-243.
- Geisler, W. S. (1999). Motion streaks provide a spatial code for motion direction. *Nature* *400*, 65-69.
- Guazzelli, A., Bota, M., and Arbib, M. A. (2001). Competitive Hebbian learning and the hippocampal place cell system: modeling the interaction of visual and path integration cues. *Hippocampus* *11*, 216-239.
- Hasselmo, M. E., Hay, J., Ilyn, M., and Gorchetchnikov, A. (2002). Neuromodulation, theta rhythm and rat spatial navigation. *Neural Netw* *15*, 689-707.
- Hines, M. L., and Carnevale, N. T. (1997). The NEURON simulation environment. *Neural Comput* *9*, 1179-1209.
- Itti, L. (2004). Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Trans Image Process* *13*, 1304-1318.
- Izhikevich, E. M., Gally, J. A., and Edelman, G. M. (2004). Spike-timing dynamics of neuronal groups. *Cereb Cortex* *14*, 933-944.
- Krekelberg, B., Dannenberg, S., Hoffmann, K. P., Bremmer, F., and Ross, J. (2003). Neural correlates of implied motion. *Nature* *424*, 674-677.
- Krichmar, J. L., and Edelman, G. M. (2002). Machine psychology: autonomous behavior, perceptual categorization and conditioning in a brain-based device. *Cereb Cortex* *12*, 818-830.
- Krichmar, J. L., and Edelman, G. M. (2005). Brain-based devices for the study of nervous systems and the development of intelligent machines. *Artif Life* *11*, 63-77.
- Krichmar, J. L., Nitz, D. A., Gally, J. A., and Edelman, G. M. (2005a). Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task. *Proc Natl Acad Sci U S A* *102*, 2111-2116.
- Krichmar, J. L., and Reeke, G. N. (2005). The Darwin Brain-Based Automata: Synthetic Neural Models and Real-World Devices, In *Modeling in the Neurosciences: From Biological Systems to Neuromimetic Robotics*, G. N. Reeke, R. R. Poznanski, K. A. Lindsay, J. R. Rosenberg, and O. Sporns, eds. (Boca Raton: Taylor & Francis), pp. 613-638.
- Krichmar, J. L., Seth, A. K., Nitz, D. A., Fleischer, J. G., and Edelman, G. M. (2005b). Spatial

- navigation and causal analysis in a brain-based device modeling cortical-hippocampal interactions. *Neuroinformatics* 3, 197-221.
- McKinstry, J. L., Edelman, G. M., and Krichmar, J. L. (2006). A cerebellar model for predictive motor control tested in a brain-based device. *Proc Natl Acad Sci U S A*.
- Medina, J. F., Carey, M. R., and Lisberger, S. G. (2005). The representation of time for motor learning. *Neuron* 45, 157-167.
- Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16, 1936-1947.
- Morris, R. (1984). Developments of a water-maze procedure for studying spatial learning in the rat. *J Neurosci Methods* 11, 47-60.
- O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res* 34, 171-175.
- Pinsky, P. F., and Rinzel, J. (1994). Intrinsic and network rhythmogenesis in a reduced Traub model for CA3 neurons. *J Comput Neurosci* 1, 39-60.
- Prescott, T. J., Montes Gonzalez, F. M., Gurney, K., Humphries, M. D., and Redgrave, P. (2006). A robot model of the basal ganglia: Behavior and intrinsic processing. *Neural Netw* 19, 31-61.
- Schultz, W., Dayan, P., and Montague, P. R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593-1599.
- Seth, A. K. (2005). Causal connectivity of evolved neural networks during behavior. *Network* 16, 35-54.
- Seth, A. K., McKinstry, J. L., Edelman, G. M., and Krichmar, J. L. (2004). Active sensing of visual and tactile stimuli by brain-based devices. *International Journal of Robotics and Automation* 19, 222-238.
- Song, S., Miller, K. D., and Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat Neurosci* 3, 919-926.
- Sporns, O., and Alexander, W. H. (2002). Neuro-modulation and plasticity in an autonomous robot. *Neural Netw* 15, 761-774.
- Thierry, A. M., Gioanni, Y., Degenetais, E., and Glowinski, J. (2000). Hippocampo-prefrontal cortex pathway: anatomical and electrophysiological characteristics. *Hippocampus* 10, 411-419.
- Wolpert, D., Miall, R., and Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences* 2, 338-347.
- Worgotter, F., and Porr, B. (2005). Temporal sequence learning, prediction, and control: a review of different models and their relation to biological mechanisms. *Neural Comput* 17, 245-319.
- Yu, A. J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron* 46, 681-692.

# Developmental Robotics:

an emerging paradigm for intelligent agents

Mark H. Lee

Department of Computer Science  
University of Wales, Aberystwyth, UK

# Aims and agenda

- Consider various approaches
- Review concepts and inspiration
- Illustrate with a case study
- Observations
- Persuade you that development is necessary for embedded learning systems

# Issues

- Why development?
- Why infants?
- Why robots?



# Some current approaches

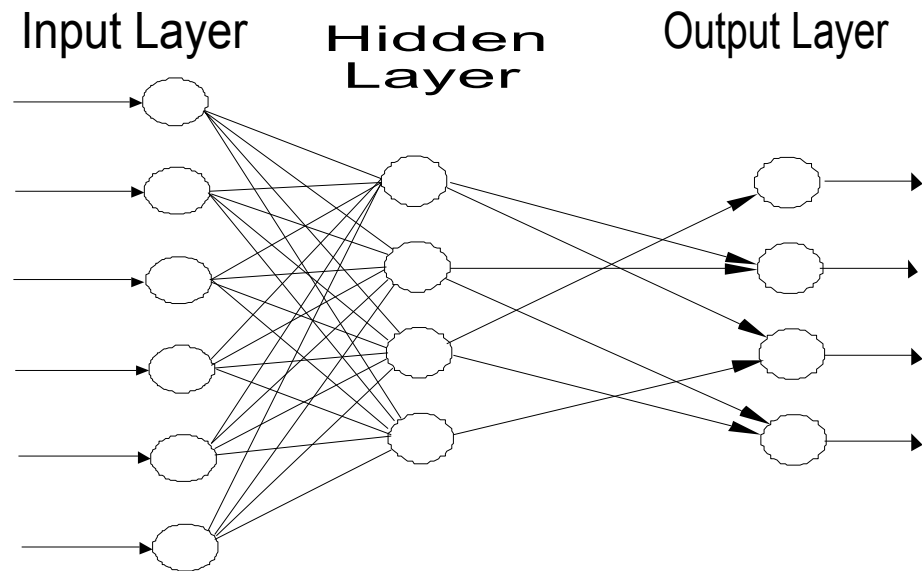
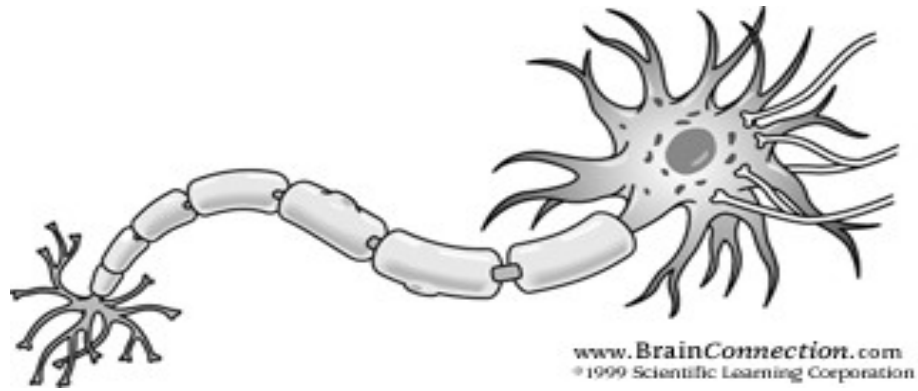
- Cognitive Science
- Connectionism
- Cognitive robotics
- Developmental approach
- Mechanism mapping
- Mechanism mining.

# Cognitive science

- Memory, reasoning, perception, language
- Computing paradigms
- Rule-based, Soar, ACT
- Modular, architectures, structures
- Tends to impose computing models.

# Connectionism (1)

- A challenge to the idea of separate process and memory
- Mainly artificial neural networks (ANNs)
- Highly parallel, adaptive, fast
- But “model-free statistical function estimators”.



# Connectionism (2)

- Large training phases - unrealistic
- Supervised - unrealistic
- Black box (attractive for some)
- Newer neuron models very detailed
- Population simulations impressive
- Dogma of neural models.

# Cognitive robotics

- Logical foundations - e.g. situation calculus
- Knowledge based - but decision theoretic, e.g. “Expectation and Feedback as Hypothetico-Deduction”
- Strictly logical - exclusively so
- Not really robotics - ignores sensory/motor
- Not really cognitive - ignores sensory/motor
- High level, abstract, and symbolic.

# Psychological (developmental) approach

- High level - but behavioural data
- Reflect bio/psycho constraints
- Abstract computational models
- Can map onto neural substrate
- Synthesis process - vis Braitenberg.

# Mechanism mapping

- Top-down
- Scientific analysis
- Artificial -> biological
- Validation against biology
- Analysis, refinement cycle.



# Mechanism mining

- Bottom up
- Inspired invention
- Biology -> mechanism
- Biological constraints - guidance
- Synthesis, simulate cycle.

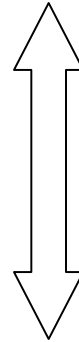
Inspiration

# Early Infant Development



Raw material

development →



Highly sophisticated

Processes to support  
cognitive development

# Alan Turing

“In the process of trying to imitate an adult human mind we are bound to think a good deal about the process which has brought it to the state that it is in.”

A.M. Turing, *Mind*, 59, 433-460, 1950.

# Turing quotes (1)

“Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain.”

“We have thus divided our problem into two parts. The child programme and the education process. These two remain very closely connected.”

“Opinions may vary as to the complexity which is suitable in the child machine.”

# Jean Piaget's four stages of human cognitive development

**Sensorimotor(0-2):** not capable of symbolic representation.

**Preoperational(2-6):** Egocentric, unable to distinguish appearance from reality; incapable of certain types of logical inference.

**Concrete operational(6-12):** capable of the logic of classification and linear ordering.

**Formal operation(12-):** capable of formal, deductive, logic reasoning.

# Infant stages

**1 month** - stare at bright objects.  
Hands normally closed but, if open,  
grasps when palm touched.

**3 months** - visually very alert,  
gaze follows toy. Hand regard,  
clasp/unclasp. Holds toy but not  
eye coordinated.

# Infant stages

**6 months** - visually insatiable, follows adults/toys, stares at small objects and tries grasp with both hands. Palmar grasp. Searches ineptly when toy lost. Takes all to mouth.

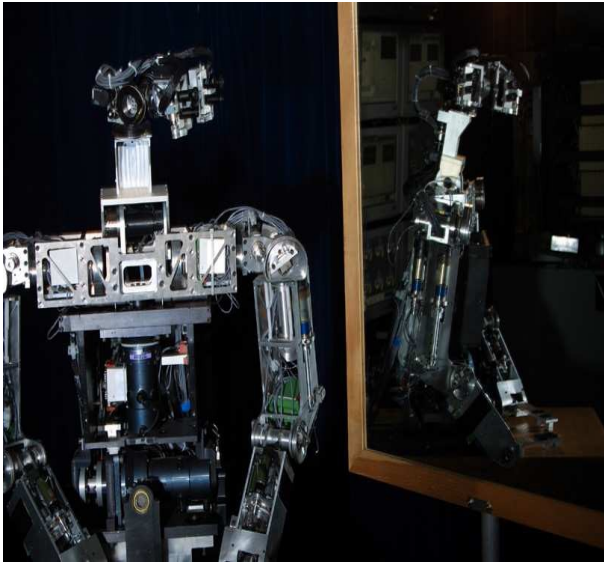


# Developmental Robotics

- Multidisciplinary research area
- Mostly inspired by developmental psychology
- Considerable emphasis on sensory-motor interaction
- Embodied (simulation frowned on)
- aka: epigenetic, life-long learning ...

# Approach

# Examples of systems used in robotic developmental learning



Cog, MIT, USA



Infanoid, CRL, Japan



BabyBot, Genoa, Italy



SAIL, MSU,  
USA



DVL, UWA, Wales

# Embodiment

“... , cognition depends upon the kinds of experiences that come from having a body with particular perceptual and motor capabilities that are inseparably linked and that together form the matrix within which reasoning, memory, emotion, language, and all other aspects of mental life are embedded.”

E. Thelen, *Infancy*, 1(1), 3-28, 2000

# DVL project (EPSRC)

## Developmental Learning Algorithms for Embedded Agents

- Psychology, rather than neuroscience
- Abstract models, as far as possible
- But biologically compatible ...
- Assumptions - explicit and compatible with psychological data
- But not psychological modelling - aim is algorithms for robotics.

# Constraints

- Staged growth of competence
- Constraints are important
- Constraints are helpful !
- Many forms of constraint:
  - Physical - morphology, mechanical, motor
  - Internal - cognitive, sensory, neural, maturational
  - Environmental - external, scaffolding, social.

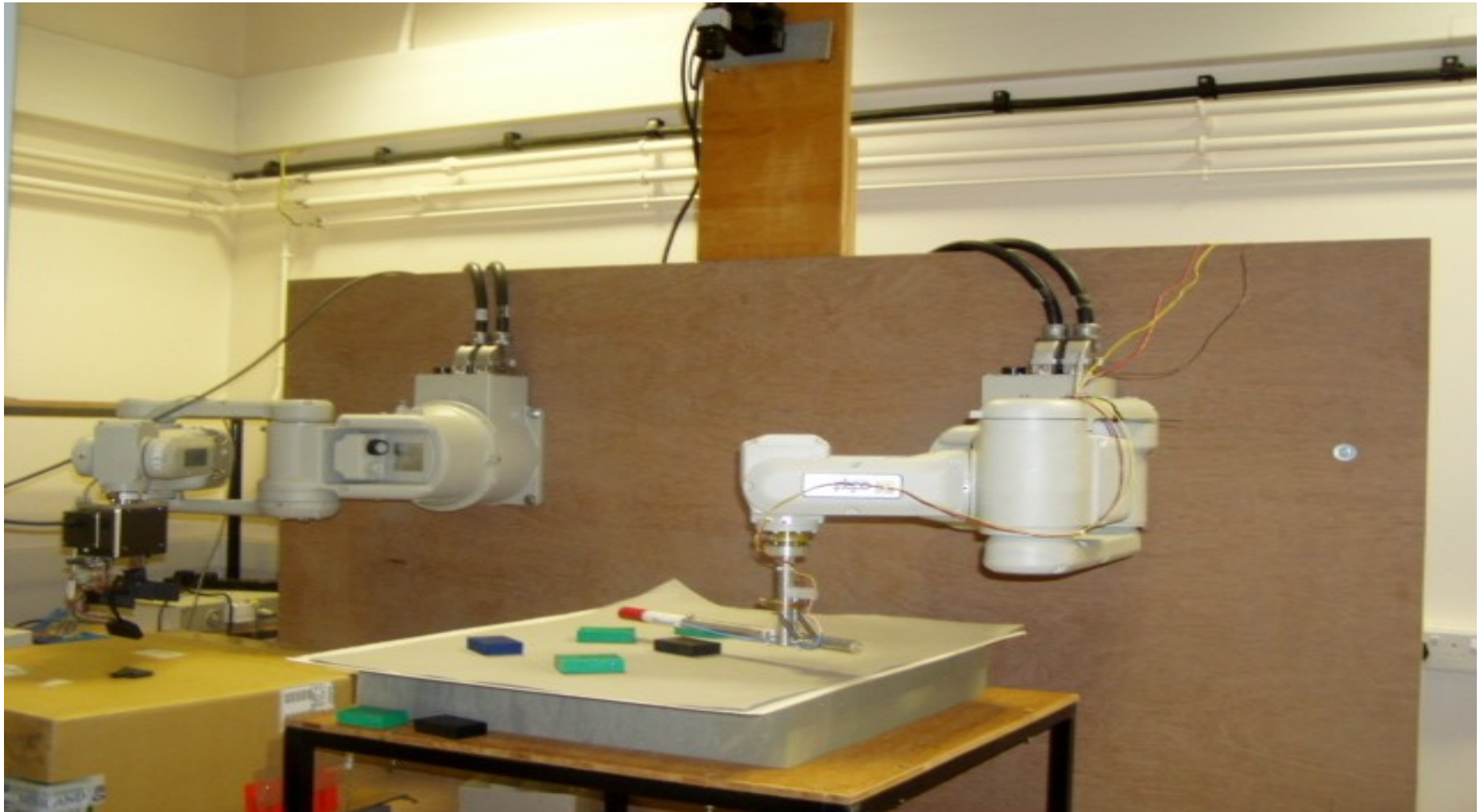
# Issues to investigate

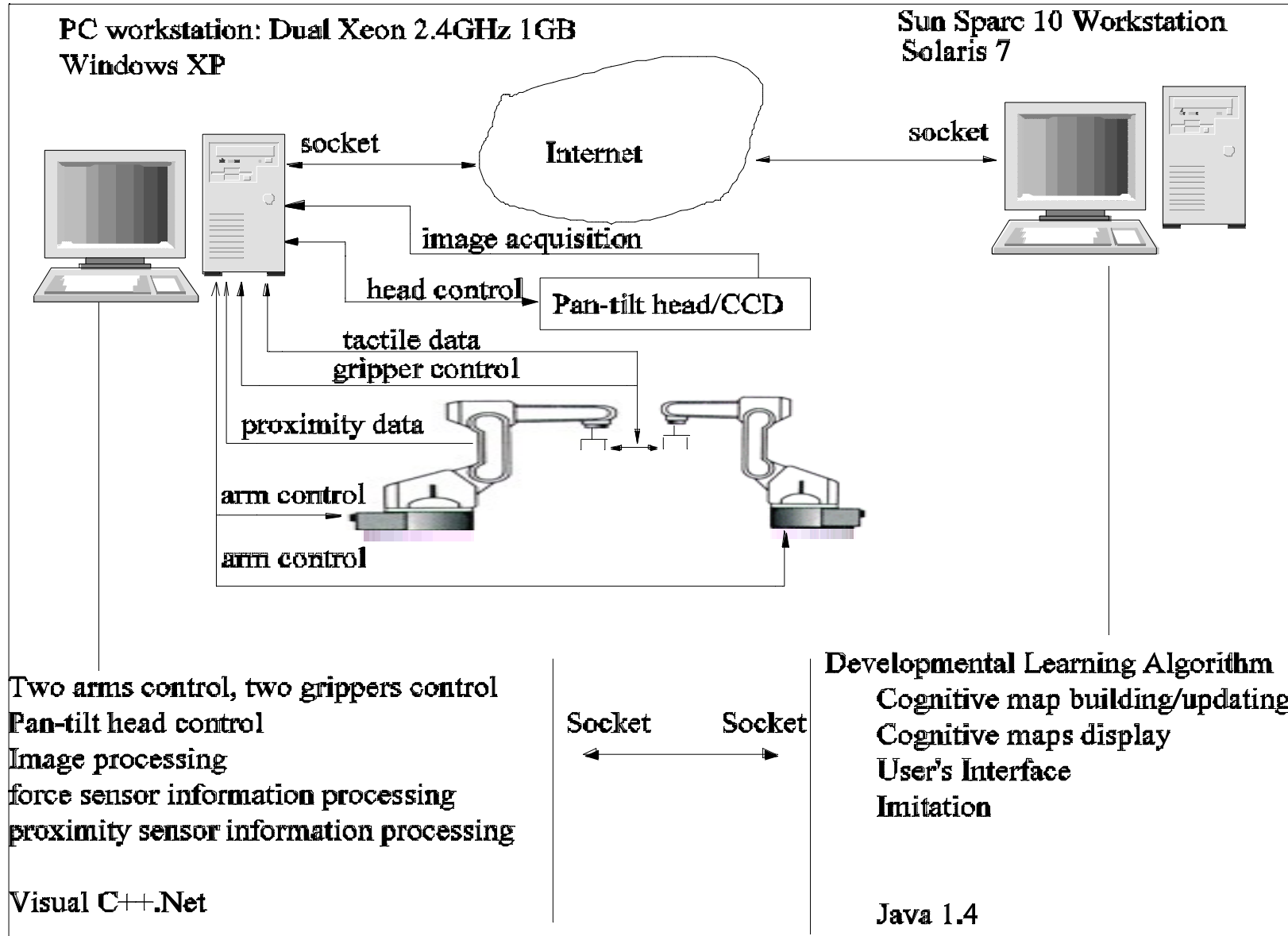
- Proprioception encoding - how can space be learned?
- Motor control - how can new actions develop?
- Coordination - intra modal and cross-modal
- Constraint schedules - how should constraints be exploited?

# Experiments



# Our Experimental Developmental Learning System





# Sensory-motor spaces

- Two Arms, each with:
  - Motor drives at the joints,
  - Proprioceptive sensing of joint angles
  - Tactile sensing of object contact
- One Eye, with:
  - Retinal axes
  - Foveal feature extraction
  - Motor pan and tilt drives.

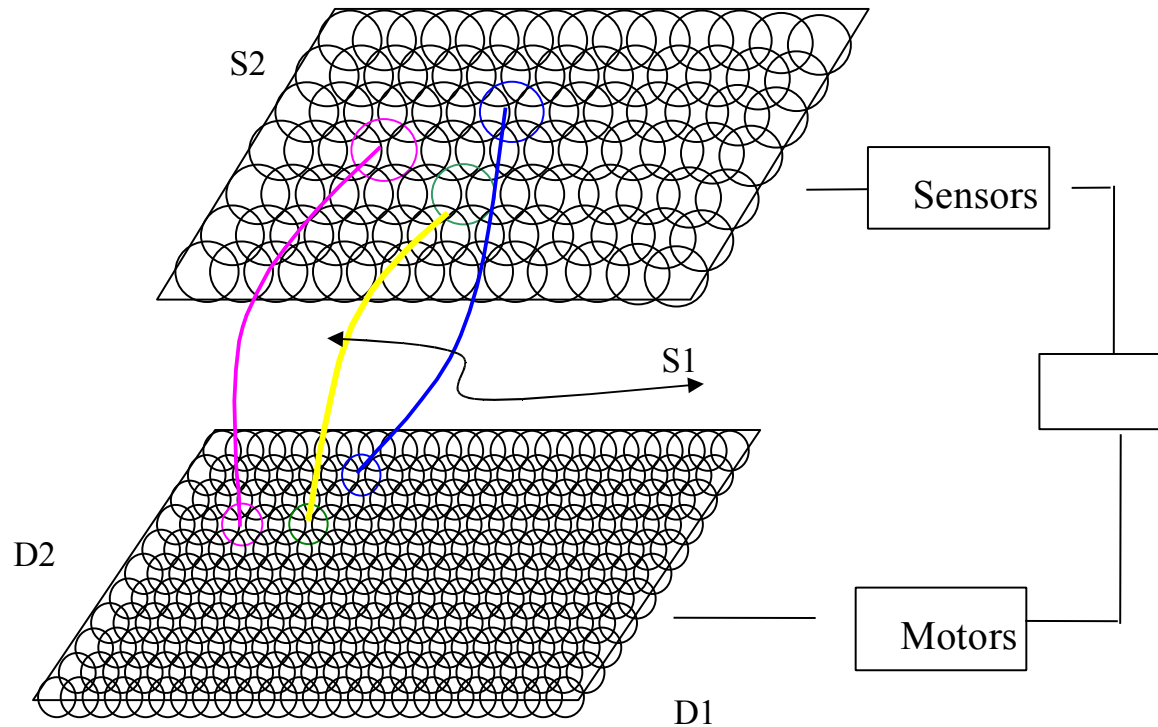
# Experimental variables

- Internal/environmental constraints
- Proprioception encoding schemes
- Proprioception resolution
- Novelty/habituation parameters.

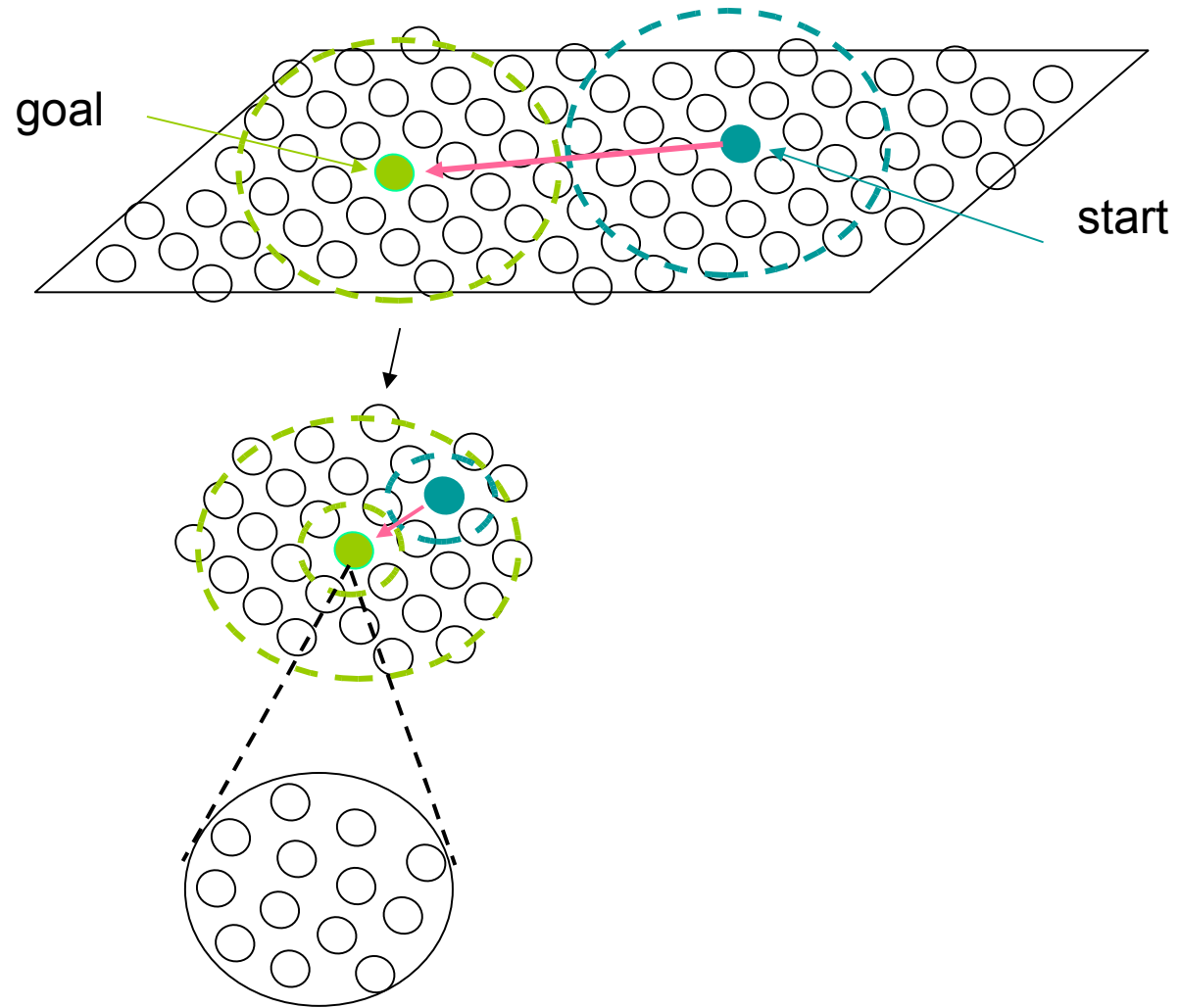
# S-M mappings

- Mappings as a computational substrate for sensory-motor learning
- Based on **fields** - overlapping patches of S-M space. Each has stimulus data, excitation levels, habituation values
- Global signals are summations of field values across a map.

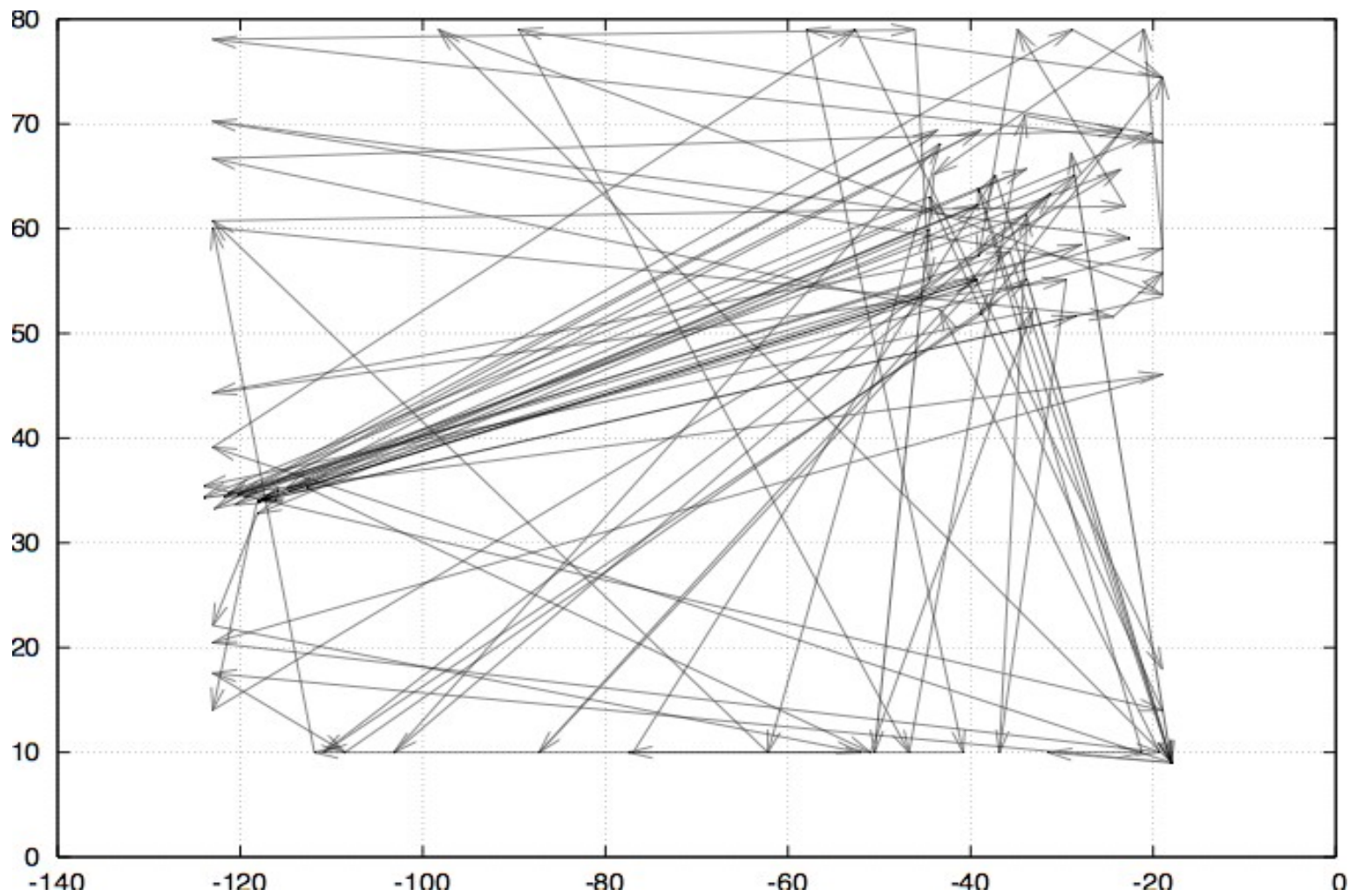
# Sensory-motor mapping system



# Variable field sizes and hierarchical maps



# Early actions

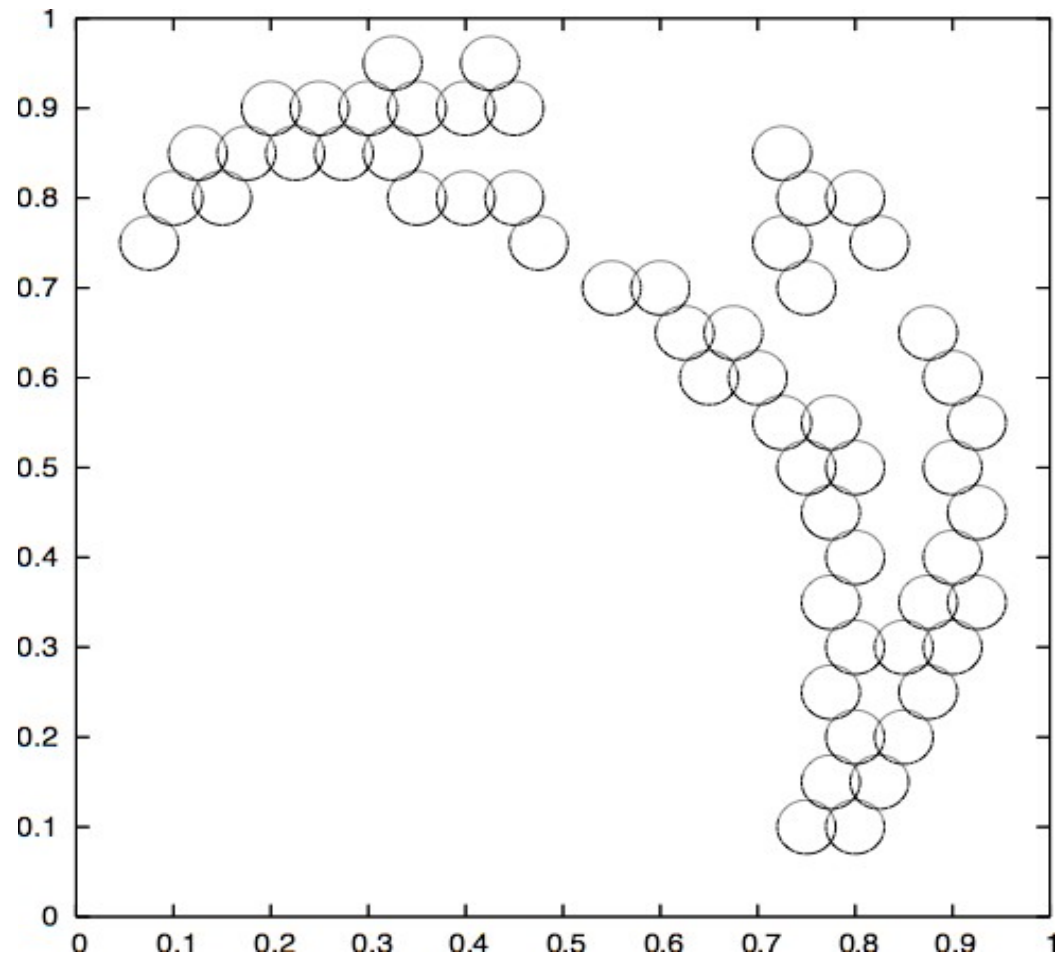




# Proprio-motor correlation

- While this is happening, the mapping system records the correlation between the D and S values
- When the map is fully/partially developed it will associate changes in sensory locations with motor acts.

# Growth of fields



# Behaviour types

1. “blind groping” actions mainly directed at the body area
2. more groping but at the boundary limits
3. unaware pushing of objects out of the local environment
4. limb movements stop upon object contact
5. repeated cycles of contact and movement, i.e. “touching” of detected objects
6. directed touching of objects and sequences of objects.

# Staged development

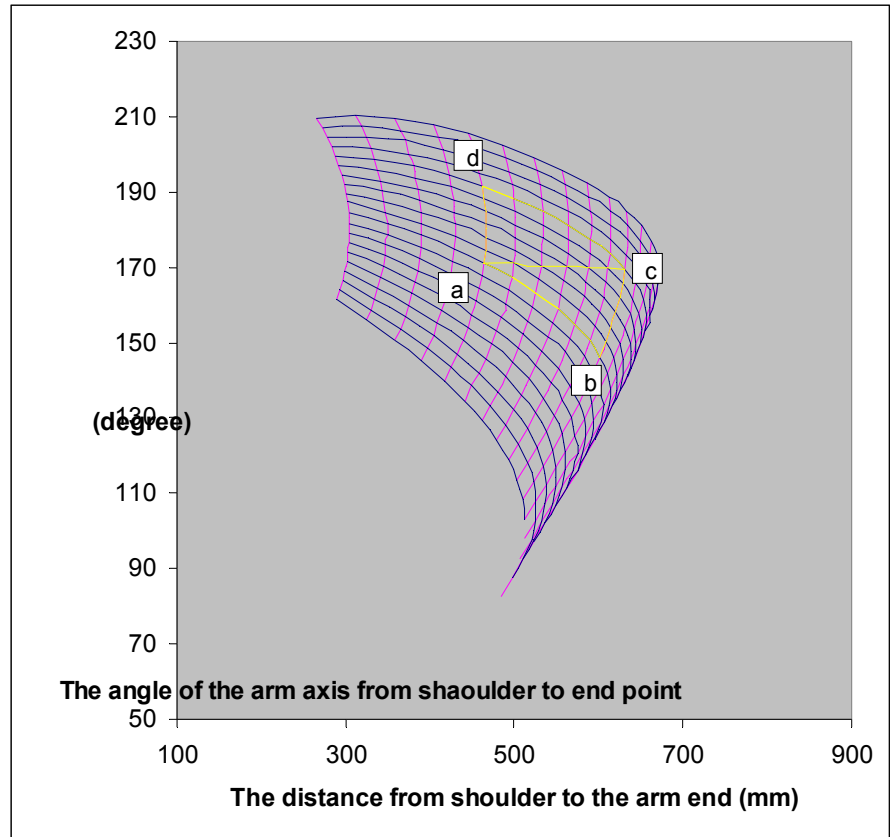
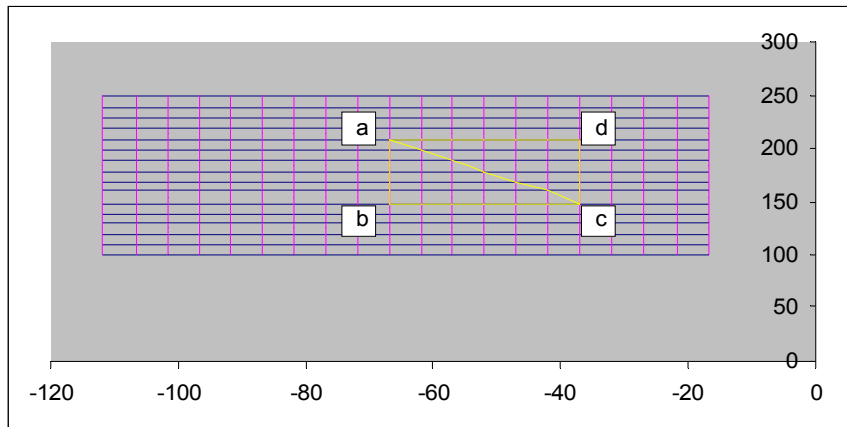
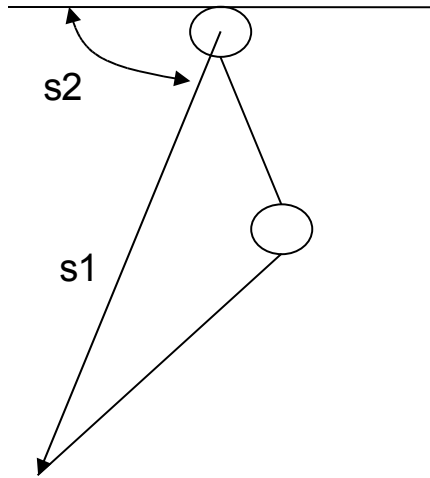
- Hand “grope” creates local space map
- Hands contact objects - sensitive grope
- Eye stimulation creates local visual map
- Eye sees hands - hand fixation
- Hands follow eye fixations
- Grasping of objects.

# Observations

# Proprioception encoding

- 4 schemes tested:
  - joint, shoulder, body, Cartesian
- None critical: but body and Cartesian common for both arms, and eye
- Muscle spindle stretch better than joint receptors.
- Mix of receptors ideal for spatial encoding  
(cf. Bosco *et al*, J. Neurophysiol. 2000)

# Shoulder encoding



# Arm kinematics

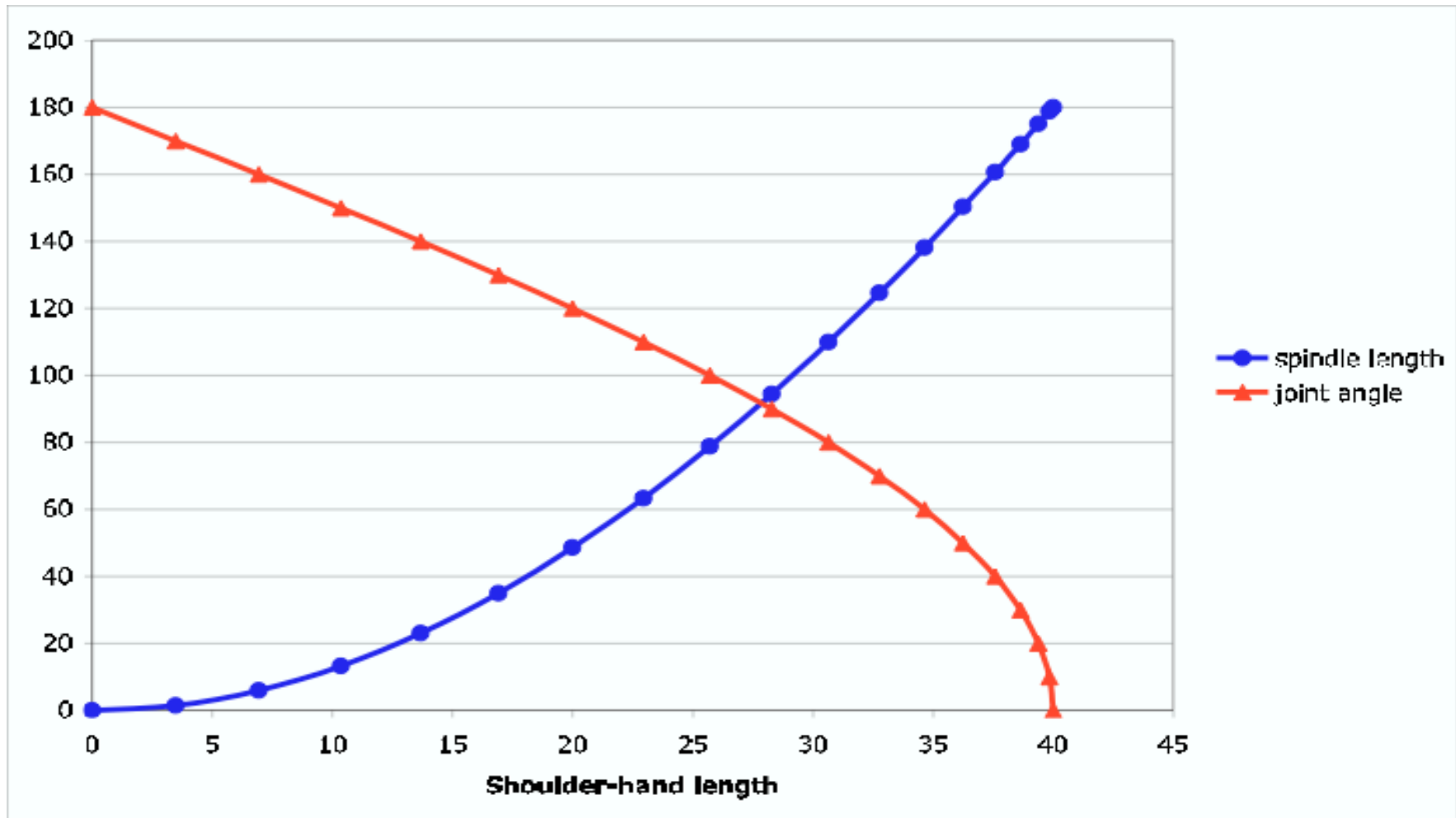
$$\text{Reach} = \sqrt{l_1^2 + l_2^2 + 2l_1l_2\cos\theta_2}$$

$$\text{Angle} = \theta_1 - \arctan(l_2\sin\theta_2 / l_1 + l_2\cos\theta_2)$$

for limb lengths,  $l_1$   $l_2$  and joint angles,  $\theta_1$   $\theta_2$



# Joint or muscle?

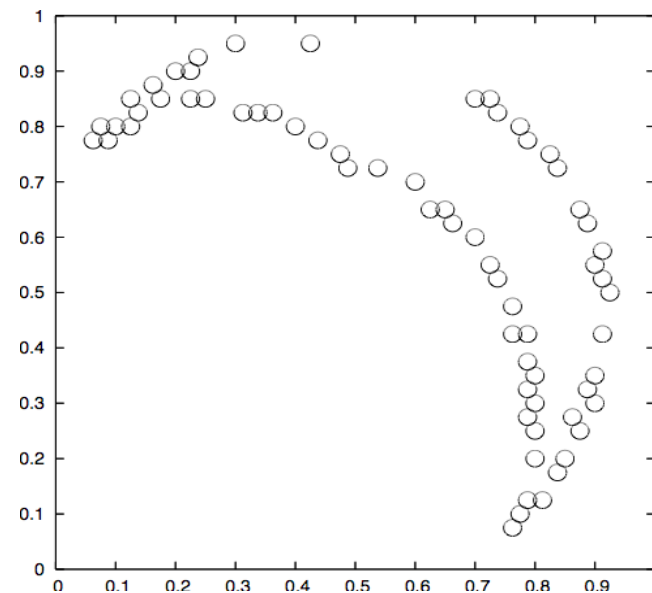
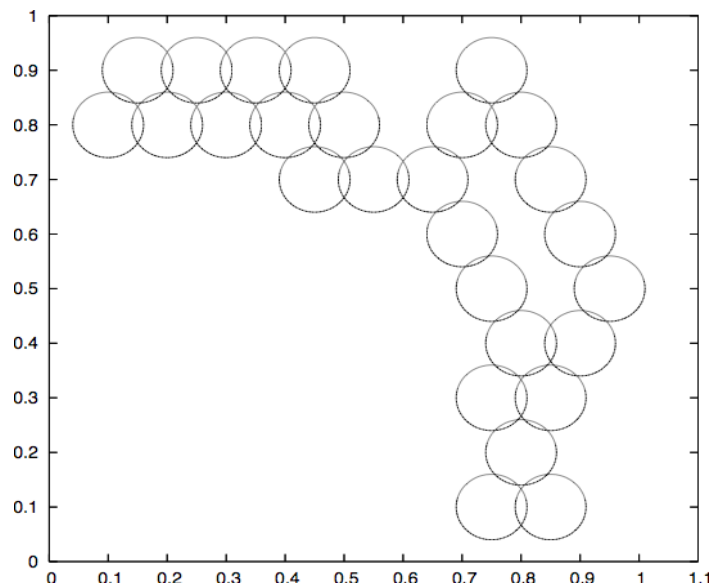
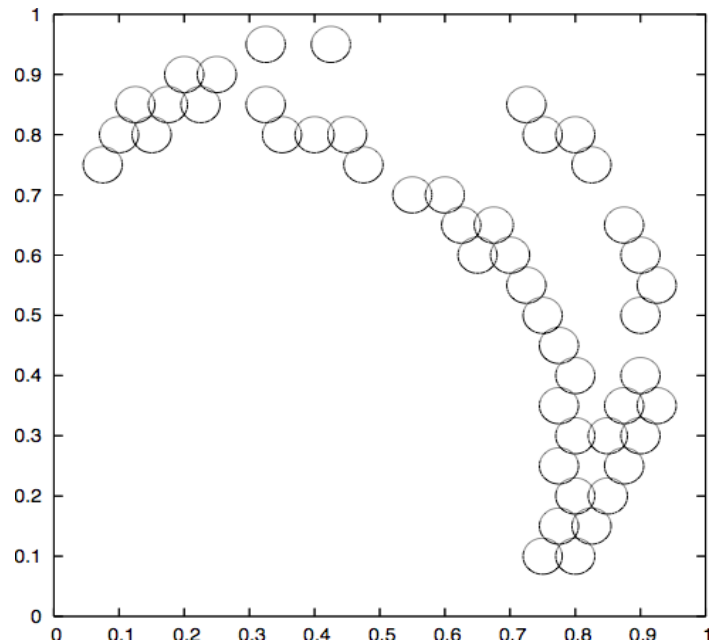


# Morphology is important

- The physical structure of the agent is a key determining factor.
- Sensory structure determines capacity and capability of sensing modalities.
- Mobility, dexterity, effect, all depend upon appropriate anatomy or hardware.

