ESSLLI - 2000
Twelfth European Summer School
in Logic, Language and Information
http://www.folli.uva.nl/Esslli/2000/esslli-2000.html

6-18 August 2000

Workshop Title:

Architectures for intelligent language users

(How to turn philosophers of mind into engineers and *vice versa*)

WEEK 2: 14-18 August

AARON SLOMAN

http://www.cs.bham.ac.uk/~axs/
A.Sloman@cs.bham.ac.uk

Topics for discussion

IDEAS DEVELOPED IN COLLABORATION WITH:

Steve Allen, Luc Beaudoin,
Darryl Davis, Brian Logan,
Catriona Kennedy, Matthias Scheutz,
Ian Wright, and others in the

COGNITION AND AFFECT PROJECT
SCHOOL OF COMPUTER SCIENCE
THE UNIVERSITY OF BIRMINGHAM
(Partly funded by the Leverhulme Trust)

See our papers and tools http://www.cs.bham.ac.uk/research/cogaff/

http://www.cs.bham.ac.uk/research/poplog/freepoplog.html

(Including the SIM_AGENT toolkit)

POSSIBLE MAIN TOPICS

- What is a language user? How many kinds are there? How many kinds of language are there?
- Generalising notions like ‘language’, ‘syntax’, ‘semantics’, ‘pragmatics’, ‘communication’
- The variety of uses of languages (of various kinds) in complex information processing systems (natural and artificial). What are information processing systems?
- The idea of the architecture of a system. Varieties of architectures. Implementation levels vs functional decomposition.
- Architecture-based systems of concepts (e.g. the periodic table of the elements, or formula-based concepts of chemical compounds).
- Languages (internal and external) as biological phenomena, products of biological evolution. Biological precursors for mechanisms supporting linguistic capabilities.
- The inherently multidisciplinary nature of the study: philosophy, psychology, brain-science, ethology, social science, linguistics, evolution, logic, mathematics, computer science, software engineering, AI...
- Relevance to science, to engineering, to philosophy
- What are the still unsolved problems?

NOTE: The rest of this document is a disorganised collection of notes, not a systematic overview of the workshop.

SUMMARY OF DAY ONE
part 1

MIND $\Longleftrightarrow$ BRAIN

VIRTUAL MACHINE $\Longleftrightarrow$ PHYSICAL MACHINE

The first relation $\Longleftrightarrow$ is often referred to as "supervenience", the second as "implementation", or "realisation", or "support", well understood intuitively by software engineers.

Contrast philosophical theories about what exists (ontology):
- **dualism (various kinds)**
- **epiphenomenalism (a type of dualism)**
- **monism (various kinds, e.g. neutral, material, mental)**
- **pluralism**

A complication: *qualia*

- SOMETHING TO DO WITH "EXPERIENCE"
- WHAT IT IS LIKE TO BE ...
- THE FIRST PERSON VIEWPOINT....
- PERHAPS NOT EXPLICABLE

Summary Part 2

Question: where do virtual machines fit in?

- **Virtual machines have causal powers**
- **Virtual machines are ubiquitous (social, economic, physical)**
  (is there an ultimate, “bottom level” reality?)

Types of virtual machines in computers

- **Purely internal, low level (e.g. bit patterns)**
- **Abstract data-structures (e.g. lists, trees, numbers)**
- **Semantic content: information about the environment**
  (e.g. flight controller, plant controller)

Summary Part 3 Virtual machines can have different mechanisms

- **They can be purely reactive (sheepdog)**
- **They may be deliberative (SHRDLU: blocks world)**
- **They may have a mixture**

Reactive systems merely respond, internally or externally, to conditions (internal or external).

They cannot describe, contemplate, evaluate, non-existent states of affairs or actions.

Deliberative architectures are able to represent, evaluate, compare *non-existent* states of affairs, actions, processes

Later we'll investigate requirements for this in some detail.

It requires some sort of language, or notation, e.g. for solving the river-crossing possible, or the triangle and square puzzle.

How many sorts of language are there?

At least:
- **Logical/Fregean**
- **Pictorial, analogical**
- **Programming languages**
- **Neural nets**

Different kinds of syntax. Different kinds of semantics.

WHAT IS A LANGUAGE USER?

SOMETHING THAT MAKES USE OF SOME SORT OF **structured medium** TO STORE, MANIPULATE, OR COMMUNICATE **information.**

What's a structured medium?

- **Something within which entities with parts and relationships can be created.**
- **New instances can be created as needed.**
- **The instances can be deleted transformed (modified) or extended.**
- **There are many types of medium, supporting different sorts of entities: continuous/discrete, flat/hierarchical(Fregean), linear/multi-dimensional/graph-structured...**
- **The entities in the medium may be**

  enduring (e.g. ink marks on paper)
  transient (e.g. speech, semaphore signals), or
  somewhere in between (e.g. marks on a sea shore).

EXAMPLES

sequences of 1s and 0s
sequences of letters, spaces punctuation marks
sequences of phonemes, morphemes
maps
photographs
bead and wire molecular models
computer programs
bit patterns in a computer
list structures in a computer (e.g. in Lisp virtual machine)
vectors defining points in a phase space
a single value that can vary linearly
natural phenomena:

- **animal traces in a forest,**
- **storm indicators**
- **perceived structure of an animal or plant**

ALL INTERESTING ONTOLOGIES INVOLVE 'STRUCTURED MEDIA'

STRUCTURES, ARCHITECTURES, SYNTAX ARE EVERYWHERE

Generalizing the concept of “syntax”

Different media have different sorts of variability.

SYNTAX **can be thought of as:** TYPE OF VARIABILITY

For anything that changes there is a ‘space of possibilities’. Such spaces can have different topologies. Our familiar notion of syntax refers to locations in a space of possible structures (e.g. possible parse trees).

But familiar examples are special cases of a more general notion.

A discrete medium supports separate syntactic CATEGORIES **or** TYPES
(possibly at different levels of size and levels of abstraction).

When it is continuous there may or may not be (fuzzy) THRESHOLDS**, defining (fuzzy) categories. E.g.:**

1. **The medium of ink marks on paper is continuous, but we have conventions for dividing possible marks into sub-categories, e.g. letters of an alphabet.**
2. **There are many “naturally” occurring, biologically evolved, categorisations of structures, e.g. animal structures, plant structures, and also some of their behaviours, as perceived by them and by other organisms.**
3. **Phonemes**

NB: A ‘medium’ may support a wider range of variability than a particular syntactic space actually uses.

This is used in learning, development, evolution, where a syntactic space is extended.

Where does syntax come from?

The type of variability (syntax) is not an absolute property of the structures or the medium.

It is RELATIVE ***to a USER (actual or possible).***

SYNTAX IS IN THE EYE OF AN (ACTUAL OR POSSIBLE) BEHOLDER
which could be an animal, or a machine,
or a PART of one.

More on syntax

1. The notion of syntax of an entity X makes no sense except in relation to a possible or actual observer O which attributes or assigns a syntax to X.
   (NB X could be a process, or something extended in time)

2. Note that X could be a part of O, or an aspect of O's behaviour

3. The syntax *synt* assigned by O to X
   syntax(O, X)

is defined in terms of how O locates X in a space of possibilities.

This implies that O could attribute a different syntax to something else. I.e. in general, if X /= Y, then often, but not always:
   syntax(O, X) /= syntax(O, Y)

Also if X is an enduring object, O could assign it a different syntax at a different time, in which case we may need a time parameter
   syntax(O, X, t)

I'll ignore that below.

4. The space of possibilities within which X locates O may have any sort of topology: e.g. the possibilities may be points on a continuous line, a discrete line (an ordered set), a tree, a graph, an N-dimensional vector space, etc.

The space need not be homogeneous. E.g. some points may have far more neighbours than others.

5. Where the space is continuous, O may divide it into "chunks" using thresholds of some sort. (Or fuzzy chunks with fuzzy thresholds). That will define a new discrete syntax, derived from the previous continuous syntax.

This is just one of many ways in which a new syntax can be derived from an old one.

Another is by abstraction over a discrete syntax, e.g. grouping collections of items with different syntax into a single new category. If done repeatedly this can lead to a hierarchical, multi-level syntax. (Characters, words, phrases, clauses, sentences, paragraphs, stories, etc.)

6. O and X do not uniquely determine a syntax, since O may use different kinds of space in different contexts, or for different purposes. E.g. sometimes a derived syntax is more useful than a more basic syntax.

So we could use different expressions for the different syntaxes that O assigns:

syntax1(O, X)
syntax2(O, X)

7. The syntax X attributes to O could be *explicitly* represented in X (e.g. a label, a set of measures, a parse tree, a chart) or it could be *implicit* in the way X relates to O. E.g. the Pop-11 compiler never builds parse trees for Pop-11 expressions, unlike most compilers.

Even where the syntax attributed by O to X is implicit, the *processes* in O associated with perception of or manipulation of X may have a syntax explicitly or implicitly attributed to them by O. (E.g. a plan for creating X.)