# Sensory resolution

(re Westermann & Mareschal, Infancy, 2004)



# Observations (1)

- Spontaneous action can be useful for gaining information, but can also be an indication of reduced learning activity
- Motor noise (and other noise) can be beneficial in early learning
- Low accuracy/resolution can be beneficial in early learning.

# Observations (2)

- Proprioception may be more important for supporting vision than previously thought. In particular, non-visual reaching can be developed prior to visually guided grasping.  
(re: Clifton et al, "Is visually guided reaching a myth?", Child Development, 1993)

# Observations (3)

- The S-M learning is all relative - based on changes - no absolute values are needed  
A given reset location provides a reference that anchors the maps - (but this could be altered later).

The future...

# Discovered structure

- Self movement - motor control, S-M coordination, spatial limits
- Object contact - static environment, spatial structure
- Loss of contact - dynamic environment.



# S-M mappings

- Maps have many advantages - partial maps effective, gross to fine scale management, generalisation, cross-modal action.
- Also supports rehearsal, planning and imagined action.

# Importance of constraints

- Scaffolding
- Bandwidth reduction
- Degrees of freedom reduction

## Constraint lifting

- Main early learning mechanism
- Triggered by plateaus in activity
- Constraints used: tactile/vision/map scale.

# Importance of Play

- Very prevalent in primates
- Role? - rehearsal, practice, exploration
- Exhaust plateaus before next stage?
- Test out all constraints?
- Growth of imagination.

# Return to Issues

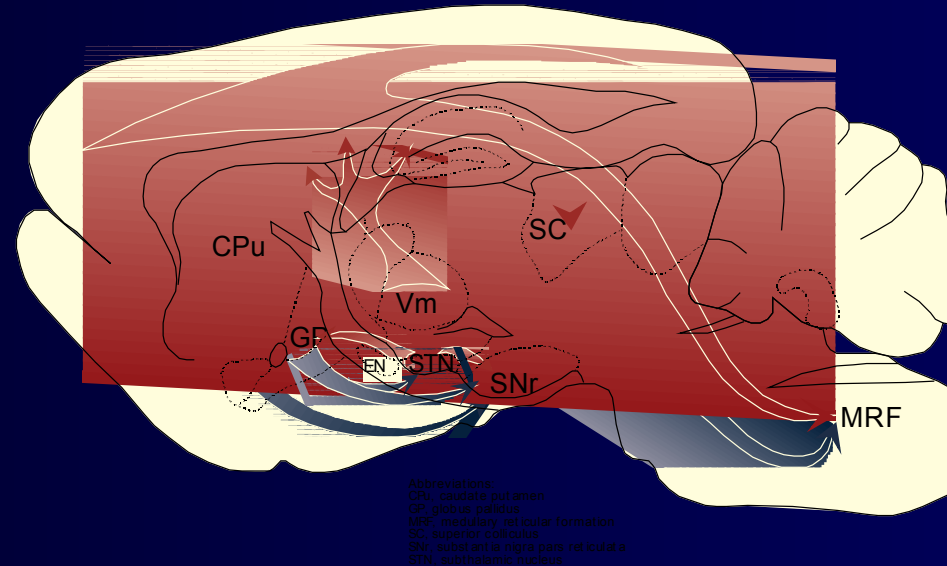
- Why development?
  - Behaviour based
  - Transitions between skill stages
- Why infants?
  - Stage  $n+1$  argument
- Why robots?
  - Easiest way to embodiment

# Summary - the important bits

- Behaviour-based
- Development is essential for learning
- Infancy is a very important developmental period
- Psychology is where the data can be found
- Simplicity of mechanisms
- Synthesis, test cycle
- “law of uphill analysis and downhill invention”
- Most of this is under-rated or under-investigated.

# Is it just a question of priority ?

Inspiration from the vertebrate basal ganglia



wellcome trust

Peter Redgrave

Neuroscience Research Unit,  
Dept Psychology,  
University of Sheffield, UK

EPSRC



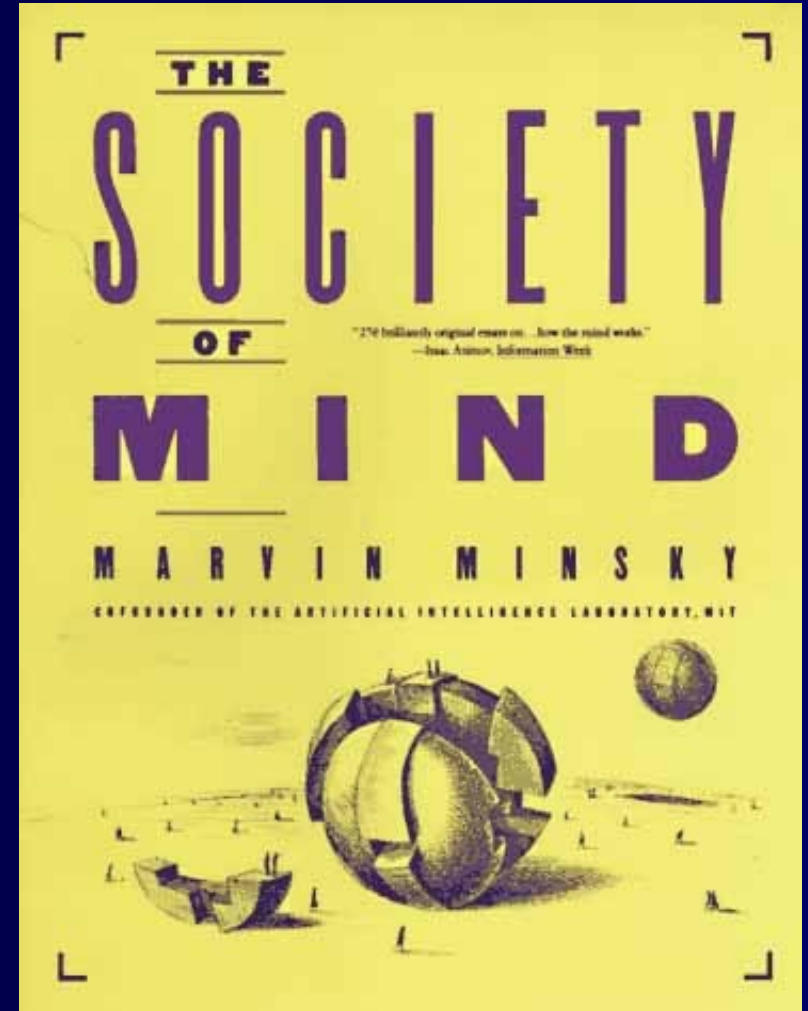
# Overview

- Selection - a fundamental computational problem
- Basal ganglia as a biological solution – looped architecture
- Evolution of competing functional systems – layered architecture
- Subcortical loops through the basal ganglia
- Cortical/subcortical competitions – a basis for irrational behaviour
- Adaptive function(s) of the basal ganglia

# A general architecture for a multifunctional system

...including the brain

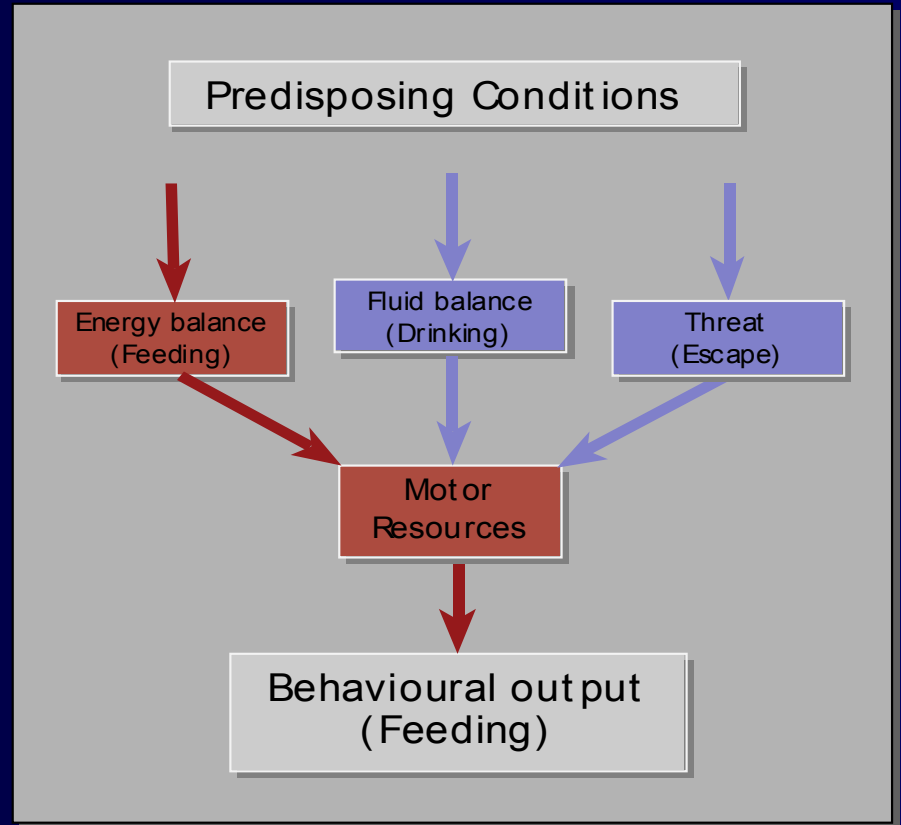
- Largely independent parallel processing functional units
- Each with:
  - specialised sensory input
  - specific functional objectives
  - specialised physiological and behavioural output





# The Selection Problem

- Multiple functional systems
- Spatially distributed
- Processing in parallel
- All act through final common motor path



At any point in time which system should be permitted to guide motor output (behaviour)?

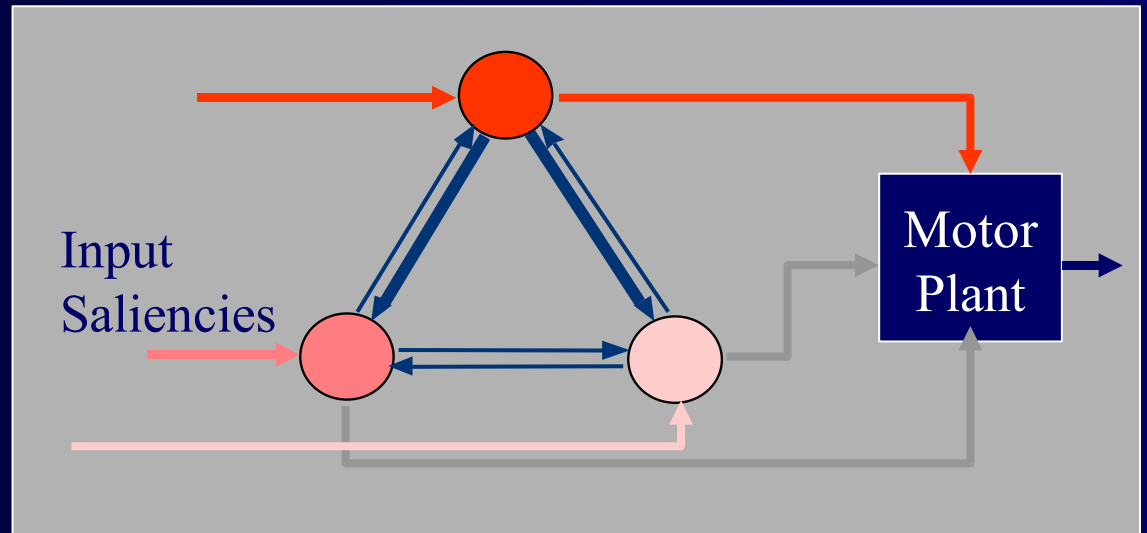
# Parallel processing within sensory representations



# Theoretical Solutions

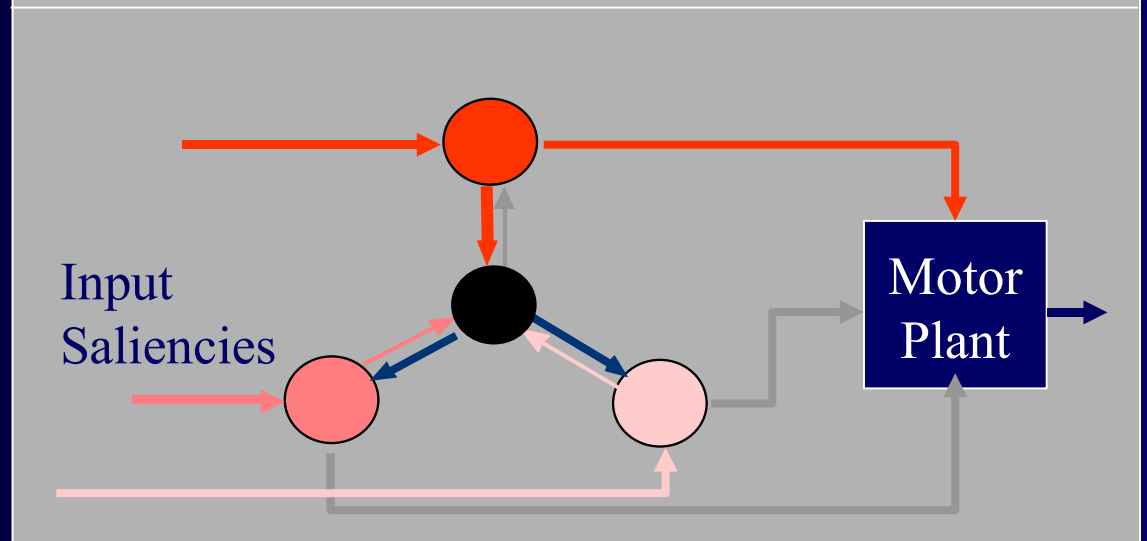
- Recurrent reciprocal inhibition

- Selection an emergent property
- Positive feedback
- Winner-take-all



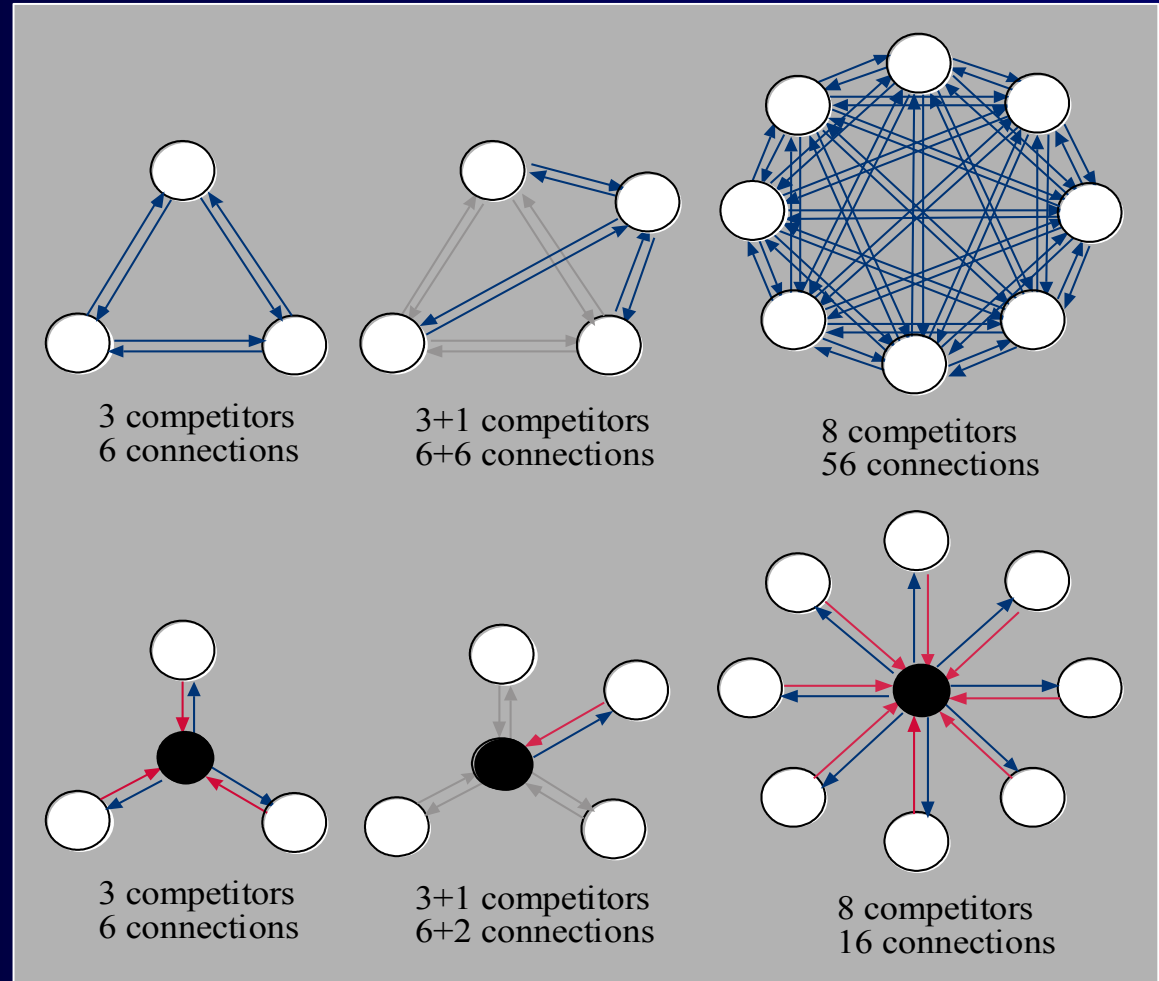
- Centralised selection

- Localised switching
- Dissociates selection from perception and motor control



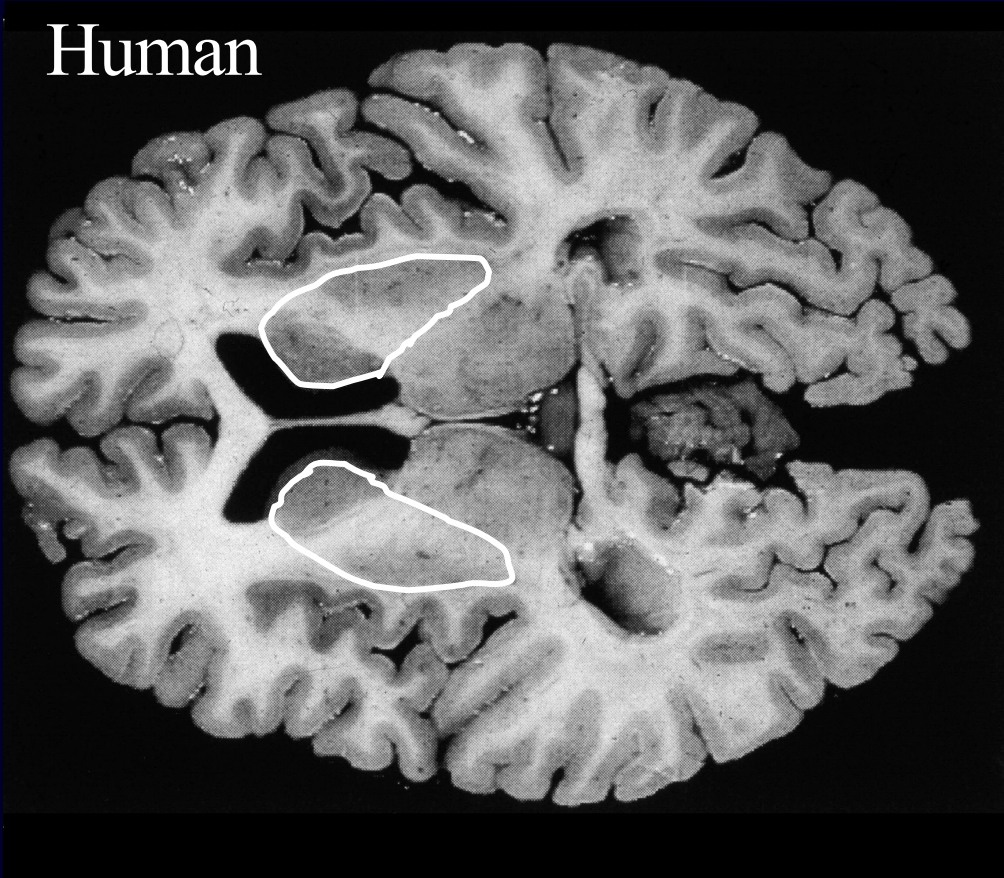
# Problems of Scale

- Recurrent reciprocal inhibition
  - Each additional competitor increases connections by  $n(n-1)$
- Centralised selection
  - Each additional competitor adds 2 further connections

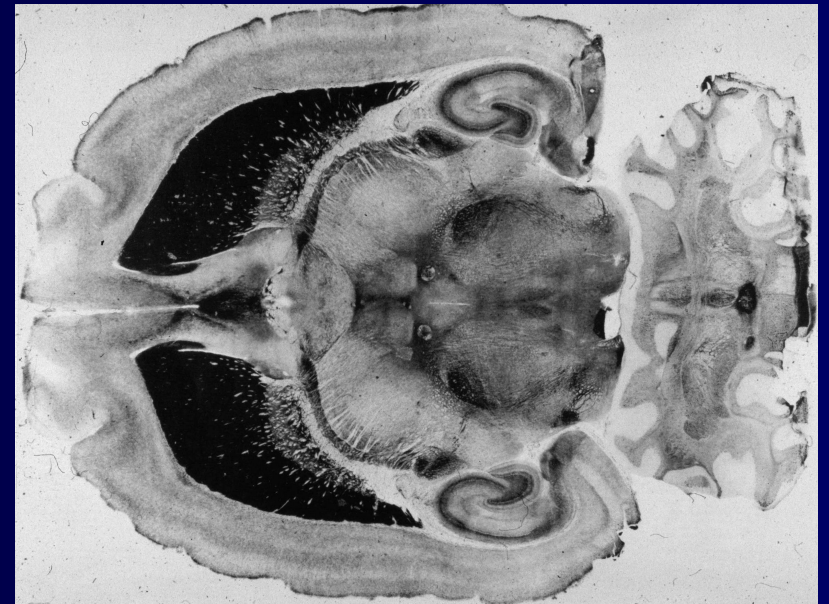


# Basal Ganglia: a biological solution to the selection problem

Human

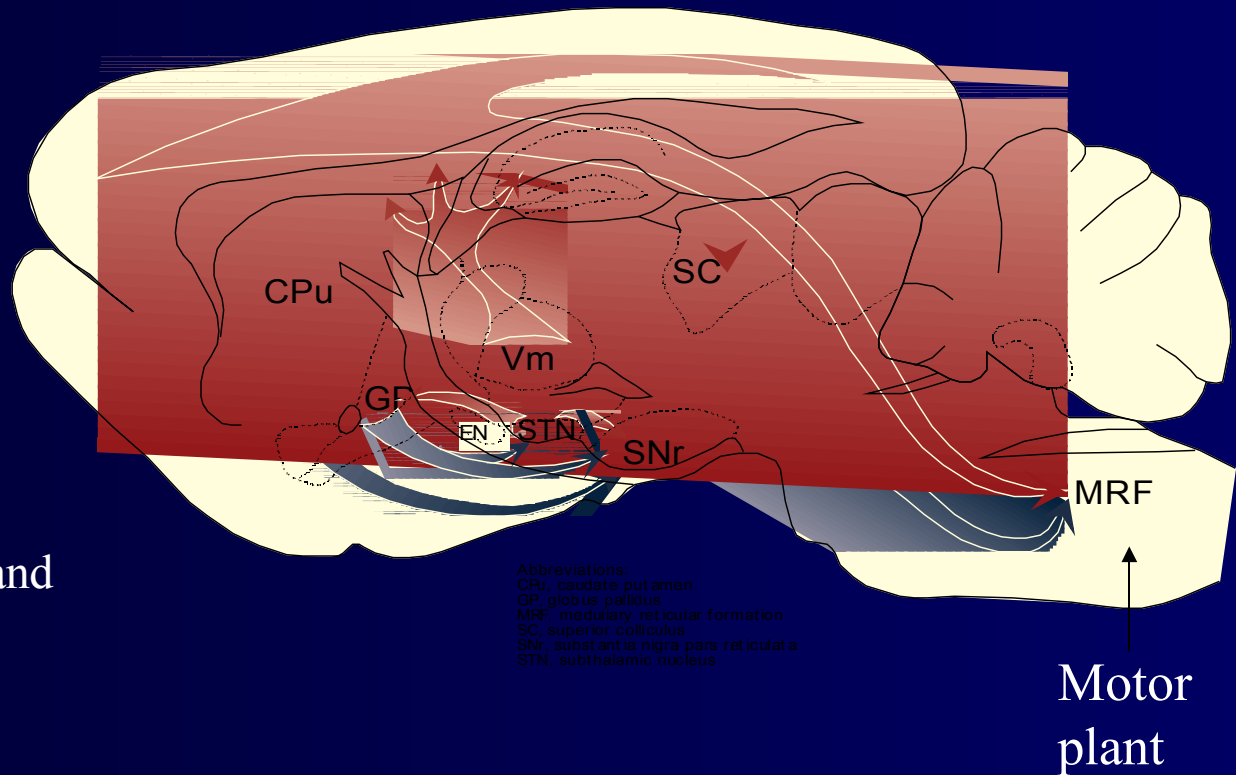


Rat



# External command systems and the basal ganglia

- External command systems
  - Cortical
  - Limbic
  - Midbrain
- Command inputs
  - Sensory
  - Cognitive
  - Affective
- Command outputs
  - Converge on brainstem and spinal motor generators
- Links with basal ganglia
  - Phasic excitatory inputs
  - Tonic inhibitory output

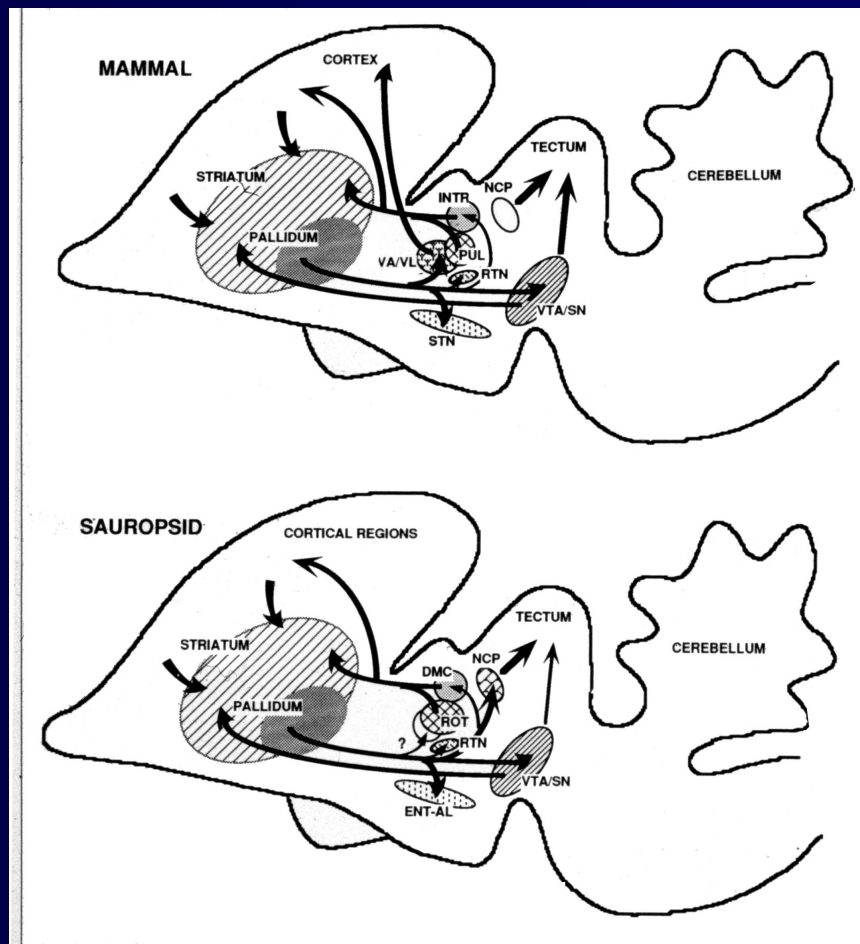


# Evolutionary conservatism

“The basal ganglia in modern mammals, birds and reptiles (i.e. modern amniotes) are very similar in connections and neurotransmitters, suggesting that the evolution of the basal ganglia in amniotes has been very conservative.”

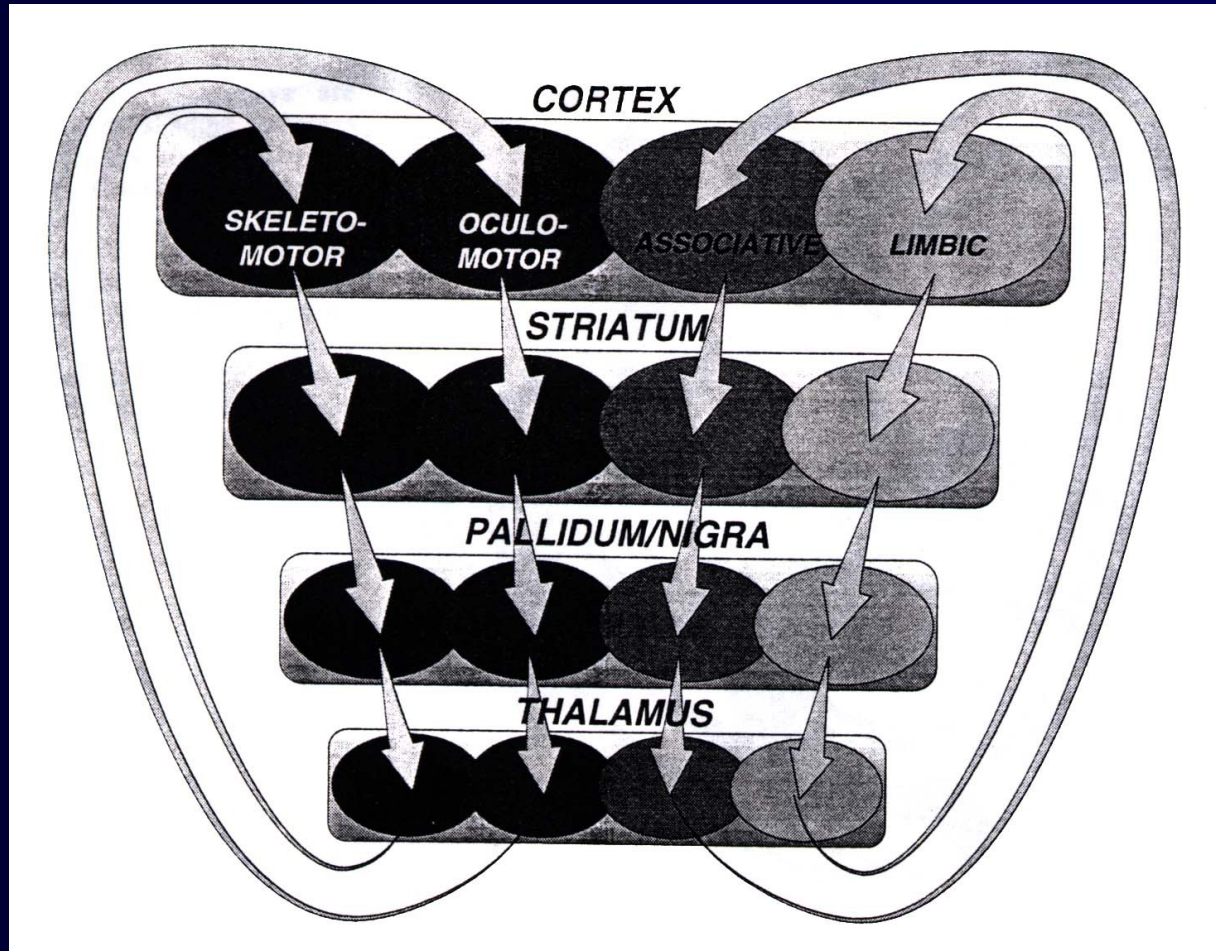
Medina, L and Reiner, A.

*Neurotransmitter organization and connectivity of the basal ganglia in vertebrates: Implications for the evolution of basal ganglia.* Brain Behaviour and Evolution (1995) 46, 235-258



**Fig. 5.** Schematic drawings of sagittal sections through the brains of a mammal and a sauropsid (i.e., birds and reptiles), showing the basic connections involved in the circuitry of the basal ganglia in both amniotic groups. Abbreviations: DMC = Avian and reptilian dorsomedial thalamic complex; ENT-AL = reptilian entopeduncular nucleus, and avian anterior nucleus of the ansa lenticularis; INTR = mammalian midline-intralaminar nuclei; NCP = nucleus of the posterior commissure in reptiles and mammals, and lateral spiriform nucleus in birds; PUL = mammalian laterodorsal-pulvinar complex and medial geniculate nucleus; ROT = avian and reptilian nucleus rotundus and avian nucleus ovoidalis/reptilian nucleus medialis; RTN = reticular thalamic nucleus; STN = subthalamic nucleus; VA/VL = ventral anterior and ventral lateral nuclei; VTA/SN = ventral tegmental area and substantia nigra.

# Basal Ganglia Architecture :Cortically based loops

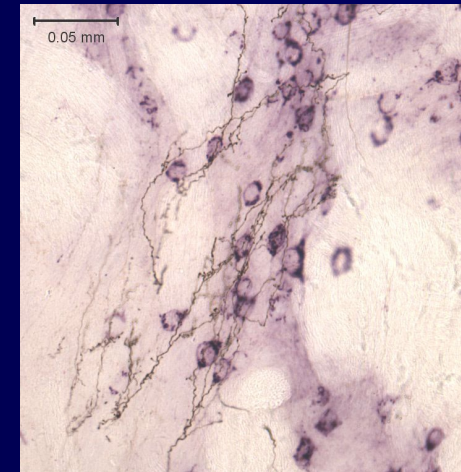
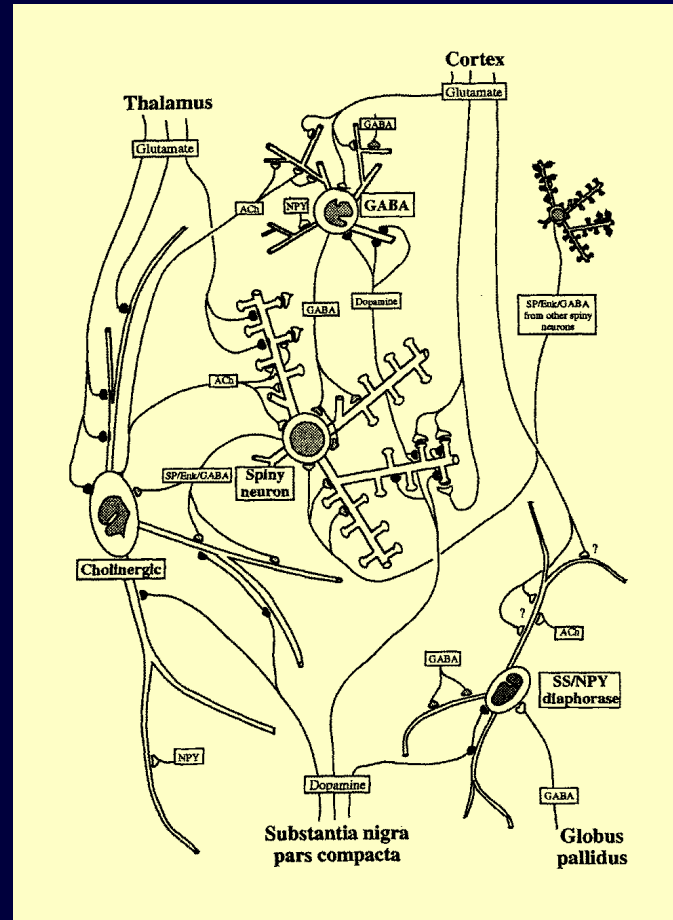
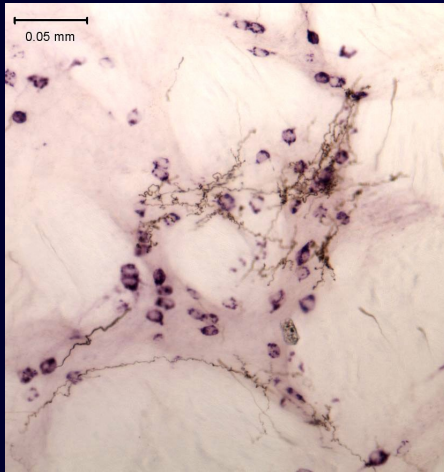


Alexander, G. E., M. R. DeLong, et al. (1986). "Parallel organization of functionally segregated circuits linking basal ganglia and cortex." *Ann. Rev. Neurosci.* 9: 357-381.