The syntax of explicit syntactic descriptions

8. Where the representation of the syntax of X is explicit in O, it is another entity X' which may have an syntax attributed to it by O. This will generally be different from the syntax attributed to X. I.e.

SYNTAX(DESCRIPTION(SYNTAX(X)))

is different from

SYNTAX(X)

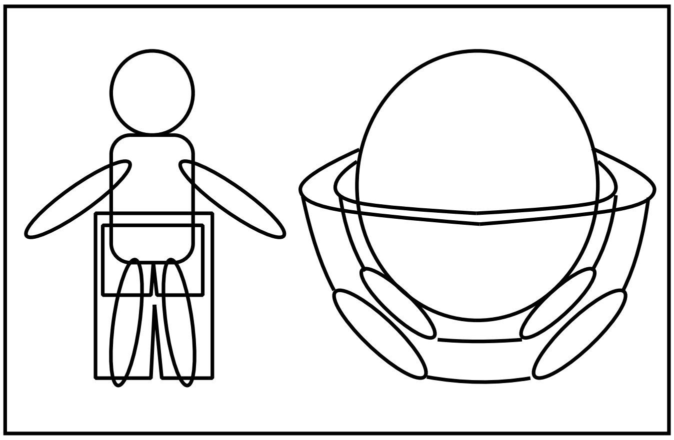
E.g. O may build two syntactic descriptions of X, S1, and S2, then compare them in order to decide which is better.

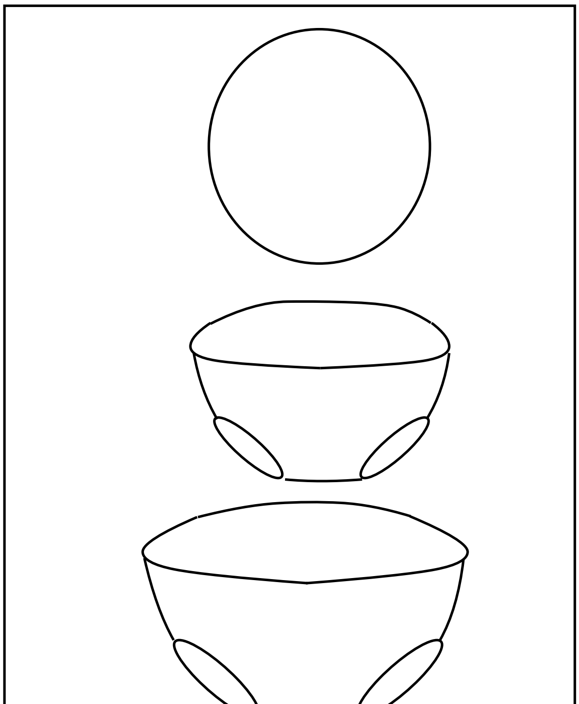
This may include building a third syntactic description of the relation between S1 and S2.

OUR ABILITY TO THINK METAPHORICALLY MAY BE AN OUTCOME OF THIS SORT OF THING: SEEING AN A AS A B, GENERALLY INVOLVES NOTING STRUCTURAL RELATIONSHIPS AND MAKING USE OF THEM.

BUT THE BASIC CAPABILITIES MAY HAVE EVOLVED MUCH EARLIER FOR THE PURPOSES OF PERCEIVING THE ENVIRONMENT IN A USEFUL WAY.

In how many different ways can Mr Bean remove his underpants without removing his trousers?





Concurrent mutual syntactic attribution

9. It is possible for another observer O2, to attribute a syntax to structures and processes in or produced by observer O1.

In some cases O1 and O2 can observe each other. Where the results lead to behaviour interesting feedback loops can result.

O2 and O1 may both be parts of the same larger system, e.g. a mind in which some components notice and respond to processes in other components.

Observers vs modes of processing

10. It may be that two or more observers, O1, O2, ... attribute the same type of syntax to objects X1, X2, X3, ... in a certain class, since they can cope with the same range of variation in the same way[*].

Then we can abstract from the observer and talk about the common syntax they all attribute to the objects in different contexts:

syntax1(X1), syntax1(X2), ...

syntax2(X1), syntax2(X2), ...

In this case we are treating syntax1, syntax2, as implicitly referring not to an *observer* but to a *mode of processing* that could be used by any (suitably equipped) observer, O1, O2, ...

---

[*]The points made here are inherently ambiguous because of the ambiguity of the notion "in the same way".

This notion of "sameness" can be defined extensionally or intensionally, and either way there are many fuzzy issues.

However, for any agreed notion of sameness there is an instance of this paragraph!

What is the benefit of having syntactic capabilities?

11. If O attributes a syntax syntax1(X) to X, then this may give O a new set of capabilities.

I.e. O may be able to DO **new things as a result.**

Some of them involve interpreting X as an enduring information bearer,

E.G. PICTURES, MAPS, SENTENCES,

Others involve interpreting X as a control signal – from O to something else, from something else to O, from O to O.

GETTING OTHER PEOPLE, OR PARTS OF YOUR BODY, OR SOME MACHINE, TO DO WHAT YOU WANT OR NEED TO HAVE DONE.

Others involve doing something with X itself

E.G. MANIPULATING IT, STORING IT, COPYING IT, USING IT AS AN INSTRUMENT, REPAIRING IT, EXPLAINING IT, ETC.

Some uses involve deriving more information by manipulating X: e.g. doing calculations, doing logical deductions.

For all those purposes you need to be able to perceive, and think about the structure of X, i.e. its syntax.

NOTE:

By building a taxonomy of such syntax-based capabilities we may hope to derive a well defined-ontology replacing/refining old unclear notions of meaning, semantics, pragmatics, etc., which will typically turn out to be special cases, some more useful than others.

How is it done?

12. How is syntactic attribution done?

Simple examples of computer programs that analyse structures of various sorts (linguistic, pictorial) have been demonstrated previously.

The ability to attribute a syntax to entities may be trivial in some cases

> **e.g. simple measurement, or a boolean detector**

or very sophisticated in others,

> **e.g. because it requires analysis and description at different levels of abstraction.**

As Chomsky, and others, have noted:

Different architectures support different syntactic capabilities.

What syntactic attributions an organism or machine can make, and what it can DO with its syntactic attributions, *depends on its architecture.*

Various special cases have been well studied

> **e.g. analysing the role of architectures with stacks for coping with arbitrarily nested structures in dealing with context free grammars.**

But why restrict ourselves to architectures for coping with a discrete, linear input stream?

THAT IS JUST A SPECIAL CASE.

We don't in general know what the range of syntax-processing architectures is,

e.g. what sorts can cope with non-linear structures, such as images, or changing scenes, or various kinds of continuous structures.

The need for new models of the system interface: New models of syntax-analysing architectures

Systems that can take in simultaneously a

- LARGE
- MULTI-DIMENSIONAL,
- CONSTANTLY CHANGING,

collection of data

- FROM MULTIPLE SOURCES,

  E.G. MULTIPLE RETINAL CELLS, TACTILE SENSORS, PROPRIOCEPTIVE SENSORS, AUDITORY SENSORS,

  POSSIBLY WITH CONTINUOUSLY CHANGING "READINGS",

and can process them all CONCURRENTLY

are not necessarily well modelled by

- SIMPLE,
- SINGLE STATE

automata, taking in streams of discrete symbols that have no intrinsic structure
(*as 2-D and 3-D input arrays have, for example*)

They are probably also not well modelled by "single state" dynamical systems, no matter how high the dimensionality.

DEDICATED, SPECIAL-PURPOSE PERCEPTUAL ARCHITECTURES ARE REQUIRED.

Multi-level concurrent syntax processing architectures

Some kinds of syntactic analysis, such as those involved in visual perception, require architectures that can process several levels of abstraction in parallel.

E.g.

- **finding edge-features, region fragments**
  MAY INVOLVE *searching* FOR GOOD WAYS TO SEGMENT
  OR GROUP THINGS LOCALLY
- **finding larger scale structures: lines, junctions, regions**
  USUALLY AMBIGUOUS: REQUIRING MORE SEARCH, OR SPECIAL
  SEARCH-AVOIDANCE MECHANISMS: E.G. CONSTRAINT PROPAGATION
- **finding relationships between these structures**
  AND SELECTING THE *useful* ONES AMONG MANY OTHERS
- **finding 3-D interpretations**
  E.G. DIFFERENT SORTS OF EDGES, DIFFERENT LOCAL
  SURFACE FEATURES
- **detecting larger scale 3-D structures**
- **seeing 3-D structures as *animate***
- **seeing motion (4-D structures)**

All of these may require a lot of “top-down” or “knowledge-driven” or “goal-driven” processing.

Evolution of syntactic processing capabilities

The mechanisms described above probably evolved long before external language as we know it.

Perhaps human speech processing uses modified versions of such multi-level concurrent analysis and interpretation mechanisms.

Many of the mechanisms are specifically geared to the syntactic structures accessible in the input

(E.G. 2-D, OR SMOOTHLY CHANGING, OR NETWORK-STRUCTURED)

They are not GENERAL PURPOSE **sensors and sensory processors.**

It is also worth noting that just because the capabilities are constantly *present* it does not follow that they are constantly *"turned on"*

THIS IS A KEY TO UNDERSTANDING THE NATURE OF ATTENTION.