# Repeating microcircuitry across territories

- External inputs
  - Cerebral cortex
  - Limbic system
  - Brainstem via thalamus



- Input functions
  - Cognitive
  - Affective
  - Sensorimotor

# Cortical loop: a specific example

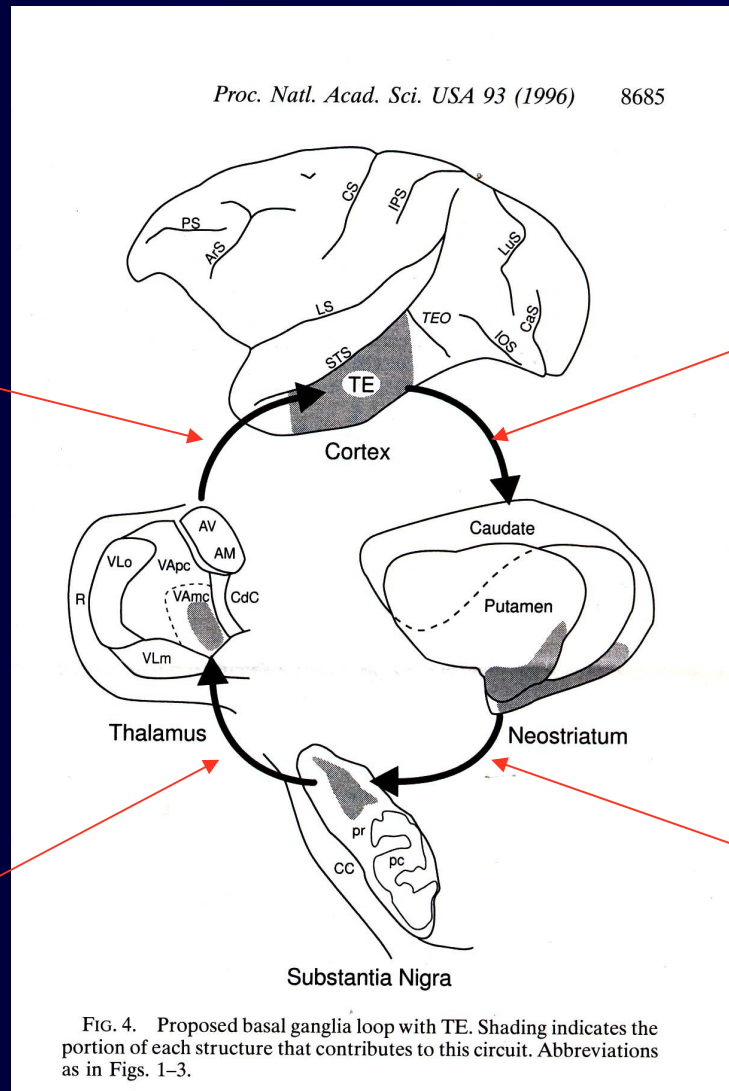
Proc. Natl. Acad. Sci. USA 93 (1996) 8685

Phasic/  
Disinhibitory  
(Positive  
Feedback)

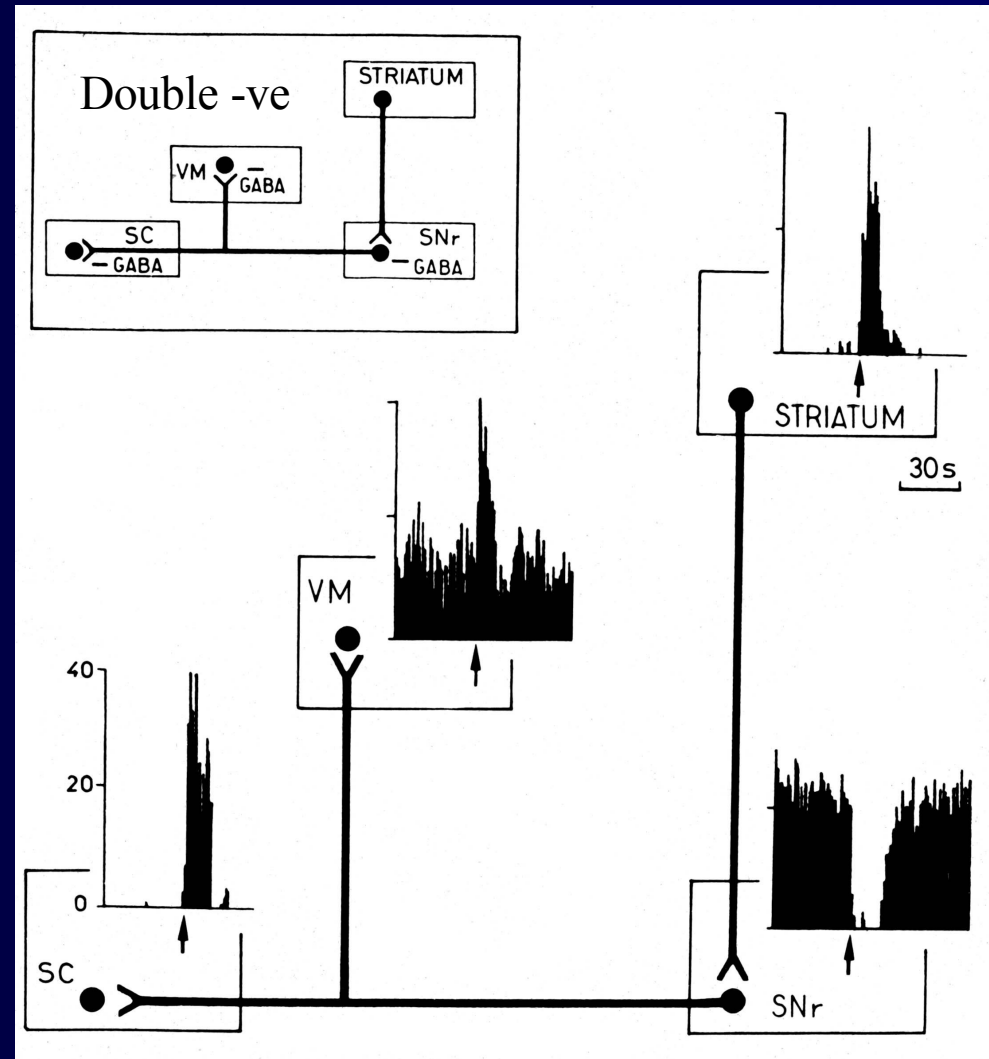
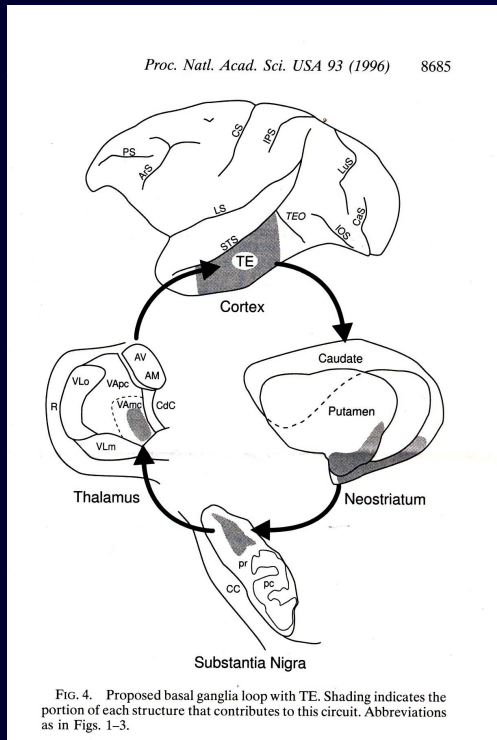
Phasic/  
excitatory

Tonic/  
inhibitory

Phasic/  
inhibitory



# Disinhibitory output

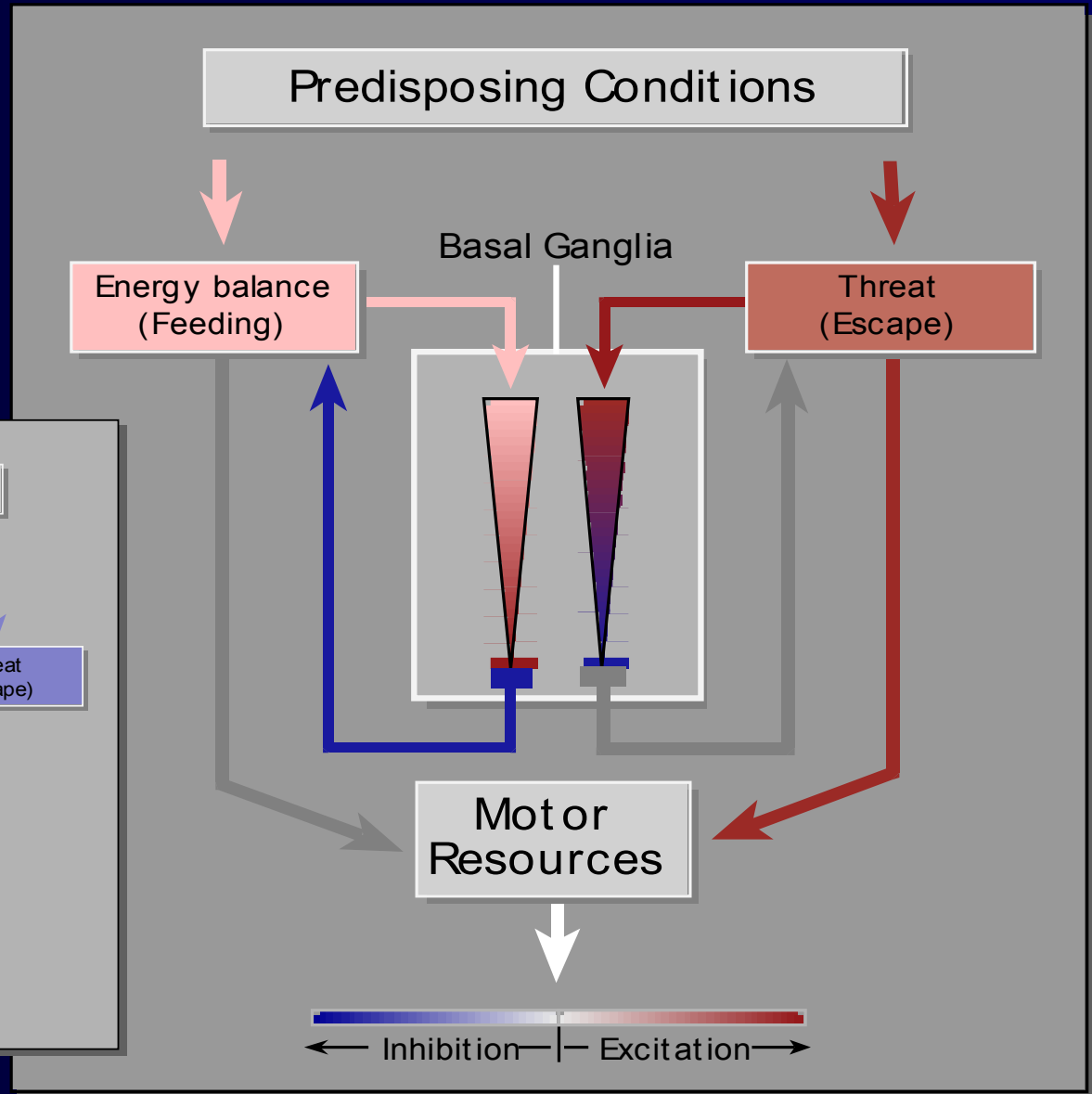
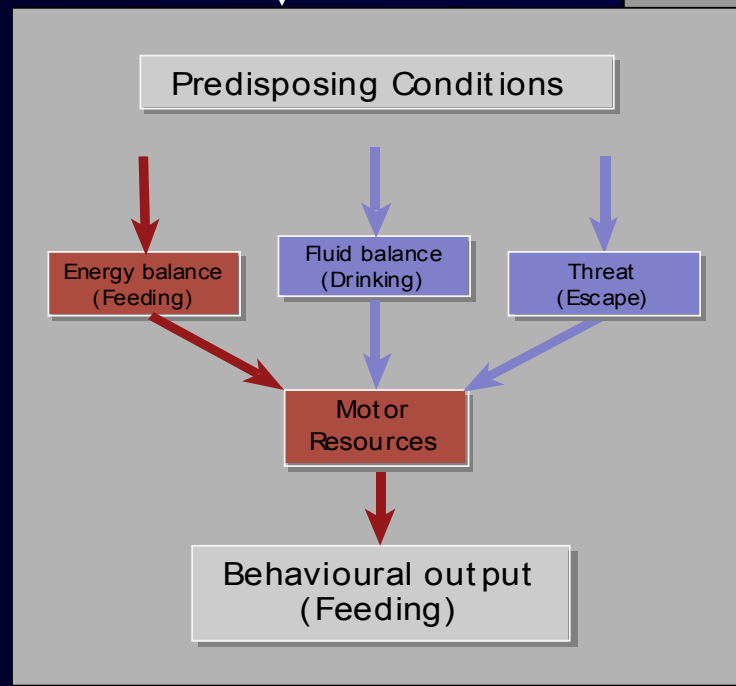


Chevalier, G. and J. M. Deniau (1990). "Disinhibition as a basic process in the expression of striatal functions." Trends Neurosci. **13**: 277-281.

# Selection by inhibition and disinhibition

Potential resolution →

The Selection Problem



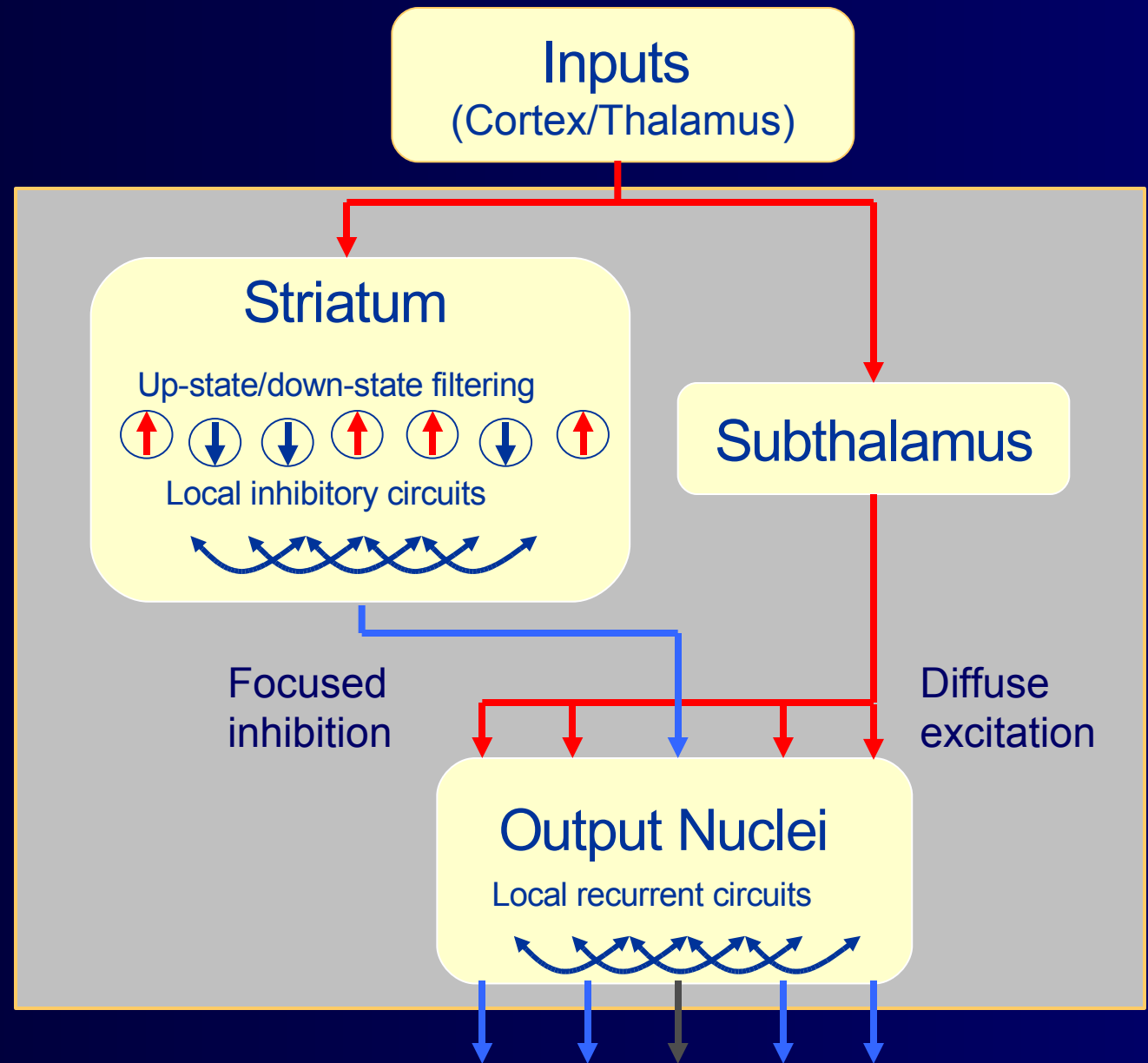
# Serial Selection in the Basal Ganglia

1) Up-down states of medium spiny neurones

2) Local inhibition in striatum

3) Diffuse/focused projection onto output nuclei

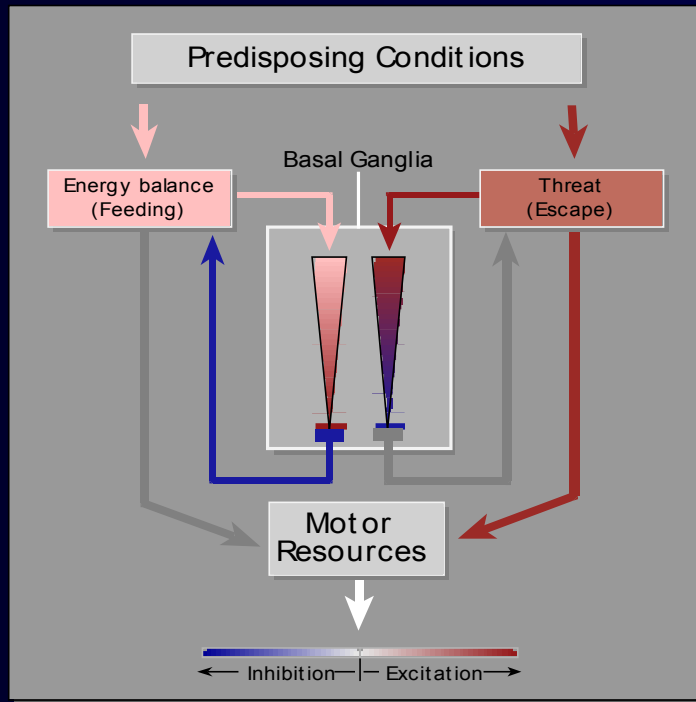
4) Recurrent inhibition in output nuclei



# Basis for selection

- Relative levels of input salience in competing channels
  - Common currency for evaluating priority
- Determined by
  - Evolution...inputs from different command modules varies across species
  - Individual experience...reinforcement learning
- Implemented by
  - Differences in relative levels of afferent activity
  - Different weights of contact in different channels

# Qualitative model:



Gurney, K., T. J. Prescott, et al. (2001). "A computational model of action selection in the basal ganglia. I. A new functional anatomy." *Biol Cybern* 84: 401-410.

# Analysis

## Analytic equilibrium solution (Kevin Gurney)

Model neurons - leaky integrators with piecewise linear output

striatum - control pathway

$$H[c_i - \epsilon/w_s(1 - \lambda)] \equiv H_i^{\uparrow}(\lambda)$$

$$x_i^{e-} = m^- [w_s(1 - \lambda_e)c_i - \epsilon] H_i^{\uparrow}(\lambda_e)$$

striatum - selection pathway

$$x_i^{g-} = m^- [w_s(1 + \lambda_g)c_i - \epsilon] H_i^{\uparrow}(-\lambda_g)$$

STN

$$x_i^+ = m^+(w_t c_i + \epsilon' - w_g y_i^e) H_i^{+\uparrow}$$

$$H_i^{+\uparrow} = H(w_t c_i + \epsilon' - w_g y_i^e)$$

GPe

$$\tilde{a}_i^e = w^- (\delta X^+ - x_i^{e-}) + \epsilon_e$$

$$y_i^e = m^e \tilde{a}_i^e H(\tilde{a}_i^e)$$

GPi/SNr

$$\tilde{a}_i^g = w^- (\delta X^+ - x_i^{g-}) - w_e y_i^e + \epsilon_g$$

$$y_i^g = m^g \tilde{a}_i^g H(\tilde{a}_i^g)$$

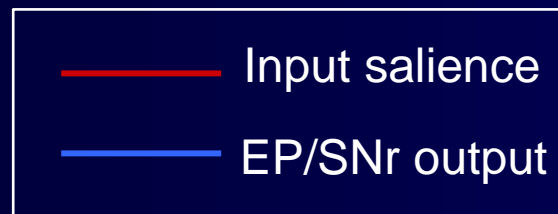
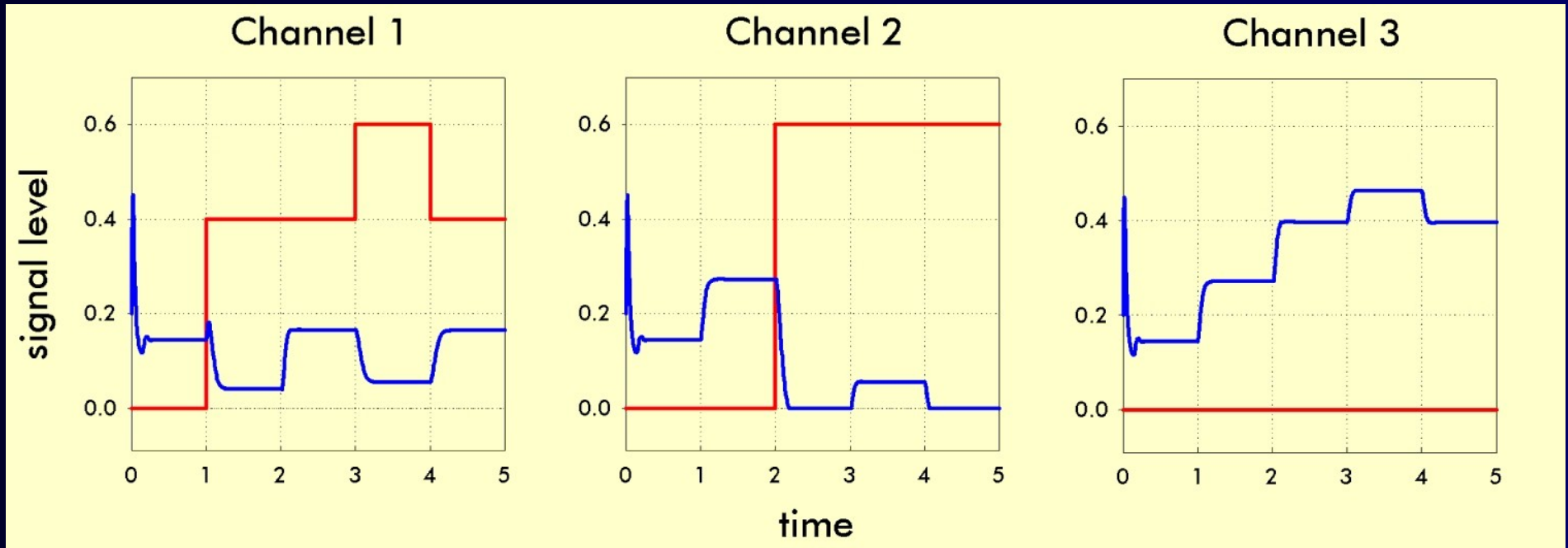
Solving for STN excitation

$$X^+ = \frac{n}{1 + \delta w_g w^- n \phi_{m,s}} \{ w_t \phi_{*,s} \langle c \rangle_*^s + \phi^s \epsilon' + \phi_{q,s} w_g w^- [(1 - \lambda_e) w_s \langle c \rangle_{q,s} - \epsilon] - w_g \phi^s \epsilon_e \}$$

....

# Network and spiking model simulations

Dynamic switching between channels on basis of changes in input salience

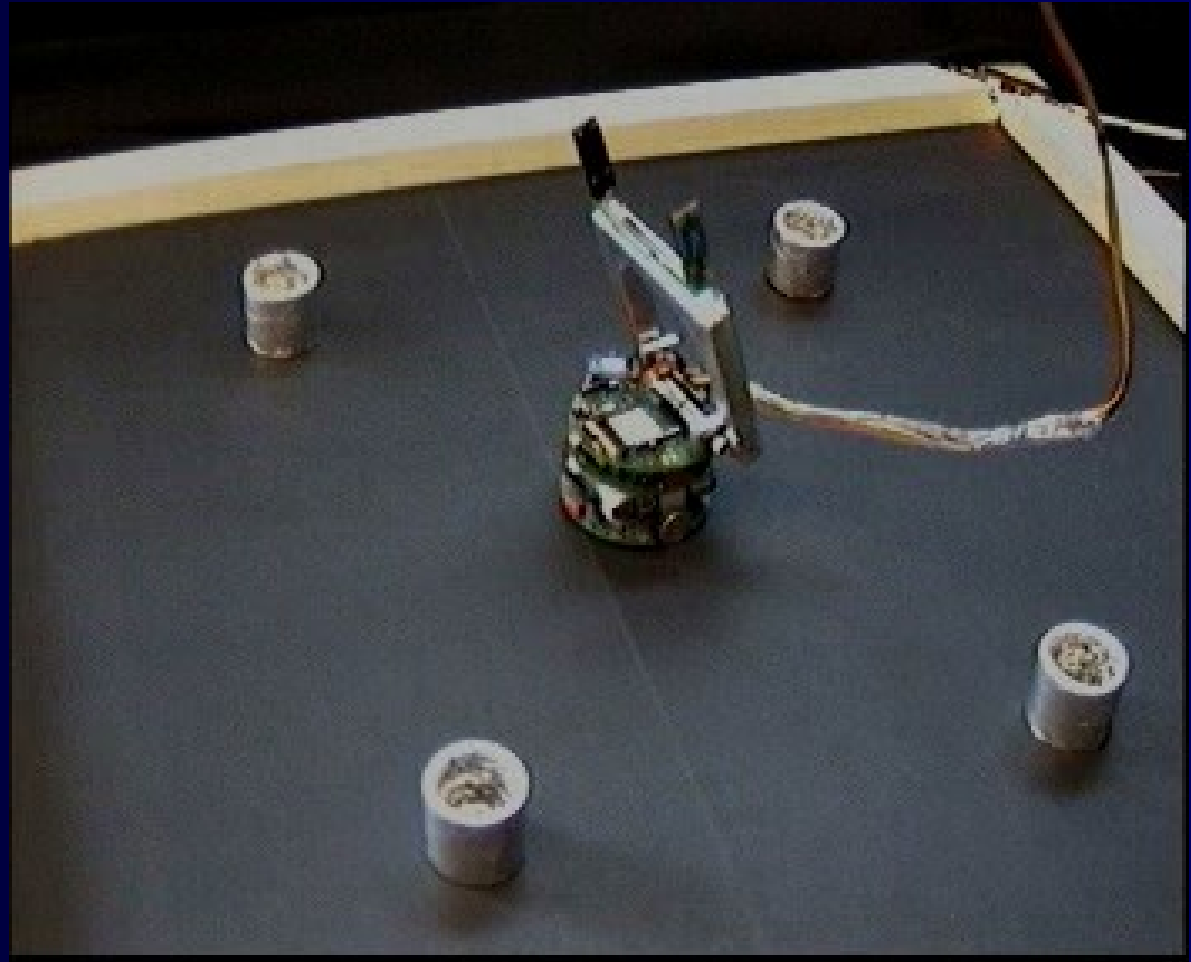


Gurney, K., T. J. Prescott, et al. (2001). "A computational model of action selection in the basal ganglia. I. A new functional anatomy." *Biol Cybern* **84**: 401-410.



# Robot Action Selection

- Motivations
  - Hunger
  - Fear
- 5 behavioural sub-systems
  - Wall seek
  - Wall follow
  - Can seek
  - Can pick-up
  - Can deposit
- 8 Infra-red sensors detect
  - Walls
  - Corners
  - Cans
- Gripper sensors detect
  - Presence/absence of can



# Conclusions

- Uniquely, selection hypothesis of basal ganglia architecture confirmed in analysis, simulation and control of robot action selection
- Represents a generic task performed in all functionally segregated territories of the basal ganglia
  - Selection of overall behavioural goal (limbic)
  - Selection of actions to achieve selected goal (associative)
  - Selection of movements to achieve selected actions (sensorimotor)
- Consistent with early development and evolutionary conservation
- Explains basal ganglia ‘involvement’ in so many tasks

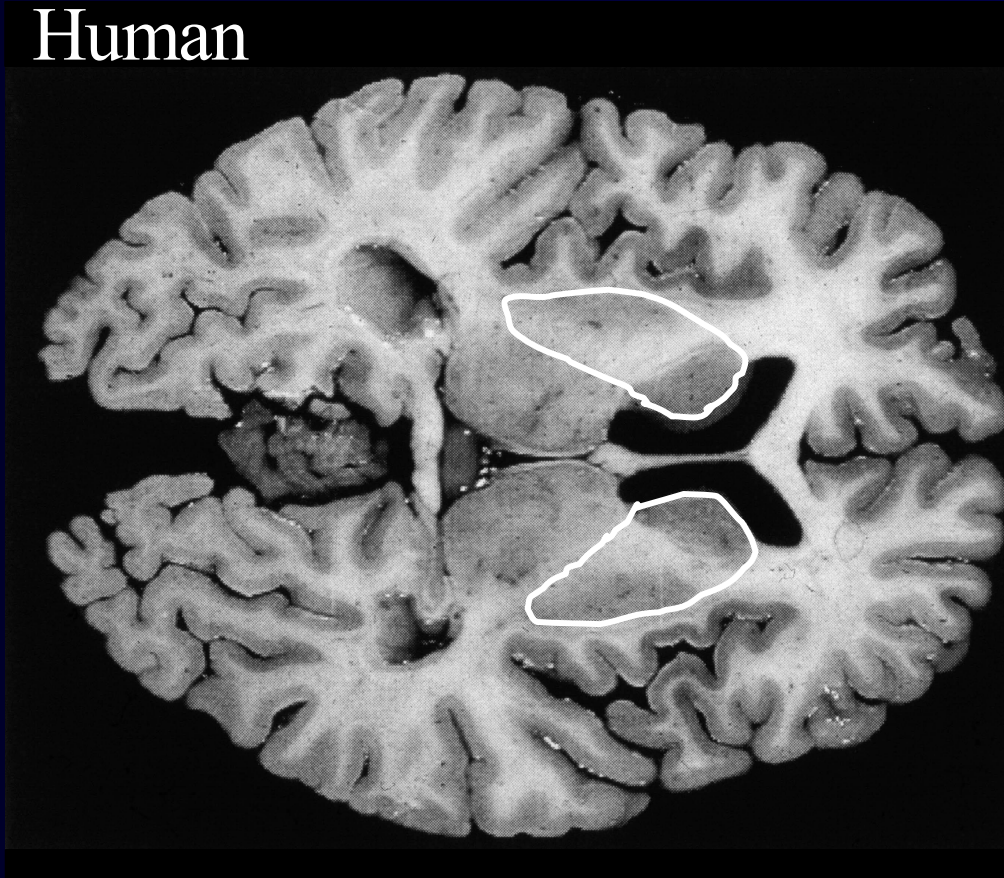
# Implications

- If the basal ganglia are operating as a central selection mechanism, what follows ?
  - Is “selective attention” a higher level description of currently selected (winning) channels ?
  - How does the evolutionary status of external command systems affect selection ?
  - What is the role of the central selector in adaptive behaviour ?

# The basal ganglia may have be conserved

.... unlike cerebral cortex and cerebellum  
the basal ganglia have not increased in  
relative size with brain development

Human

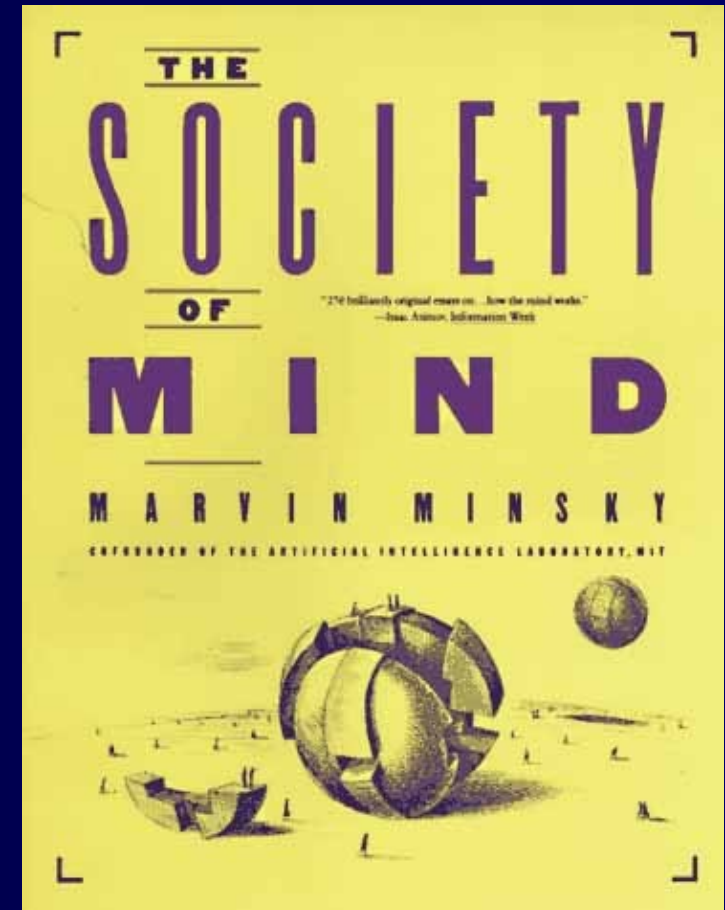


Rat



...but the competing systems certainly haven't

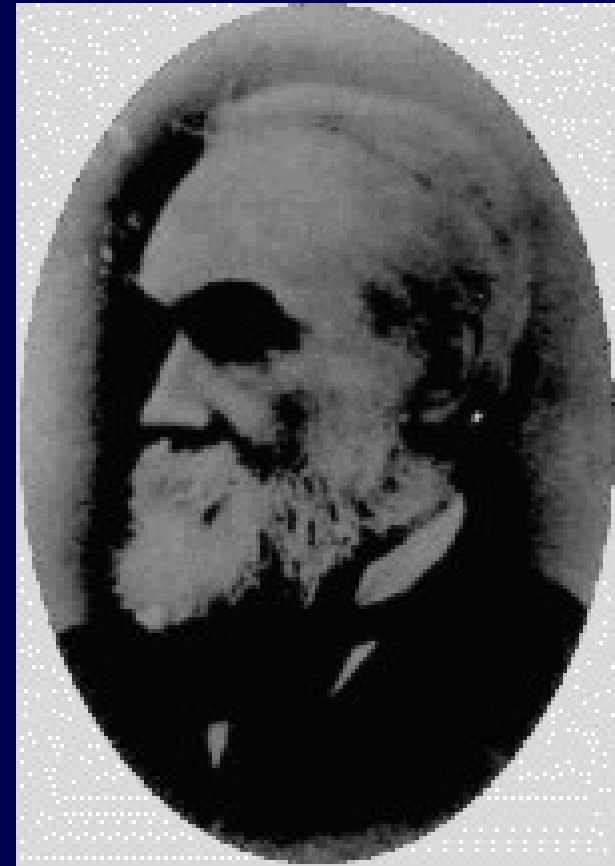
- How have functional units developed during evolution ?
  - Early systems simple solutions
  - Later components added to provide increasingly sophisticated solutions
  - ....to the same problems



## Layered architecture: not a new idea

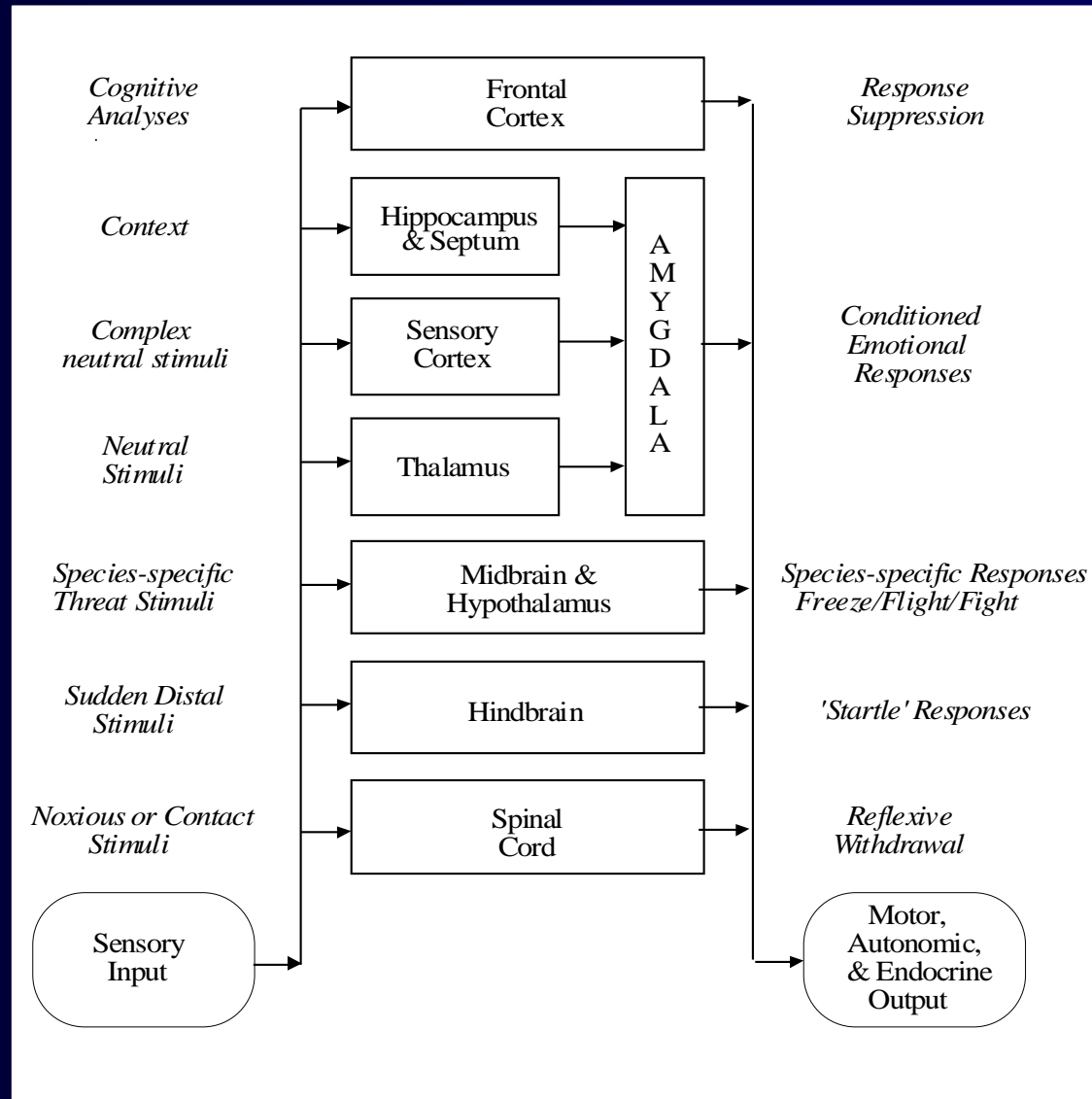
“That the middle motor centers represent over again what all the lowest motor centers have represented, will be disputed by few. I go further, and say that the highest motor centers (frontal lobes) represent over again, in more complex combinations, what the middle motor centers represent.”

From “The evolution and dissolution of the nervous system” (1884)



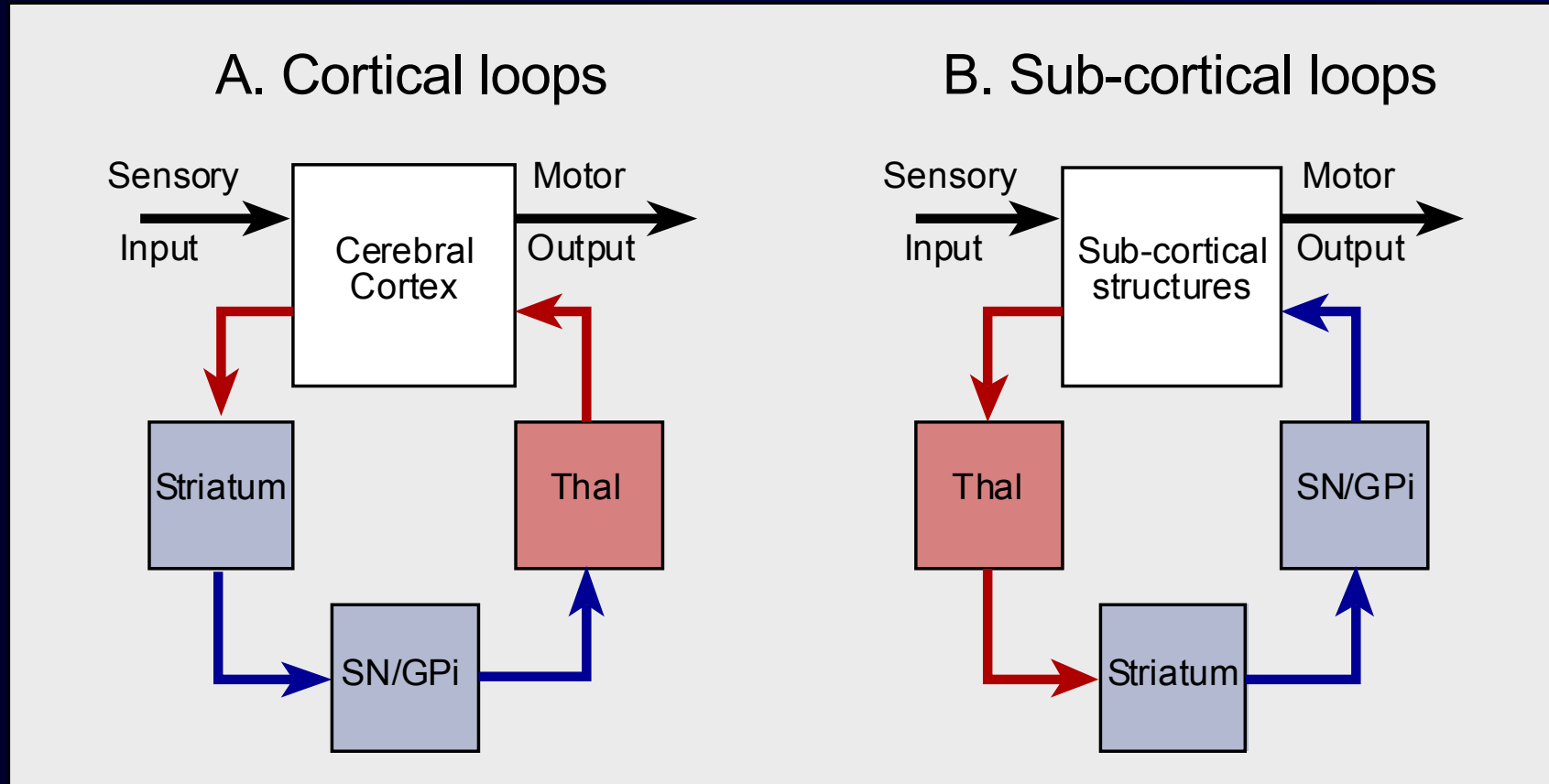
**John Hughlings Jackson**  
**1835-1911**

# Increasing sophistication across the neuraxis



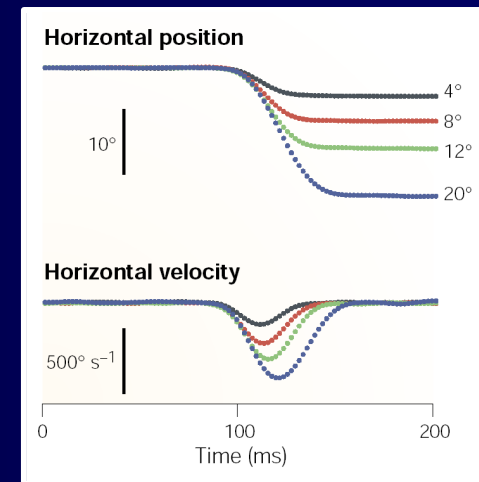
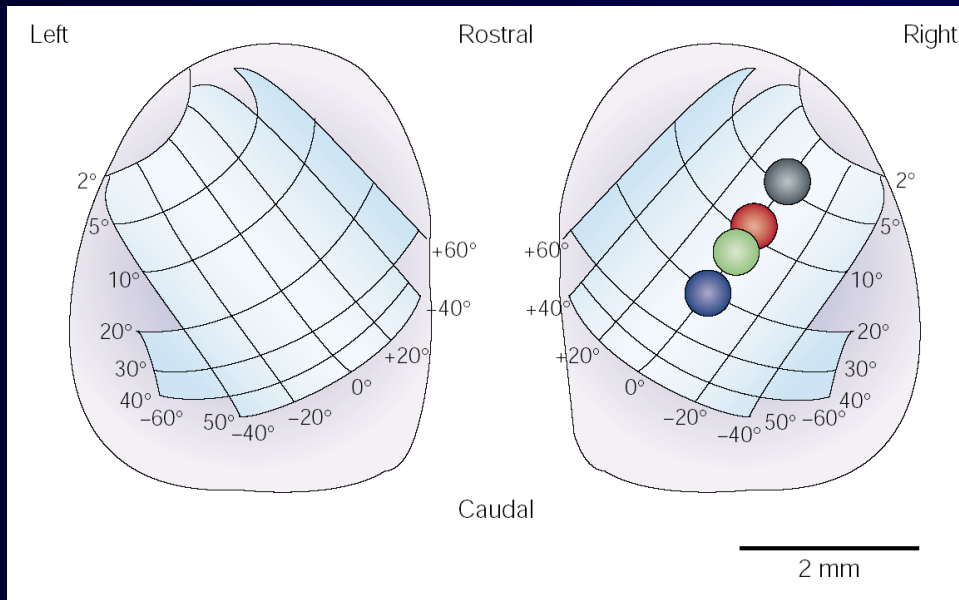
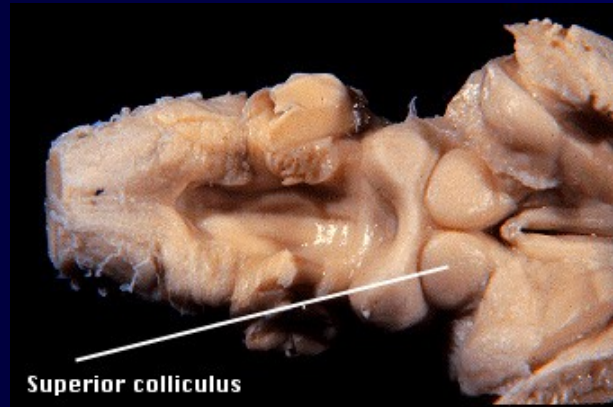
# How was selection done before cortical loops ?