An intelligent organism may have a whole armoury of mechanisms for attributing syntactic structure to perceived objects and selectively turn them on and off in the light of

- **current needs**
- **the nature of the current environment**
- **current knowledge or beliefs**

E.g. looking for your hammer in your garage when you think it is on one of a particular group of shelves.

TOWARDS SEMANTICS: WHAT IS MEANING?

WRONG QUESTION: *HOW MANY DIFFERENT KINDS OF MEANING ARE THERE?*

Additional architectural features are required if a syntactic characterisation is to be usable in association with a semantic interpretation.

What additional features?

We should explore the variety of types of architectures for organisms and robots with semantic capabilities, in order to get good ideas.

THEY ARE ALL CONTROL SYSTEMS.

A crucial common enabling feature:

The ability to use any kind of representation (neural, sybolic, map-like, or whatever) to represent something requires the ability to have something like *internal* sensors and motors.

I.e. it requires the ability to sense and to manipulate and use internal information-bearing structures.

Computers have internal sensory motor capabilities, for operating on bit patterns.
These can provide the infrastructure for far more sophisticated virtual machines with semantic capabilities

Compare the ways in which a computer can use bit patterns to refer

- **to memory locations (including special registers)**
- **to the contents of memory locations**
- **to instructions to be performed.**

Out of such simple internal "sensory motor" capabilities we know wondrous and diverse new capabilities can emerge.

- **compilers and interpreters for new languages**
- **operating systems**
- **plant control systems, flight control systems**
- **email systems and many internet capabilities**
- **rich virtual physical reality systems**
- **complex virtual machines with mind-like capabilities**

What is an information user (manipulator)?

VERY hard to define in general terms.

It's a type of "control system".

- **Something which reacts to *abstract* or *syntactic*, not just *physical*, properties, of physical structures**
- **in a context-sensitive way**
- **in a manner that preserves, achieves, prevents, modifies states (internal or external)**
- **where the states, and the mechanisms producing and reacting to them, form some kind of enduring (but possibly developing) system**
- **which has an "integrated" collection of capabilities related to the reactions, not all of which need be constantly active (they are activated only under certain conditions)**
- **some of which are concerned with producing/manipulating/ storing/using entities in a structured medium**
- **where some of the entities may be external to the user and some within the user.**
- **where some of the reactions, or behaviours, are external and some internal.**

'Reacting to syntax' means having reactions (possibly internal) that vary according to syntactic, not physical sub-categories, or properties.

THAT'S ALL VERY VAGUE.

IT COULD BE SEEN AS A FIRST CRUDE ATTEMPT TO CAPTURE A NOTION THAT COVERS A WIDE RANGE OF ORGANISMS AND ALSO MANY INTELLIGENT, MORE OR LESS AUTONOMOUS MACHINES.

I.e. things with minds, of many kinds, including very simple limiting cases (thermostats, micro-organisms) and also very rich and complex minds, like humans.

They are all instances of designs in the same huge, ill-understood 'design space'. (See later)

DEGREES AND KINDS OF INTEGRATION

A language-using capability could be disconnected from other capabilities.

E.g. there could be a module that reads in sentences and spews out parse trees, but does nothing with them.

Or a module that reads in questions and on the basis of pattern matching uses them to spew out answers.

The answers might be derived from a database of information.

All this requires no grasp of semantics (as normally understood): it could all be purely syntactic, with no understanding of what questions are for, why they need answers, nor any understanding of the content of the questions or the answers, as in Eliza.

WHAT MORE IS NEEDED IF THE SYSTEM IS TO 'UNDERSTAND'?

It is often assumed that the answer is: *Integration with sensors and motors*

But that is, at most, a requirement for understanding sentences that refer to things in the physical environment.

Understanding questions about arithmetic, or group theory, or the nature of logic, might be done without *any* current or past connection with physical sensors and effectors.

Moreover, what sorts of links to sensors and motors would suffice for a grasp of semantic content? Think of simple cases:

- **where a "toy" robot obeys commands.**
- **where Eliza occasionally uses a sensor reading in part of the answer**
- **where Eliza makes a face on a screen** SMILE **if the word "happy" turns up or** FROWN **if the word "angry" occurs.**

WHAT SORTS OF INTEGRATION ARE POSSIBLE BETWEEN SYNTACTIC CAPABILITIES AND OTHER KINDS?

HOW MANY VARIETIES OF INTEGRATION ARE THERE?

UNDERSTANDING THE PROBLEM IS OFTEN HARDER THAN FINDING SOLUTIONS

We often *think* we know what the problem is when we don't really.

E.g. What is perception for?

- **To learn what is out there (Marr)?**
- **Also: To grasp possibilities! (Gibson – affordances)**

More generally: think of perception as a biological process, with biological functions, integrated with the organism's needs and capabilities: it is NOT a mathematical reverse-camera (Marr).

VISION CAN BE USED SIMULTANEOUSLY FOR POSTURE CONTROL, FOR ROUTE SELECTION, FOR UNDERSTANDING A MECHANISM, FOR SEEING HOW TO..., FOR ENJOYMENT OF THE VIEW...

VISUAL INFORMATION GOES IN PARALLEL TO DIFFERENT SUB-MECHANISMS WHICH PROCESS IT DIFFERENTLY, EXTRACTING QUITE DIFFERENT KINDS OF INFORMATION, GEARED TO DIFFERENT NEEDS, DIFFERENT CAPABILITIES, AND USING DIFFERENT BACKGROUND KNOWLEDGE.

E.G. SPECIALISED TASKS LIKE MUSICAL SIGHT-READING.

"All perception is controlled hallucination" (Helmholtz).

(That goes for perception of syntactic structure also)

What sorts of things can humans and other animals learn?

Not just rules, or new weights in a neural net.

Many different things:

NEW NOTATIONS

NEW CONCEPTS, OR ONTOLOGIES

NEW GENERALISATIONS,

NEW MODES OF REASONING,

NEW PREFERENCES, VALUES, TASTES,

NEW GEOGRAPHICAL INFORMATION,

NEW FACES,

NEW MOTOR SKILLS,

NEW MUSICAL STYLES,

NEW WAYS OF LEARNING ....

WHAT ABOUT NEW SUB-ARCHITECTURES?
NEW ARCHITECTURAL LINKS?
(TRAINED SPORTING/ATHLETIC REFLEXES.)

**What sort of architecture can facilitate all this?
Learning theorists often think there's just one kind of learning: he who makes a hammer thinks everything is a nail.**

Likewise: What is language for? To communicate with?

NO! ... Well, that and other things.

Communication covers just a subset of types of uses of a subset of types of language.

What other things?

WHAT IS LANGUAGE FOR?
NOT JUST COMMUNICATION.

Languages (of many kinds) can be used

- To think with
- To desire, intend, deliberate, plan,
- To wonder why, wonder whether, wonder how...
- To play
- To create (many types of things)
- To perceive
- To do mathematics (not necessarily on paper, etc.)
- To rehearse what you may want to say or do later
- To remember (store information for future use)
- To reminisce or recall
- To categorise your own internal states, for yourself (called meta-management, below)
- To admonish, encourage, or deceive yourself
- For many forms of control

Several common mistakes about language.

1. Ignoring the varieties of internal information processing in organisms: in our pre-human ancestors, in other animals, in humans doing other things than verbal communication.

2. Investigating language, its functions, its structures, its semantic content, the processing mechanisms, without asking how linguistic mechanisms engage with other components of an integrated mental architecture, including cognitive, motivational, perceptual, learning, deliberating, problem-solving, planning and acting, capabilities.

3. (Corollary): It is a mistake to attempt to design and implement mechanisms that do language processing (e.g. resolving anaphoric reference, interpreting indirect speech acts, planning what to say and how to say it) *as if* these were tasks to be performed by a linguistic mechanism *in isolation*: ignoring all the other components of the architecture.

The contexts of linguistic processing

EXAMPLE:

IF GENERAL PURPOSE SKILL-LEARNING MECHANISMS CAN INDUCE THE FUNCTION OF INDIRECT SPEECH ACTS OR METONYMIC EXPRESSIONS IN A CONTEXT, THEN THERE MAY NOT BE ANY NEED FOR PURELY LINGUISTIC MECHANISMS TO DERIVE THE INTERPRETATION (MOST OF THE TIME, ANYWAY).

So you can't investigate language processing without investigating perception, problem solving, learning.
E.g. we can learn a 'sub-language' during a conversation.

Without a broad survey of connections between linguistic and other capabilities, we will be confused about, for instance what "meaning" is, how to define "semantics".

We need to ask some hard questions:

- **What other mechanisms are there in animal (robot, softbot) minds, which use or engage with various kinds of 'linguistic' mechanisms?**
- **Which biological mechanisms were precursors of linguistic mechanisms? E.g. planning mechanisms need to be able to cope with 'nested' symbolic structures in various ways.**

  Sensory mechanisms need to be able to communicate with various cognitive, motivational, and motor mechanisms: how many of these forms of communication use mechanisms providing infrastructure for overt linguistic communication?