## Subcortical loops through the basal ganglia

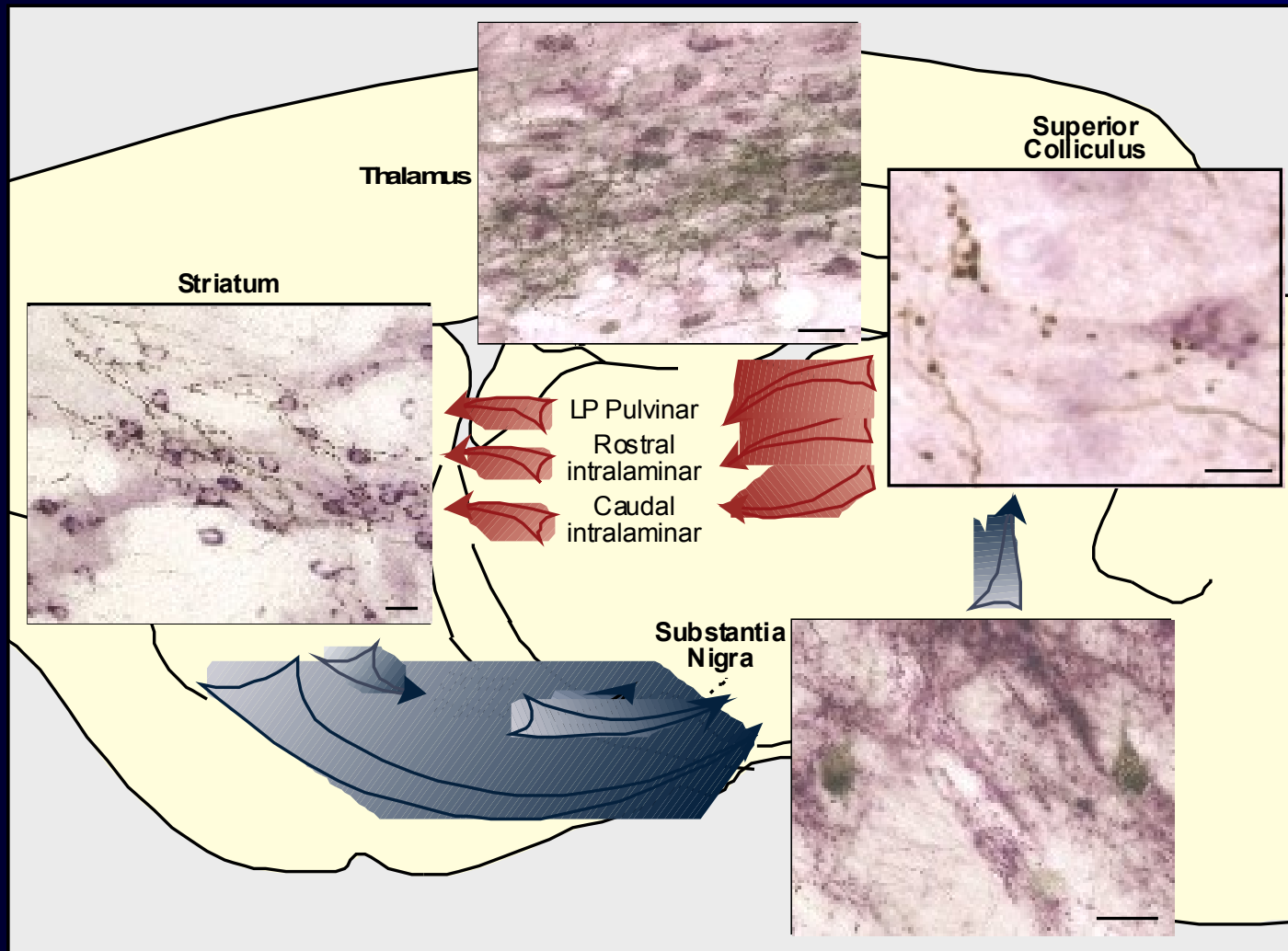




# Midbrain superior colliculus



# Subcortical loops from the superior colliculus



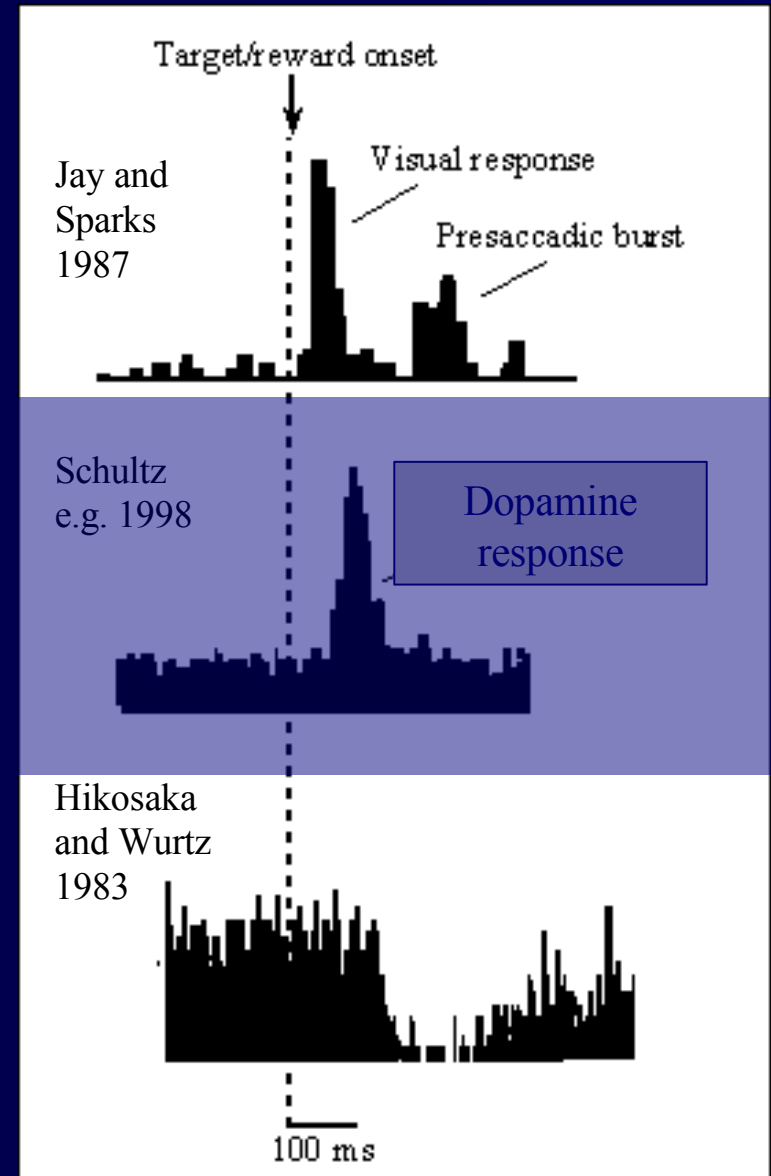
# Parallel processing sensory representations



# Signal timing in the superior colliculus

- Unexpected visual stimuli elicit sensory and motor responses in the superior colliculus:
  - short latency sensory reaction (~40 ms)
  - longer latency (<150 ms) pre-saccadic motor burst temporally associated with orienting

- Activity in basal ganglia output nuclei :
  - at 120ms+ nigrotectal disinhibition releases the orienting motor response in the colliculus



# Cortical and subcortical command systems

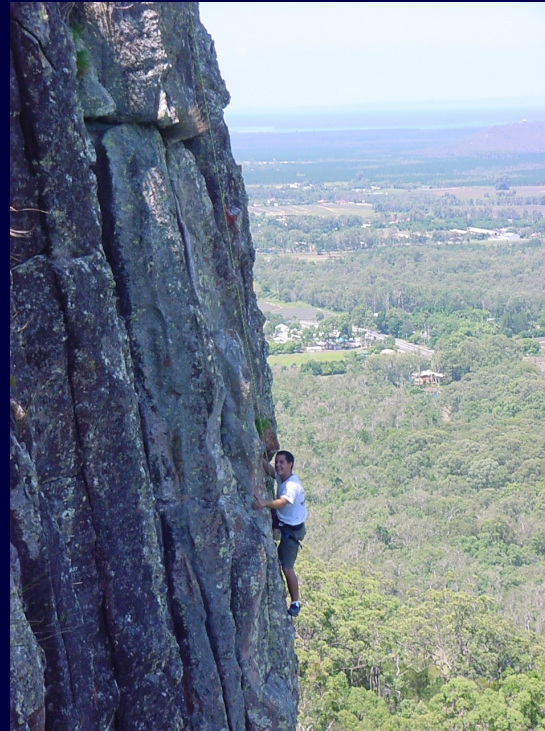
## Architecture for rational/irrational behaviour

- Cortical representations (bids) often based on more sophisticated sensory analyses and models of action consequences
- Subcortical representations heavily dependent on immediate sensory events
- What happens when they go head-to-head in the basal ganglia ?
  - ...depends on relative input salience

# Cortical/subcortical competition ?

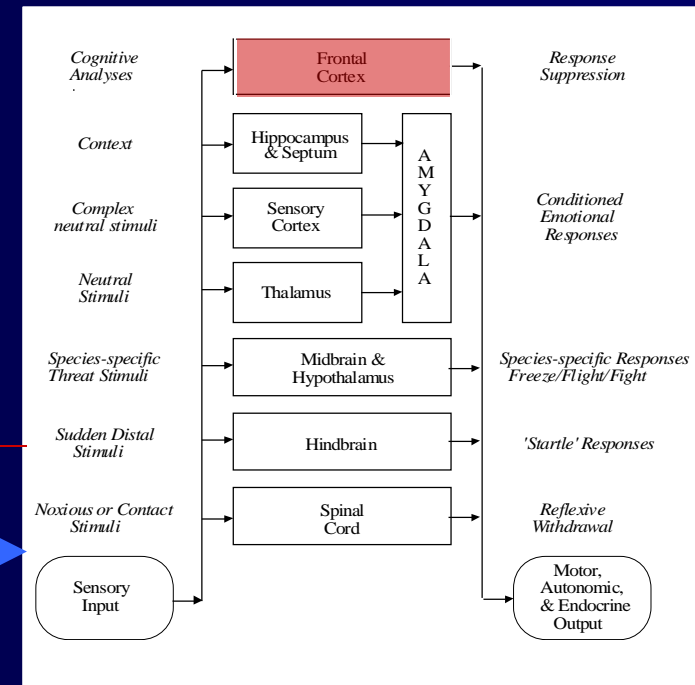
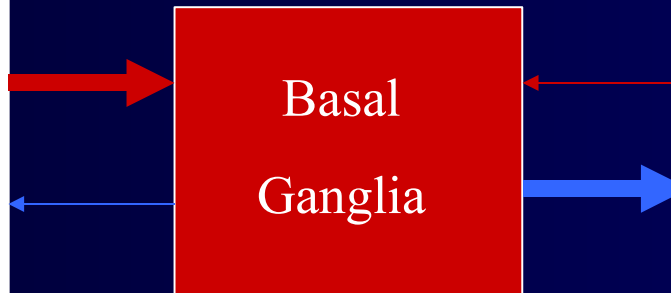
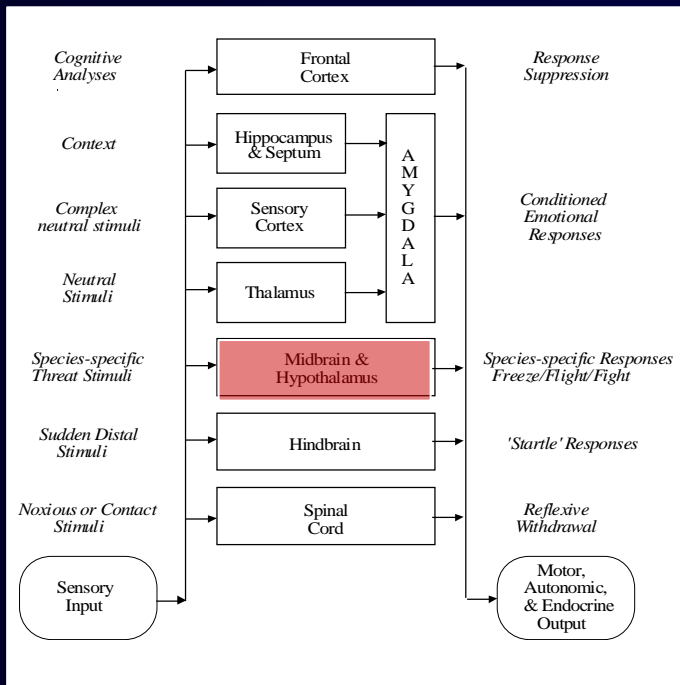
- Subcortical system

- Slow optic flow in lower visual field
- Defense reaction



- Cortical system

- Knowledge of rope strength
- ...go for it !



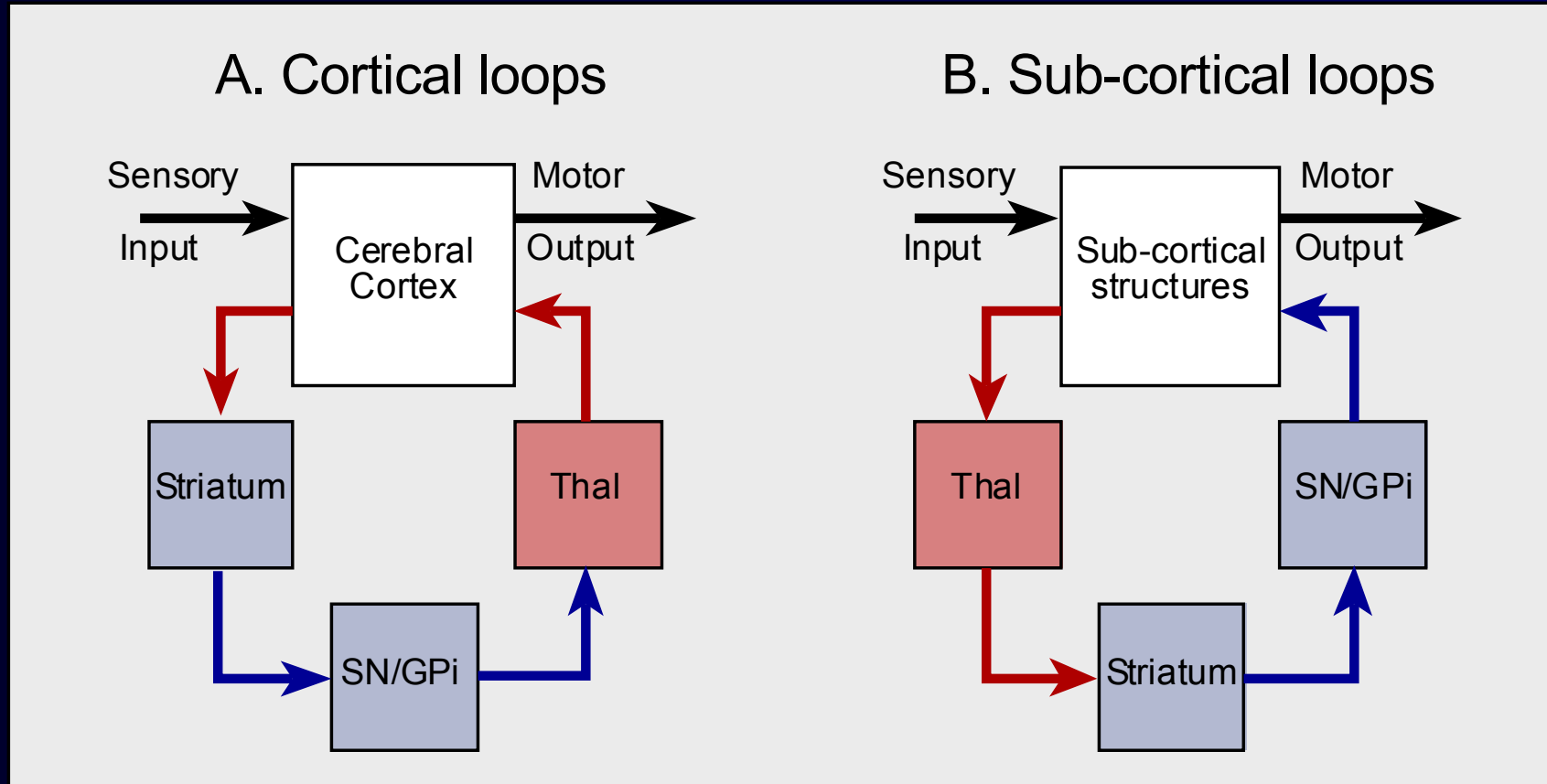
# Examples of (cortical) loosers

- **Phobias**
  - Specific trigger stimuli known to be harmless
  - Elicit uncontrollable fear and defensive reactions
- **Anxiety-panic attacks**
  - Situations known not to be dangerous
  - Incapacitating anxiety in absence of specific triggers
- **Post-traumatic stress disorders**
  - Current circumstances unrelated to traumatic event
  - Irrelevant stimuli evoked flash-backs which elicit uncontrollable fear and defensive reactions
- **Addictions**
  - Knowledge of detrimental effects of drug dependence explicit
  - Often powerless in the face of drug/food/sex related sensory stimuli
- **Head versus heart**
  - Situations where we should know better



# Cortical and subcortical loops

## An architecture for understanding such conflicts





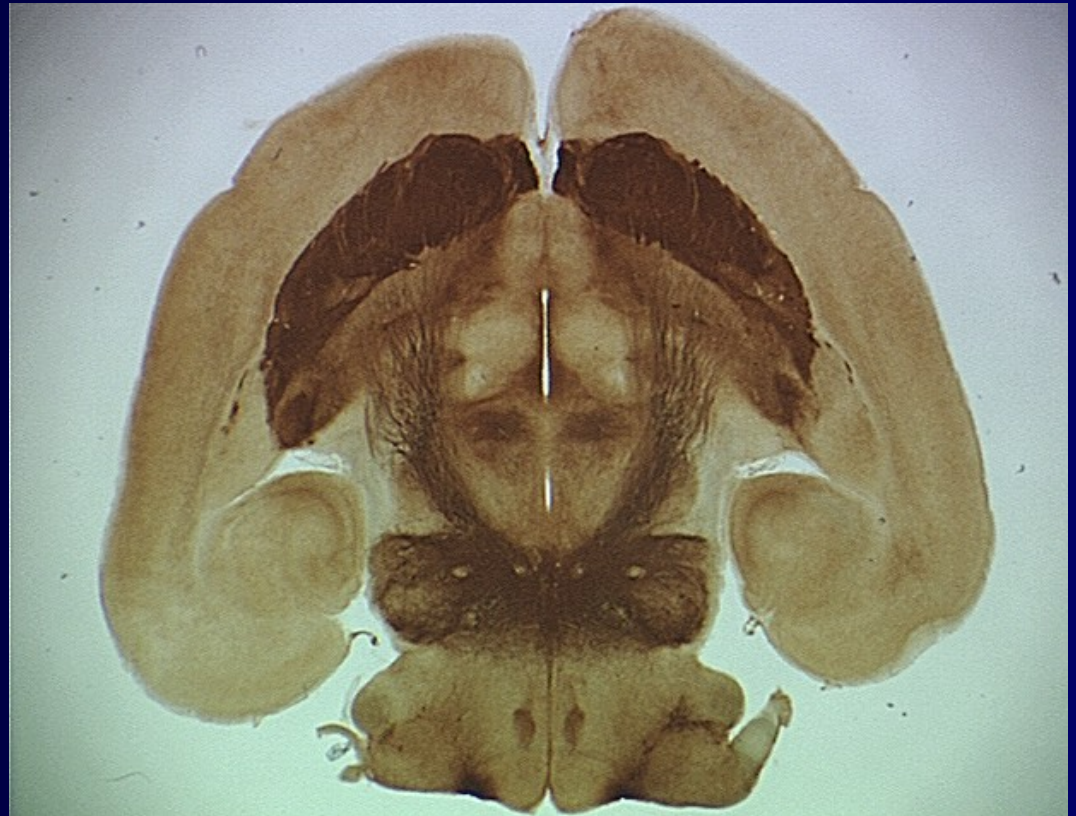
# Adaptive selection

For action selection to adapt with experience, must be responsive to reinforcement consequences of action-outcome contingencies

Ascending dopaminergic systems in rat brain

- Selective adjustment of afferent signals by reinforcement outcome
- and/or adjustment of input weights of reinforced channels
- The role of dopamine in reinforcement learning

Picture by Wes Chang  
(Gallo center San Francisco)



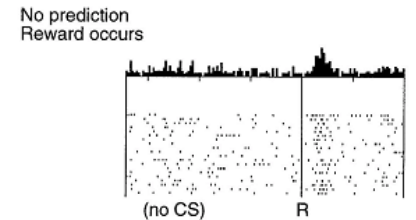
# Dopaminergic neurones sensitive to reward

- Phasic short-latency sensory response

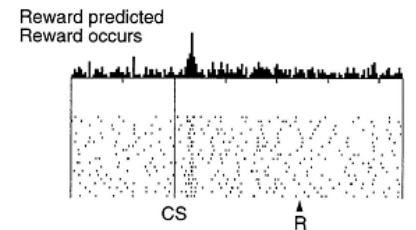
- Short latency (70-100ms)
- Short duration (~ 100ms) burst of impulses

Schultz W. *J. Neurophysiol.*  
(1998)

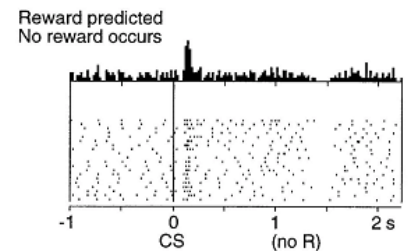
Unexpected reward



Reward-predicting stimulus



Unexpected reward omission



- Schultz (1998) – signals reward prediction error

- Shares many characteristics of ‘r’ in Temporal Difference algorithms
- Used to adjust response probabilities in associative learning

# Phasic dopamine unlikely to signal reward prediction error

- Elicited by unpredicted biologically salient stimuli
  - Salient by virtue of:
    - novelty (independent of reward value)
    - association with reward
    - intensity
    - physical resemblance to reward related stimuli
- Response homogeneity
  - 100ms latency 100ms duration response constant across:
    - species
    - experimental paradigms
    - sensory modality
    - perceptual complexity of eliciting events
- Response latency (~100ms)
  - Precedes gaze shift that brings event onto fovea...

# The latency constraint

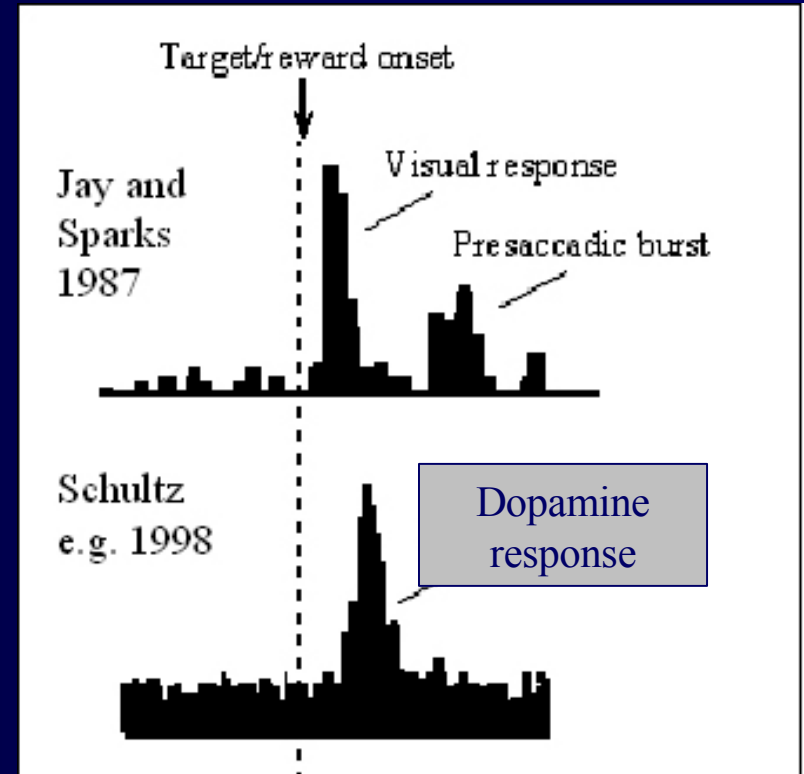
Unexpected visual stimuli elicit sensory and motor responses in superior colliculus:

- sensory response (~40 ms)
- motor response (<150 ms)

Phasic DA responses occur before foveating eye-movements

70-100ms after stimulus onset

- **Conclusion:** anomaly of having brain's main reinforcement learning systems relying on reward identification done by pre-attentive, pre-saccadic stimulus processing



## So how was it for you ?

“We also noticed that DA neurons typically responded to a visual or auditory stimulus when it was presented unexpectedly, but stopped responding if the stimulus was repeated; a subtle sound outside the monkey’s view was particularly effective.”

Takikawa Y, Kawagoe R, Hikosaka O. 2004. A possible role of midbrain dopamine neurons in short- and long-term adaptation of saccades to position-reward mapping. *J Neurophysiol* 92(4):2520-2529.

If phasic dopamine isn't signaling  
reward prediction error....  
what is it signaling ?

# Essential characteristics of the phasic dopamine signal

- A striking resemblance to the Temporal Difference reinforcement error term
  - .....suggests it is critically associated with reinforcement learning
- It is precisely timed
  - .....involved in a process where the timing of the reinforcement signal is critical

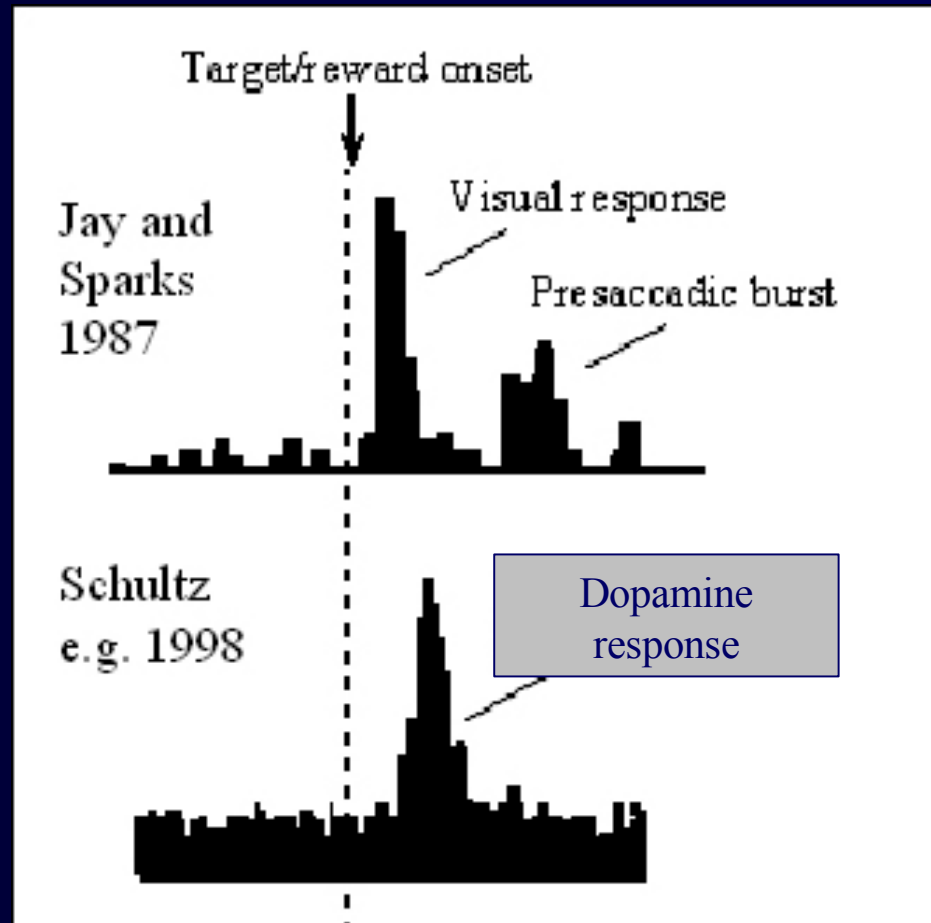
But more information needed

# Prior Questions

- What is the source of the short latency sensory (visual) input to dopamine neurones ?
- What signals does the timed dopamine response interact with in target regions of the basal ganglia ?



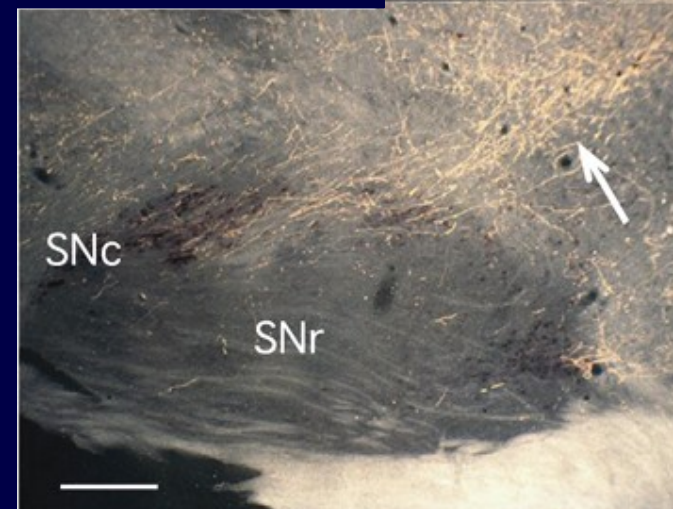
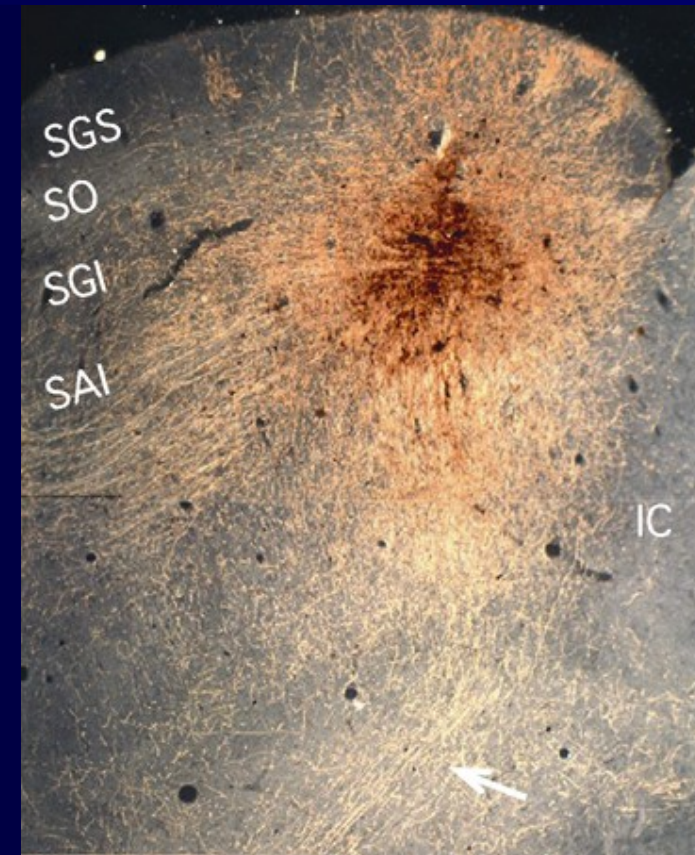
# Response latencies suggest the superior colliculus



# Colliculus as the source of visual input: I

## Anatomical Evidence

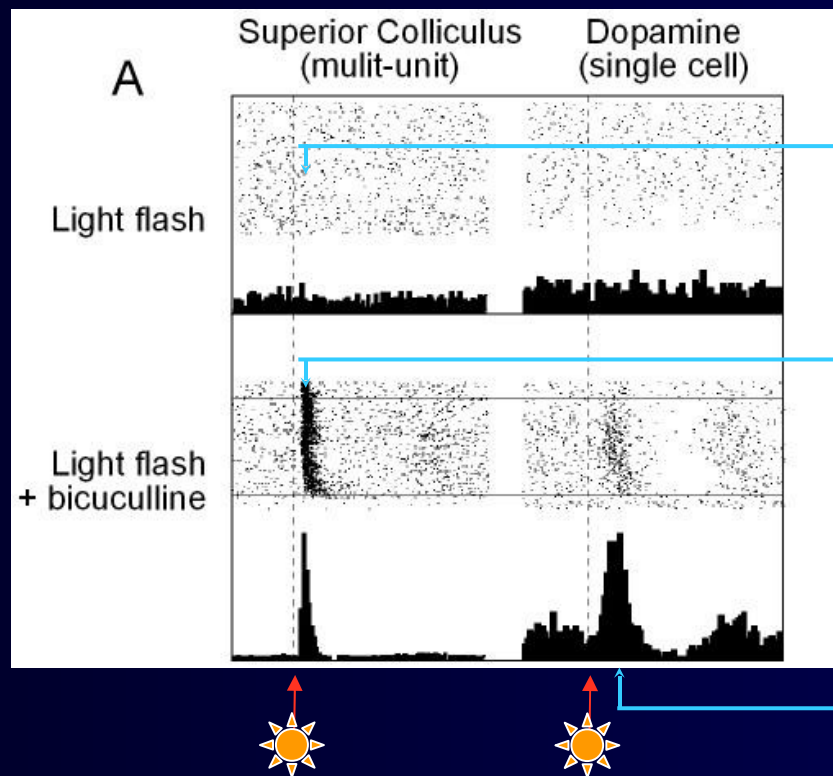
- The Tectonigral projection
- Direct pathway discovered from superior colliculus to substantia nigra pars compacta



Comoli, et al. (2003). *Nature Neurosci* 6: 974-980.

# Colliculus as the source of visual input: II

## Electrophysiological Evidence

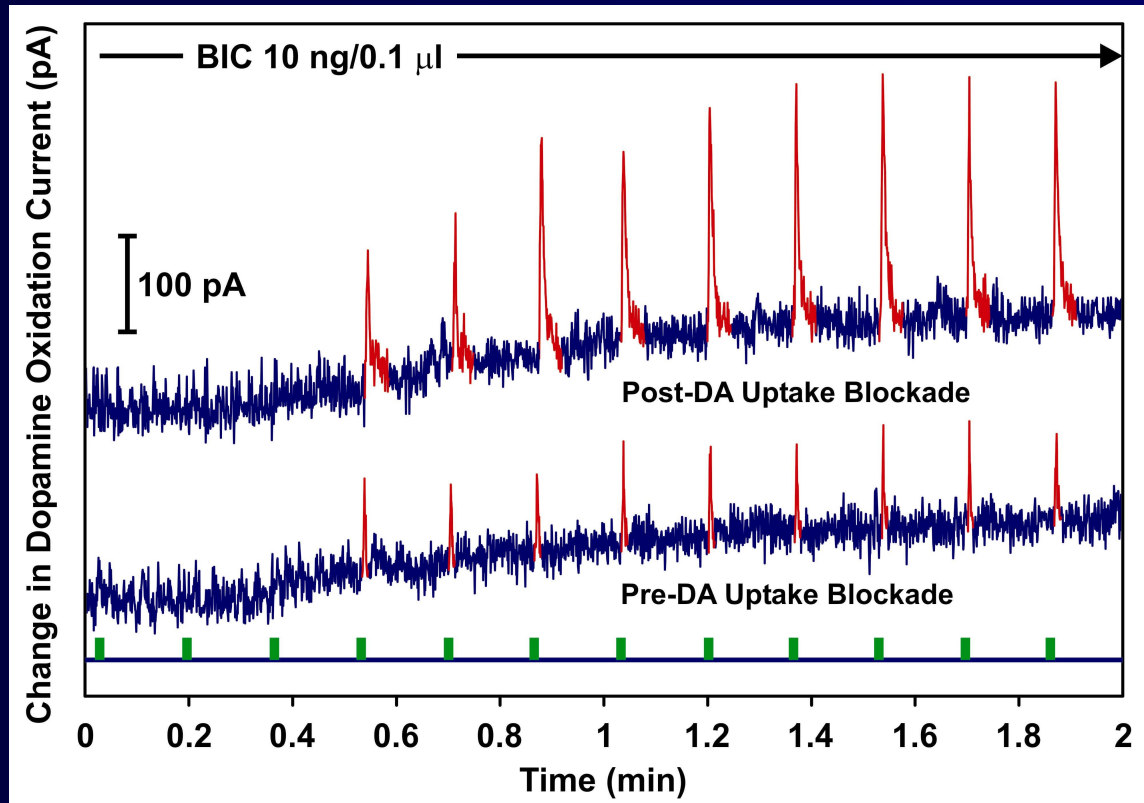


- Pre-drug baseline
  - No flash-evoked response in deep SC or DA cells
- After BIC into deep SC
  - local neurones responsive to light
- When SC cells ‘see’ so do DA cells
  - Excitatory responses: 17/30 (56.6%)

# Colliculus as the source of visual input: III

## Electrochemical Evidence

- No release to light without collicular bicuculline
- 10-40ng bicuculline in 100-400nl into colliculus elicited light response
- Amplitude and duration of response increased by selective DA re-uptake blocker *Nomifensin*

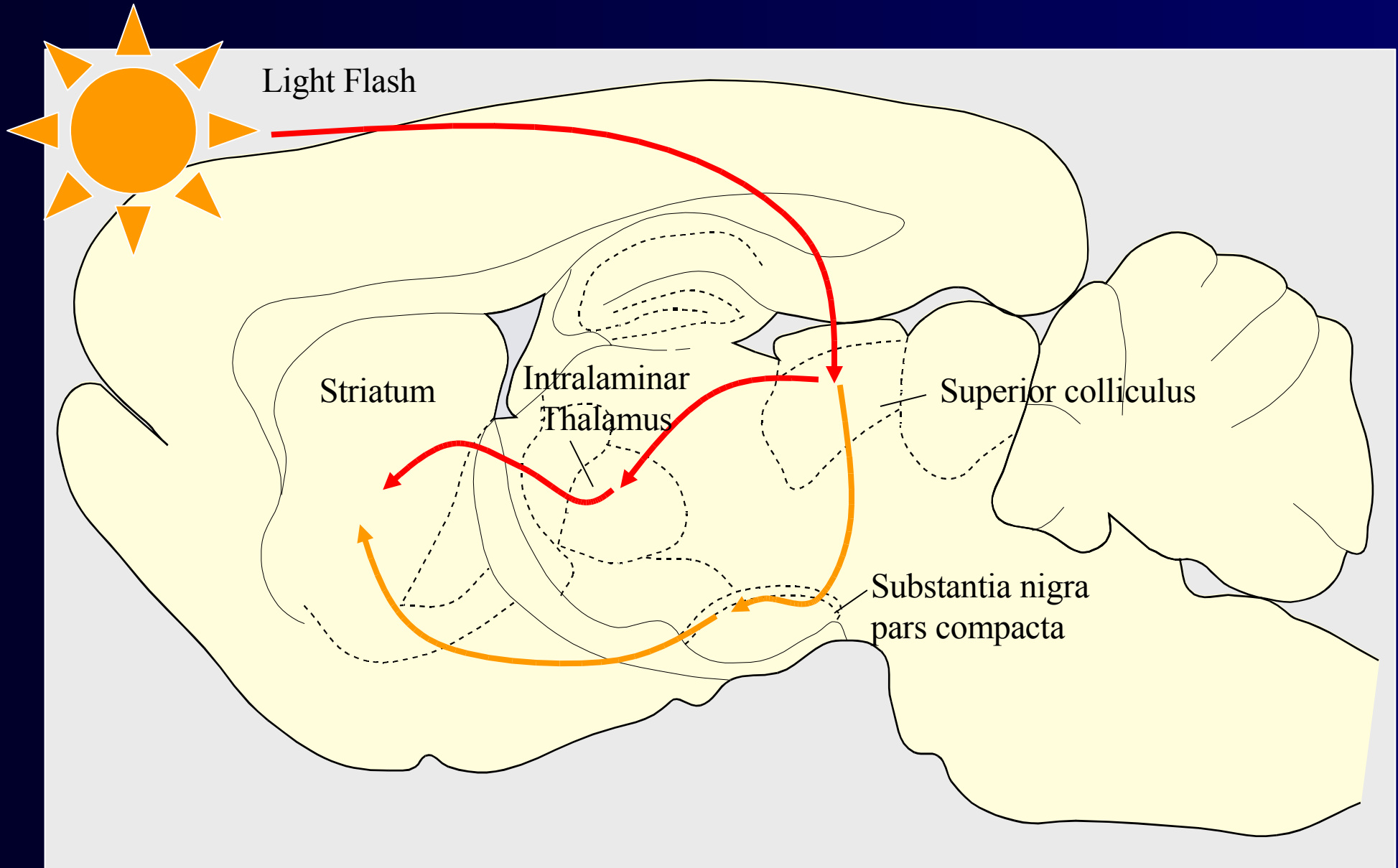


## Question:

What signals are present in the target regions at the time of the phasic dopamine input ?

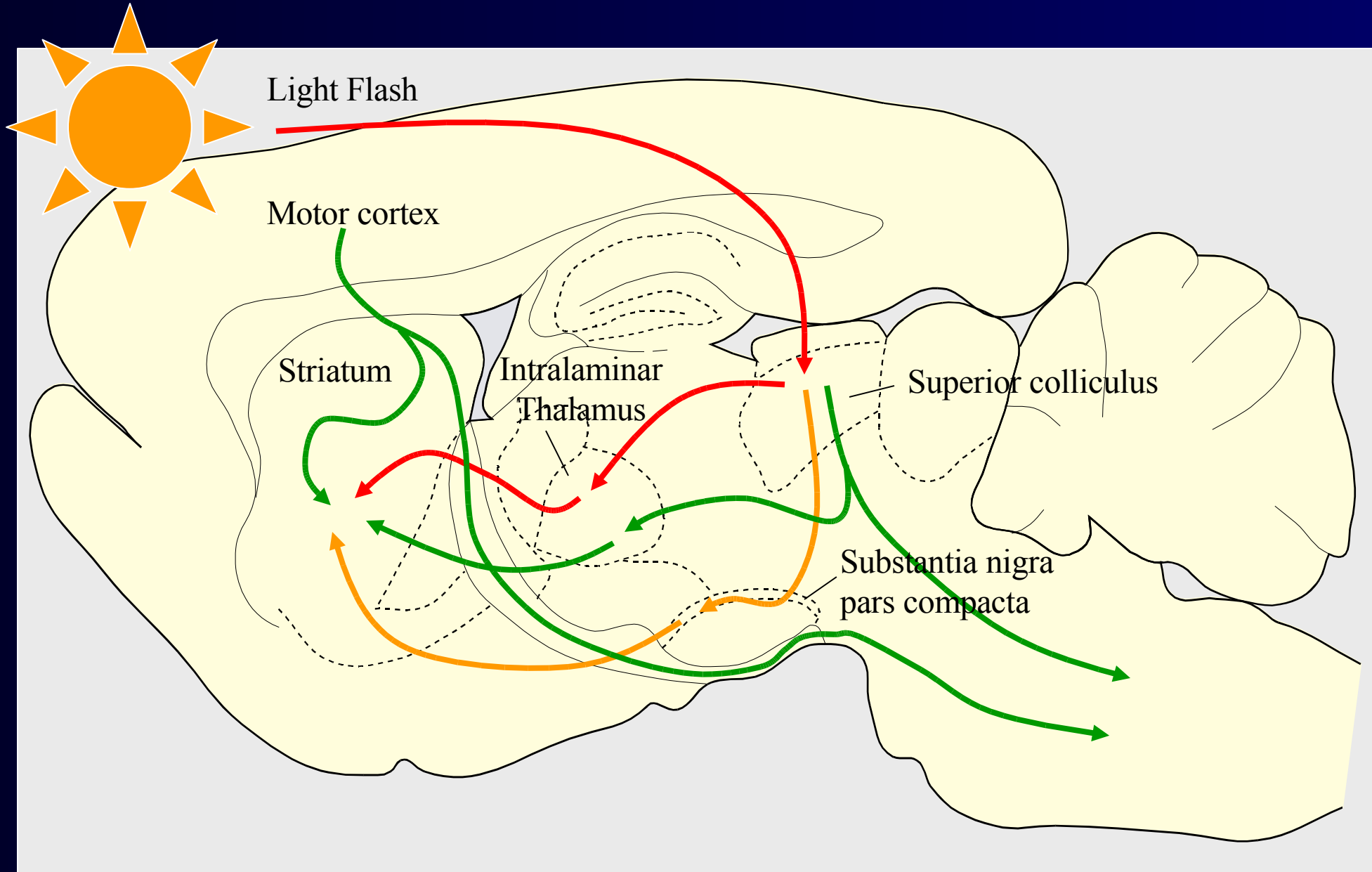
- **1<sup>st</sup> Signal** – a separate representation of the sensory event that fired off the dopamine signal

# Sensory inputs to the striatum



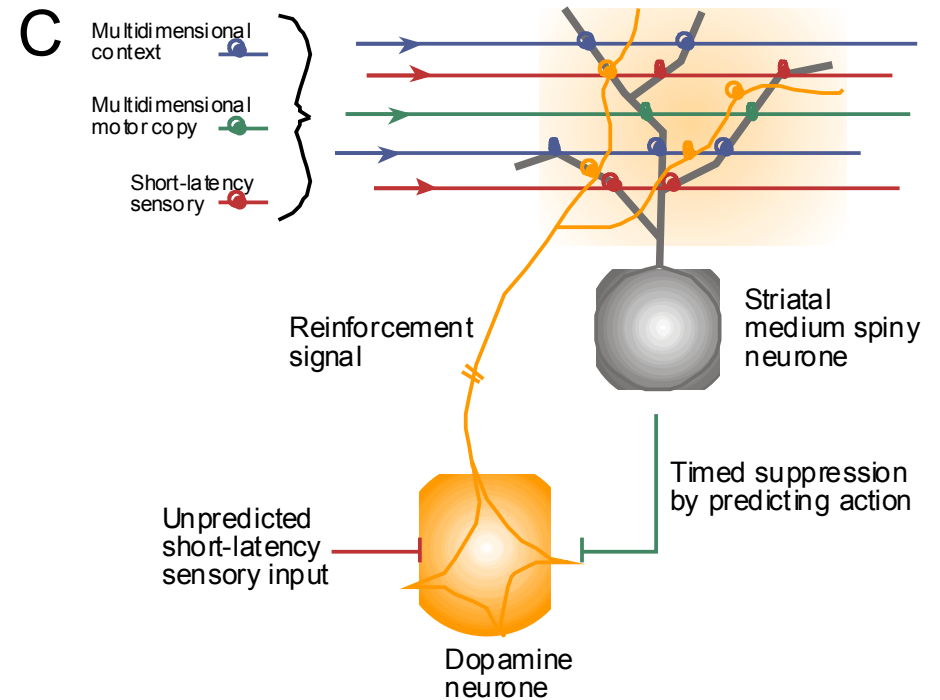
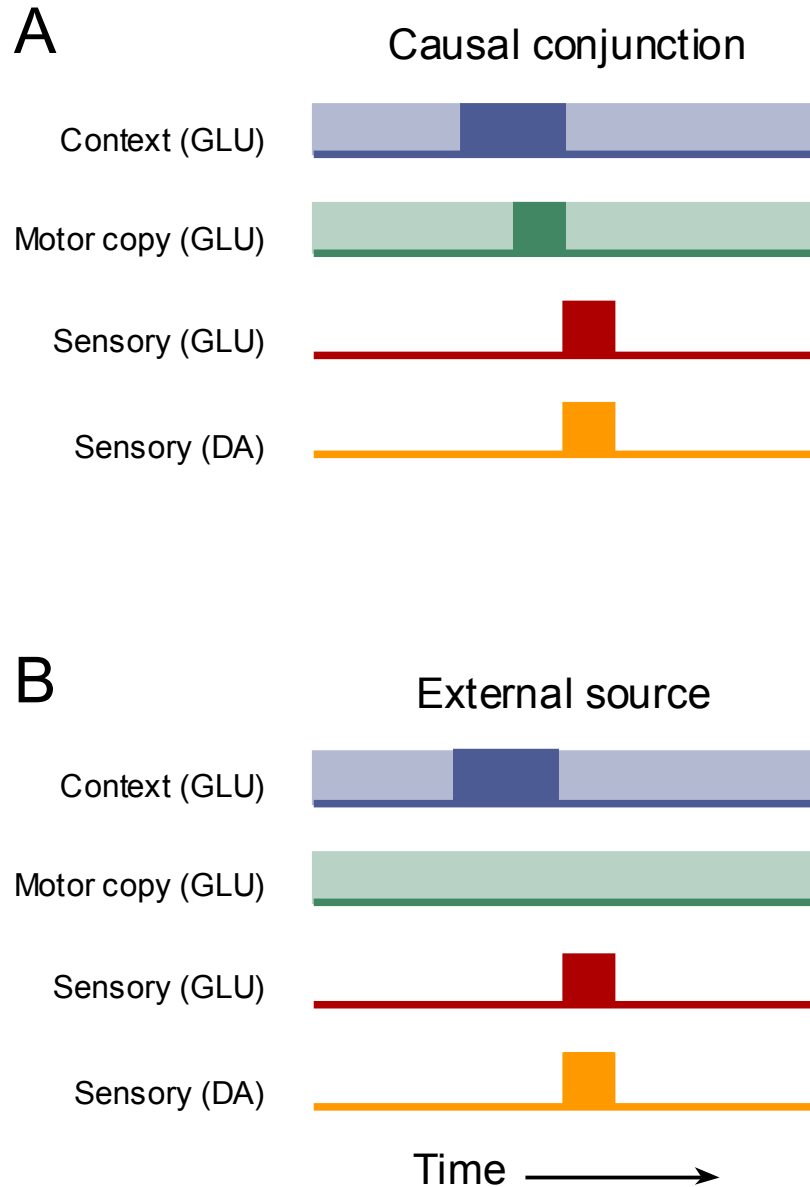
- 2<sup>nd</sup> Signal – a running efference copy or corollary discharge of ongoing motor commands

# Motor inputs to the striatum: Efference copy

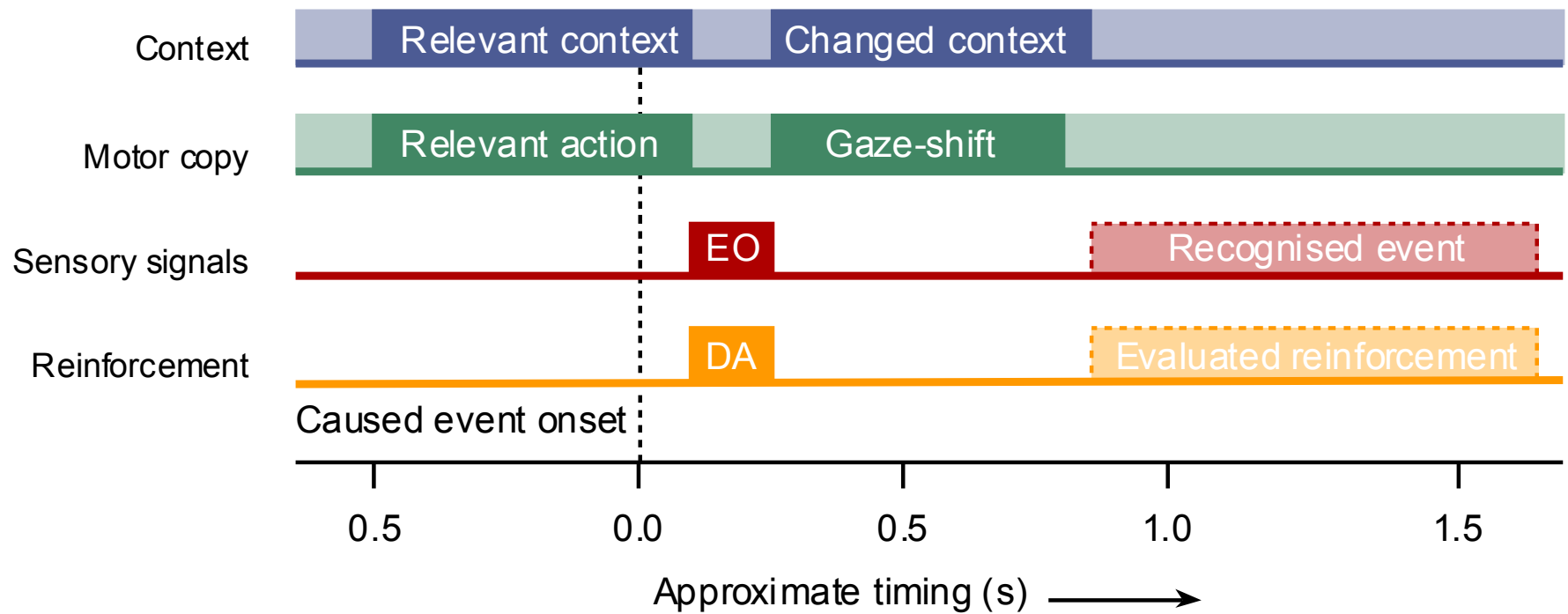




# Causal Contingencies



# Why a short latency reinforcement signal is essential



What-action-caused-the-event learning

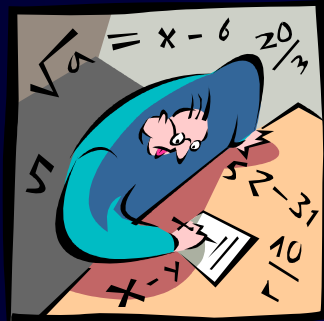
# Conclusions

- Multifunctional systems must have effective solution(s) to the selection problem
- The basal ganglia appear to provide a biological solution deemed adequate for > 400M years
- Distribution of competitors across different levels of the neuraxis can lead to competition between systems of different evolutionary status
- Analysis of basal ganglia functional architecture suggests intrinsic reinforcement properties could operate to determine agency

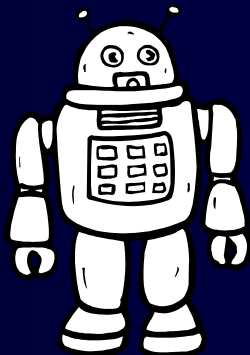
# The Team




- Biology
  - Veronique Coizet
  - Eliane Comoli
  - Ellie Dommett
  - Paul Overton



- Computation
  - Kev Gurney
  - Mark Humphries



- Robotics
  - Tony Prescott
  - Jon Chambers



# Cognition and Consciousness: Is There a Fundamental Link?

Murray Shanahan  
Imperial College London  
Department of Computing



# Overview

- Three related issues
  - Neural parallelism
  - Modular theories of mind
  - The frame problem
- Global workspace theory (GWT)
- Applying GWT to the three issues



# First Issue

# Neural Parallelism

- An animal's nervous system is massively parallel
- Massive parallelism surely underpins human cognitive prowess
- So how are the massively parallel computational resources of an animal's central nervous system harnessed for the benefit of that animal?
- How can they orchestrate a coherent and flexible response to each novel situation?
- What is their underlying architecture?
- Nature has solved this problem. How?





# Second Issue

# Modular Theories of Mind (1)

- Many cognitive scientists advocate modular theories of mind (Gardner, Tooby & Cosmides, Fodor, Mithen, Carruthers)
- The mind comprises (or incorporates) an assemblage of distinct specialist *modules*
- Fine-grained horizontally modular theories (eg: Tooby & Cosmides) posit specialists for particular behaviours (eg: foraging)
- More coarse-grained vertically modular theories (eg: Fodor) posit specialists for certain input and output processes (eg: parsing, low-level vision)

# Modular Theories of Mind (2)

- In addition to the specialist modules, all modular theories demand (for humans) some overarching faculty, central system, super-module, meta-representational facility, or whatever
- This addition is capable, when required, of transcending modular boundaries to produce flexibly intelligent behaviour rather than an automatic, preprogrammed response to a novel situation
- But nobody has a very convincing account of this

# Third Issue

# The Frame Problem (1)

- The frame problem originated in classical AI

How can we formalise the effects of actions in mathematical logic without having to explicitly enumerate all the trivial non-effects?

- This is tricky, but was more-or-less solved in the mid 1990s
- Our concern is the wider interpretation given to the frame problem by philosophers, notably Dennett and Fodor

# The Frame Problem (2)

- Fodor's version:

How do *informationally unencapsulated* cognitive processes manage to select only the information that is relevant to them without having to explicitly consider everything an agent believes ?

- A cognitive process is *informationally unencapsulated* if it has the potential to draw on information from any domain
- Analogical reasoning is the epitome of informational unencapsulation

# Computational “Infeasibility”

- Fodor claims that informationally unencapsulated cognitive processes are computationally infeasible

“ The totality of one’s epistemic commitments is *vastly* too large a space to have to search ... *whatever* it is that one is trying to figure out. ”

(Fodor, 2000)

- Fodor believes that this is a fatal blow for cognitive science as we know it because it entails we cannot find a computational explanation of the human mind’s “central systems”

# Fodor's Modularity of Mind

- The mind's *peripheral* processes are special purpose, do things like parsing and low-level vision, and are computational
- The mind's *central* processes are general purpose, do things like analogical reasoning, are informationally unencapsulated, and (probably) *aren't computational*

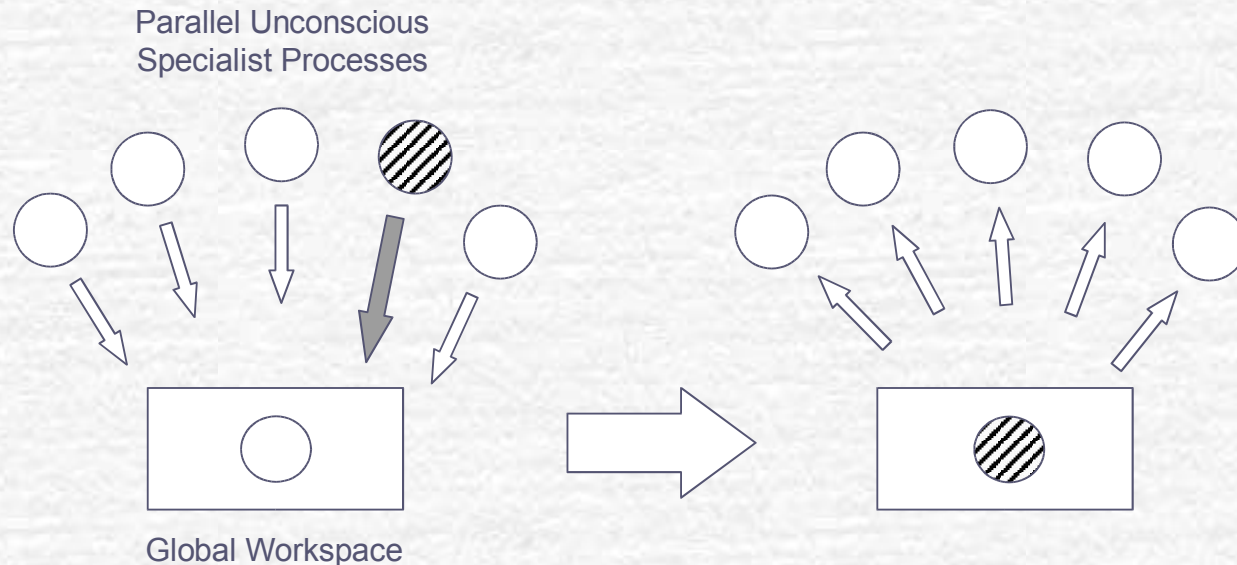
“... it probably isn't true that [all] cognitive processes are computations. ... [so] it's a mystery, not just a problem, what model of the mind cognitive science ought to try next. ” (Fodor, 2000)





# The Solution

# Global Workspace Architecture



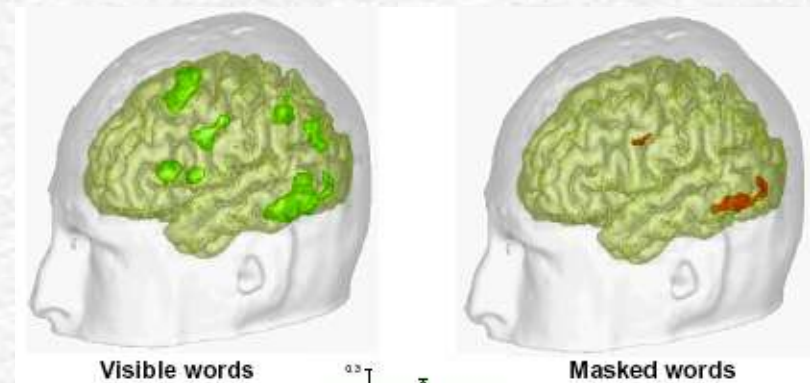
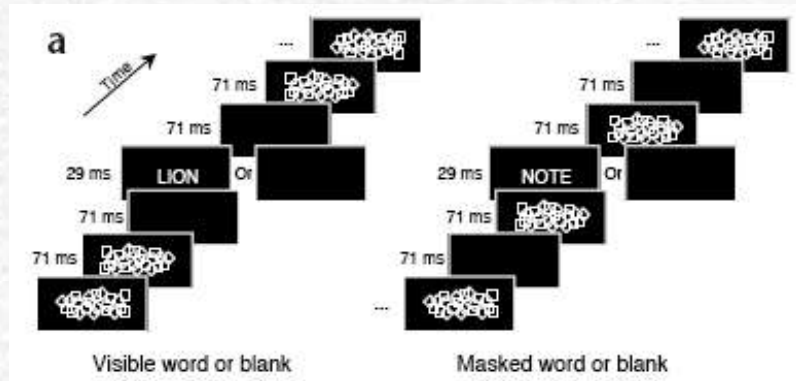
- Multiple parallel *specialist* processes compete and co-operate for access to a *global workspace*
- If granted access to the global workspace, the information a process has to offer is *broadcast* back to the entire set of specialists

# Conscious vs Non-Conscious

- Global workspace theory (Baars) hypothesises that the mammalian brain instantiates such an architecture
- It also posits an empirical distinction between conscious and non-conscious information processing
- Information processing in the parallel specialists is non-conscious
- Only information that is broadcast is consciously processed

# Empirical Evidence

- Contrastive analysis compares and contrasts closely matched conscious and unconscious brain processes
- Dehaene, *et al.* (2001)
  - Imaged subjects being presented with “masked” words
  - Masked and visible conditions compared

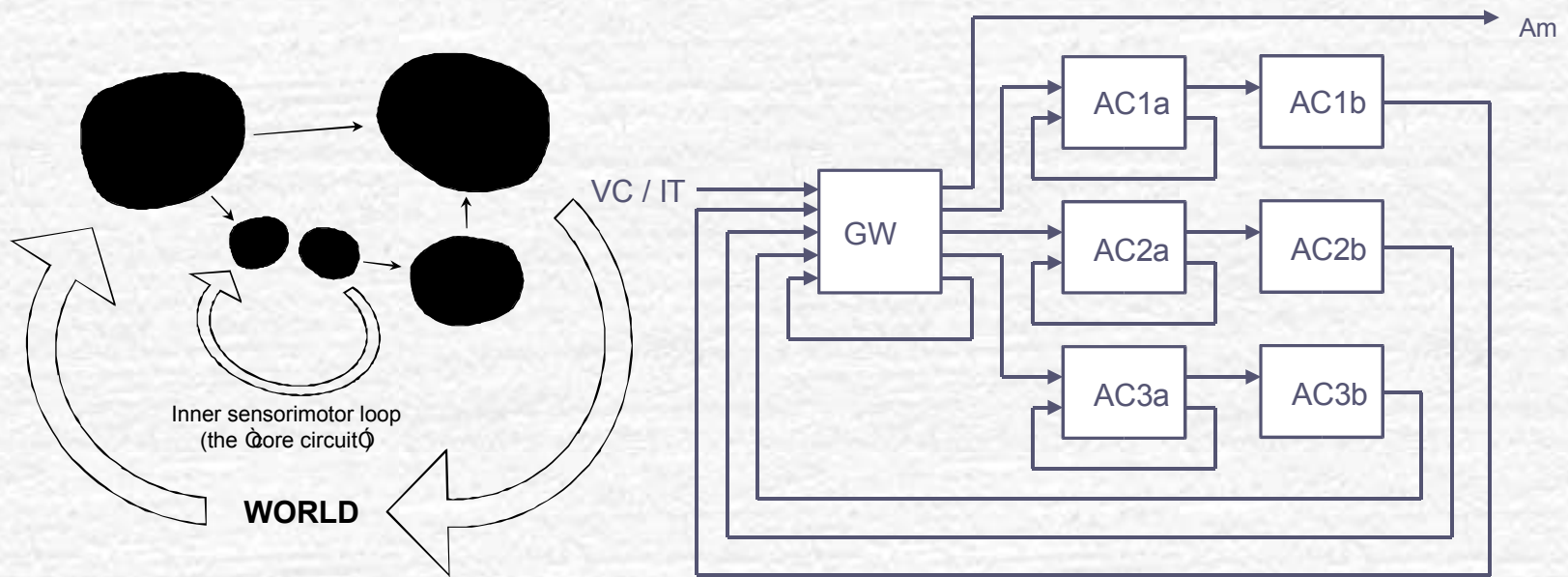


- Such experiments suggest that conscious information processing recruits widespread brain resources while unconscious processing is more localised

# Embodiment

- According to GWT, only something that instantiates a global workspace architecture is capable of conscious information processing
- But this is a necessary not a sufficient condition
- I have argued (Shanahan, 2005) that the architecture must direct the actions of a spatially localised body using a sensory apparatus fastened to that body
- This allows the set of parallel specialists a shared viewpoint, from which they can be indexically directed to the world and fulfil a common remit

# Combining GWT with Internal Simulation



This “core circuit” combines an internal sensorimotor loop with mechanisms for broadcast and competition, and thereby marries the *simulation hypothesis* (Cotterill, Hesslow) with *global workspace theory* (Baars)

# Applying the Solution

# Serial from Parallel / Unity from Multiplicity

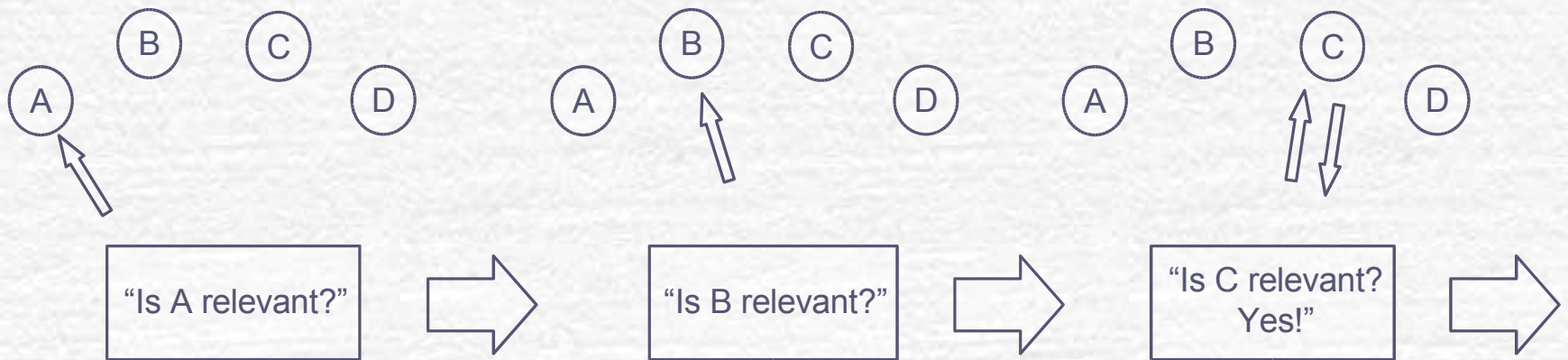
- The global workspace architecture harnesses the power of massively parallel computation
- The global workspace itself exhibits a *serial* procession of states
- Yet each state-to-state transition is the result of filtering and integrating the contributions of huge numbers of *parallel* computations
- The global workspace architecture thereby distils unity out of multiplicity
- This is perhaps the essence of consciousness, of what it means to be a singular, unified subject



# GWT and the Frame Problem (1)

- Both Fodor and Dennett seem to have a strictly serial architecture in mind when they characterise the frame problem

Peripheral Processes (Modules)



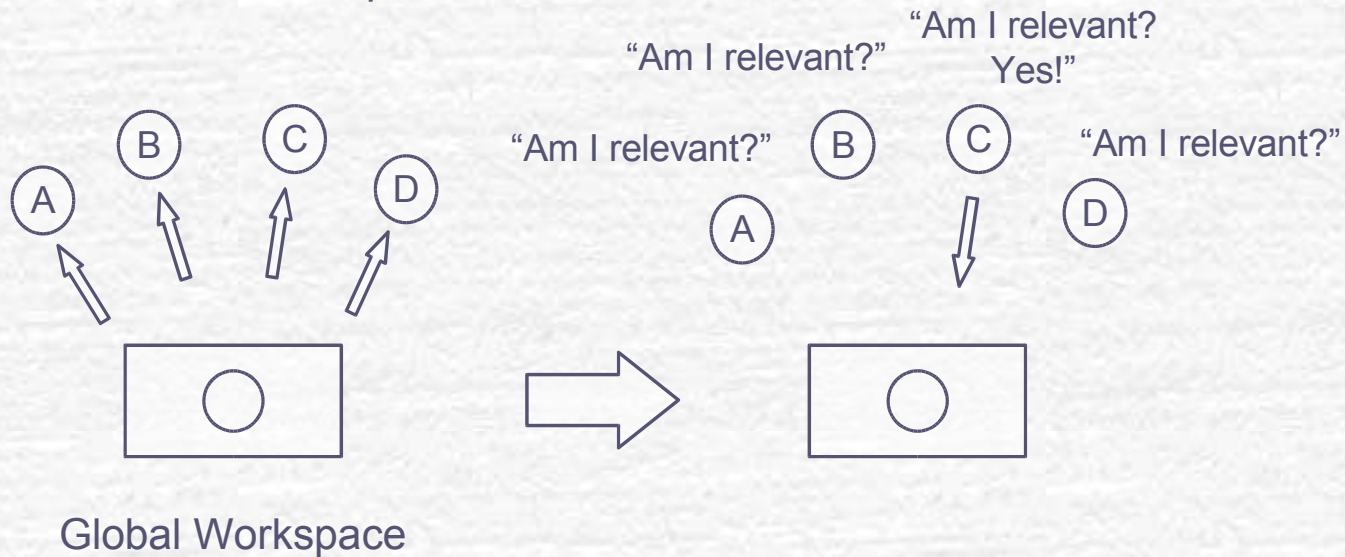
Central Processes

- This certainly looks computationally infeasible

# GWT and the Frame Problem (2)

- But global workspace architecture offers a parallel alternative

Parallel Unconscious Specialists



- In the context of an appropriate parallel architecture, the frame problem looks more manageable

# Analogical Reasoning (1)

- Analogical reasoning is informational unencapsulation in its purest form

“Analogical reasoning depends precisely upon the transfer of information among cognitive domains previously assumed to be irrelevant ” (Fodor)

- Computational models of analogical reasoning distinguish between
  - *retrieval* – the process of finding a potential analogue in long-term memory for a representation already in working memory – and
  - *mapping* – the subsequent process of finding correspondences between the two

# Analogical Reasoning (2)

- Retrieval is the locus of the frame problem in analogical reasoning
- The most psychologically plausible computational model is currently LISA (Hummel & Holyoak), which mixes serial and parallel computation, and also fits a global workspace architecture very closely



# Reinventing Modularity

- The global workspace architecture can be appropriated by any of the modular theories of mind
- It potentially supplies the means of transcending modular boundaries required to realise human-level, flexible, creatively intelligent cognition
- Its application to the frame problem in general, and to analogical reasoning in particular, is an example of this

# Conclusion

*Is there a fundamental link between cognition and consciousness?*

There is plentiful support for an affirmative reply

So perhaps an understanding of cognition has to go hand-in-hand with an understanding of consciousness

# References

Shanahan, MP. & Baars, B.J. (2005). Applying Global Workspace Theory to the Frame Problem, *Cognition* 98 (2), 157–176.

Shanahan, MP. (2005). Global Access, Embodiment, and the Conscious Subject, *Journal of Consciousness Studies* 12 (12), 46–66.

Shanahan, MP. (2006). A Cognitive Architecture that Combines Inner Rehearsal with a Global Workspace, *Consciousness and Cognition*, in press.

**Towards A Theory of Vision:  
Requirements for a robot with human child-like or  
crow-like visual and learning capabilities.**

**Aaron Sloman**

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

School of Computer Science, The University of Birmingham

**With help from colleagues on the CoSy project**

<http://www.cs.bham.ac.uk/research/projects/cosy/>

(Maria Staudte at DFKI kindly commented on an early draft)

And others, including Jackie Chappell, Biosciences, Birmingham.

These slides will be accessible from from symposium web site. See also

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

<http://www.cs.bham.ac.uk/research/projects/cosy/papers/>



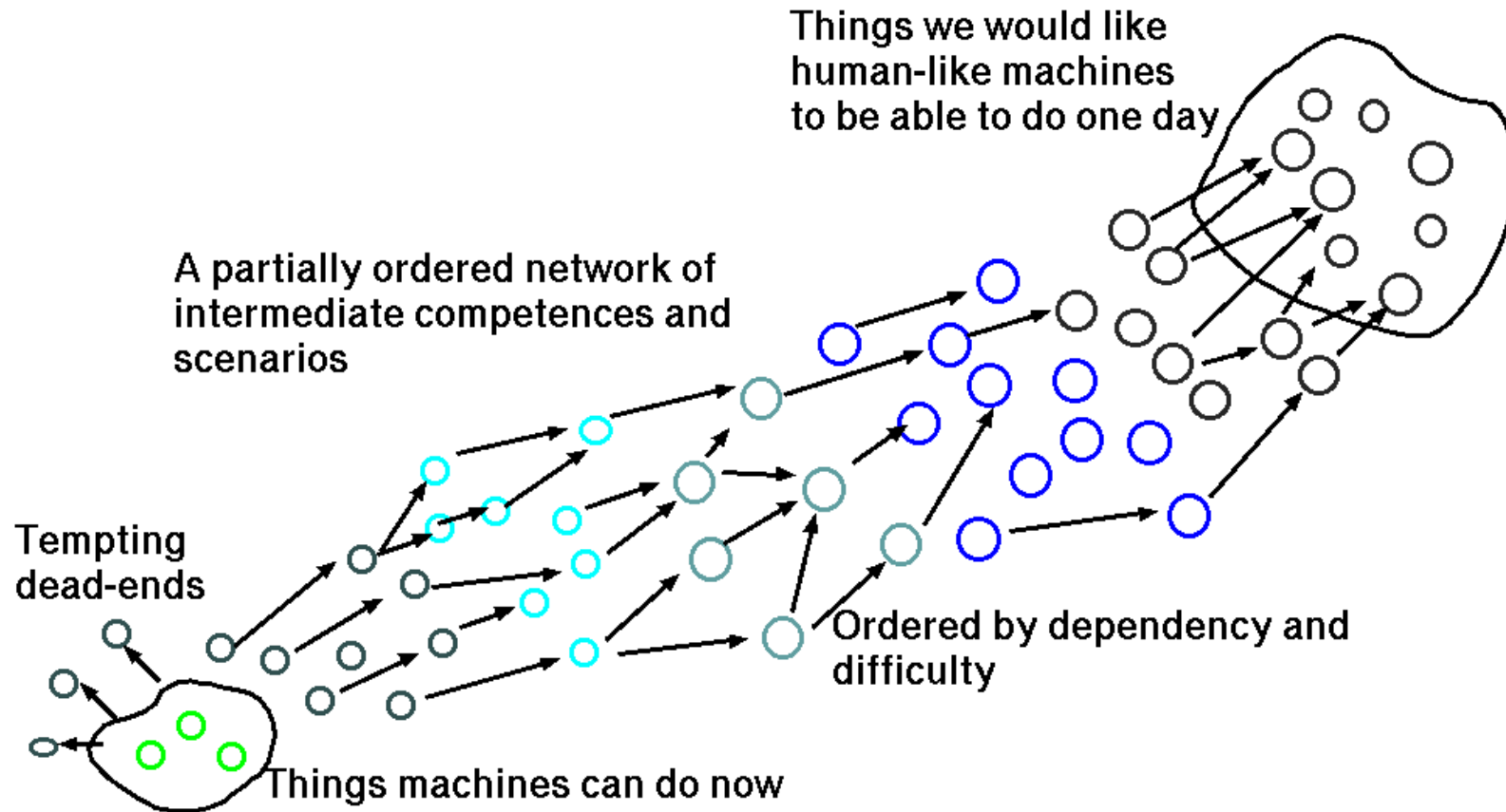
## 1 Some background

We start with some background to put the research in vision into the larger context of an attempt to specify **requirements** for future robots with human-like capabilities.

This is of interest both because it may help with the practical goal of designing more intelligent and more useful, flexible, companionable robots and also because it may help us understand human beings better by understanding the requirements for designs that explain what we can do.

The latter is my personal aim: I am more a biologist (and philosopher) than an engineer, but the engineering methodology is required for doing biology, psychology and philosophy well.

# Steps towards a research roadmap



**Forward chaining research asks: how can we improve what we have already done?**

**Backward chaining research asks: what is needed to achieve our long term goals?**

**See the introduction to GC5 in the booklet and on the web: researchers don't put nearly enough effort into analysing requirements.**

**Many of the hardest tasks are concerned with seeing 3D motion and affordances**

# Doing science

---

We need to move beyond 'Here's my architecture'.

For real scientific knowledge we need to have a theory about the space of possible designs and how design options relate to task requirements.

This leads to the idea of studying relations between

- **design space** (space of possible designs), and
- **niche space** (space of possible sets of requirements).

# REQUIREMENTS FOR ANIMAL & ROBOT VISION

---

**Vision is a process involving multiple concurrent simulations at different levels of abstraction in (partial) registration with one another and sometimes (when appropriate) in registration with visual sensory data and/or motor signals.**

**Max Clowes: Vision is controlled hallucination.**

**We add: multi-level controlled hallucination.**

The theory has different facets, which link up with many different phenomena of everyday life as well as experimental data, and with a host of problems in philosophy, psychology (including developmental and clinical psychology), neuroscience, biology and AI (including robotics).

It raises new questions for AI, psychology, neuroscience and others.

Example: watch this video of child playing with a toy train set.

[http://www.cs.bham.ac.uk/~axs/fig/josh34\\_0096.mpg](http://www.cs.bham.ac.uk/~axs/fig/josh34_0096.mpg)

# **Perceiving structures vs perceiving affordances**

## **Structures**

**things that exist, and have relationships, with parts that exist and have relationships**

## **Affordances (positive and negative)**

**processes that could or could not (sometimes conditionally could or could not) be made to exist by the agent, with particular consequences for the perceiver's goals, preferences, likes, dislikes, etc.:**

**modal, as opposed to categorical, types of perception.**

- **Betty looks at a piece of wire and (maybe??) sees the possibility of a hook, with a collection of intervening states and processes involving future possible actions by Betty.**
- **The child looks at two parts of a toy train remembers the possibility of joining them, but fails to see the precise affordances and is mystified and frustrated: presumably he sees parts and structural relationships because he can grasp and manipulate them in many ways. But he appears not to see some affordances.**
- **Seeing affordances seems to be related to being able to run simulations of unseen but possible processes in registration with the scene.**

**How specialised are the innate mechanisms underlying the abilities to learn categories, perceive structures, understand affordances, especially structure-based affordances.**

**Beware the tabula rasa trap: millions of years of evolution were not wasted!**

# We should not consider only human competence

---

Humans are a result of billions of years of evolution producing many different solutions to the problems of coping with a complex environment.

## Betty the hook-making New Caledonian crow.

Give to google: betty crow hook:  
You'll find a link to the Oxford Zoology lab, with videos of Betty making hooks in different ways.

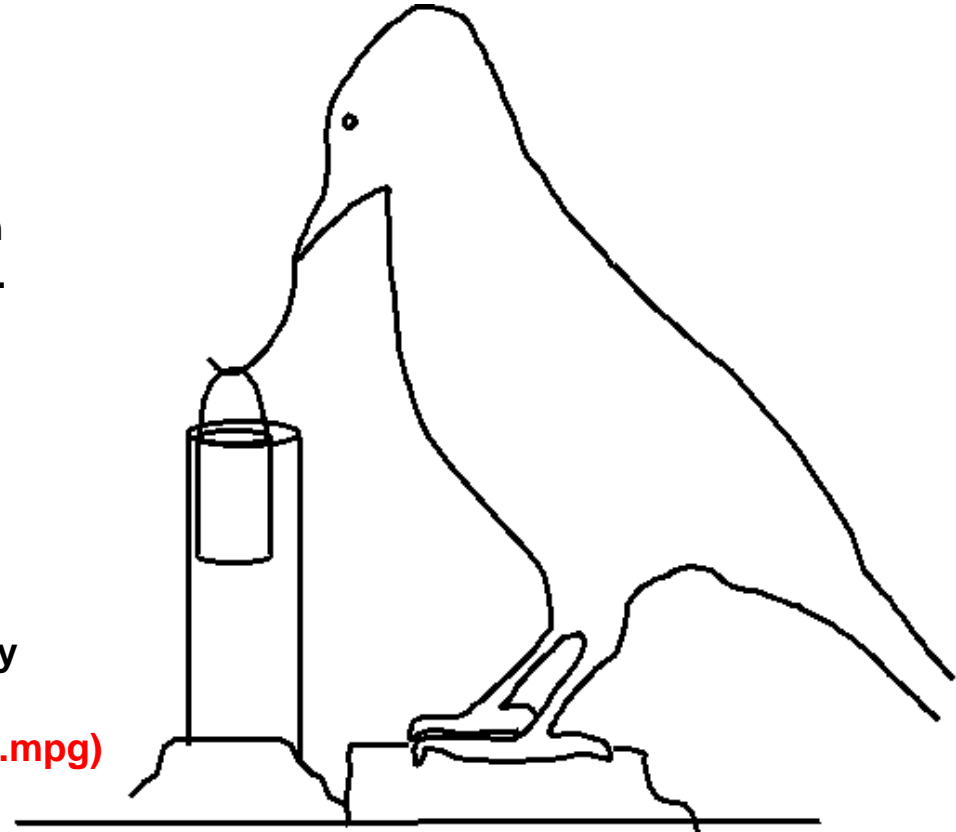
She **appears** to have a deep understanding of structure, process and causation.

See the video here:

<http://news.bbc.co.uk/1/hi/sci/tech/2178920.stm>

Contrast the 18 month old child attempting unsuccessfully to join two parts of a toy train by bringing two rings together

([http://www.cs.bham.ac.uk/~axs/fig/josh34\\_0096.mpg](http://www.cs.bham.ac.uk/~axs/fig/josh34_0096.mpg))



Does Betty see the possibility of making a hook before she makes it?

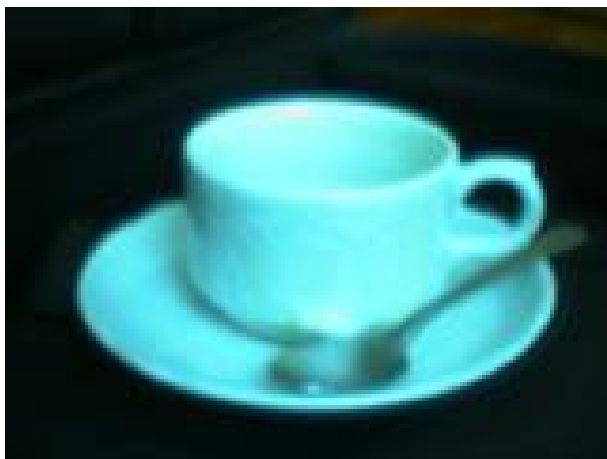
She seems to. How?

# Some tasks for a crow-challenging robot?

## UPDATING THE BLOCKS WORLD

Using a two-finger gripper, what actions can get

from this:



to this:



and back again?

### Or with saucer upside down?

Unfortunately even perceiving and representing the initial or final state (e.g. as something to copy) seems to be far beyond the capabilities of current AI vision systems, let alone thinking about possible actions to transform one to the other – e.g. the angle of approach required to grip cup or saucer or spoon in a particular location, e.g. the left-most point of the saucer's rim, or the tip of the teaspoon's bowl.

# Some tasks for a crow-challenging robot? (2)

Consider how, prior to the action, the agent has to

- identify parts of objects, or parts of parts, e.g. the edge of the handle, or the far edge of the handle or a certain portion of the edge of the saucer
- see and understand their shapes and relationships
- identify possible actions: grasping **this thing here** from **this direction**  
**Could such deliberative premeditation use the action schema (operator) with approximate, qualitative parameters instead of the more definite actual parameters that would be used if the action were performed?**
- think about various effects of actions, including changing effects of continuous processes

NOTE: there are problems here partly analogous to problems of reference and identification in language, except that the mode of reference is not linguistic and what is referred to typically cannot be expressed in language because it is anchored in non-shared structures and processes.

(Internal 'attention' processes are partly like external pointing processes: virtual fingers. )



# Compare Freddy the 1973 Edinburgh Robot

Some people might say that apart from wondrous advances in mechanical and electronic engineering there has been little increase in sophistication since the time of Freddy, the 'Scottish' Robot, built in Edinburgh around 1972-3.

Freddy II could assemble a toy car from the components (body, two axles, two wheels) shown. They did not need to be laid out neatly as in the picture.

However, Freddy had many limitations arising out of the technology of the time.

E.g. Freddy could not simultaneously see and act: partly because visual processing was extremely slow.

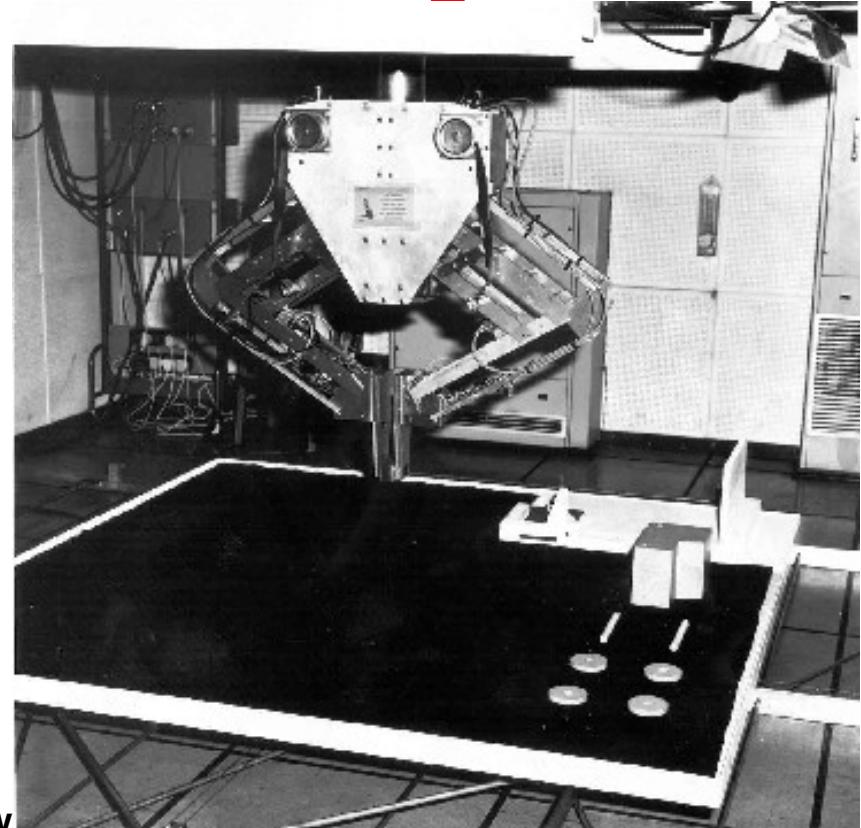
Imagine using a computer with 128Kbytes RAM for a robot now.

There is more information on Freddy here

<http://www.ipab.informatics.ed.ac.uk/IAS.html>

<http://www-robotics.cs.umass.edu/pop/VAP.html>

In order to understand the limitations of robots built so far, we need to understand much better exactly what animals do: we have to look at animals (including humans) with the eyes of (software) engineers.



# Perception of shape is not shape-reconstruction

---

What sort of 3-D interpretation is required depends on what it is to be used for.

Shape perception in computers is often demonstrated by giving the machine one or more images, from which it constructs a point-by-point 3-D model of the visible surfaces of objects in the scene (sometimes using laser range-finders).

This achievement is then demonstrated by projecting images of the scene from new viewpoints.

But there is no evidence that any animal can do that and very few humans (e.g. some artists) can produce accurate pictures of viewed objects using a new viewpoint, whereas many graphics engines do it.

Human/animal understanding of shape, including having information relevant to action and prediction, is very different from having a point by point 3-D model

The point of perception is not making images: the results must be useful for action – e.g. building nests from twigs, peeling and dismembering food in order to get at edible parts, escaping from a predator, making a tool, using a tool.

A 'percept' constructed by the perceiver needs to include information about what is happening, what could happen and what obstructions there are to various kinds of happening (positive and negative affordances).

These happenings are of many different kinds, so different kinds of information must be synthesised from sensory information (influenced by prior knowledge, prior ontologies, prior goals).

## 2 A vision system has to be part of a larger architecture

In my 'Talks' directory

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

there are several presentations on architectures, and on a conceptual framework called **CogAff** for thinking about types of architectures supporting multiple kinds of functionality operating in parallel.

As an example of the application of the CogAff framework, it is conjectured that the human architecture has a kind of complexity illustrated very sketchily in the next slide.

There's no time to explain this now, but many of the features of vision that are mentioned in the rest of this presentation depend on the fact that vision has multiple functions because visual mechanisms are connected to many different subsystems in the architecture.

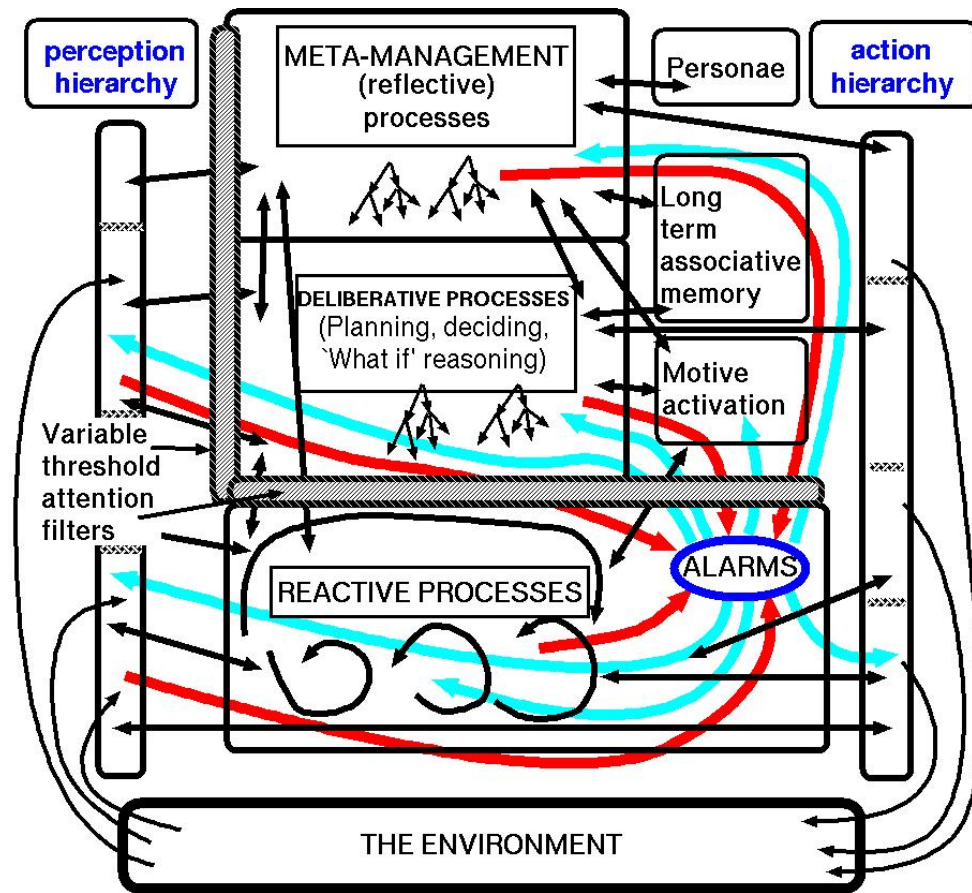
Because of this, evolution produced multilayered vision systems (and multilayered action systems) that, at least in humans, do not have a fixed structure but can be extended during learning and development, including learning to read, learning to understand abstract diagrams, and learning to see new kinds of affordances, e.g. the properties of hooks.

# A hypothetical Human-like architecture: H-CogAff (See <http://www.cs.bham.ac.uk/research/cogaff/>)

This is an instance (or specialised sub-class) of the architectures covered by a generic schema called “CogAff”.

Many required sub-systems are not shown.

Different kinds of process simulation may go on in different parts of the architecture – some very old and widely shared, some relatively new and found in very few species.



# A Shift of View

---

- For many years I assumed (like many other people) that if we could understand perception of static scenes we could later deal with motion.
- I also thought (as explained below) that perception of a static scene involved forming (static) descriptions of its contents (at different levels of abstraction), and that a theory of perception of motion might later be derived from that.

# A Shift of View

---

- For many years I assumed (like many other people) that if we could understand perception of static scenes we could later deal with motion.
- I also thought (as explained below) that perception of a static scene involved forming (static) descriptions of its contents (at different levels of abstraction), and that a theory of perception of motion might later be derived from that.
- Then I learnt about Gibson's theory of affordances, which made it necessary to relate perception of static scenes to the *possibility of* (and constraints on) actions and their consequences that are not occurring, but might occur.
- For a while I assumed that a theory of perception of affordances could be tacked onto a theory of perception of structure by representing the perceived affordances as collections of something like condition-action rules associated with various parts of a scene.