  Memory is a way for an organism to communicate with itself at another time. How much of that apparatus is relevant to communication between organisms?
- **What kinds of *syntax* (structural variability), internal *pragmatic* functions, and *semantic* capabilities preceded the evolution of overt linguistic communication.**

Don't assume that the semantics of linguistic structures depend only on how they relate to the operation of perceptual and motor systems.

There are also *internal* architectural information needs.

AI used to be mainly about representations and algorithms

Now questions about architectures are seen to be equally (or more) important

WHY?

- **We need to know how to put diverse components together in a working system.** WHAT DOES 'WORKING' MEAN?
- **It's very likely that we cannot understand the evolution of mind without understanding the *co-evolution* of components.**
- **We now have a small set of ideas about architectures (including our 'Cogaff' architecture, described below)**
- **And a variety of tools for investigating and using them Soar, Cogent, PRS, ACT-RPM, Jack, Sim_agent .....**
- **But the space of architectures is huge and ill defined.**

  We have a lot more exploring to do if we wish to understand its properties.

WILL FORMAL MODELS OF SYSTEMS TELL US WHAT WE NEED TO KNOW?

If they are too complex we cannot understand them without building and playing with working instantiations. We have to think like engineers to understand our models.

NOTE:
Turing machines are largely irrelevant. They are discussed far more by people who attack AI than by people who do AI.

A Turing machine is not a good model for an arbitrary information processing architecture.

Why? See
http://www.cs.bham.ac.uk/~axs/misc/turing-relevant.html

We need some good organising ideas.

Many people produce architecture diagrams, and then tell stories about how they work.

But we need to look for good organising principles.

We also need to identify CONSTRAINTS **to narrow our search.**

Obvious constraints:

- **being physically possible**
- **being tractable/feasible**
- **what is implementable on biological mechanisms**
- **being suited to the functional requirements**
  BUT WHAT ARE THEY?
- **more subtle constraint: “what is evolvable”.**

(Beware of fashionable constraints: ‘groundedness’, ‘embodiment’...)

Deep understanding will not come from studying ONE case e.g. a typical adult human mind!

We need to explore alternatives, understand trade-offs.

Let's look at neighbourhoods

- **in design space**
- **in niche space**

and learn from their similarities and differences.

We need to understand different types of *trajectories* through these spaces, in evolution, in individual development, in learning, in cultural change, ...

We need to understand the interactions between the trajectories, i.e. *the many feedback loops* in co-evolution.

We need to understanding architectures not only for individuals, but for sub-mechanisms and for larger structures:

FAMILIES, TEAMS, PAIRS FIGHTING, ECONOMIC SYSTEMS, ECO-SYSTEMS.

No bit of this will be fully understood without putting it in the context of the rest.

Some philosophy of science

NB: Don't assume that good scientific theories need to be empirically falsifiable.

What is more important is explanatory power: having rich, diverse, objectively derivable consequences.

The consequences may be of the form: *X can occur*, or *X can exist*.

Such propositions are not empirically falsifiable.

> ***Universally quantified propositions are falsifiable, not existentially quantified propositions.***
>
> ***A statement about what is necessary can be falsified, but not a statement about what is possible.***

But the unfalsifiable statements can be supported by examples!

Even ONE **well-attested case demonstrates that there is something that needs to be explained.**

The most important advances in science are not discoveries of LAWS **but** EXTENSIONS OF ONTOLOGIES**.**

E.g. there are electrons, protons, electric fields, valences, genes, grammars, parsers ....

See chapter 2 of THE COMPUTER REVOLUTION IN PHILOSOPHY **(1978)**

Some important themes

1. Biological minds are evolved control systems

- **They control many things in parallel:**

  PERCEPTUAL PROCESSES ***(assigning syntax to the environment)***
  MOTOR PROCESSES ***(using the syntax of possible actions)***
  LEARNING (OF MANY KINDS)
  MOTIVATION
  MOODS
  EMOTIONS
  LANGUAGE PROCESSES
  BODILY FUNCTIONS
  LONG TERM GOALS, PREFERENCES, VALUES, STANDARDS, ETC.

- **Most of what is going on is unconscious, so don't expect introspection to be very informative: it is one tool among many.**

  MOST OF WHAT'S MENTAL IS UNCONSCIOUS.

- **Use the design stance not the intentional stance**
- **Don't expect the tools and concepts of control engineers to be adequate,**

  .... nor those of any other single discipline

2. All evolution is co-evolution, including:

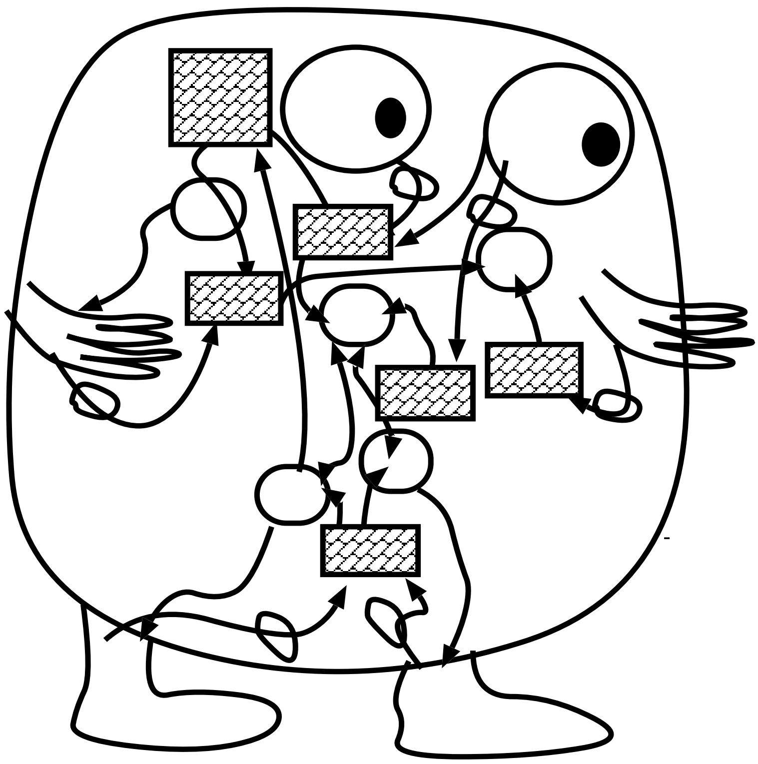
- **co-evolution** BETWEEN **types of organisms**
- **co-evolution** WITHIN **organisms (compare Popper 1976)**
  You are an ECO-SYSTEM of mind (not just a SOCIETY of mind)

For most of AI and Cognitive Science (and philosophy and brain science) one of the main reasons why progress is slow is that we don't yet know what the problems are.

There is much conceptual confusion, which can be reduced by exploring architecture-based concepts: architecture-based mental ontologies.

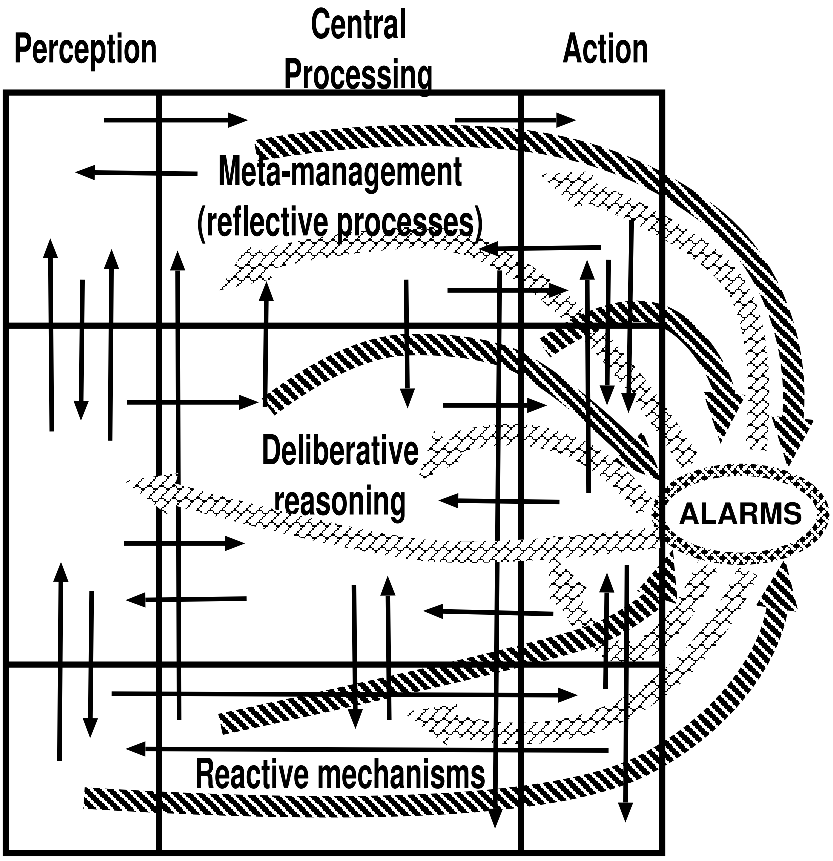
WHAT SORT OF ARCHITECTURE CAN ACCOUNT FOR SUCH PHENOMENA?