# A Shift of View

---

- For many years I assumed (like many other people) that if we could understand perception of static scenes we could later deal with motion.
- I also thought (as explained below) that perception of a static scene involved forming (static) descriptions of its contents (at different levels of abstraction), and that a theory of perception of motion might later be derived from that.
- Then I learnt about Gibson's theory of affordances, which made it necessary to relate perception of static scenes to the *possibility of* (and constraints on) actions and their consequences that are not occurring, but might occur.
- For a while I assumed that a theory of perception of affordances could be tacked onto a theory of perception of structure by representing the perceived affordances as collections of something like condition-action rules associated with various parts of a scene.
- In retrospect it seems silly to have forgotten that vision evolved in organisms embedded in a dynamically changing environment – so its primary function must be not to discover **what exists** in the environment, but **what is happening** in the environment, including the perceiver's movements and actions.
- Add the observation that what is happening, and what is potentially important to an organism, is not a unique process, but a collection of processes at different levels of abstraction, e.g. a wave moving horizontally towards the shore and millions of molecules mostly moving roughly up and down in the same place.

### **3 Example: A child playing with a train-set on the floor**

The video mentioned above shows a child about three and a half years old doing things with a train set that surrounds him as he sits in the middle, turning this way and that, pointing at things behind him to answer questions, pushing the train through a tunnel, changing his position to replace a tree knocked down by the back of his head when he puts his head down to look through the tunnel.

[http://www.cs.bham.ac.uk/~axs/fig/josh\\_tunnel.mpg](http://www.cs.bham.ac.uk/~axs/fig/josh_tunnel.mpg) (5MB)

[http://www.cs.bham.ac.uk/~axs/fig/josh\\_tunnel\\_big.mpg](http://www.cs.bham.ac.uk/~axs/fig/josh_tunnel_big.mpg) (15MB)

(High resolution version.)

My claim that this child is running various simulations of things going on in the environment begs the question: ‘What kind of thing is a simulation?’

My provisional answer is that anything that is capable of usefully representing a process can be called a simulation for present purposes, even if it is a static structure accessed sequentially.

Later I’ll say more about what I do and do not mean.

NOTE: I am not making any use of Grush’s distinction between ‘emulation’ and ‘simulation’, though it is possible that it will turn out that what I mean by ‘simulation’ is what he means by ‘emulation’.



# Snapshots from tunnel video

A child playing with his train illustrates the theory.



- The child clearly knows what's going on in places he cannot see.
- He can point at and talk about something behind him that he cannot see.
- When he turns to continue playing with the train he knows which way to turn and roughly what to expect.
- When the train goes into the tunnel and part of it becomes invisible, he does not see the train as being truncated, and he expects the invisible bit to become visible as he goes on pushing.
- He sees the whole train as one thing while part of it is hidden in the tunnel.
- What is the role of **vision** in all of this? Frequently sampling the environment?

**Vision is concerned with what is and is not happening in the environment – that's potentially of relevance to the perceiver: ongoing situations and processes.**

# Tunnel vision

---

Think about the child playing with and talking about his toy train, with track, tunnel and other things on the floor around him.

How many different levels of abstraction occur in

- the processes he needs to perceive,
- the processes he needs to use in controlling his actions,
- the processes he needs to think about, explain, modify, predict, ...

Is there a sharp division between

- seeing geometric structures, relationships, changes and
- seeing causal and functional relations?

Is there a sharp distinction between what the child sees as **caused by his action**, and what he sees as merely **happening in the environment**?

Could the same mechanisms represent both?

Compare

- Movement of the truck he is holding and pushing
- Movement of the truck adjacent to the one he is pushing
- Movement of the trucks in the tunnel that cannot be seen
- Reappearance of the front of the train from the far end of the tunnel

**We return to perception of causal relations later.**

# Background

---

- **There are many views of the nature and function(s) of vision, including the following:**
  - Vision produces information about physical objects and their geometric and physical properties, relationships in the environment.  
(Marr and many others.)
  - Much recent work treats vision as a combination of recognition, classification and prediction – the latter sometimes used in tracking  
(often using classifications arbitrarily provided by a teacher, rather than being derived from the perceiver's needs and the environment).
  - Vision controls behaviour (Obviously true?)
  - Behaviour controls perception, including vision. (W.T.Powers)
  - Vision is unconscious inference (Helmholtz)
  - Vision is controlled hallucination (Max Clowes) **Pretty close**
  - Grush on Emulation theory of representation (BBS 2004)
- **I'll try to present phenomena that require a richer deeper theory.**

It will be evident that the new theory uses many of the above ideas, and assembles them with some new details. Some of the ideas are criticised.

# Relationship with CoSy project

---

**A change of view came while I was working on the CoSy project**

<http://www.cs.bham.ac.uk/research/projects/cosy/>

I have been thinking about many of the problems for many years, but what made things click into place recently was examining very closely the perceptual and representational requirements for a robot manipulating 3-D objects on a table-top, e.g. watching a hand picking up a cup, or assembling a meccano model.

**Try thinking about it yourself!**

**Using one or two hands, perform simple, everyday actions on cups, spoons, scissors, paper, string, a handkerchief, nuts and bolts, tin-openers, your food, a sweater you put on or remove ....**

**and watch very, very closely.**

**How can your brain represent the information you use, including**

- all the things and processes you see, as complex 3-D objects move while changing their shapes and mutual relationships,
- what you anticipate,
- your recollection of what just happened,
- your thoughts about what would have happened if you, or someone else, had done something different?

**PERHAPS YOU WILL INVENT THE SAME THEORY.**

# The theory is not totally new

---

## There are many precursors of different kinds:

Some old philosophical theories of minds as idea-manipulators.

Kant's *Critique of Pure Reason* (1780) (Including his theory of mathematical knowledge)

Helmholtz: perception is unconscious inference

Kenneth Craik in 1943 (animals use predictive models)

Ulric Neisser and others (1960s): theories of vision as analysis by synthesis, and hierarchical synthesis.

Karl Popper (our hypotheses can die in our stead)

William T Powers: Behaviour controls perception.

Lots of control engineering using 'predictive' models.

Max Clowes: Vision is controlled hallucination

David Hogg's work on perceiving a walking person (1983)

My own work in the 1970s on multi-level perception and visual reasoning

Work by Tsotsos on motion perception.

Roger Shepard and others on mental rotation tasks.

Steve Kosslyn on imagery

Phil Johnson-Laird on reasoning with mental models

JJ Gibson on perceiving affordances (and his earlier ideas about 'perceptual systems')

Minsky's *Society of Mind* and other work.

Arnold Trehub: (1991) *The Cognitive Brain*

Alain Berthoz (2000) *The Brain's sense of movement*,

Murray Shanahan AISB 2005

Philippe Rochat, 2001, *The Infant's World*,

R. Grush, 2004, The emulation theory of representation: ... BBS, 27,

**And probably more: but does any combine all the elements proposed here?**

#### **4 From structures (in the Popeye system) to processes**

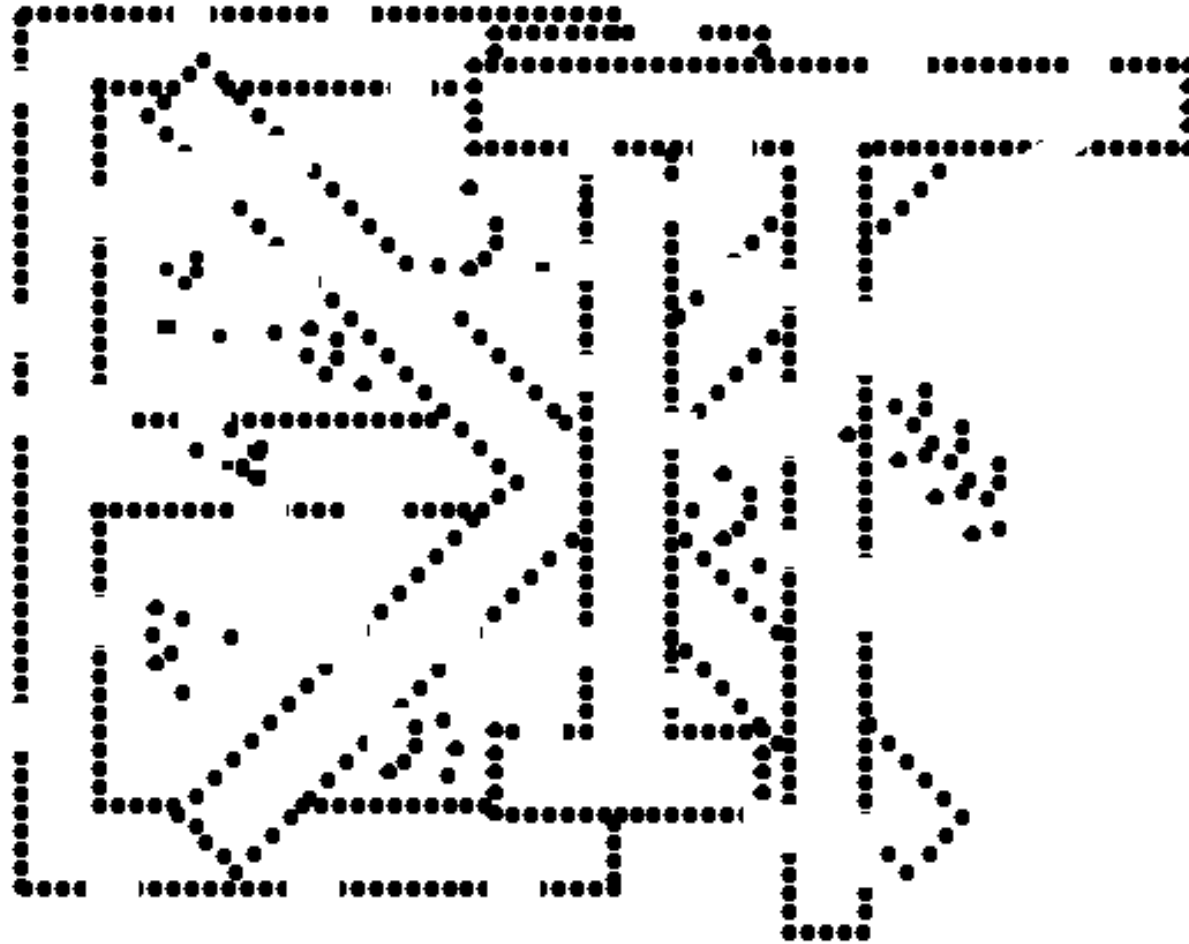
**About 30 years ago a project at Sussex University explored some aspects of the theory that perception of complex and noisy structures could be facilitated by a visual architecture in which processes at different levels of abstraction, concerned with different ontologies, ran concurrently with a mixture of bottom up and top down control, including top-down control of attention.**

**But there was nothing in this about perceiving **processes** at different levels of abstraction, as is proposed here. Yet some of the ideas remain relevant.**

# How do we process noisy pictures? (1)

---

DO YOU SEE A WORD?



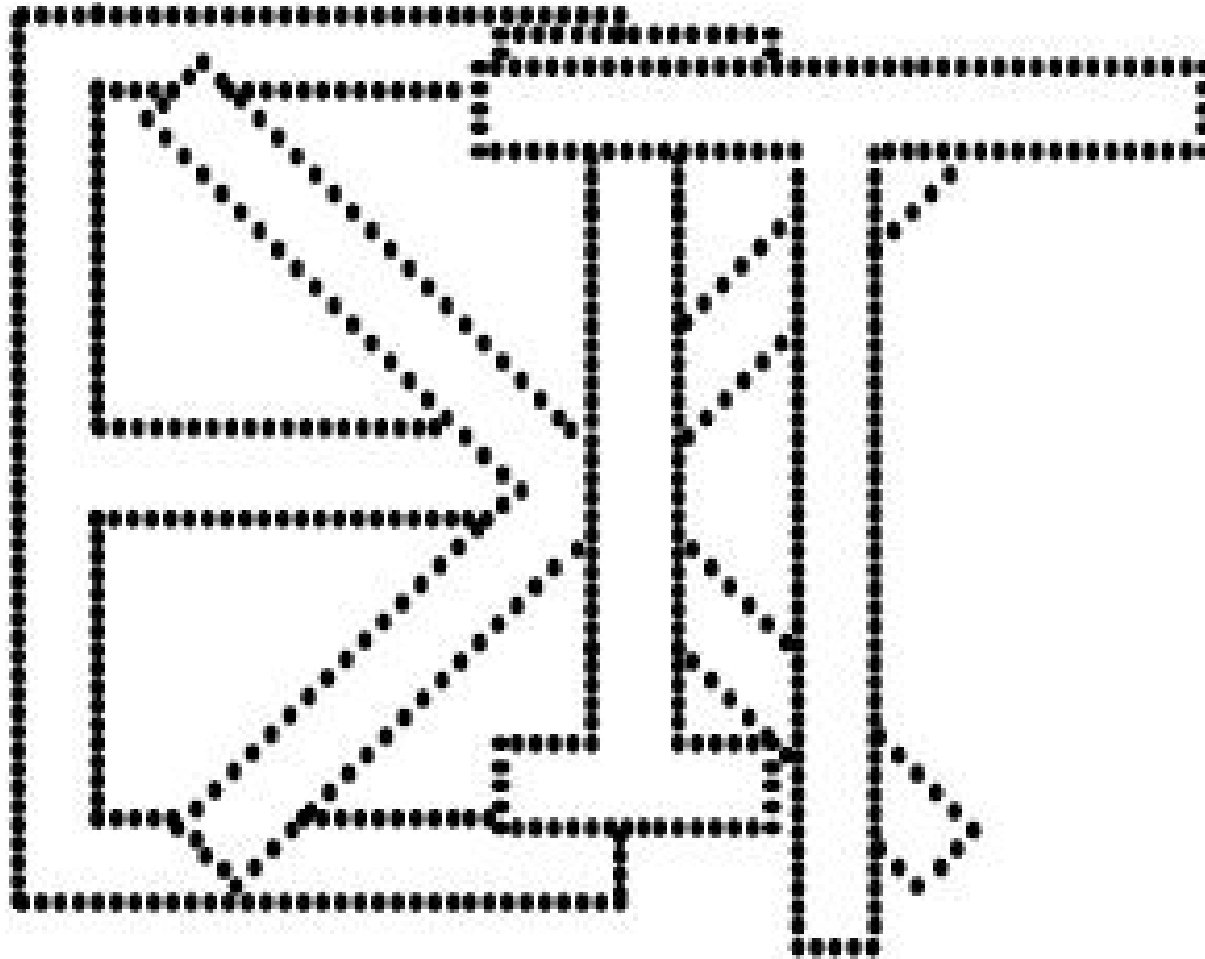
# How do we process noisy pictures? (2)



# How do we process noisy pictures? (3)

---

DO YOU SEE A WORD?



# Multiple levels of structure perceived in parallel

Old conjecture: We process different layers of interpretation in parallel.

Obvious for language. What about vision?

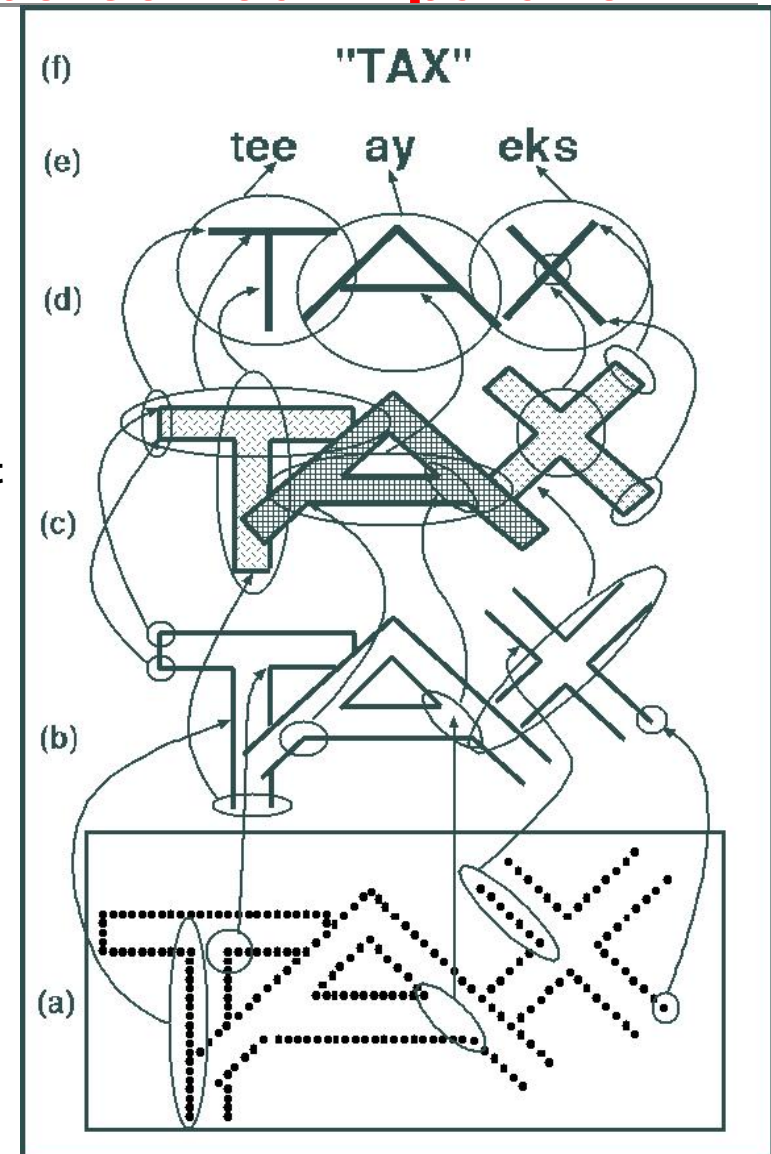
Concurrently processing bottom-up and top-down helps constrain search. There are several ontologies involved, with different classes of structures, and mappings between them – so the different levels are in ‘partial registration’.

- At the lowest level the ontology may include dots, dot clusters, relations between dots, relations between clusters. All larger structures are **agglomerations** of simpler structures.
- Higher levels are more abstract – besides **grouping** (agglomeration) there is also **interpretation**, i.e. mapping to a new ontology.
- Concurrent perception at different levels can constrain search dramatically (POPEYE 1978) **(This could use a collection of neural nets.)**
- Reading text would involve even more layers of abstraction: mapping to morphology, syntax, semantics, world knowledge

From *The Computer Revolution in Philosophy* (1978)

<http://www.cs.bham.ac.uk/research/cogaff/crp/chap9.html>

Replace all that with concurrent multi-level processes – using different process-ontologies.



## Seeing a cup at multiple levels of abstraction

---



**Many levels of structure and of affordances.**

**The identification of 'objects' is not fixed by the environment: e.g. thinking about different places to grasp.**

**But even that is not all that goes on in vision**

# From Structures to Processes

---

In the light of earlier observations we can replace the idea that

1. seeing involves multi-level structures in partial registration using different ontologies,

with the claim that

2. seeing involves multi-level process-simulations in partial registration using different ontologies, with rich (but changing) structural relations between levels.

## NOTE:

After developing these ideas I found that Philippe Rochat's book *The Infant's World* claims on pages 103-7 that there is evidence that even at 4 months infants are capable of 'dynamic imagery', used to predict trajectories of objects when they pass out of view.

# The walking man

---

- Shortly after the work on Popeye was done, David Hogg was a PhD student in the same department working on motion perception.

*D. Hogg. Model-based vision: A program to see a walking person. **Image and Vision Computing**, 1(1):5–20, 1983.*

- His well known ‘walking man’ system was an early example of what I am now talking about: his model-based interpretation of a video of a walking man amounted to a simulation of a walker, partly controlled by the changing image data, and partly controlled by the dynamics of the model.
- Despite being his supervisor I did not appreciate the full significance of that work till now.

I think he also did not see the full significance of what he had done: he described the system as showing how to use a model to interpret an image, rather than claiming to show how to interpret a sequence of images as representing **a process**.

- **Making Popeye see dotty images of moving overlapping laminas forming different words at different times would have been a very different task ....**

# **How to see a static scene as a process**

---

**If all this is right, our ability to see processes is used even when we look at a static scene:**

**it's just that then the process is one in which nothing changes.**

- **But if something started changing we would see it, using the same mechanisms as were previously perceiving the static configuration.**
- **A static scene is just a special kind of process, in which nothing changes.**
- **Whether the things change or not the system has to be prepared for many possibilities.**
- **Thus perception of a static structure already involves perception of possibilities for motion (mostly latent: the simulation capabilities may be turned on if motion occurs, and left dormant otherwise).**

**This could be seen as a minimal notion of affordance.**

# The importance of concurrency

---

Besides emphasising the importance of **processes** as being the content of what is perceived (i.e. not just static structures), we are also emphasising the importance of **concurrency**, namely the perception as involving multiple perceived processes, some at the same level of abstraction, some at different levels of abstraction

- Perceived concurrency is involved in various human and animal activities involving two or more individuals engaged in fighting, dancing, mating, playing games, performing music, etc.
- Doing this well implies a need to be able to keep track of (partly by running simulations?) the actions of others at the same time as planning and performing one's own actions.
- What are the evolutionary precursors of this, e.g. in hunting animals and prey of hunting animals, including parents defending young from predators?
- Concurrency is also important in social learning, since many social interactions are concurrent rather than simply based on turn-taking: e.g. dancing, old friends embracing, lifting or pushing a heavy article, and mating.
- **Conjecture:** our architecture evolved to support at least three sorts of concurrency:
  - Perceiving multiple concurrent external processes
  - Representing the same process at different levels of abstraction
  - Different concurrent actions in an individual, such as walking (including posture control), working out where to walk, discussing philosophy with a companion, using different parts of the information-processing architecture.

# Liberation from the here and now

---

## CONJECTURE

The same mechanisms (or similar mechanisms produced using evolution's 'duplicate then differentiate' strategy) can be used

(a) Without using sensor-specific or motor-specific representations

(b) In relation to things that are not currently perceived

– Past

– Remote

– Future

**Contrast:**

**multi-modal integration vs a-modal abstraction**

**Contrast:**

**Learning about (intra-somatic) sensorimotor contingences**

**VS**

**Learning about objective (extra-somatic) condition-consequence contingencies**



# **We are not insects**

---

The vast majority of animals (microbes, crustaceans, fish, reptiles....) may be able to get by with much less powerful and flexible perceptual systems.

They may always be involved in control of **current** actions (including quite sophisticated dynamical systems with predictive control mechanisms – using feedforward loops – e.g. flying insects).

But what humans and a few other species goes far beyond that: and much research on vision and robotics, including some research in neuroscience(?), does not take account of the requirements – e.g. to be able to remember what you did, to understand what went wrong, to think about what someone else may do, to plan several steps ahead....

**Theories of insect intelligence may not be adequate for chimps, cheetahs and crows, let alone humans.**

**Note: all this may be unfair to some insects.**

**(And what about the octopus?)**

# Simulation capability exceeds behavioural capability

If human brains (and perhaps others) can construct and run simulations of processes of many kinds, there is no need for each one to be **closely** related **either** to the specific motor system that would be used to produce such processes **or** to the sensory systems that would be used to perceive such a process.

After all, we can perceive many processes we cannot produce, e.g. waterfalls – and we shall later give examples of perceiving and thinking about ‘vicarious affordances’, i.e. affordances for others.

So we have an ability to experience and appreciate processes that are richer and more complex than anything we can produce using our own bodies.

- **Evolution apparently ‘discovered’ the benefits of structural and causal disconnection between representation and thing represented, long ago (in a subset of animals only?): can we replicate this in our designs?**
- **Compare**
  - the ability of a prey animal to think about what a predator might do
  - the ability of a composer to think up a multi-performer composition, and specify it in a musical score.
  - the ability of a general to prepare orders for various concurrently active platoons.
  - the ability of some programmers to design, implement, and debug programs involving concurrent processes (e.g. operating systems).

# So....

---

**Many current theories of embodied cognition ignore the extent to which evolution discovered the power of disembodied cognition for a small subset of species**

Infant humans seem to have minds with learning and developmental capabilities that can use a variety of different bodily forms available from infancy to achieve a common adult humanity.

For example, consider the thalidomide babies born limbless in the 1960s, and the artist Alison Lapper, celebrated here

<http://www.ldaf.org/pages/dail/dailarticles.htm#lapper>

<http://www.mymultiplesclerosis.co.uk/misc/alisonlapper.html>

She can clearly see many structures and processes that can be seen by people with normal arms and legs.

# **Reminding the audience: relevant things you probably know**

---

**There are many aspects of our everyday experience that people may or may not notice that seem to involve this ability to run some sort of simulation of environmental processes.**

**So this is not really a theory that's new to you, even if you previously never thought about it.**

- E.g. when you see something moving behind an opaque object you don't see the moving object as being truncated – you see it as having a hidden portion that continues to move (like the child in the video pushing his train into a tunnel), and typically you know roughly where the hidden parts are as the motion continues (though of course stage conjurers can fool us because we are not infallible).**
- Many cartoons and jokes depend on our ability to run simulations derived from the information presented, e.g. pictorially or verbally.**
- Doodles depend on this ability too. In fact many/most(?) forms of visual art do.**

**Some cartoons showing 'snapshots' of extended processes follow. Some project into both future and past, some only one or the other.**

## 5 Cartoons and miming

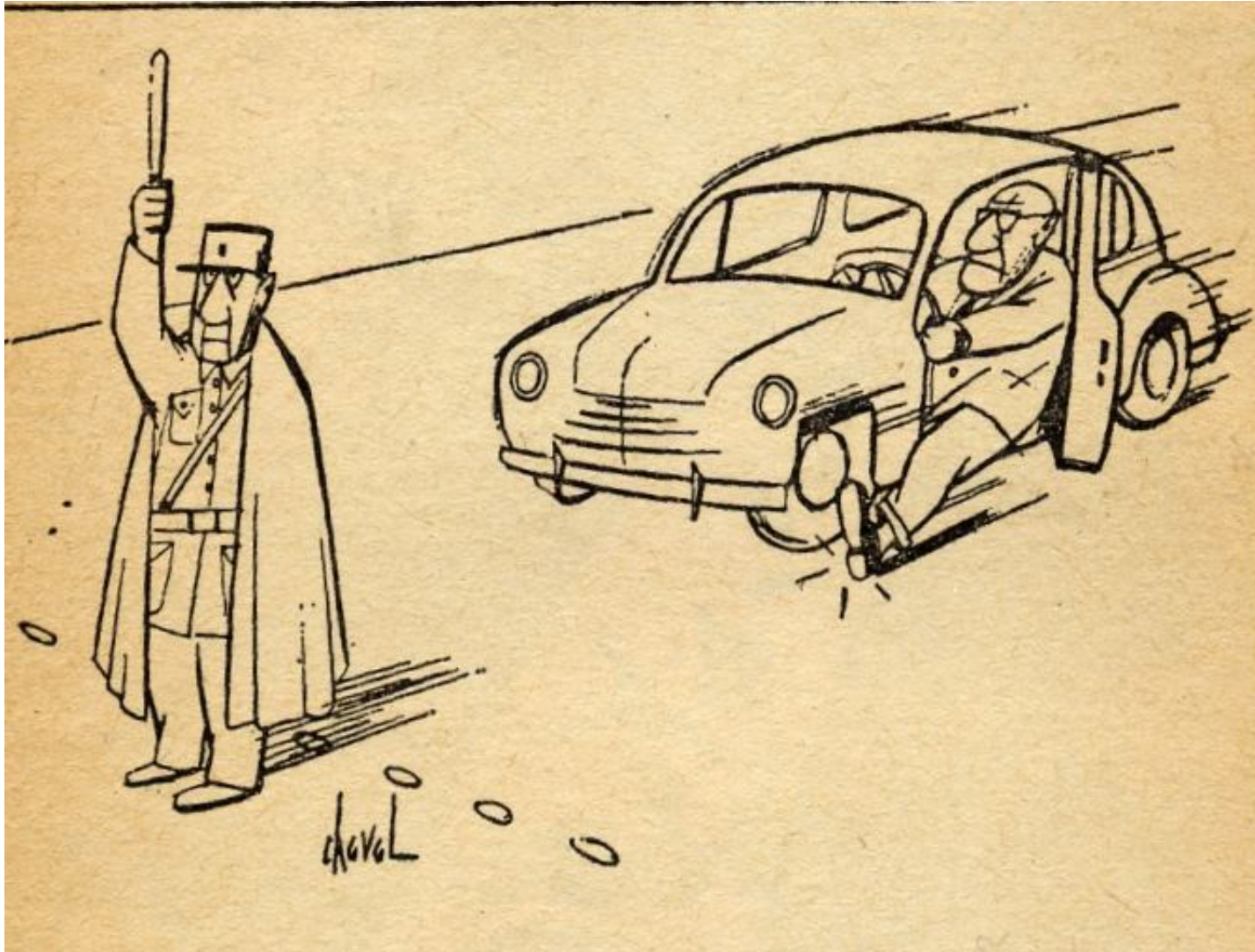
# 'French Cartoons' Published 1955

Ed William Cole and Douglas McKee, : Panther Books

---

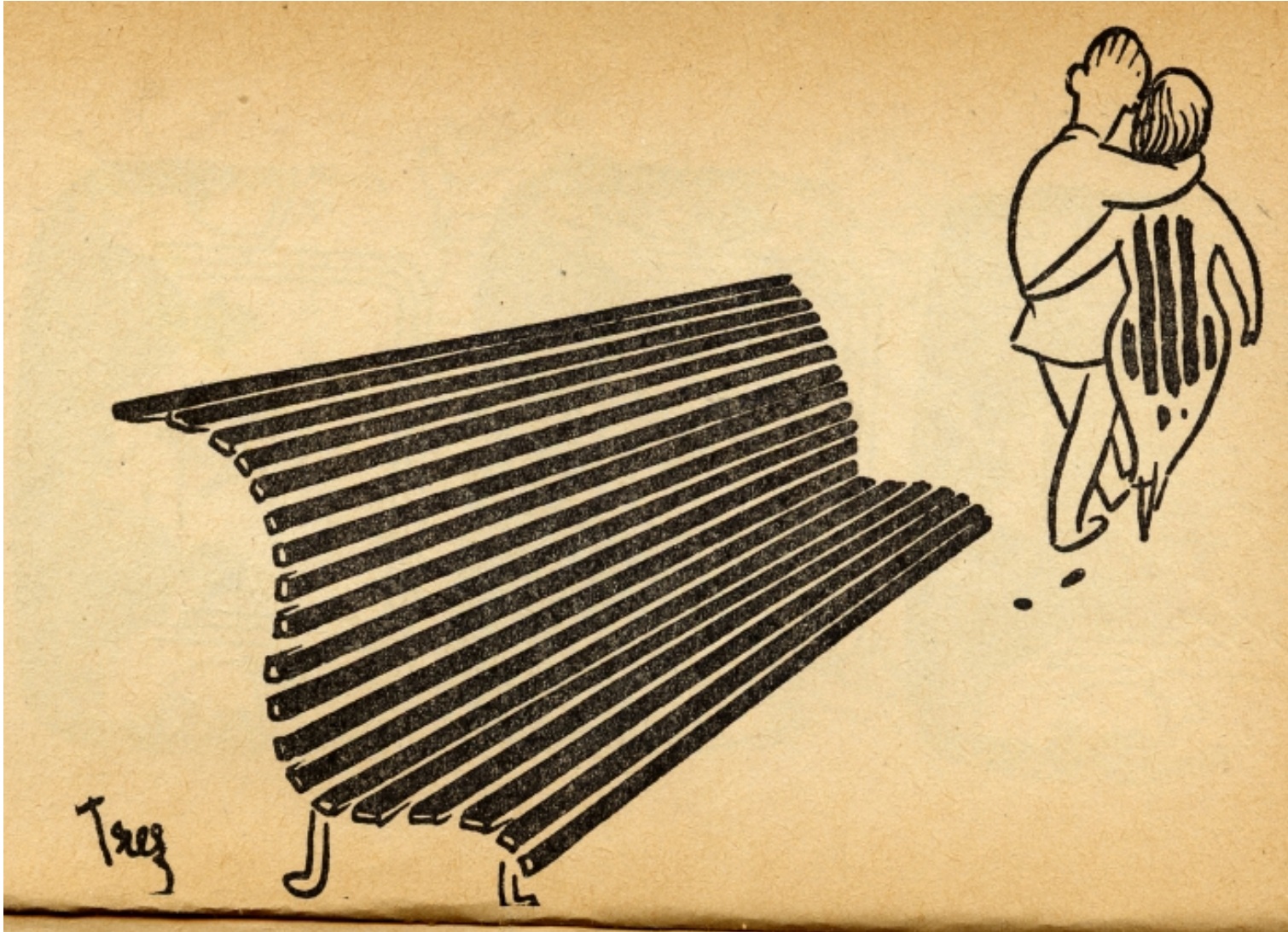
**When you look at the cartoons that follow, what past and future processes come to mind and how do they relate to details of the scene?**

# Mostly future



Another kind of footbrake ???

# Mostly past



Understanding the picture involves 'running a simulation' but at a high level of abstraction with many details of the previous history left out.



# We produce joke actions also

Using a tennis ball and badminton shuttlecock to simulate eating an ice-cream – he never actually licked the ball.

We often use external simulations, including gestures, diagrams, working models. However most of our examples below will be cases of purely internal simulation.

Perhaps a major function of play in young mammals is developing simulation capabilities through learning about different things to simulate (as opposed to developing motor skills, muscles, etc.)



Evolution (and processes in individual development) somehow gave us the ability to make use of either **internal** or **external** objects, when running simulations. My 1971 IJCAI paper claimed that reasoning with diagrams is essentially the same thing whether done **on paper** or **in the mind**.

**Brain mechanisms for this are still waiting to be discovered.**

(See the interesting discussions in BBS paper and commentary by R.Grush, 2004 – found after much of this had been written).

# Sensorimotor vs Condition consequence contingencies

---

Insects may be restricted to learning conditional probabilities relating total sensory and motor signal-sets.

In some cases that would be combinatorially explosive – e.g. all the ways of perceiving grasping, whether done with mouth, or left hand or right hand, or two hands holding an object.

An organism that can abstract from all the intra-somatic sensorimotor details and represent extra-somatic relationships between surfaces and their consequences (e.g. if something moves) **independently** of how movements are produced or sensed has a great advantage in generality and economy.

That includes being able to perceive and think about actions done by others: perceiving vicarious affordances.

**So mirror neurons should have been called ‘abstraction neurons’.**

**CONJECTURE:** this ability to represent objective structures and processes, a kind of disembodiment, was a major evolutionary development.

**It’s not clear whether that is present at birth – though much else is.**

## **6 Perceiving causation**

**Two kinds of causation: Humean (probabilistic, evidence based) and Kantian (deterministic: based on hypothesised structures)**

# Perceiving causation

---

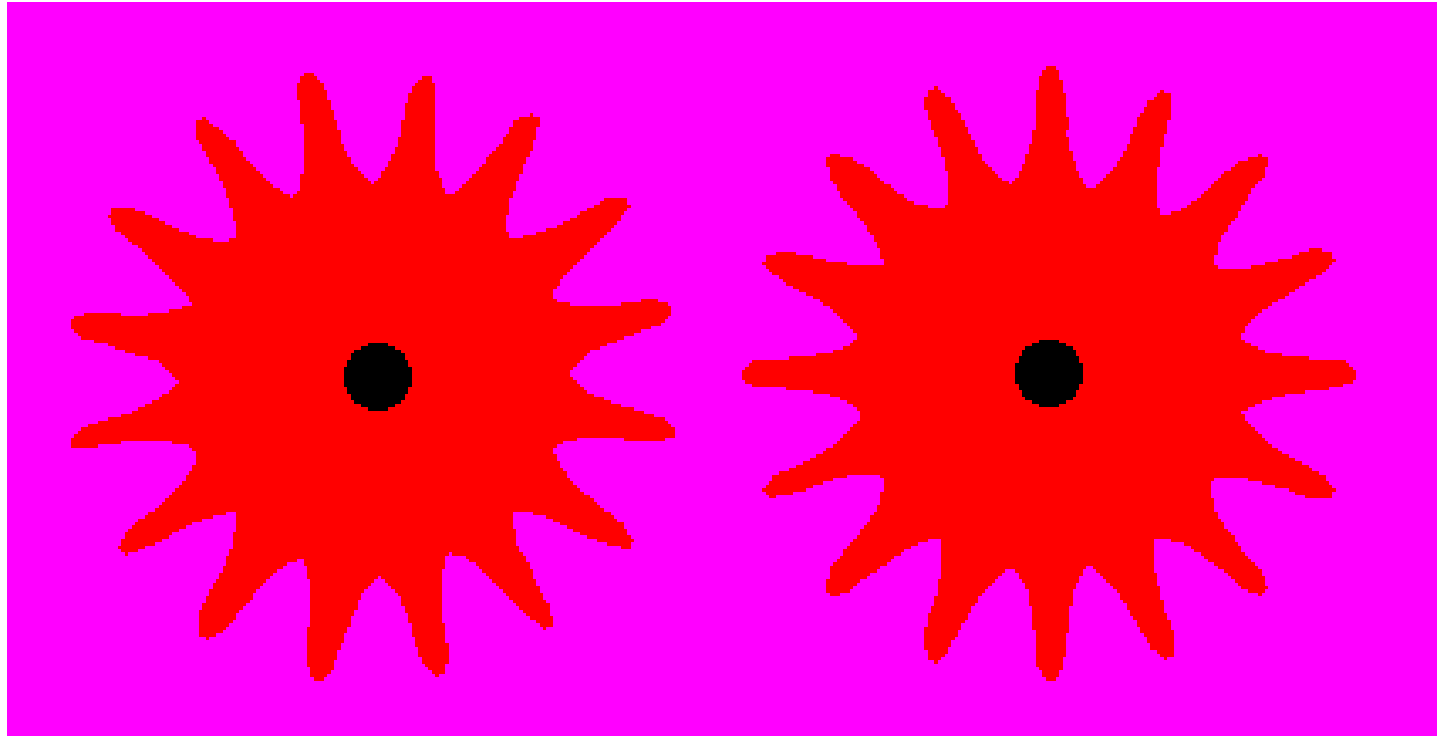
**Our ability to perceive moving structures, and our meta-level ability to think about what we perceive, is intimately bound up with perception of causation and affordances.**

**In some cases the causal relations are inherent in what is seen, whereas in others they involve invisible structures and processes: but the same key idea is used in both cases.**

**Illustrations follow.**

# Invisible, Humean, causation – mere correlation

Two gear wheels attached to a box with hidden contents.  
Here we do not perceive causation: we infer it from statistics.



Can you tell by looking what will happen to one wheel if you rotate the other about its central axis?

You can tell by experimenting: you may or may not discover a correlation.

Compare experiments reported by Alison Gopnik in her invited talk at IJCAI'05, Edinburgh July 2005

# Visible, intelligible, Kantian, causation

Two more gear wheels:

Here you (and some children) can tell 'by looking' how rotation of one wheel will affect the other.

NB The simulation that you do makes use of not just perceived shape, but also **unperceived constraints**: rigidity and impenetrability. These constraints need to be part of the

perceiver's ontology and integrated into the simulations, for the simulation to be deterministic.

Visible structure does not determine all the constraints: we also have to learn about the nature of materials, to see what is happening, and understand causation.

**We need to explain how brains and computers can set up and run simulations involving multiple concurrent changes of relationships, subject to varying constraints determined by context.**

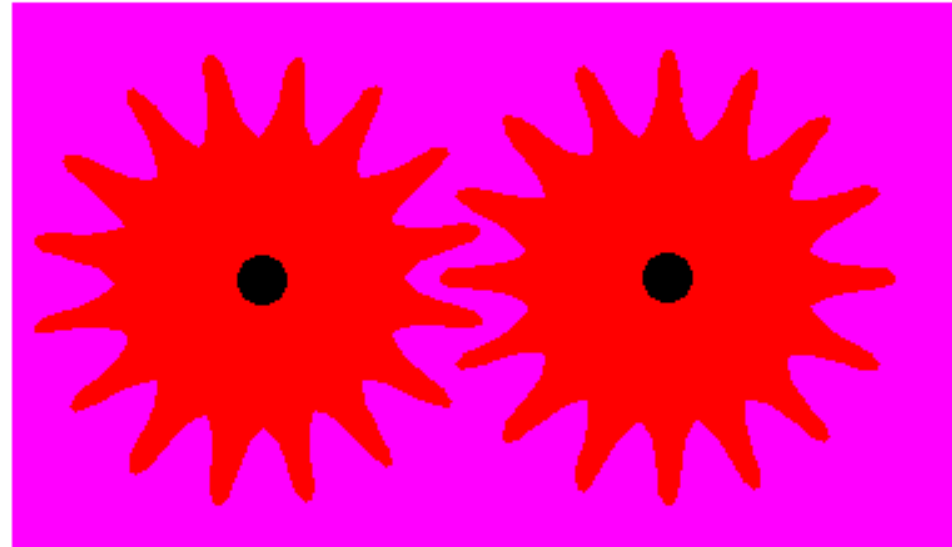
These ideas are developed in two online documents

<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0506>

COSY-PR-0506: Two views of child as scientist: Humean and Kantian

<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0601>

COSY-DP-0601 Orthogonal Competences Acquired by Altricial Species (Blanket, string and plywood).



# Humean and Kantian Causation

---

- When the only way you can find out what the consequence of an action will be is by trying it out to see what happens, you may acquire knowledge of causation based only on observed correlations. This is ‘Humean causation’ – David Hume said there was nothing more to causation than constant conjunction, and this is now a popular view of causation: causation as statistical (often represented in Bayesian nets).
- However if you don’t need to find out by trying because you can see the structural relations (e.g. by running a simulation that has appropriate constraints built into it) then you are using a different notion of causation: Kantian causation, which is deterministic and structure-based.
- I claim that as children learn to understand more and more of the world well enough to run deterministic simulations they learn more and more of the Kantian causal structure of the environment.
- Typically in science causation starts off being Humean until we acquire a deep (often mathematical) theory of what is going on: then we use a Kantian concept of causation.
- **This requires learning to build simulations with appropriate constraints.**

For more on this see this talk

<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0505>

**COSY-PR-0506: Two views of child as scientist: Humean and Kantian**

## **7 Geometry-based causation**

**Perceiving causation in changing geometric structures.**

**We can often see and understand consequences of motion of one part of a structure, including being able to predict effects on other parts.**

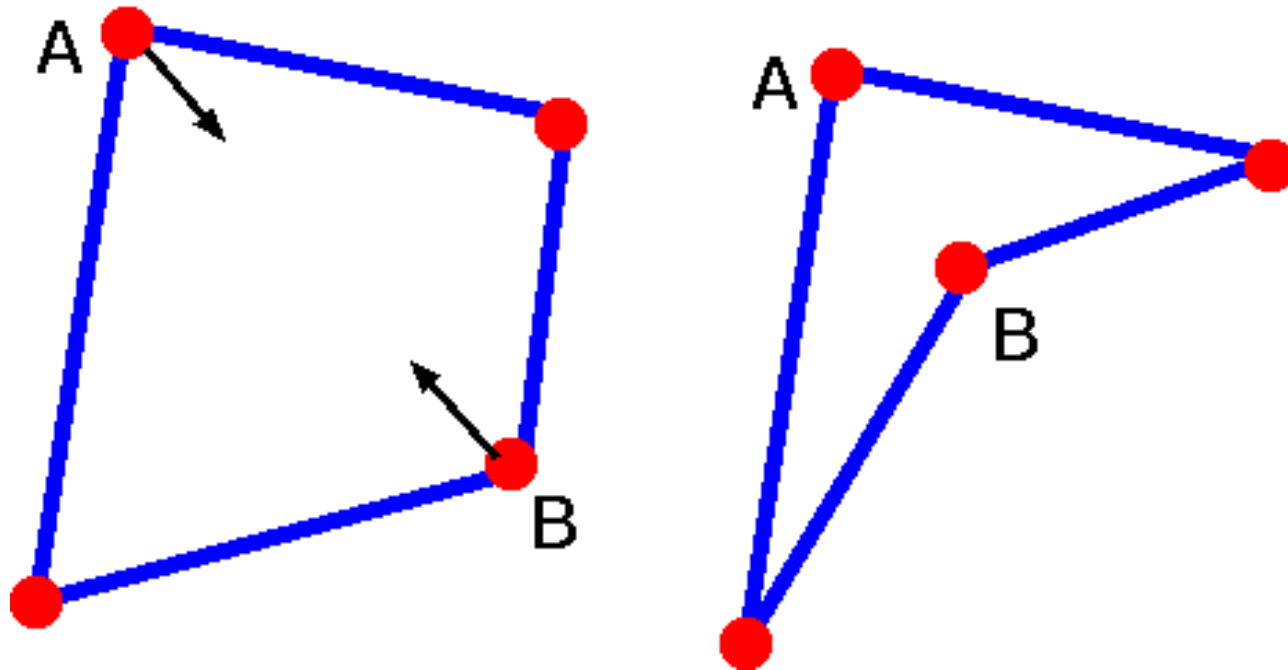
**But not when the structures are too complex, or have too many degrees of freedom.**

**Every kind of human competence has fairly low complexity limits, even though humans are enormously flexible in deploying and combining their competences.**



# Simulating motion of rigid, flexibly jointed, rods

On the left: what happens if joints A and B move together as indicated by the arrows, while everything moves in the same plane? Will the other two joints move together, move apart, stay where they are. ???



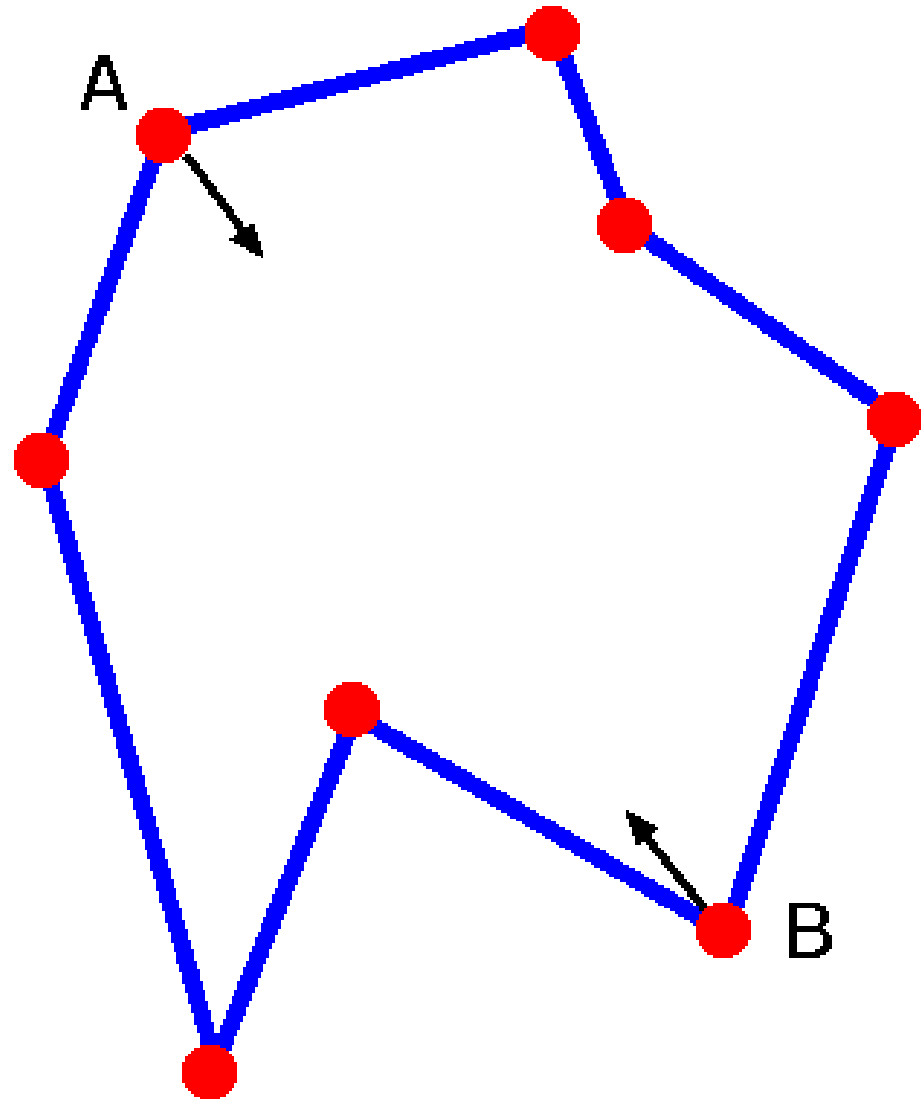
- What happens if one of the moved joints crosses the line joining the other two joints?
- We can change the constraints in our simulations: what can happen if the joints and rods are not constrained to remain in the original plane?

## Multiple links: how we break down

Can you tell how the other rods will move, if A and B are moved together and all the rods are rigid, but flexibly jointed?

There are not enough constraints. In this case our causal reasoning merely allows us to think about a range of options, though it is not easy. Unlike simpler linkages, most people will not be able to see whether the continuum of possible processes divides into clearly distinct subsets except (perhaps) by spending a lot of time exploring.

As situations get more complex, human abilities to simulate degrade rapidly: our understanding of Kantian causation tends to be limited to relatively simple, deterministic cases, though we can learn to grasp more complex structures and processes – up to a point. Perhaps intelligent artificial systems will have similar limitations.



# The moral?

---

- Processes we can imagine, see, think about are not necessarily related to what our own bodies can do: the importance of embodiment is currently being grossly oversold.
- Humans do not scale up, though we do ‘scale out’ – many different competences are available that can be combined in different ways.

**How they are acquired, represented, stored, accessed and combined, is largely unknown.**

**(That’s one difference between what I am saying and ‘global workspace theory’, which doesn’t address those questions.)**

## **8 Multi-modal perception of causation**

**We can combine information from different senses to produce a running simulation of what is going on.**

**(As Grush (2004) points out.)**

**In some case what is represented in the simulation is not sensed at all, until some time after the simulation starts.**

# Mixed mode input to an integrated simulation

- What you hear, like what you see, can be a process occurring in the environment, for instance hearing someone moving round you when your eyes are shut.
- If you are sitting in a room with a door opening into a corridor, subtle aspects of the changing sound of footsteps (which you process unconsciously) may produce a percept of an unseen person moving to the door, so that you know when he will become visible – a device used often in movies.
- Likewise when you see the unseen person's shadow changing.
- So the process you **hear** occurring and the things you **see** occurring may exist in the same integrated simulation — which is just as well since they exist in the same spatial environment.
- Likewise what a dentist sees and feels with the probe as she looks into the patient's mouth need to be in the same perceived part of the world, and when you use a hand to feel the underside of the table you are looking at **you see and feel the same table**.
- If you push a pencil up through a hole in the table you see and feel the same moving pencil.

# Sensory modality and mode of representation

- **Sensory modality driving a simulation need not determine the nature of the percept.**
- **A unitary, amodal, percept of a process can be driven by input from diverse sensory modalities – e.g. seeing, hearing, feeling the same thing happening.**
- **What is simulated does not determine the nature of the medium used to implement the simulation, as long as it has a rich enough structure and appropriate mechanisms to create, modify, access and use the contents.**
- **Examples of what the simulation might be include:**
  - a set of variables with changing values driven by sensory data
  - a database of logical assertions along with insertions and deletions driven by sensory data
  - a hybrid mechanism – logical assertions with equations linking changing variables, as can happen in some spreadsheets,
  - a spatially structured changing model,
  - a stored ‘script’ for the process with a pointer moving through the script at a rate determined by sensory input,
  - it may use a powerful form of representation that we have not yet thought of though evolution discovered it long ago.
- **Whatever form of representation is used, currently known brain mechanisms do not seem to support the required functionality.**

# Visual reasoning about something unseen

**An example of disconnection between simulation and sensory data.**

**If you turn the plastic shampoo container upside down to get shampoo out, why is it often better to wait before you squeeze?**

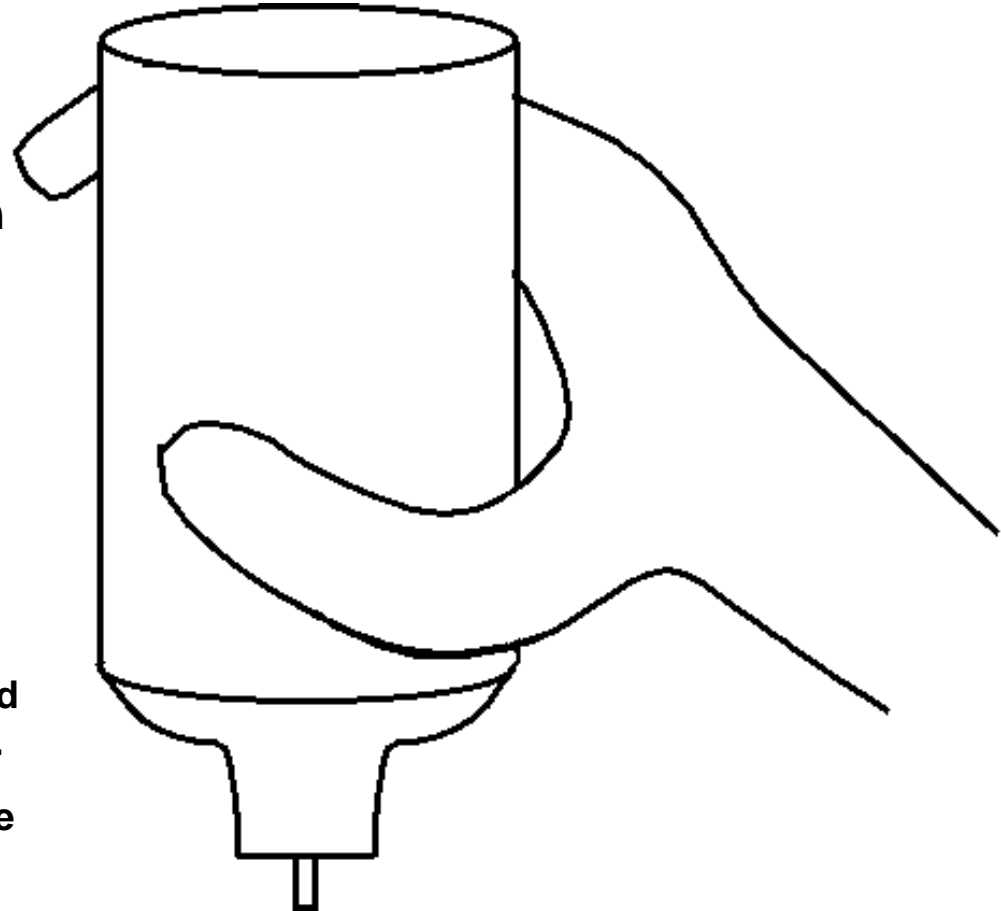
**In causal reasoning we often use runnable models that go beyond the sensory information: part of what is simulated cannot be seen – a Kantian causal learner will constantly seek such models, as opposed to Humean (statistical) causal learners, who merely seek correlations.**

**Note that the model used here assumes uncompressibility rather than rigidity.**

**Also, our ability to simulate what is going on explains why as more of the shampoo is used up you have to wait longer before squeezing.**

**Sometimes we run the wrong simulation if we don't understand what is going on.**

**Like the person who suggested that you have to wait for the water from the shower to warm the air in the container.**



## 9 Many distinct competences have to be learnt

The competences described above are not all present at birth, though some of the mechanisms required to acquire them are (while other learning mechanisms have to be produced by learning).

They are not **pre-configured** by genetic mechanisms, like innate abilities or innate latent genetically-determined competences that emerge long after birth (e.g. sexual competences, or migration in some birds).

The learnt, meta-configured competences need powerful bootstrapping mechanisms.

See

A. Sloman and J. Chappell (2005), The Altricial-Precocial Spectrum for Robots, *Proceedings IJCAI'05* pp. 1187–1192.

<http://www.cs.bham.ac.uk/research/cogaff/05.html#200502>

A. Sloman and J. Chappell (2005), Altricial self-organising information-processing systems, *AISB Quarterly*, 121, Summer 2005, pp. 5–7,

<http://www.cs.bham.ac.uk/research/cogaff/05.html#200503>

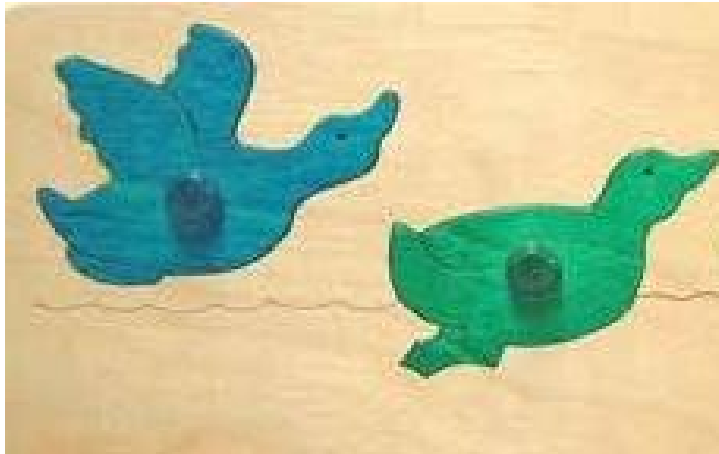
What the bootstrapping mechanisms achieve is extremely dependent on what is in the environment (including the culture), which is why altricial species with many meta-configured competences can differ enormously in what they know and can do, unlike precocial species, in which most competences are pre-configured, like deer which run with the herd soon after birth.

The examples that follow indicate some of what a child has to learn to see, before it can control its actions so as to achieve its goals, like inserting a puzzle piece where it belongs.



# We cannot do it all from birth

The causal reasoning we find so easy is difficult for infants.



A child learns that it can lift a piece out of its recess, and generates a goal to put it back, either because it sees the task being done by others or because of an implicit assumption of reversibility. At first, even when the child has learnt which piece belongs in which recess there is no understanding of the need to line up the boundaries, so there is futile pressing.

Later the child may succeed by chance, using nearly random movements, but the probability of success with random movements is **very** low. (Why?)



Memorising the position and orientation **with great accuracy** will allow toddlers to succeed: but there is no evidence that they have sufficiently precise memories or motor control. Eventually a child understands that unless the boundaries are lined up the puzzle piece **cannot** be inserted. Likewise she learns how to place shaped cups so that one goes inside another or one stacks rigidly on another.

These changes require the child to build a richer ontology for representing objects, states and processes in the environment, and that ontology is used in a mental simulation capability. **HOW?**

Stacking cups are easier partly because of symmetry, partly because of sloping sides: both reduce the uniqueness of required actions, so the cups need less precision and are easier to manage.

# Learning ontologies is a discontinuous process

- The process of extending competence is not continuous (like growing taller or stronger).
- The child has to learn about **new kinds** of
  - objects,
  - properties,
  - relations,
  - process structures,
  - constraints,...
- and these are different for
  - rigid objects,
  - flexible objects,
  - stretchable objects,
  - liquids,
  - sand,
  - mud,
  - treacle,
  - plasticine,
  - pieces of string,
  - sheets of paper,
  - construction kit components in Lego, Meccano, Tinkertoy, electronic kits...

**I don't know how many different things of this sort have to be learnt, but it is easy to come up with many significantly different examples.**

# CONJECTURE

---

## In the first five years

- a child learns to run at least hundreds, possibly thousands,
- of different sorts of simulations,
- using different ontologies
- and different kinds of constraints on possible motions  
with different materials, objects, properties, relationships, constraints, causal interactions.
- and throughout this learning, perceptual capabilities are extended by adding new sub-systems to the visual architecture, including new simulation capabilities

Some more examples are available in

<http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0601>

COSY-DP-0601 Orthogonal Competences Acquired by Altricial Species (Blanket, string and plywood).

## 10 Much of what is learnt is about kinds of stuff

Human children (and presumably also chimpanzees, nest building-birds and members of other altricial species) learn many things about the environment by playful exploration, using a collection of special-purpose mechanisms developed by evolution for the task.

Part of what they learn concerns **the behaviour of various kinds of physical stuff** in the environment, including

- kinds of material like:
  - sand, water, mud, straw, leaves, wood, rock,
  - and in our culture also: things like paper, cloth, cotton-wool, plastic, aluminium foil, butter, treacle, velcro, meal, concrete, glue, mortar,
  - various kinds of food (meat, fish, vegetable matter, peanut-butter, etc.)
- kinds of components that can be combined to form larger objects including:
  - lego, meccano, tinker-toy, Fischer-technik, and many more,
  - including, for nest-building birds, twigs, leaves, etc.

‘Behaviour’ of such things includes their responses to being folded, crushed, picked up, thrown, twisted, chewed, sucked, pressed together, compressed, stretched, dropped, and also the properties of larger wholes containing them.

The variety of kinds of stuff and kinds of behaviour should not be thought of as a **continuum**, e.g. something that might be form a vector space parametrised by a collection of real-valued parameters. Rather there are qualitative and structural differences important in many sub-ontologies that have to be learnt separately (even if some precocial species have precompiled subsets).

**A few examples follow: you can probably think of many more.**

# Cloth and Paper

---



**You have probably learnt many subtle things unconsciously about the different sorts of materials you interact with (e.g. sheets of cloth, paper, cardboard, clingfilm, rubber, plywood).**

**That includes learning ways in which you can and cannot distort their shape.**

**Lifting a handkerchief by its corner produces very different results from lifting a sheet of printer paper by its corner – and even if I had ironed the handkerchief first (what a waste of time) it would not have behaved like paper.**

**Most people cannot simulate the **precise** behaviours of such materials but we can impose constraints on our simulations that enable us to deduce consequences.**

**In some cases the differences between paper and cloth will not affect the answer to a question, e.g. the example on the slide about folding a sheet of paper, below.**

# What do you know about cloth and paper?

There are probably many things you know about cloth and (printer) paper that you have never thought about, but implicitly assume in your reasoning about them, including imagining consequences of various sorts of actions.

## Common features

- Both have two 2-D surfaces, one on each side.
- Both have bounding edges.
- Both can be made to lie (approximately) flat on a flat surface.
- Both can be smoothly pressed against a cylindrical or conical surface, but not a spherical (concave or convex surface)
- To a first approximation neither is stretchable, in the sense that between any points P1 and P2 there is a maximum distance that can be produced between P1 and P2, if there is no cutting or tearing.
- Both can be cut, torn, folded, crumpled into a ball....

## Differences

- most cloth can be slightly stretched (though some is very stretchy)
- Paper folded and creased tends to retain its fold, cloth often doesn't (there are exceptions, especially if heat is applied).
- Paper folded and not creased tends to return to its flatter state. It is more elastic.
- Paper folded once can stand upright resting on either a V-shaped edge or a pair of parallel edges.
- Paper is rigid within its plane (three collinear points remain collinear while the paper lies flat).

**NOTE:** tissue paper is somewhere in between.

# Contributors to simulation features

---

- We have so far seen that both shape and material can contribute to features of a simulation, including the constraints on what can and cannot change and what the consequences of change are.

- Another thing that can be important is **viewpoint**.

E.g. viewpoint can interact with opacity of materials, as well as with the mathematics of projection from 3-D to 2-D.

# Sometimes a simulation includes a viewpoint

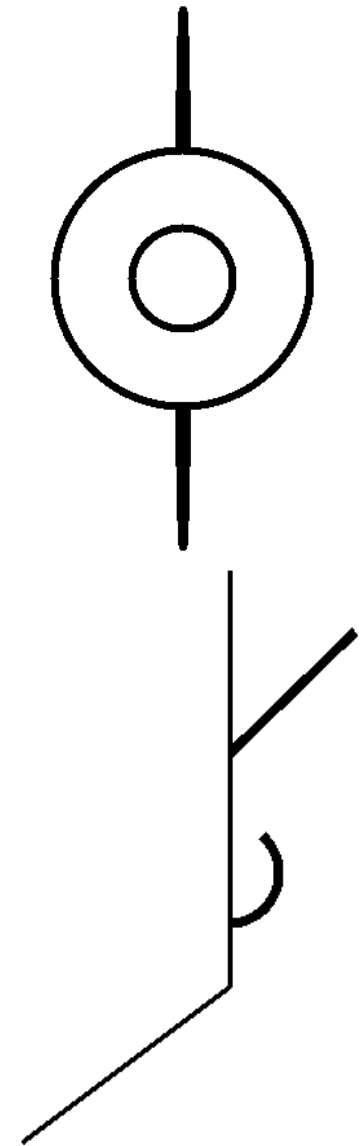
**Doodles illustrate our ability to generate a simulation (possibly of a static scene) from limited sensory information (sometimes requiring an additional cue, such as a phrase ('Mexican riding a bicycle', or 'Soldier with rifle taking his dog for a walk').**

**In both of these two cases the perceiver is implicitly involved: one involves a perceiver looking down from above the cycling person, whereas the other involves the perceiver looking approximately horizontally at a corner of a wall or building.**

**In both cases the interpretation includes not only what is seen but also occluded objects: the simulation depends on knowing about opacity.**

**This does not imply that we have opaque objects in our brains: merely that opacity is one of the things that can play a role in the simulations, just as rigidity and impenetrability can.**

**The general idea may or may not be innate, but creative exploration is required to learn about the details.**





## We can see things from more than one viewpoint

- **Vicarious affordances:** a parent watching a child needs to be able to see what is and is not possible in relation to the child's needs, actions, possible intentions, etc. (It is also useful to be able to perceive a potential predator's affordances.)
- This may include such things as visualising the scene from the child's viewpoint, including working out what the child can and cannot see – and the possible consequences of the child seeing some things and not seeing others.
- Some people can draw pictures of how things look from some other place than their current location.
- This ability to contemplate the world from multiple viewpoints, not just one's own current viewpoint, is essential for planning, since at some future state in the plan one's location and orientation could be very different from what it is now, yet it still needs to be reasoned about in extending the plan.
- The ability to perceive and use information about 'vicarious' affordances (affordances for others) and the ability to perceive affordances for oneself in the past (e.g. thinking about a missed opportunity) or future (planning to use opportunities that have yet to be created) may use the same mechanisms **because both are disconnected from current viewpoint.**

**Could that be the main point of substance behind all the fuss about “mirror neurons”? They should have been called **abstraction neurons.****

# Seeing things from the viewpoint of your hand

## The importance of hand-eye uncoordination!

- The evolution of body-parts for manipulation that can move independently of a major sensor perceiving what's happening (hands vs beak or mouth) had profound implications for processing requirements.
- Most animals are restricted to doing most of their manipulation with a mouth or beak, which cannot move much without the eyes moving too.
- If your eyes move as your gripper moves, because they are closely physically connected, then the sensorimotor contingencies linking actions and their sensory consequences will have strong, useful regularities that can be learnt and used.
- If a gripper can move independently of the eyes then the variety of relationships between actions and sensed consequences explodes.

The explosion can be reduced by modeling action at a level of abstraction removed from sensory changes: e.g. by representing actions as altering 3-D structures and processes (including subsequent actions), independently of how they are sensed.

- The mapping between sensory data and what is perceived becomes very indirect, and there may need to be several intermediate layers of interpretation: perception becomes akin to constructing a structured theory to explain complex data. (Compare the 'dotty picture' example, above.)
- This is one of many reasons for NOT regarding perception as simply concerned with detecting sensorimotor contingencies.

# Seeing from no particular viewpoint

---

**Dealing with a changing scene perceived by a moving observer may, for some purposes, require a representation of what is happening that is viewpoint independent as well as being modality independent.**

# Sensorimotor vs action-consequence contingencies

## Two evolutionary 'gestalt switches'?

The preceding discussion implies that during biological evolution there was a switch (perhaps more than once) from

insect-like understanding of the environment in terms of **sensorimotor contingencies** linking internal motor signals and internal sensor states (subject to prior conditions),

to

a more 'objective' understanding of the environment in terms of **action-consequence contingencies** linking changes in the environment to consequences in the environment,

followed by

a further development that allowed a **generative** representation of the principles underlying those contingencies, so that novel examples could be predicted and understood, instead of everything having to be based on statistical extrapolation.

To be more precise, it was an **addition** of a new competence rather than a **switch**

One of the major drivers for this development could be evolution of body parts other than the mouth that could manipulate objects and be seen to do so.

However the cognitive developments were not **inevitable** consequences: e.g. crabs that use their claws to put food in their mouth do not necessarily use the more abstract representation.

## **11 No good theories about shape perception exist**

**A huge amount of work on machine vision totally ignores shape and is concerned only with recognition, classification, prediction, or tracking, more or less treating the world as two-dimensional.**

**However there are some attempts to get machines to perceive shape.**

**Unfortunately these mostly seem to use inadequate requirements for shape perception. E.g. using vision and laser-scanning or whatever, to produce a detailed 3-D model of space occupancy which can be given to computer graphics programs to project images from any viewpoint in different lighting conditions may be very useful for many applications (e.g. medical imaging, and computer games) this does not give the computer a kind of understanding of shape that is required for manipulating objects.**

# Structures vs combinations of features

---

It is important to understand the difference between

- **Categorising**
- **Perceiving and understanding structure.**

You can see (at least some aspects of) the structure of an unfamiliar object that you do not recognise and cannot categorise: e.g. you probably cannot recognise or categorise this, though you see it clearly enough.

```
  Oooo
  Oooooo-----+
  OOOooooOOO    +
  |oooOOOooo-----+
  +-----+
```

## What is seeing without recognising?

There's a huge amount of work on visual **recognition** and **labelling** e.g. statistical pattern recognition. (Using totally arbitrary collections of benchmark images.)

But does that tell us anything about perception of structure?

Much work on vision in AI does not get beyond categorisation.

There is some work that attempts to identify structure from visual images, but the form in which structure is represented is merely a volumetric model, which may be very suitable for generating graphical displays from different viewpoints, but does not include any **understanding of the structure by the computer** – it leaves the main representational problems unsolved.

**There is something even more subtle and complex than perception of structure.**

# How many non-human species?

**Betty the hook-making New Caledonian crow.**

Give to google: betty crow hook:  
You'll find a link to the oxford zoology lab, with videos of Betty making hooks in different ways.

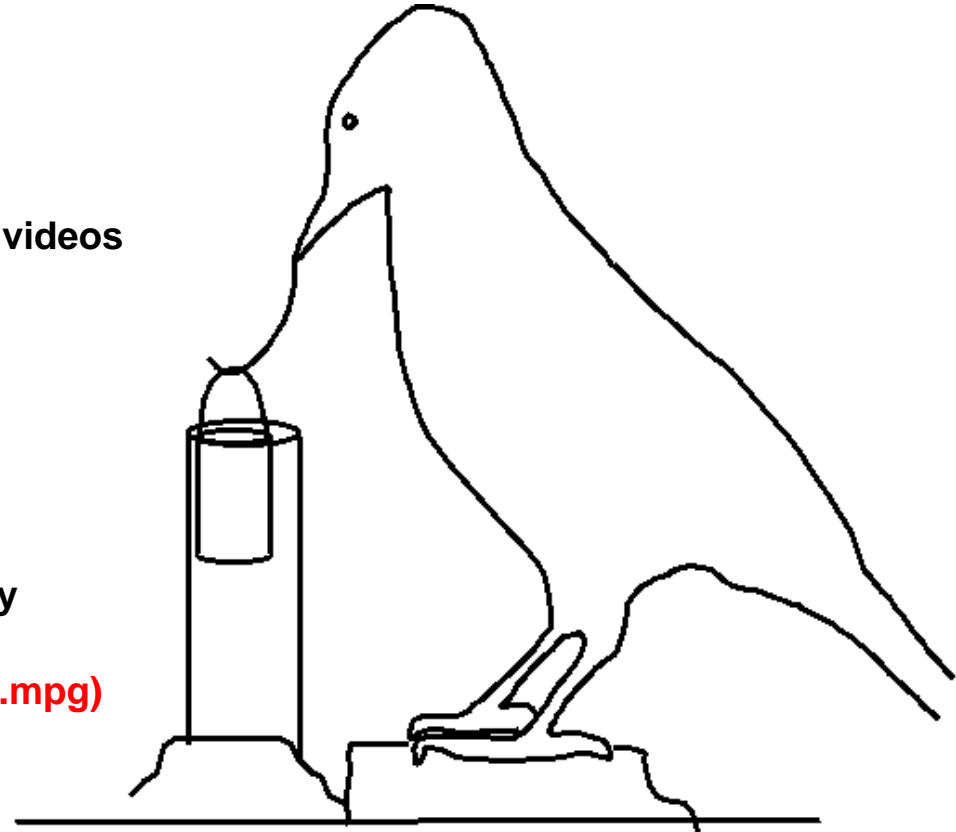
She **appears** to be a Kantian causal reasoner.

See the video here:

<http://news.bbc.co.uk/1/hi/sci/tech/2178920.stm>

Contrast the 18 month old child attempting unsuccessfully to join two parts of a toy train by bringing two rings together

([http://www.cs.bham.ac.uk/~axs/fig/josh34\\_0096.mpg](http://www.cs.bham.ac.uk/~axs/fig/josh34_0096.mpg))



**Does Betty see the possibility of making a hook before she makes it?**

**She seems to. How?**

# Understanding how hooks work

---

- Betty seems to understand how hooks work when she uses hooks to lift a basket of food out of the glass tube.
- The depth of understanding seems even greater when she demonstrates her ability to make hooks from straight pieces of wire in several different ways. I have also seen her make a hook from a long thin flat strip of metal.
- The behaviour is clearly not random trial and error learning behaviour: she seems to know exactly what to do, even though she does things in slightly different ways, e.g. making hooks using different techniques.
- Note that in Betty's environment far more distinct motions are possible than in the multi-rod linkage a few slides back: how does she confidently select a course through the continuum of continua?  
**The answer cannot simply be: by running a simulation, because the simulation might have the same problem of under-determination.**
- A young child does not start off understanding how a hook and a ring can interact in such a way as to allow the hook to pull the ring and what it is attached to.
- At some stage that (Kantian) understanding develops.  
But I don't think anyone knows how – even if some psychologists know when.
- The next slide points to a video showing a child who has not yet got there.

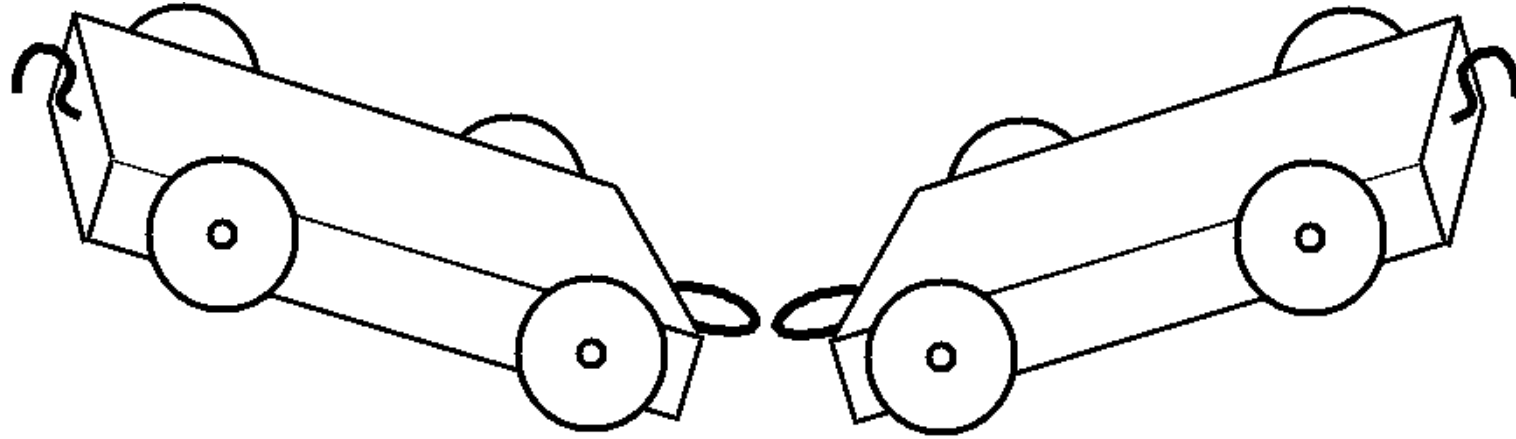


## **12 A child can appear less competent than a crow**

**We next show a video of a 19 month old child who is competent in many ways but seems to fail to understand how a hook and ring are used to join up a toy train.**

# Defeating a 19 Month old child

---



See the movie of an 19-month old child failing to work out how to join up the toy train – despite a lot of visual and manipulative competence also shown in the movie.

- [http://www.jonathans.me.uk/josh/movies/josh34\\_0096.mpg](http://www.jonathans.me.uk/josh/movies/josh34_0096.mpg)  
4.2Mbytes
- [http://www.jonathans.me.uk/josh/movies/josh34\\_0096\\_big.mpg](http://www.jonathans.me.uk/josh/movies/josh34_0096_big.mpg)  
11 Mbytes

The date is June 2003, when he was 19 months old. (Born 22 Nov 2001)

A few weeks later he had no problem joining up the train.

Was he a Humean causal learner or a Kantian causal learner?

I suspect the latter, but specifying the simulation model developed by a learner who understands hooks and rings will not be easy.

### **13 Running 2-D or 3-D simulations to answer questions**

**Perhaps the child who fails to join up the train does not understand because he has not yet learnt to simulate processes in which a hook and a ring form a connection that is useful for pulling.**

**Why not? Why are some competences innate, and some learnt. Why are some learnt very early and some only later.**

**Maybe we still have to understand the dependency relations between hundreds, or thousands, of sub-competences.**

**There are many problems we can solve, by running 2-D or 3-D simulations.**

**Some examples follow.**

# Simulating potentially colliding cars

---



The two vehicles start moving towards one another at the same time.

The racing car on the left moves much faster than the truck on the right.

Whereabouts will they meet – more to the left or to the right, or in the middle?

Where do you think a five year old will say they meet?

# Five year old spatial reasoning



The two vehicles start moving towards one another at the same time.

The racing car on the left moves much faster than the truck on the right.

Whereabouts will they meet – more to the left or to the right, or in the middle?

Where do you think a five year old will say they meet?

One five year old answered by pointing to a location near 'b'

Me: Why?

Child: It's going faster so it will get there sooner.

What is missing?

- Knowledge?
- Appropriate representations?
- Procedures?
- Appropriate control mechanisms in the architecture?
- A buggy mechanism for simulating objects moving at different speeds?

# Mr Bean's underpants

This paper (from a conference on thinking with diagrams in 1998)

<http://www.cs.bham.ac.uk/research/cogaff/00-02.html#58>

discusses how we can reason about whether Mr Bean (the movie star) can remove his underpants without removing his trousers.

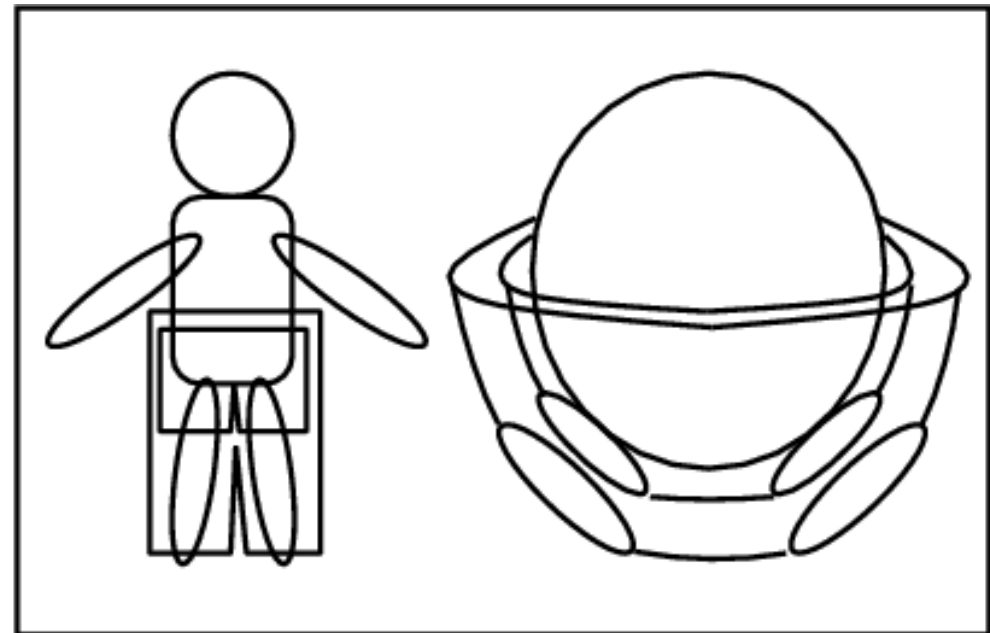
People often don't see all the possibilities at first.

The paper discusses how changing the simulation to a topologically 'equivalent' one can help us count the possible ways to perform the task.

Children can learn to perform such actions (as party tricks) physically long before they can reason with the mental simulations.

**What changes as the simulation ability develops?**

In part it seems to require an introspective ability to understand the nature of the simulations we use.



**See**

Jean Sauvy & Simonne Sauvy *The Child's Discovery of Space, From Hopscotch to Mazes: an Introduction to Intuitive Topology* (Translated P.Wells 1974).

# KANT'S EXAMPLE: 7 + 5 = 12

Kant claimed that learning that  $7 + 5 = 12$  involved acquiring *synthetic* (i.e. not just definitionally true) information that was also not *empirical*. I think his idea was related to the simulation theory of perception – but I am guessing.

You may find it obvious that the equivalence below is preserved if you spatially rearrange the twelve blobs within their groups:

$$\begin{array}{r} \text{ooo} \\ \text{ooo} \\ \text{o} \end{array} + \begin{array}{r} \text{o} \\ \text{o} \\ \text{ooo} \end{array} = \begin{array}{r} \text{oooo} \\ \text{oooo} \\ \text{oooo} \end{array}$$

Or is it?

How can it be obvious?

Can you see such a general fact?

How?

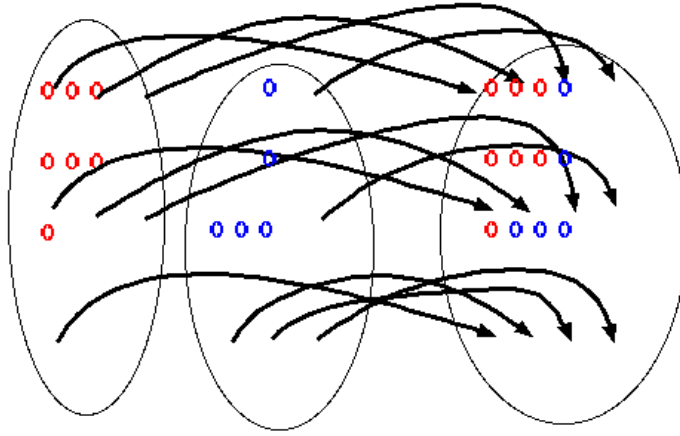
What sort of equivalence are we talking about?

I.e. what does “=” mean here?

Obviously we have to grasp the notion of a “one to one mapping”.

That **can** be defined logically, but the idea can also be understood by people who do not yet grasp the logical apparatus required to define the notion of a bijection — if they have a way of thinking about the consequences of motion of the blobs.

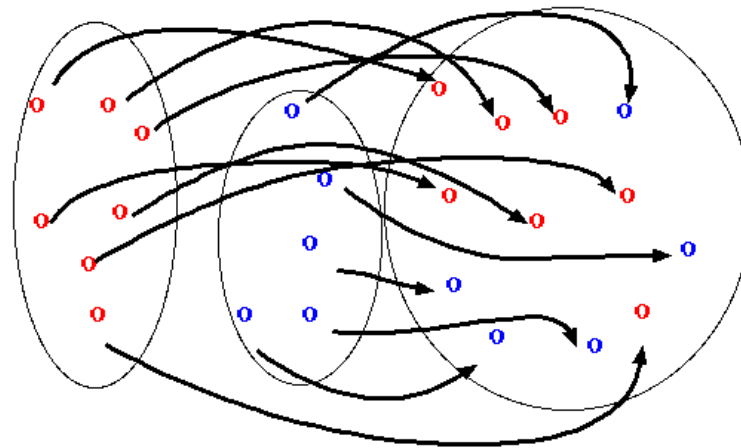
# SEEING that $7 + 5 = 12$



Then rearrange the items, leaving the strings attached.

Is it 'obvious' that the correspondence defined by the strings will be preserved even if the strings get tangled by the rearrangement?

Join up corresponding items with imaginary strings.



Is it 'obvious' that the same mode of reasoning will also work for other additions, e.g.  $777 + 555 = 1332$

Humans seem to have a 'meta-level' capability that enables us to understand why the answer is 'yes'. This depends on having a model of how our model works – e.g. what changes and does not change if you add another pair of objects joined by a string.

But that's a topic for another occasion.



## **14 What the simulation theory does and does not say**

**So far I have given many examples, and talked very vaguely about perception and reasoning as involving various kinds of simulations, using different ontologies with different sorts of constraints, different viewpoints, etc.**

**But the theory is easily misunderstood – and also still has many gaps.**

**I'll now try to make it a little more precise, including saying what I am NOT claiming.**

# The concurrent simulation theory in more detail

- Different simulations of the same scene may be used in different sub-mechanisms running simulations at different levels of abstraction and serving different functions.
- Some parts of simulations may **go beyond sensory data**, e.g. including unobserved sub-mechanisms (Kant)
- Some of the processes are **continuous** some **discrete**.
- The continuous and discrete processes may both have **different levels of resolution**.
- There may be **gaps** in the simulation at all levels (for different reasons)
- **Mode of processing can change dynamically**: parts of the simulation may be selected for more detailed processing, or type of processing can be changed.
- Seeing static scenes involves running **simulations in which nothing happens** – though many things could happen (cf. seeing affordances).
- The mechanisms originally evolved to support perceptual and motor control processes but became detachable from that role in humans and can be used to think about things that could never be observed,  
e.g. search spaces, high-dimensional spaces, infinite sets, including operations on transfinite ordinals (move all the odd numbers after the even numbers and reverse their order).  
See my paper 'Diagrams in the mind' 1998  
<http://www.cs.bham.ac.uk/research/cogaff/96-99.html#38>

# Development of perceptual sub-systems

---

The ability to run these simulations is not static, and may not even exist at birth:

- Visual capabilities described here develop in part on the basis of developing architectures for concurrent simulations and in part on the basis of learning new types of simulation, with appropriate new ontologies and new forms of representation.
- The initial mechanisms that make all of this possible must be genetically determined (and there may be limitations caused by genetic defects).
- But the *contents* of the abilities acquired through various kinds of learning are heavily dependent on the environment – physical and social, and on the individual's history. Some innate content is needed for bootstrapping.
- For instance someone expert at chess or Go will see (slow-moving!) processes in those games that novices do not see.
- Expert judges of gymnastic or ice-skating performance will see details that others do not see.
- An expert bird-watcher will recognize a type of bird flying in the distance from the pattern of its motion without being able to see colouring and shape details normally used for identification.

A deeper theory would explain the variety of types of changes involved in such developments: including changes in ontologies used, in forms of representation, and perhaps also in processing architectures.

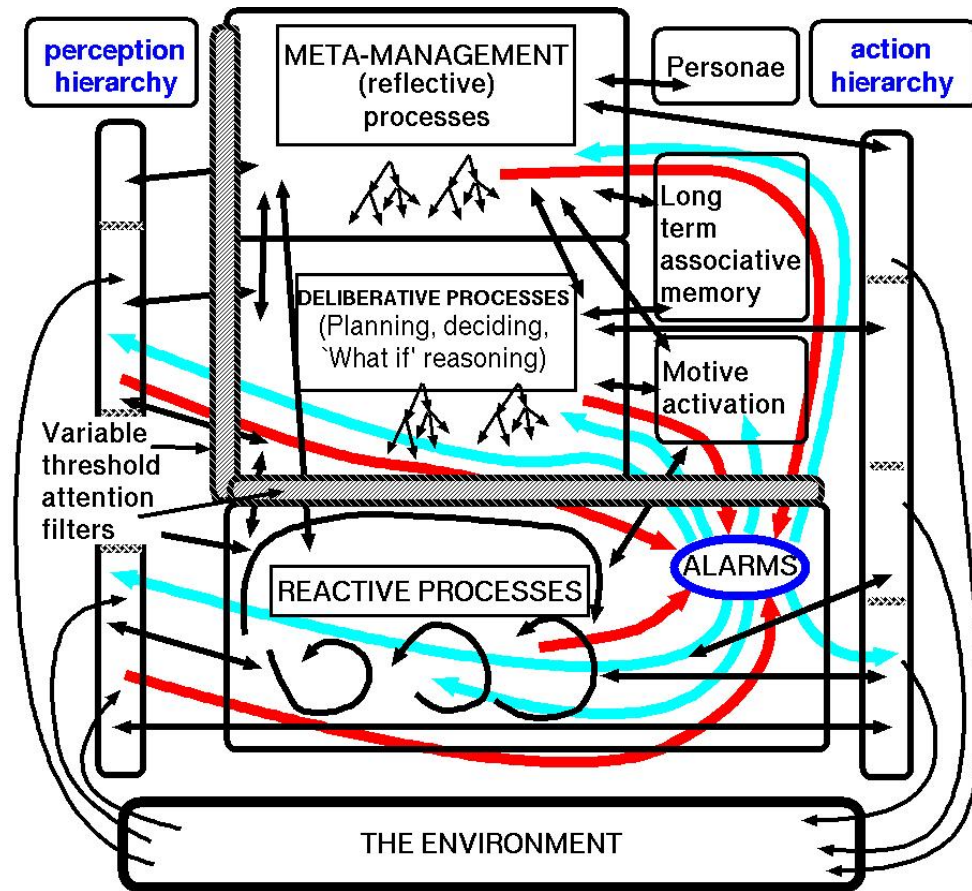
These will be changes in virtual machines implemented in physical brains.

# A hypothetical Human-like architecture: H-CogAff (See <http://www.cs.bham.ac.uk/research/cogaff/>)

This is an instance (or specialised sub-class) of the architectures covered by a generic schema called “CogAff”.

Many required sub-systems are not shown.

Different kinds of process simulation may go on in different parts of the architecture – some very old and widely shared, some relatively new and found in very few species.



(This is an illustration of some recent work on how to combine things: much work remains to be done. This partly overlaps with Minsky's *Emotion machine* architecture.)

For more details, see the presentations on architectures here

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

# Seeing intentional actions

---

Seeing a person or animal or machine doing something may involve a richer ontology than is required for seeing physical things moving under the control of purposeless physical forces.