COULD IT BE AN UNINTELLIGIBLE MESS?



Yes, in principle.
However, it can be argued that evolution could not have produced totally non-modular yet highly functional brains. **Problems about search apply as much to evolution as to engineering design.**

The (Birmingham) 'CogAff' Architecture (A partial view)



This view of the architecture is motivated by superimposing

(a) the 'triple tower' (input-central-output) view

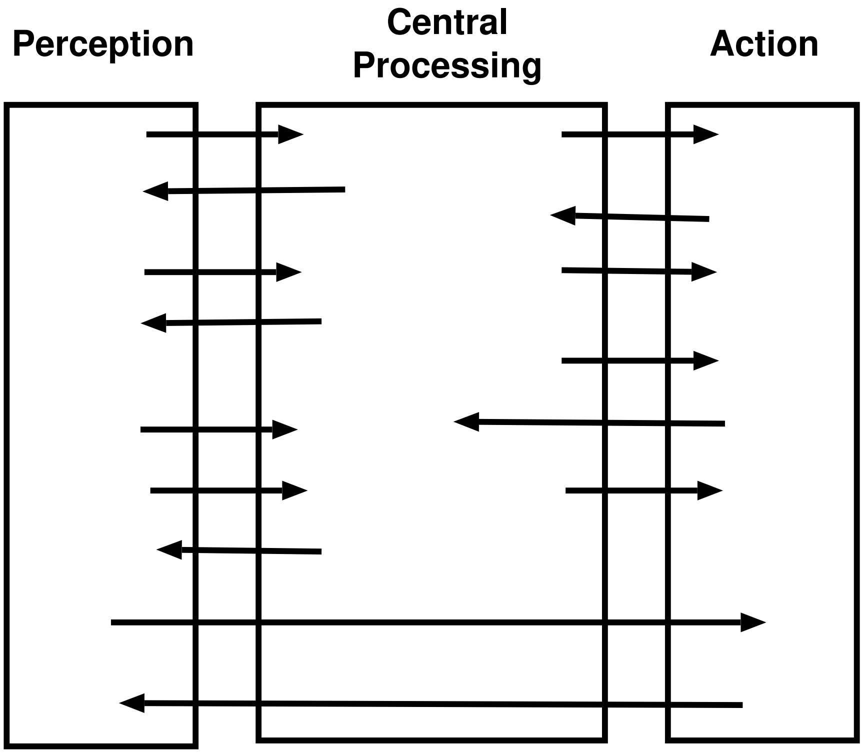
and

(b) 'triple layer' (three stages of evolution) view

and adding an alarm mechanism.

Missing additional components are described later.

The “triple tower” View



Note that perceptual capabilities can include sophisticated syntax-attributing capabilities.

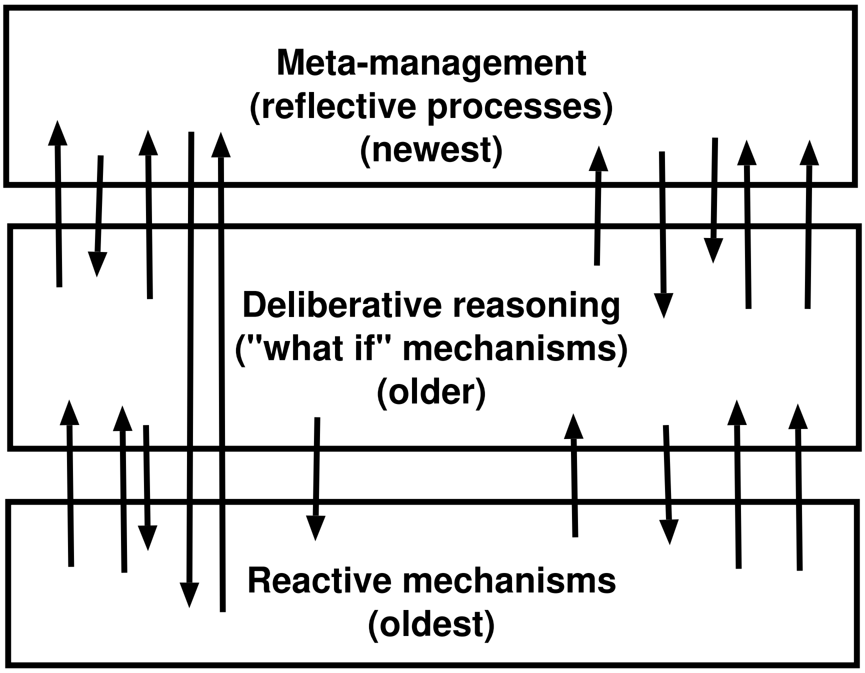
There are many variants of such models: (NILSSON, ALBUS)

We need to understand the design options and the requirements.

Systems can be “nearly decomposable”.

Boundaries between sub-systems can change with learning and development.

ONE OF MANY LAYERED VIEWS



Many variants, but with different subdivisions and interpretations of subdivisions

Compare: “triune brain”:
reptilian, old mammalian, new mammalian.

(See Models of Models of Mind paper in DAM symposium proceedings, in Cogaff directory.)

WHAT SORTS OF LAYERS

Different principles of subdivision for layers:

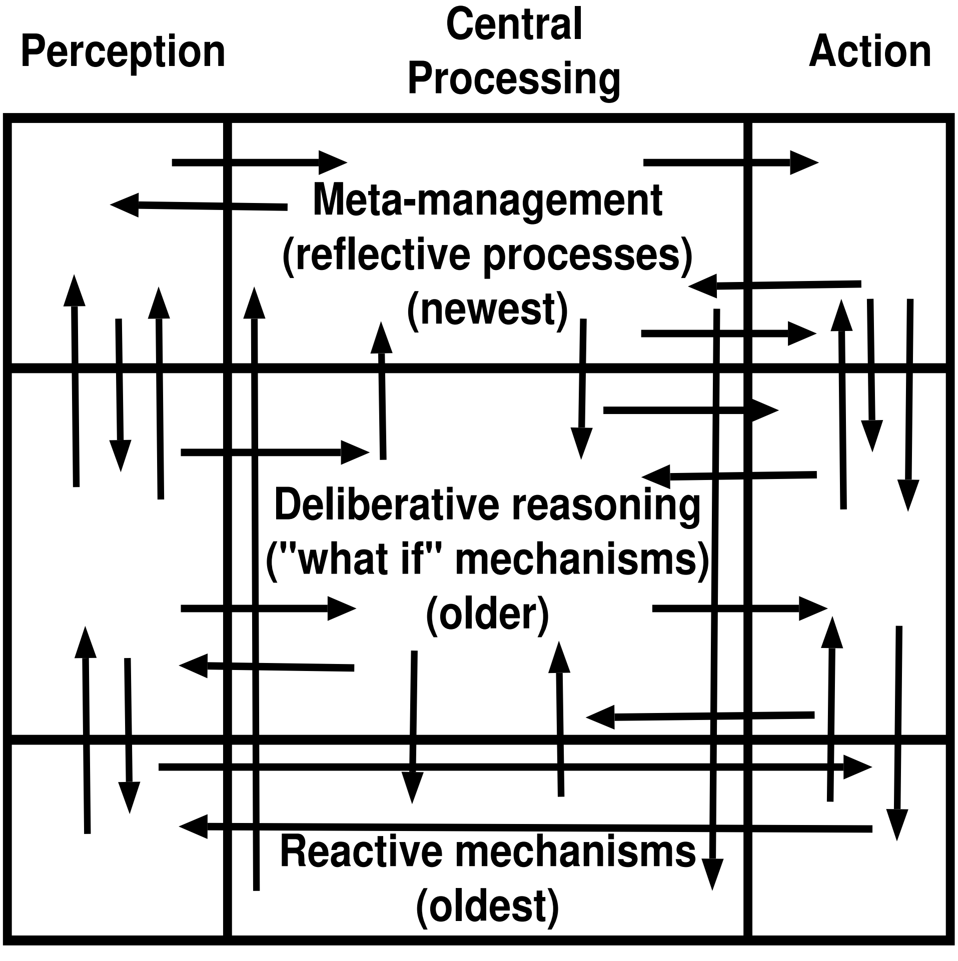
- **control-hierarchy,**
- **information flow (data upwards, control downwards?)**
  (e.g. the 'Omega' Ω model of information flow)
- **abstraction,**
- **sophistication of processing**
- **evolutionary**

Or some combination.