- If you see a marble rolling down a slope occasionally changing direction or bouncing into the air as a result of surface irregularities or stones in its path, your simulation may include changes of position, speed and direction of motion, all consistent with what you know about physical objects.
- If you see a person walking down a slope occasionally moving to one side and picking things off bushes, you will see not only physical motion, but **the execution of an intention**, possibly several intentions, e.g. getting to something at the bottom of the slope, collecting biological specimens, and eating berries.
- One of the things a child has to learn to do is interpret perceived motion in terms of inferred goals, plans and processes of plan execution. Thus the simulations run when intentional actions are perceived may include a level of abstraction involving **plan execution**.

For a recent discussion see Sharon Wood, 'Representation and purposeful autonomous agents' *Robotics and Autonomous Systems* 51 (2005) 217-228

<http://www.cogs.susx.ac.uk/users/sharonw/papers/RAS04.pdf>

- When several individuals are involved, there may be several concurrent, interacting, processes with different intentions and plans to simulate. Learning to understand stories beyond the simplest sequential narratives requires learning to do this. (Contrast coping with 'flashbacks'.)

## 15 What I am NOT saying

The theory being proposed is easily misinterpreted.

The following slides attempt to explain what is **not** being said, by pointing out that some tempting interpretations of the theory are wrong.

# Disclaimers: No claim is made:

---

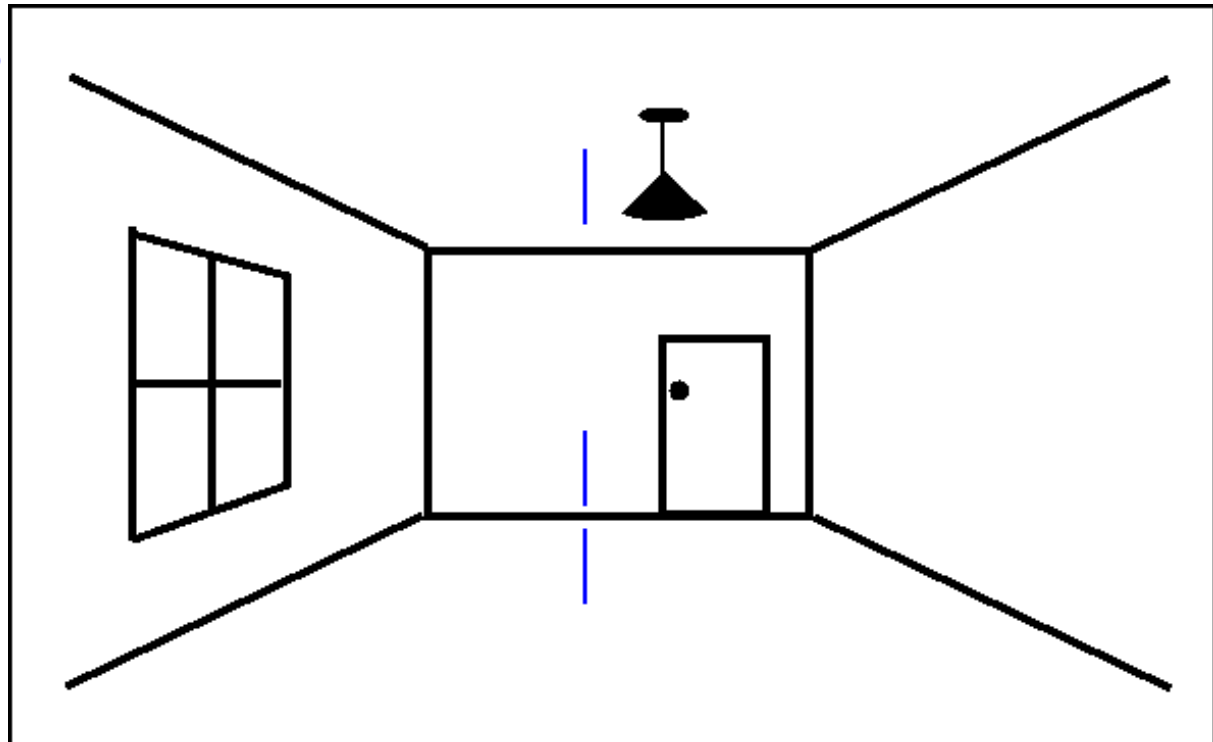
- That the simulations at any level are complete
- That they are accurate (errors, imprecision and fuzziness abound)
- That we are aware of all the simulations we are running
- That only humans can do this
- That all humans can run the same kinds of simulations
  - Different kinds of education, different kinds of training, e.g. artistic, athletic, mathematical training, playing with different kinds of toys, etc. can all produce different ontologies, representations and simulation capabilities. Even children with similar competences may get there via different routes along a partially ordered network of trajectories. **There are genetic differences too – e.g. ‘Williams syndrome’ children don’t develop normal spatial competences.**
- That it is obvious how to implement these ideas in artificial visual systems
- That the theory is compatible with any current theory of learning
- That the theory is compatible with known brain mechanisms
  - We may have to search for previously undiscovered mechanisms (including previously unknown types of virtual machines implemented in brains)
  - See Trehub’s book (*The Cognitive Brain, 1991*) for some relevant ideas.
  - There are probably lots of things I should have read but have not.
  - There is considerable overlap with the BBS paper by R.Grush (2004): The Emulation Theory of Representation.

# Isomorphism is not needed

Here's a modified version of a picture from chapter 7 of *The Computer Revolution in Philosophy*, also in the 1971 IJCAI paper.

Objects and relations within a picture need not correspond 1 to 1 with objects and relations within the scene, as is obvious from 2-D pictures of 3-D scenes.

For example: pairs of points in the image that are the same distance apart in the image can represent pairs of points that are different distances apart in 3-D space – e.g. vertically separated points on the walls, and horizontally separated points on the floor and ceiling. (And *vice versa*.)



Some pairs of parallel edges in the scene are represented by parallel picture lines, others by converging picture lines.

The small blue lines can be interpreted in different ways, with different spatial locations, orientations and relationships. On each interpretation the structure of the image remains unchanged, but the structure of the 3-D scene changes.



# MAJOR DISCLAIMER

---

**I am not claiming that simulations have to be isomorphic with what they simulate**

- As pointed out in my 1971 paper, analogical representations use relations to represent relations but they need not be **the same** relations:  
**Think of a 2-D picture of a 3-D scene (the same 2-D relation 'above' can correspond to different 3-D relations in different parts of the picture – floor, far wall, ceiling).**  
See <http://www.cs.bham.ac.uk/research/cogaff/crp/chap7.html>
- Not all simulations of spatial processes have to be spatial: it may often be simpler to use equations, for example, and psychological behavioural experiments may be wholly unable to determine which kind of implementation is used without having access to design information.
- Somehow we have developed enormously flexible ways of using mappings between one changing structure and another changing **or static** structure – it is a matter of learning what kinds of formalism with what kinds of constraints do and do not work for particular tasks.  
**E.g. programming language constructs can map onto dynamic graphical displays.**
- The ability I am talking about goes on being developed throughout life as we acquire more and more kinds of expertise.
- **That means a complete theory will have to explain that acquisition process – and no finite theory will explain all past, present and future human competence.**

# Inadequate alternative theories

Among the precursors to the theory are several that in different ways are inadequate, despite providing useful steps in the right direction.

- One general kind of inadequate theory assumes that what is perceived can be expressed as a collection of measures, sometimes called ‘state variables’, (e.g. coordinates, orientations, and velocities of objects in the scene) and that what is simulated can be expressed as continuous or discrete changes in a (possibly) large vector of state variables.
- This kind of numerical representation is inadequate because it fails to capture **the structure** of the environment, e.g. the decomposition into objects with parts, and with different sorts of relationships between objects, between parts within an object, between parts of different objects, etc.  

**People who are familiar with a particular collection of mathematical techniques keep trying to apply them everywhere instead of analysing the problems to find out what forms of representation are really required for the tasks in hand.**
- Many theories do not do justice to the diversity of functions of vision. E.g. some people seem to think the sole or main function of vision is recognition of instances of object types.
- Most theories of vision do not allow that we see not only what exists but what can and cannot happen in a given situation – affordances.
- Dynamical systems theorists have some of the right ideas but restrict ontologies and forms of representation to what physicists understand.

# Terminology

---

- Some people distinguish simulation, emulation, imagery, etc.
- What I call a simulation is **a representation of a process** that can be used for a variety of purposes, e.g. recording, predicting, tracking, explaining, controlling.
- A simulation may itself be a process, or it may in some cases be a re-usable static trace of a process, e.g. an executable plan, even a plan with loops and conditionals – with a ‘now’ pointer.
- The same process may be simulated at different levels of abstraction:  
simulations run at a high level may be very much faster than what they represent.
- Different sorts of simulations are useful for different purposes.
- A child continually learns new sorts of simulations and new uses for old sorts.
- Some running simulations can change direction, can explore options.
- Some simulations are continuous, and some discrete, and some simulated processes are continuous and some discrete.  
A continuous simulation may represent a discrete process and *vice versa*.  
It is difficult for a continuous simulation do searching, e.g. in a space of possible explanations or possible plans: discretisation makes multi-step planning feasible.
- A simulation may change in complexity and structure as it runs (e.g. simulation of development of an embryo — unlike simulations that involve a fixed dimensional state vector).
- The things that change in a simulation need not be numerical variables.
- We probably don’t yet know all the powerful ways of representing processes that evolution may have discovered and implemented in brains.
- In principle a simulation can itself be simulated (e.g. at a higher level of abstraction) – as in John Barnden’s ATT-META system. <http://www.cs.bham.ac.uk/jab/ATT-Meta/>

## **16 Re-runnable check-points**

**One of the consequences of discretisation is support for multi-step deliberation, e.g. systematic searching for a plan, including use of back-tracking.**

# Re-runnable check-points

---

- When searching for a solution to a problem we often have to explore a branching space of possibilities.
- Continuous simulations are not good tools for exploratory searching because there are always infinitely many possible branch points with infinitely many branches.
- This can be overcome by doing the searching with the aid of a discrete, more abstract, symbolic version of the simulation, and saving check-points, which can later be compared with one another.
- Ideally the check-points should be able to generate new lower-level runs of the simulation, when you back-track to a check-point.
- But for this, fully fledged deliberative mechanisms (for exploring answers to 'what if questions') could not really use simulations.
- So the development of discrete (symbolic) forms of representation was a major step for evolution. It had profound consequences including making mathematics and human language possible.

Some animals probably use discrete symbols in internal languages.

<http://www.cs.bham.ac.uk/research/cogaff/81-95#43>

# Orthogonal environment-related competences 1

A typical child about five years old has much detailed knowledge of several distinct kinds, which can be combined in different ways in perceiving, understanding and planning actions in the environment:

- many **kinds of physical stuff** with different physical properties (e.g. water, sand, mud, wood, string, rope, paper, metal, stone, plastic, human skin, cotton wool, hair, butter, treacle, plastic film, aluminium foil, various kinds of food, wind, breath, fire and many more)
- different **kinds of surface features** – flat curved, smooth, rough, sticky, slimy, wet, sharp, textured in different ways, with ridges, furrows, dents, etc. etc.  
This decomposes further into yet more orthogonal sub-spaces.
- different **shapes of whole objects**, varying in topological and metrical aspects, with both continuous and discrete sub-spaces, at different levels of abstraction,  
E.g. there are discrete differences between numbers of holes, between being symmetric or not, having a long axis or not, etc. as well as a huge variety of types of continuous variation.
- different **ways in which new, possibly more complex, wholes can be formed by combining or modifying things** (in ways that depend on their shape, material, etc.)  
We could include ‘negative’ combinations, e.g. gouging out, carving, punching a hole, to make a new shape as in sculpture.  
Other shape-making transformations include bending, twisting, etc.

(continued...)

# Orthogonal environment-related competences 2

.... Continued from previous page

- different **sorts of spatial relations** between different objects of similar or different material (e.g. containing, touching, being glued to, being hooked round, being a certain distance apart, resting on, being mixed, attracting, repelling, etc. etc.)  
There's a particularly important difference between 'rigid' containment (e.g. the streak of metal in a rock, the screw in a plank) and 'fluid' containment, e.g. water, sand or a small ball in a mug, a river flowing in its bed.
- different **kinds of force** that can be applied to things, e.g. prodding, poking, stroking, squeezing, twisting, pulling, pushing, screwing, patting, ....
- different **sorts of process** that can occur, including moving, rotating, changing shape, entering, coming out of, passing between, pushing, pulling, stretching, swaying, covering, uncovering, putting on (clothing), flocking, swarming, as well as applying forces, and changing the application of forces .....

Some of these may result from the individual's actions, some merely observed.

As remarked previously, more complex things can be observed by an individual than produced by that individual, e.g. a busy street scene, a waterfall, a football match.

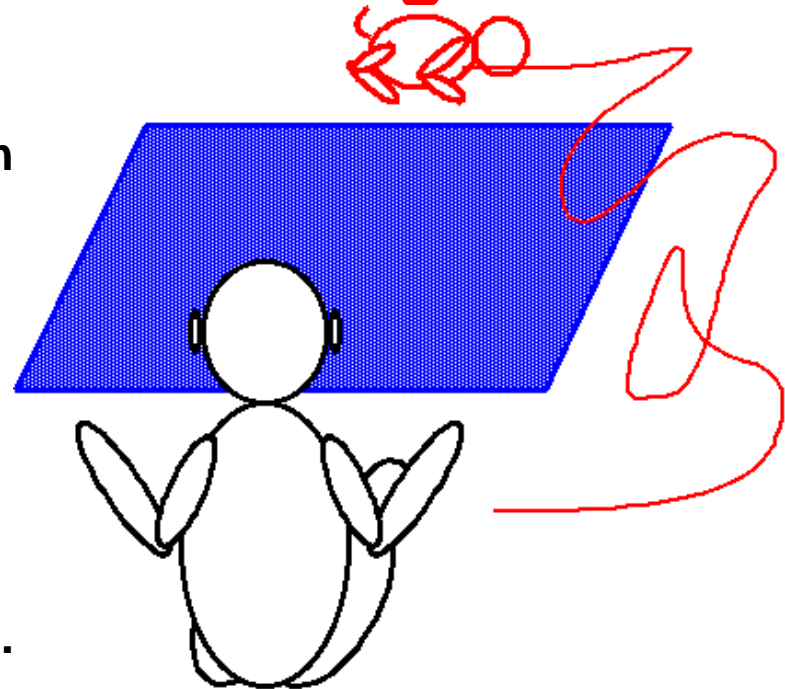
(There may also be behaviours an animal (e.g. insect) can produce that it cannot perceive because its perceptual mechanisms lack the required sophistication.)

**These lists are illustrative, not definitive or exhaustive, and do not include social abilities.**

## Example: Blanket and String

If a toy is beyond a blanket, but a string attached to the toy is close at hand, a very young child whose understanding of causation involving blanket-pulling is still Humean, may try pulling the blanket to get the toy.

At a later stage the child may either have extended the ontology used in its conditional probabilities, or learnt to simulate the process of moving X when X supports Y, and as a result does not try pulling the blanket to get the toy lying just beyond it, but uses the string.



**However the ontology of strings is a bag of worms, even before knots turn up.**

Pulling the end of a string connected to the toy towards you will not move the toy if the string is too long: it will merely straighten part of the string.

The child needs to learn the requirement to produce a straight portion of string between the toy and the place where the string is grasped, so that the fact that string is inextensible can be used to move its far end by moving its near end (by pulling, though not by pushing).

Try analysing the different strategies that the child may learn in order to cope with a long string, and the perceptual, ontological and representational requirements for learning them.



# Creativity in a physical environment

---

The different kinds of knowledge mentioned above can be combined in many different ways, including **novel** ways, in understanding what is perceived in the environment and what actions are and are not possible in different circumstances, and what the consequences of those actions will be.

We need to understand architectures and mechanisms for combining such knowledge and competences where appropriate.

Chapter 6 of *The Computer Revolution in Philosophy* attempted to analyse some of the processes about 30 years ago, but only at a high level of abstraction. <http://www.cs.bham.ac.uk/research/cogaff/crp/chap6.html>

- Sometimes competences are combined in **physical action**, using new combinations of material, tool, arrangement of parts or actions, to solve a problem; but in some cases it is done in thought (i.e. using deliberative mechanisms), as pointed out by Craik, Popper and many others.
- Precocial species, e.g. spiders, may have very specific ‘hard wired’ combinations of competence regarding specific kinds of stuff, specific spatial structures and processes; whereas humans some other altricial species are able both to **extend** knowledge within each of the categories, and to **forge new combinations** in perceiving novel scenes and performing novel actions — a meta-competence that underlies engineering, science and art.
- Such competence in pre-linguistic children and non-linguistic animals cannot depend on language, though it may be part of the basis for language, which, with other forms of cultural information-transmission (e.g. toys) enormously enhances and accelerates development.
- In a young child and in many animals the creative recombination of competence is applied in perceiving and using affordances for oneself, whereas humans later learn to see ‘vicarious affordances’, as discussed previously – essential in parents and carers watching children who may be about to hurt themselves, or may need help, or in seeing opportunities for predators who may attack one’s young.

# How much of this applies to other animals?

- Not all animals can learn these things, even if they share a lot of physical structure with humans.
- So it is likely that there are very specific, very powerful brain mechanisms involved, possibly several different mechanisms that evolved in different combinations — we are not discussing all-or-nothing capabilities.
- Even among humans there may be different combinations, e.g. Archimedes, Shakespeare, Newton, Kant, Mozart, Darwin, Turing. Picasso, Menuhin – in which case there is no such thing as **human psychology**.
- If the hundreds, or thousands, of different kinds of knowledge acquired in the first few years are stored in different parts of the brain, using different mechanisms, then different sorts of brain damage or deficiency could interfere with different sub-competences. Has anyone looked? (**E.g. Williams' Syndrome?**)
- Since most of the creative brain mechanisms evolved before human language capabilities and appear in pre-linguistic children, despite involving rich forms of semantic and syntactic competence (using internal representations), it could be that the generative (combinatorial) and extendable aspects of those pre-linguistic competences provided a foundation for the later evolution of linguistic competence.

Perhaps that is an example of the common pattern in evolution: duplication of structures or mechanisms followed by differentiation. (See the 'primacy' paper.)

# Conjecture

---

Alongside the innate **physical sucking reflex** for obtaining milk to be digested, decomposed and used all over the body for growth, repair, and energy, there is, in some species, a genetically determined **information-sucking reflex**, which seeks out, sucks in, and decomposes information, which is later recombined in many ways, growing the information-processing architecture and many diverse recombinable competences.

**Human-like robots will also need to be able to do that.**

**HOW ???**

See also <http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0601>

# Integrating Cognition, Emotion and Autonomy

**Tom Ziemke**

School of Humanities & Informatics  
University of Skövde, Sweden

**[tom.ziemke@his.se](mailto:tom.ziemke@his.se)**

April 4, 2006

# Mind vs. Life

- *"That mind requires life is often stated, and even more commonly assumed."*

(Margaret Boden's plenary abstract)

- Is it really?
- AISB 2005 machine consciousness symposium vote: only 3 out of about 30 participants thought that cognition/consciousness requires life/autopoiesis

# Embodiment?

- Different notions of embodiment emphasize:
  - Interaction (agent-environment)
  - Structural coupling (agent-environment)
  - Adaptation (to an ecological niche)
  - Physicality
  - Morphology
  - Complex interplay of morphology, neural processes and environment
  - Grounding of cognition/representation in sensorimotor processes (e.g. simulation theories)
  - Facilitation of social interaction
  - Autopoietic organization of living organisms

# Embodiment à la Murray

- "a spatially localised body using a sensory apparatus fastened to that body"
- "shared viewpoint, from which they can be indexically directed to the world"
- parallel-to-serial transition "perhaps the essence of consciousness, of what it means to be a singular, unified subject"
- Is that it?

# Organismic Embodiment?

- examples:
  - Maturana & Varela's (1980) work on autopoiesis and the biology of cognition
  - von Uexküll's (1928) theoretical biology
  - Damasio's (1994, 1999) theory of emotion and consciousness
  - Barandiaran & Moreno's (in press) notion of emotional embodiment



# Why organisms?

- Philosophy: see M. Boden's plenary talk
- Science:
  - All interesting cognitive systems we know of are organisms; it's those we want to understand
- Engineering:
  - Living organisms have a range of highly useful capacities (self-maintenance, self-repair, etc.)
    - Robots with energy autonomy, etc.
    - Autonomic computing systems
- Historical:
  - Dennett (1978) - "*Why not a whole iguana?*"
  - Brooks (1989a) - "to build *complete creatures* rather than isolated cognitive simulators"

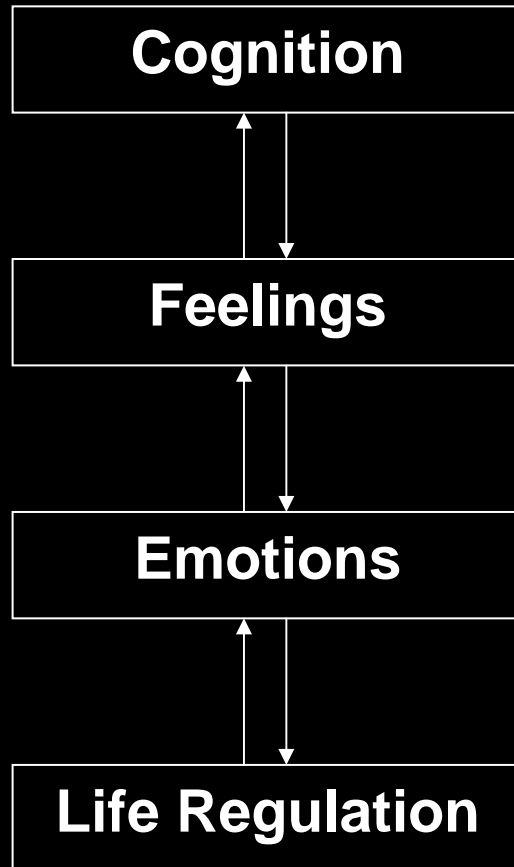
# Internal Robotics (Parisi, 2004)

- ... behaviour is the result of the interactions of an organism's nervous system with both the external environment and the internal environment, i.e. with what lies within the organism's body.
- While robotics has concentrated so far on the first type of interactions (*external robotics*), to more adequately understand the behaviour of organisms we also need to reproduce in robots the inside of the body of organisms and to study the interactions of the robot's control system with what is inside the body (*internal robotics*). (p. 325)

## Barandiaran & Moreno (in press)

- sensory-motor nervous system (SMNS) vs. nervous system of the interior (INS) (cf. Edelman, 1989)
  - INS: neuroendocrine system, autonomic nervous system, limbic system, etc.
- emotion ~ complex interplay between INS and SMNS (e.g. Damasio, 1994; Lewis, 2005)
- emotional embodiment:
  - the modulatory capacity of emotional dynamics is recruited to adaptively modify the SMNS

# Levels of Regulation (Damasio, 1999)



complex, flexible, and customized plans of response are formulated in images and may be executed as behavior

images (representations) of sensory patterns signaling pain, pleasure, and emotions

complex, stereotyped patterns of response, which include primary, secondary and background emotions

relatively simple, stereotyped patterns of response, incl. metabolic regulation, reflexes, the mechanisms behind pain and pleasure, drives and motivations

# ICEA project - motivation

- Jan 2006 – Dec 2009, 100+ person-years, 8 million euros
- “the emotional and bioregulatory mechanisms that come with the organismic embodiment of living cognitive systems also play a crucial role in the constitution of their high-level cognitive processes, and
- models of these mechanisms can be usefully integrated in artificial cognitive systems architectures, which will constitute a significant step towards truly autonomous cognitive systems that reason and behave, externally and internally, in accordance with energy and other self-preservation requirements, and thus sustain themselves over extended periods of time.”

# Consortium

- Skövde Cognition & AI Lab – Tom Ziemke (coordinator)
- Animat Lab, Paris – Jean-Arcady Meyer
- College de France – Sidney Wiener
- CNR, Rome – Baldassarre, Parisi, Nolfi
- Sheffield - Tony Prescott, Peter Redgrave
- Bristol Robotics Lab – Chris Melhuish
- BAE Systems, Bristol – Hector Figuereido
- Cyberbotics – Olivier Michel
- Hungarian Academy of Sciences – Peter Erdi
- Autonomous Systems Lab, Madrid – Ricardo Sanz

# Emotions (Damasio, 2004)

- ... emotions are bioregulatory reactions that aim at promoting, directly or indirectly, the sort of physiological states that secure not just survival, but ... [also] well-being. (p.50)
- ... emotional responses target both the body and other regions of the brain ... The responses alter the state of the internal milieu (using, for example, hormonal messages disseminated in the bloodstream); the state of the viscera; the state of the musculoskeletal system, and they lead a body now prepared by all these functional changes into varied actions or complex behaviours. (p. 51)

# Emotion (Petta, 2003)

- Emotion can be viewed as a flexible adaptation mechanism that has evolved from more rigid adaptational systems, such as reflexes and physiological drives ...
- The flexibility of emotion is obtained by decoupling the behavioral reaction from the stimulus event. The heart of the emotion process thus is not a reflexlike stimulus-response pattern, but rather the appraisal of an event with respect to its adaptational significance for the individual, followed by the generation of an action tendency aimed at changing the *relationship* between the individual and the environment. (p. 257)



# Feeling (Damasio, 1999)

- feeling = “the mental representation of the physiologic changes that occur during an emotion”
- while emotions involve bodily reactions, feelings (mental images of those reactions) allow the cognizer to temporarily decouple its cognitive processes from its immediate bodily reactions
  - e.g. anticipation of bodily reactions in the planning of behavior
  - “as if body loop” (Damasio)
    - a neural “internal simulation” (cf. Murray’s and Owen’s talks) that uses the brain’s body maps, but bypasses the actual body

# Feelings of emotion (Damasio, 2004)

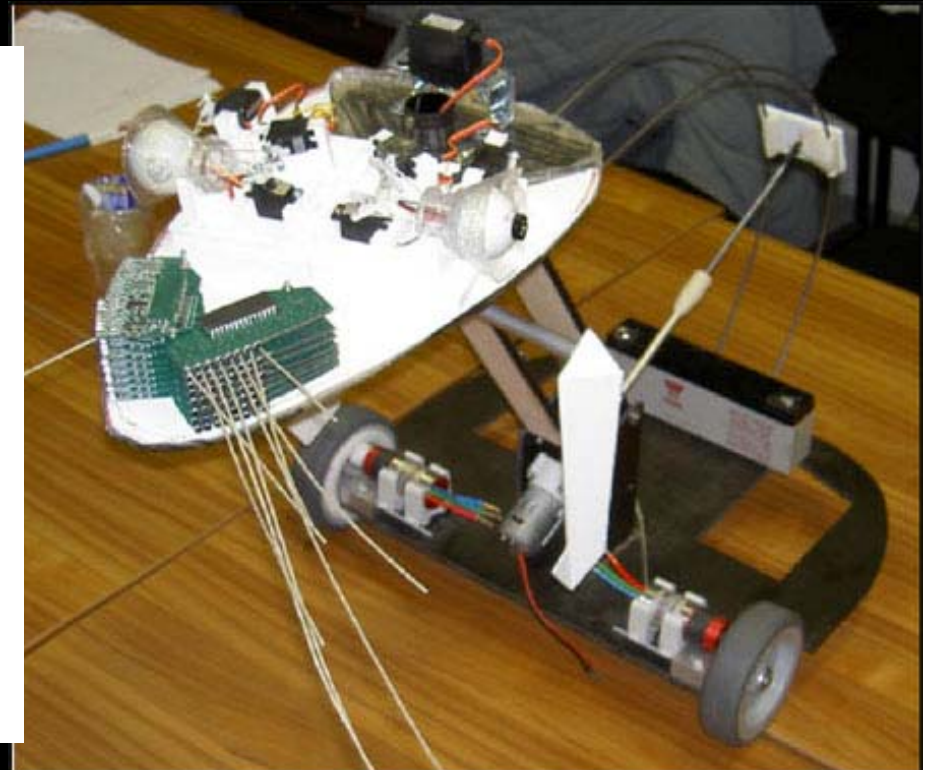
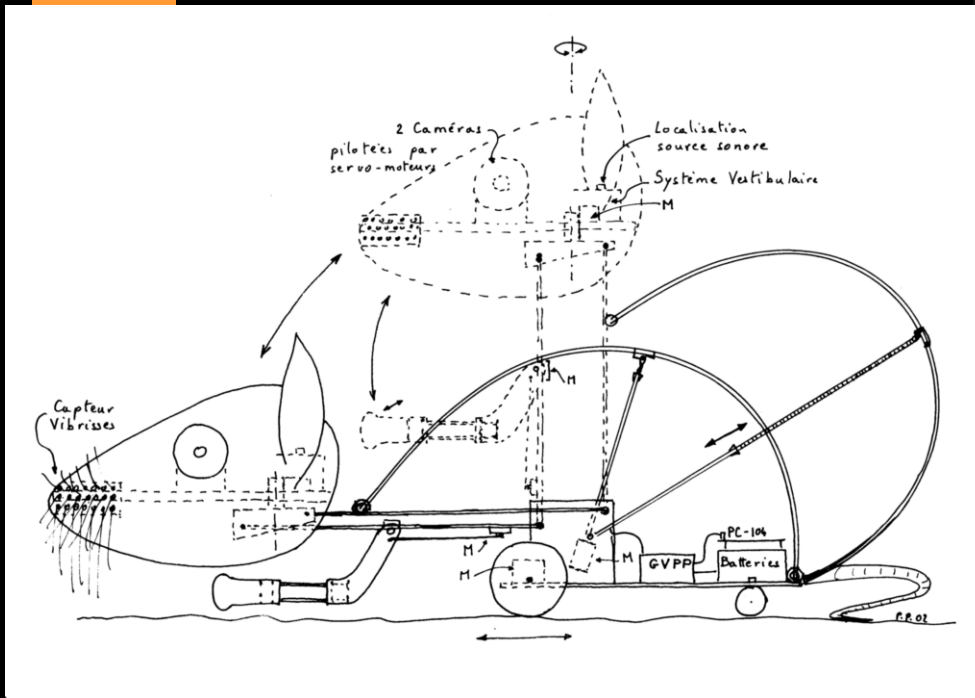
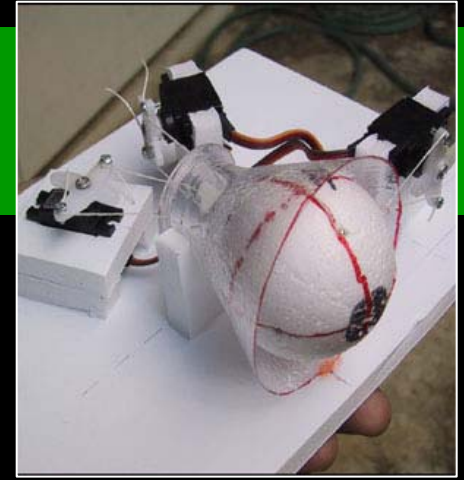
- The essence of feelings of emotion is the mapping of the emotional state in the appropriate body-sensing regions of the brain. (p. 52)
- Whereas emotions provide an immediate reaction to certain challenges and opportunities ... [t]he adaptive value of feelings comes from amplifying the mental impact of a given situation and increasing the probabilities that comparable situations can be anticipated and planned for in the future so as to avert risks and take advantage of opportunities. (pp. 56-57)

# ICEA - The rat as a starting point

- Massive literature on behavior & neurobiology
- Rather homologous to man
- Clever, intelligent, adaptive, compact
  - a model that works
- Realizable target for a four-year project
  - compared to human
- Complements existing EC-funded Cognitive Systems projects
- But: will surpass (selected) rat cognitive capacities

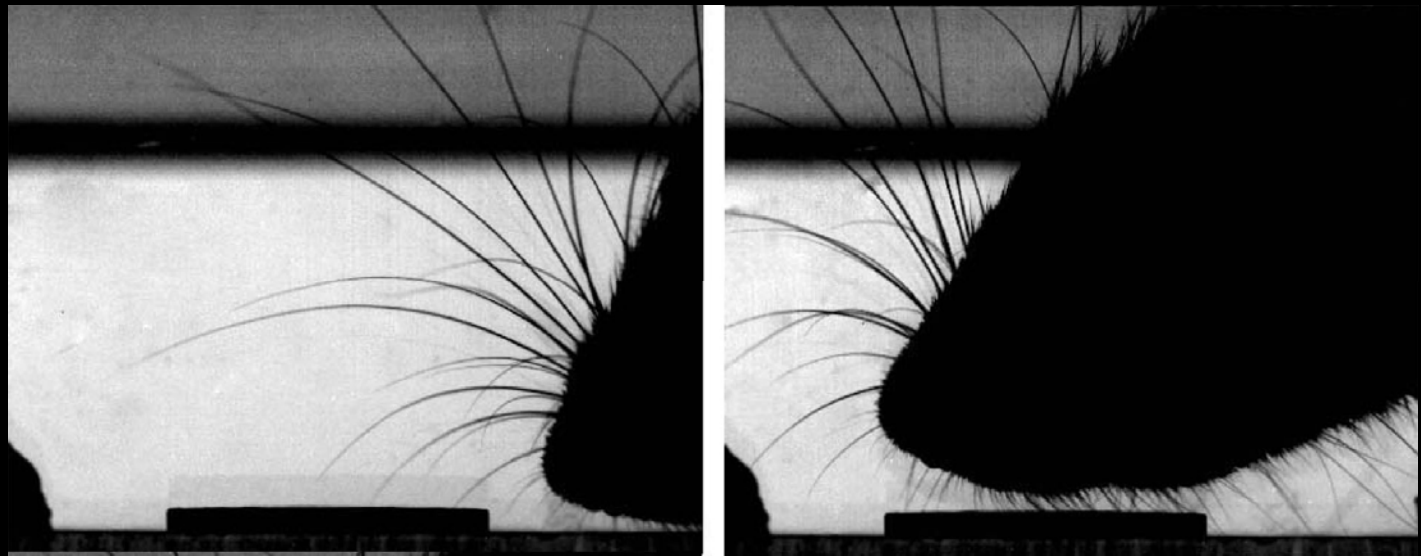
# ICEAbot

- Rat-like physical robot platform
- Builds on the Animat Lab's previous *Psikharpax* project



# Active whisking

- active touch for perception and spatial cognition
  - a neck with 3-DOF two arrays of macro-vibrissae, and
  - an array of smaller micro-vibrissae that provide a form of tactile ‘fovea’ for close-up examination of surfaces
  - based on high-speed digital videography of real rats (cf. Sheffield’s posters & demo)



# ICEAsim

- Rat-like simulation platform
  - based on Cyberbotics' Webots toolkit
  - used by all modelers in the consortium
    - Integration of models
  - based on ICEAbot
    - but with additional features: active whiskers, metabolism, etc.
- Will be made available for free to other researchers

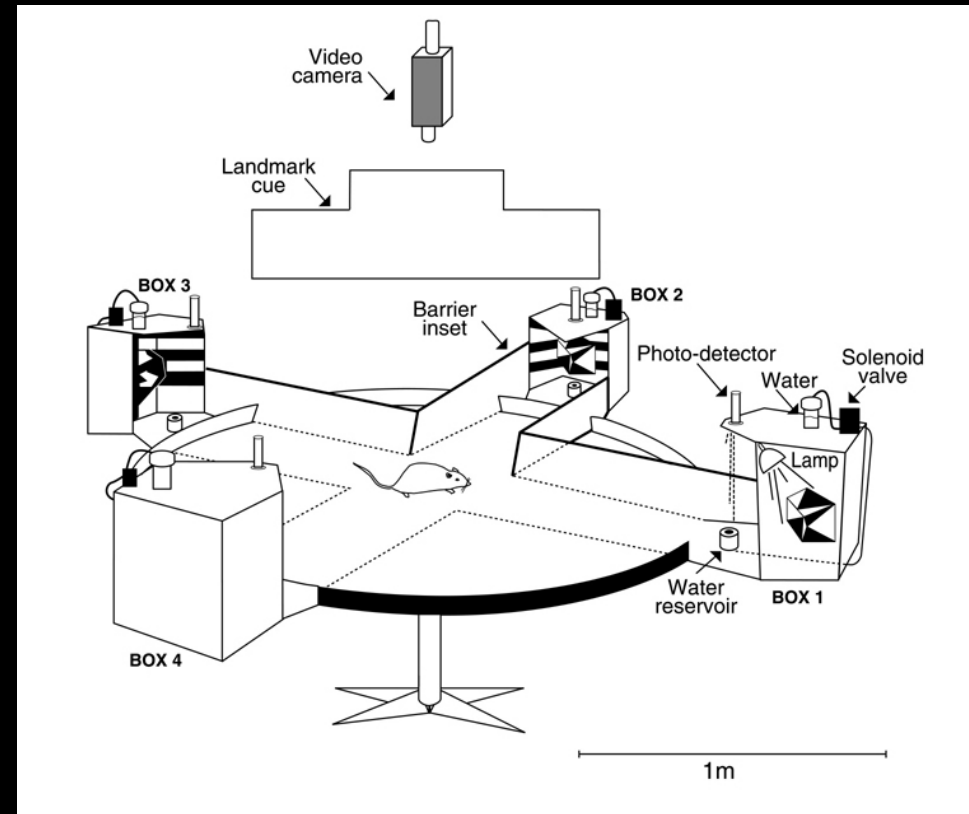
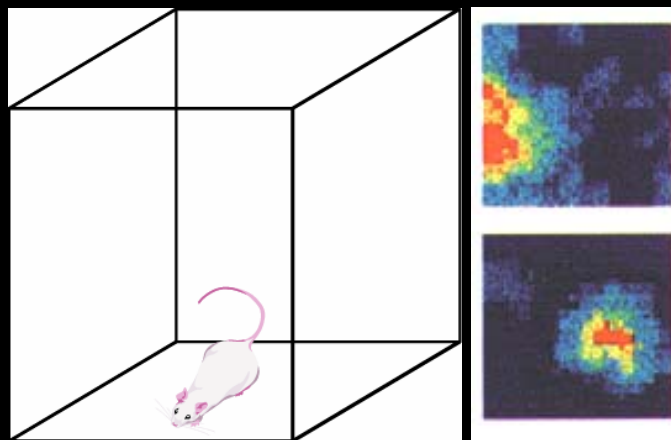


# Project elements

- overall volume: 100+ person-years of funded research
  - about 10% neurophysiology, rat experiments
  - about 80% comp. modelling, robotics, systems integration
  - about 10% theoretical integration
- alternative breakdown:
  - three main ‘chunks’, 25% each
    - central ICEA integrated robot and simulation platforms
    - motivated spatial cognition/behaviour
    - emotion-based representation/cognition
  - smaller ‘chunks’
    - layered self-defense architecture
    - energy autonomy

# Spatial behavior & cognition

- rat neurophysiology
- computational neuroscience models at different levels of abstraction



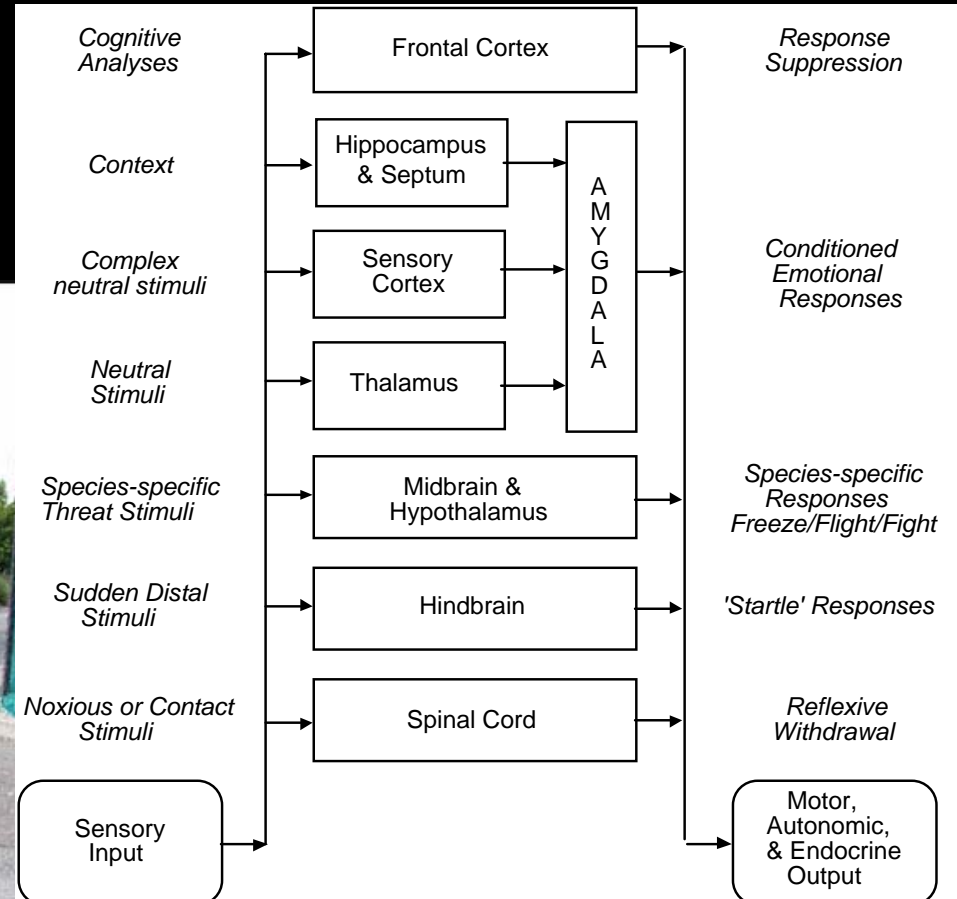


# Mammalian brain structures modeled

- *cortex* - planning, motivation, working memory, and analysis of sensory data
- *cerebellum* - anticipation, prediction
- *amygdala* - emotion and classical conditioning
- *basal ganglia* (incl. nucleus accumbens) - action selection sequencing, and reinforcement learning (operant conditioning)
- *hippocampus* - spatial and contextual memory
- *superior colliculus* - orienting
- *hypothalamus* - drives
- *brain-stem* - bio-regulation and pattern generation

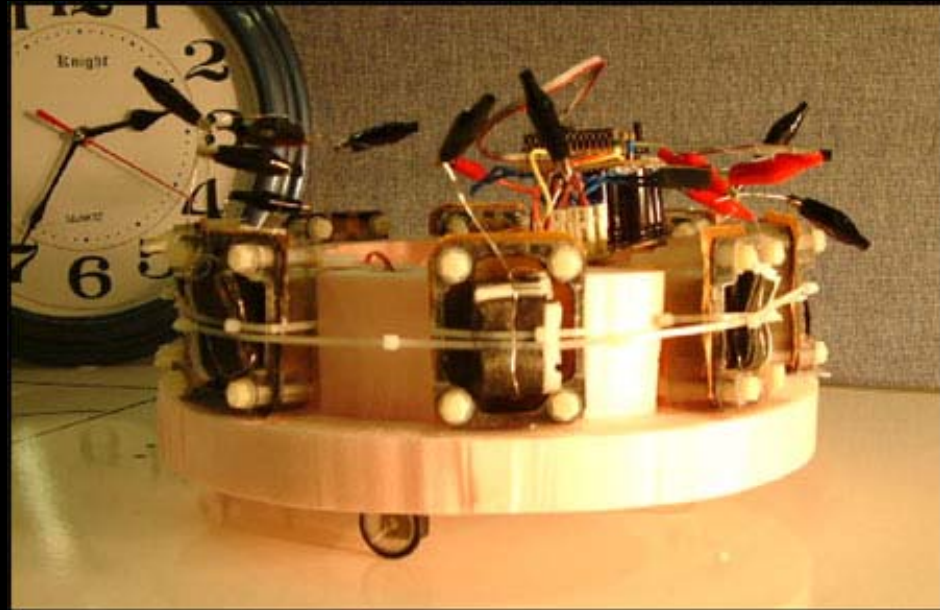
# BAEbot

- Layered defence architecture (cf. Peter Redgrave's talk) on an all-terrain vehicle



# BREADbot

- "Bio-regulation and energy autonomy with digestion"
- Based on the BRL's current work on energy autonomy using microbial fuel cells
- Coordination of internal homeostasis and effective foraging behavior



# Anticipation, imagination, planning

- starting point: simulation/emulation theories of cognition/representation as based on simulated agent-environment interaction
  - e.g. Hesslow, Grush, Barsalou
    - Thought ~ simulated sensorimotor interaction
  - Damasio's "as-if body loop"
    - e.g. anticipation of bodily/emotional reactions in the planning of behavior
  - What's the right level of abstraction?
- beyond the real rat's cognitive capacities (?)

# Integrating everything ...



Lines : 256  
Quads : 554  
Triangles : 78

... to be continued

# ICEA Summary

(in terms of Jeff's BBD principles)

- *"Incorporate a simulated brain ..."*
  - ICEA: integration of partial models of mammalian brain
- *"Active sensing and ... movement in the environment"*
  - ICEA: rat-like robots doing rat-like tasks
- *"Engage in a behavioral task"*
  - ICEA: rat-like tasks in rat-like environments
- *Categorization without a priori knowledge/instruction*
  - ICEA: in particular for abstraction/representation
- *"Adapt behavior when an important ... event occurs"*
  - ICEA: emotional appraisal, value systems, etc.
- *"Comparisons with experimental data acquired from animals ..."*
  - ICEA: based on and compared to rats