The Cogaff model emphasises the last three

COMBINING THE VIEWS: LAYERS + PILLARS = GRID

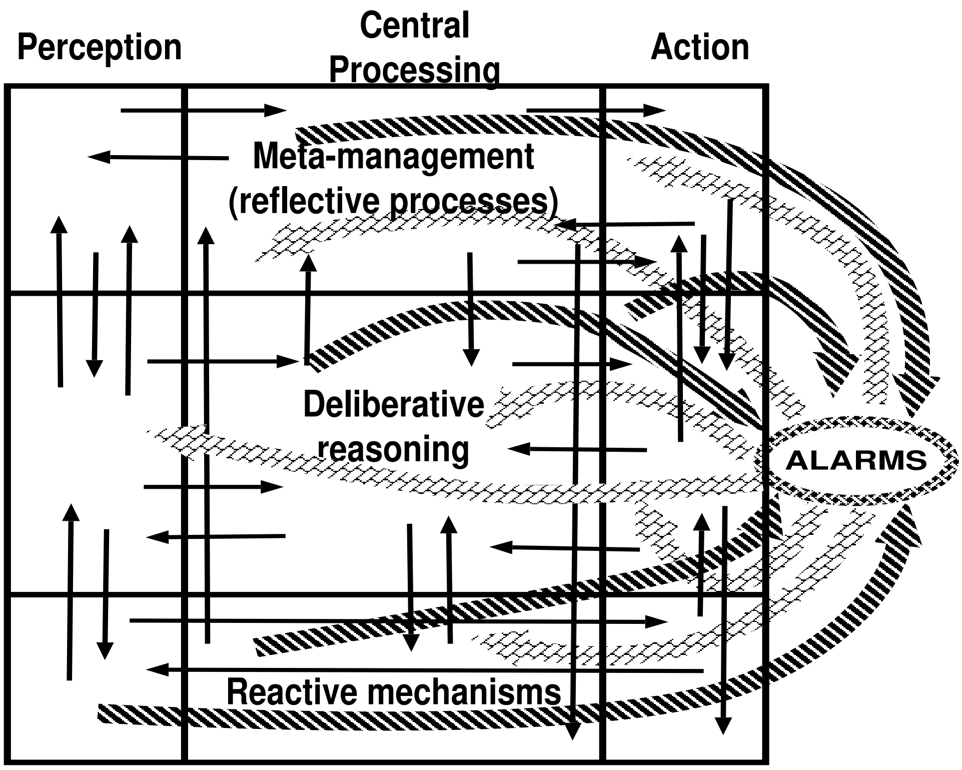
A grid of co-evolving sub-organisms, each contributing to the niches of the others.



As processing grows more sophisticated, so it can become slower, to the point of danger.

FAST, POWERFUL,
"GLOBAL ALARM SYSTEM"
NEEDED

IT WILL INEVITABLY BE STUPID!

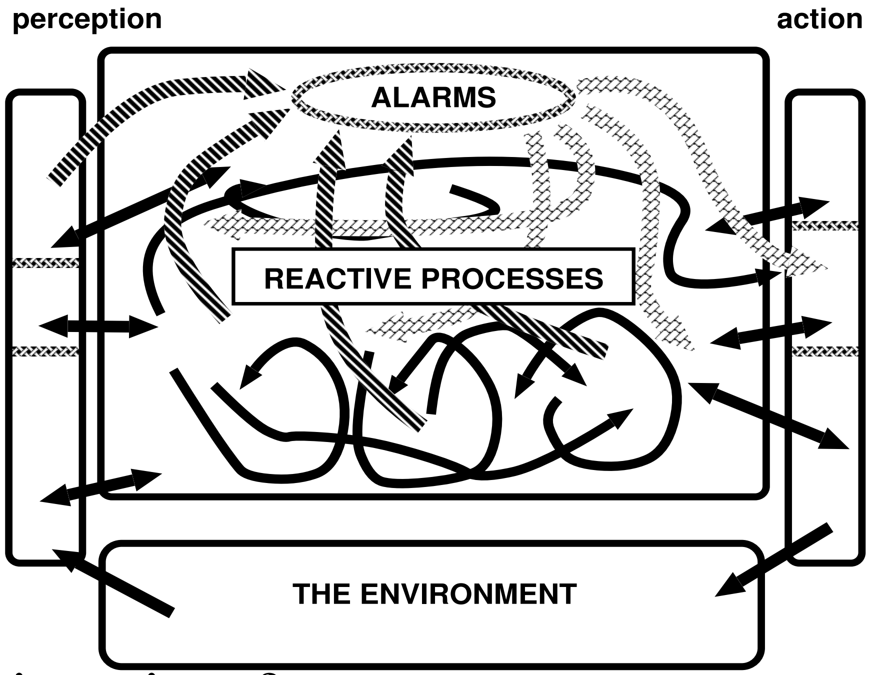


MANY VARIANTS POSSIBLE.

E.g. one alarm system or several?

(Brain stem, limbic system, ...???)

NOT ALL PARTS OF THE GRID ARE PRESENT IN ALL ANIMALS



How to design an insect?

Will a purely reactive architecture suffice?

YES, for many purposes.

Reactive systems can be arbitrarily sophisticated in their "externally observable" behaviour.

But there are trade-offs.

- **How long it takes to evolve all the behaviours**
- **Storage required**

Deliberative mechanisms are far more complex, but have different trade-offs.

ALARM MECHANISMS IN REACTIVE SYSTEMS

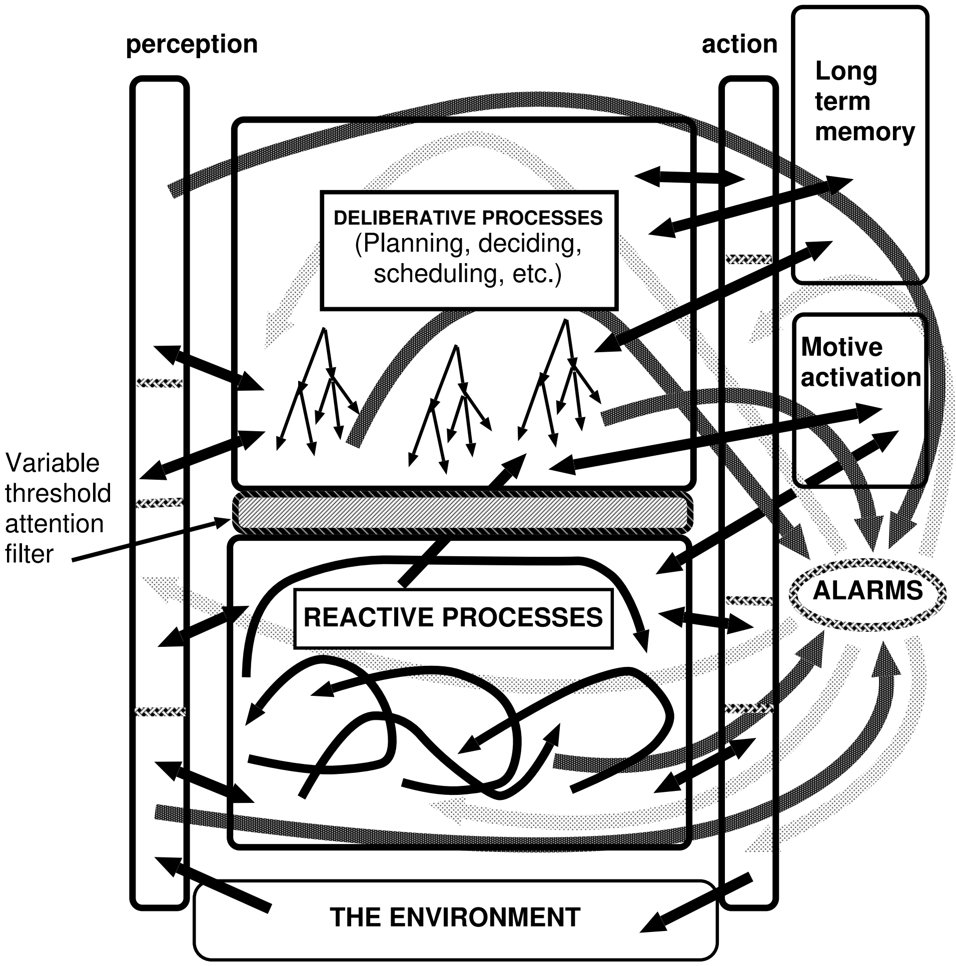
(Global interrupt/override):

- **Allows rapid redirection of the whole system, for sudden dangers or sudden opportunities**

- FREEZING
- FIGHTING, ATTACKING
- FEEDING (POUNCING)
- GENERAL AROUSAL AND ALERTNESS (attending, vigilance)
- FLEEING
- MATING
- MORE SPECIFIC TRAINED AND INNATE AUTOMATIC RESPONSES

Closely related to what Damasio and Picard call "Primary Emotions"

Reactive and deliberative layers with alarms



Resource limits in a slow deliberative layer may make some sort of attention filter necessary,

so that deliberative processes are not constantly being redirected

by new motives or other distractions from the reactive mechanism or from perceptual systems.

Prerequisites for deliberative layers.

The importance of compositional syntax

The need for temporary re-usable workspace for creating descriptions of hypothetical machines

The diverse roles of goals (motives, preferences, evaluations, ...)

The inherent sequentiality and discreteness of processes

New types of long term memory, compared with reactive systems e.g. chunked associations

THESE REQUIREMENTS COULD CAUSE EVOLUTIONARY PRESSURE FOR NEW DEVELOPMENTS IN PERCEPTUAL AND MOTOR SUBSYSTEMS:

The mind as an eco-system

What are the uses of ‘what if’ reasoning?

Explaining observed events (why did the window break)

Hypothesising invisible (e.g. occluded) parts of visible objects.

Understanding how something works

Predicting what might happen in the future (e.g. what X will do, what the consequences of Y’s (or my) actions will be).

Making plans to achieve goals in the future

Building scientific theories

Doing mathematics

Many forms of creativity

The *content* of ‘what if’ reasoning may be

- **about the future**
- **about the past**
- **about how things might have been different now**
  (I MIGHT HAVE BEEN TALKING ABOUT PROGRAMMING)
- **about a quite different context from the current physical environment**
  (WHAT IS HAPPENING AT HOME, OR IN X’S MIND)

Compare work on metaphor at Birmingham: John Barnden’s ATT-META system.

RELATIONS BETWEEN REACTIVE AND DELIBERATIVE MECHANISMS

There are many details of the relationship still to be worked out.

1. Deliberative mechanisms will be implemented in reactive mechanisms of some sort.

2. Reactive mechanisms can interrupt and disturb deliberative mechanisms, and generate goals for deliberative mechanisms, e.g. by detecting a need for food, or warmth, etc.
(ITI S NOT ALL TOP-DOWN CONTROL.)

3. Deliberative mechanisms can harness reactive mechanisms in the execution of sub-steps in plans.

4. Noticing patterns in reactive behaviours may lead to new deliberations (though this self-observation needs a form of meta-management).

5. Repeated plan execution under the control of a deliberative layer can *train* a reactive layer so that later it can perform the same actions autonomously (faster, more smoothly, with fluency, but with less scope for variation).

THIS IS A "SKILL COMPILER" MECHANISM

Although all this fits our everyday experience, many details of the enabling architecture have still to be worked out.

SOME DIFFERENCES BETWEEN THE LAYERS

1. Reactive systems can be highly parallel, very fast, and use analog circuits, e.g. for tight feedback control loops.

Deliberative systems are inherently slow, serial, discrete, knowledge-based, resource limited. (Why?)

2. Reactive systems can record current state and immediately required actions, using simple representations,

E.G. MEASUREMENTS, LABELS, VECTORS, "FLAT" DESCRIPTIONS.

Deliberative mechanisms provide 'What if' reasoning and new kinds of representation capabilities supporting more complex and variable types of syntax. (Part of the reason why a re-usable general purpose workspace is required.)

3. Learning in reactive layers is mostly restricted to changing weights in an existing architecture, and perhaps storing new associations (e.g. in neural nets).

Learning in deliberative layers can include creating new structures using an old formalism, and developing new formalisms, new ontologies, new theories

4. New plans in a reactive system come from chaining associations through repeated behaviour

Deliberative mechanisms can create a new plan in a single process prior to execution.

Meta-management adds reflective abilities

If a system can benefit from observing the environment in which it acts, it can also benefit by observing its internal states and processes.

It can learn, or improve its internal processes (e.g. deliberating, solving problems) if it can observe, categorise, evaluate, and control internal processes.

EXAMPLE: DETECTING REDUNDANCIES OR LOOPS IN PLANNING OR PROBLEM SOLVING.

Meta-management requires architectural enhancements

allowing internal processes to be "non-intrusively" monitored, evaluated and controlled

- **Will use a mixture of reactive and deliberative mechanisms**
- **May need new formalisms to represent internal states**
- **Will certainly need new concepts for classifying, describing**
- **Some of this may include evaluation of motives, as well as internal behaviours.**
- **Scope for strong cultural influences**

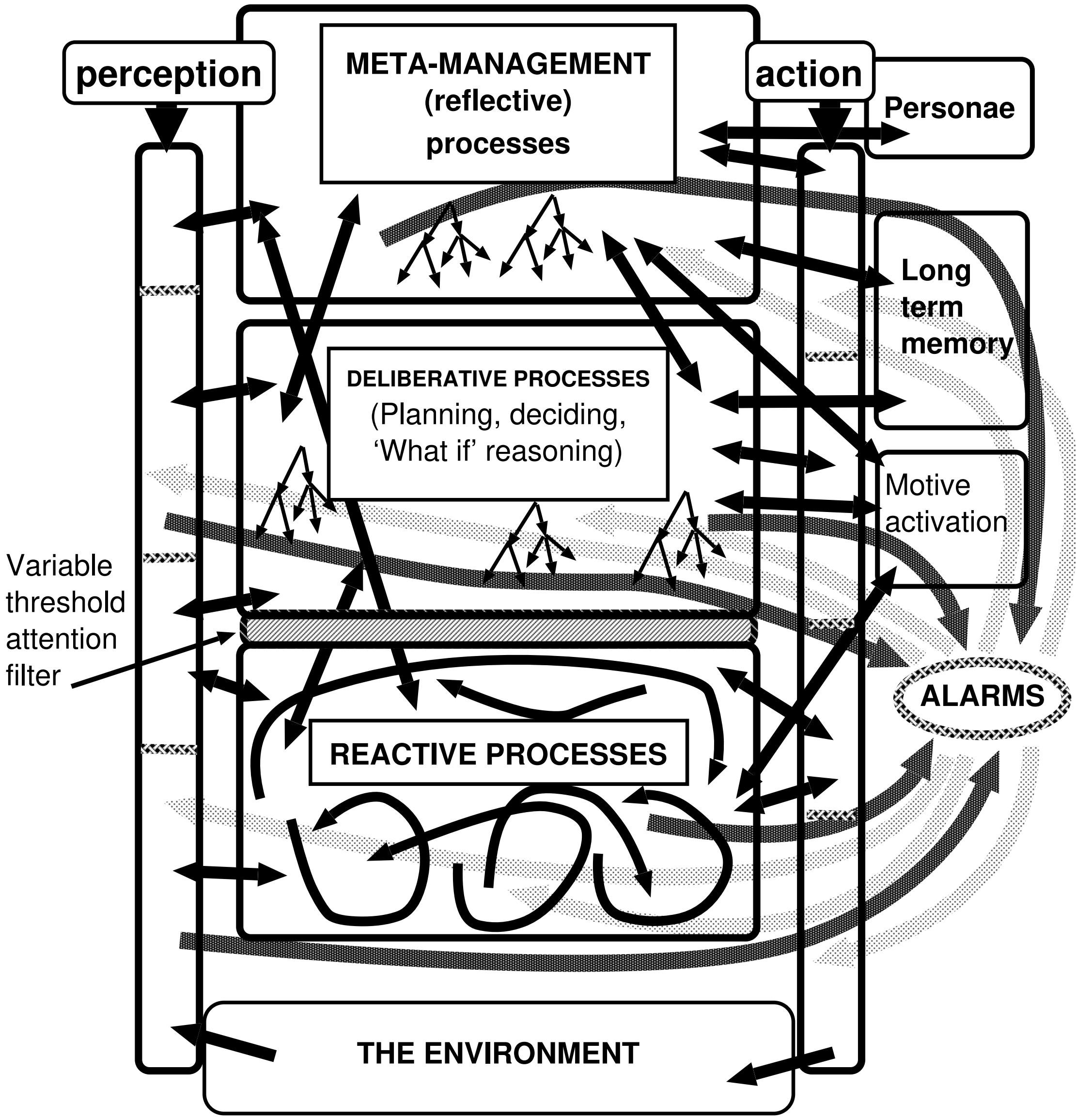
NOTE: the concept of "executive function" in clinical usage or psychological theory combines/confuses the deliberative and meta-management layers.

We know they are different because many AI systems have deliberation without meta-management.

Frontal lobe damage in humans can produce similar effects (Phineas Gage).

(Antonio Damasio: *Descartes' Error* 1994)

METAMANAGEMENT WITH ALARMS



Note the need to allow different *sub-personalities* to be "in control" at different times: variable modes of thinking, reacting, behaving.

Why does a normally pleasant and kind person sometimes exhibit road rage, or tyranny over subordinates?

METAMANAGEMENT AND SELF-CONSCIOUSNESS

Many architectures provide a type of consciousness: any perception of the environment to gain useful information is a type of consciousness, a type of awareness.

Meta-management provides a new type of self-awareness.

Or rather several types, depending on which kinds of internal states are observed.

This can include observation of intermediate databases in perceptual subsystems:

COULD THAT EXPLAIN WHAT SENSORY QUALIA ARE?

VARIETIES OF MOTIVATIONAL SUB-MECHANISMS

MOTIVATION IS NOT JUST ONE THING

Motives or goals can short term, long term, permanent.

They can be triggered by physiology, by percepts, by deliberative processes, by metamanagement.

So there are many sorts of motive generators: MG

However, motives may be in conflict, so motive comparators are needed: MC.

But over time new instances of both may be required, as individuals learn, and become more sophisticated:

Motive generator generators: MGG

Motive comparator generators: MCG

Motive generator comparators: MGC

and maybe more:

MGGG, MGGC, MCGG, MCGC, MGCG, MGCC, etc ?

THERE ARE ALSO “EVALUATORS”

Current state can be evaluated as good, or bad, to be preserved or terminated.

These evaluations can occur at different levels in the system,

and in different subsystems,

accounting for many different kinds of pleasures and pains.

(Often confused with emotions.)

They are also a crucial part of mechanisms of learning involving positive or negative reinforcement.

A full analysis of pleasure and pain is not yet possible, since details of the roles of various kinds of evaluation have yet to be worked out.

Where are the motive generators and the evaluators?

All over the system –

not just at the ‘top’ of a multi-layered processing architecture

(where some people put “the will”)

CONTRAST THE OMEGA MODEL. (Ω)

Different architectural layers support different sorts of mental phenomena and help us define

AN ARCHITECTURE-BASED

ONTOLOGY OF MIND

with diverse roles for ‘languages’

Different animals will have different mental ontologies

Humans at different stages of development will have different mental ontologies

Many of our mental concepts are DEEPLY confused and/or "cluster concepts"

Examples:

- **E.g. 'emotion', 'consciousness', 'representation', 'computation', 'understanding', 'semantics', 'freedom', 'self'.**
- **We can use architecture-based concepts to refine and extend some of them.**
  Others may be worthless (e.g. a soul that can exist independently of any physical implementation).
- **Compare physics: the architecture of matter underpins the periodic table of the elements and the structural possibilities in chemical formulae.**
- **But beware: there's not just one architecture for mind**
  We need COMPARATIVE **investigations**

I.e. collect examples of many types of real phenomena.
Try to build a theory which explains them all!

Subject to constraints from neuroscience, psychology, biological evolution, feasibility, tractability, etc.
(As explained previously)

Our theories should allow for variation across types of minds

- **variation across species,**
- **variation within species,**
- **variation within an individual during normal development**

  (INFANTS, TODDLERS, CHILDREN, TEENAGERS, PROFESSORS...)
- **variations due to brain damage**
- **variation across planets**

  (GRIEVING, INFATUATED, MARTIANS?)
- **variation across the natural/artificial divide**

  ROBOTS INTERESTED IN PHILOSOPHY

ANYONE WHO COMES UP WITH ONE ARCHITECTURE FOR MINDS HAS PROBABLY GOT IT WRONG!

Architecture-based concepts of mind

Within each architecture expect to find families of concepts where you previously thought there was one.

To generate such families,

consider the possible states and processes that can occur within the architecture

and the possible relations between them.

Look for explanatory clusters.

- **different kinds of learning —** MANY **kinds**
- **many notions of consciousness (and qualia)**
- **different sorts of beliefs, intentions, desires**
- **different types of languages, different types of syntax and semantics**
- **different sorts of emotions**

  primary, secondary, tertiary emotions (and more to come)
- **different kinds of moods, motivations, attitudes, personalities,**

COMPARE THE ARCHITECTURE OF MATTER

- **the periodic table of the elements**
- **the variety of types of chemical compounds**
- **the variety of types of chemical processes**

But there is only one physical (chemical) world whereas there are many types of minds, each supporting different collections of concepts of mentality.

AN ALARM MECHANISM
(BRAIN STEM, LIMBIC SYSTEM?)
ALLOWS RAPID REDIRECTION OF THE WHOLE SYSTEM.

Can be triggered by and can redirect reactive AND deliberative processes.

ALARMS IN A HYBRID ARCHITECTURE

- **Freezing, fleeing, arousal etc. as before**
- **Becoming apprehensive about anticipated danger**
- **Rapid redirection of deliberative processes.**
- **Relief at knowing danger has passed**
- **Specialised learnt responses: switching modes of thinking.**

Damasio & Picard:

cognitive processes trigger "secondary emotions".

We can now distinguish different sub-categories, e.g.

- ***purely central* secondary emotions**
- ***partly peripheral* secondary emotions.**

and many more perhaps

ON SOME (MISGUIDED) THEORIES, THE FORMER ARE IMPOSSIBLE!

META-MANAGEMENT AND TERTIARY EMOTIONS

Tertiary emotions (previously called "perturbances") involve interruption and diversion of thought processes.

I.e. the metamanagement layer does not have complete control.

WHY?

- **New information from other sub-systems can cause interrupts.**
- **New motives from other subsystems can cause interrupts.**
- **Global alarm signals triggered by events elsewhere can cause interrupts and re-direction.**

VARIABLE THRESHOLD INTERRUPT FILTERS CAN HELP REDUCE THESE EFFECTS.

Sometimes meta-management seems to be 'turned off', e.g when we are totally absorbed in some task.

QUESTION:
Is it essential that all sorts of emotions have physiological effects outside the brain, e.g. as suggested by William James?

NO: which do and which do not is an empirical question, and there may be considerable individual differences.

Some tertiary emotions may be purely central.

Different emotions associated with different layers

The REACTIVE **layer with** GLOBAL ALARMS **supports "primary" emotions:**

- **being startled**
- **being disgusted by horrible sights and smells**
- **being terrified by large fast-approaching objects?**
- **sexual arousal? Aesthetic arousal ?**

  etc. etc.

The DELIBERATIVE **layer enables "secondary" emotions (cognitively based):**

- **being anxious about possible futures**
- **being frustrated by failure**
- **excitement at anticipated success**
- **being relieved at avoiding danger**
- **being relieved or pleasantly surprised by success**

  etc. etc.

Note the different *linguistic* (syntactic, semantic, pragmatic) preconditions for primary emotions and for secondary emotions.

NB Not all emotions are functional: some are emergent side-effects of the operation of functional components of an architecture.

Disruptions of the third layer produce characteristically human emotions

i.e. the types poets, novelists and playwrights write about.

WE CAN EXPLAIN SOME DISPUTES AND CONFLICTING DEFINITIONS E.G. of “emotion”

Different researchers focus on different features of a very complex system.

But they are unaware of the other features.

Like the proverbial collection of blind men all trying to say what an elephant is:

- **One feels the trunk**
- **One feels a tusk**
- **One feels an ear**
- **One feels a leg**
- **One feels the tail**

etc.

They are all right — about a tiny part of reality.

We need to aim for a more comprehensive picture.

We do not yet understand much about architectures

- **how many types they are**
- **what the trade-offs are**
- **how they evolve and develop**
- **how they differ among animals**
- **how purely software architectures will differ**
- **how many kinds of learning there are**

We need architecture-based concepts of learning and development.

ADDITIONAL COMPONENTS

EXTRA MECHANISMS NEEDED

personae (variable personalities)

attitudes **standards & values**

formalisms **categories** **descriptions**

moods (global processing states)

motives **motive comparators**

motive generators (Frijda's "concerns")

Long term associative memories

attention filter **skill-compiler**

MANY PROFOUND IMPLICATIONS

e.g. for kinds of development

kinds of perceptual processes

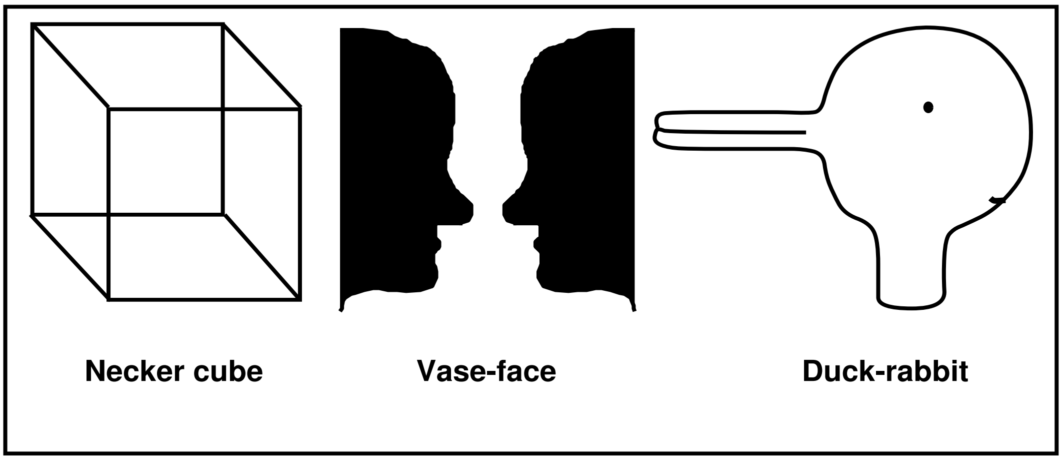
kinds of brain damage

kinds of emotions

SENSING AND ACTING 'PILLARS' CAN BE ARBITRARILY SOPHISTICATED

- **Don't regard sensors and motors as mere transducers. They can have sophisticated information processing architectures.**
- **Both inherently involve syntax as defined previously**
- **Perception and action can both be hierarchically organised with concurrent interacting sub-systems.**
- **Perception goes far beyond segmenting, recognising, describing what is "out there". It includes:**
  - **providing information about *affordances* (not Marr, but Gibson)**
  - **triggering physiological reactions (e.g. posture, sexual responses)**
  - **evaluating what is detected,**
  - **triggering new motivations**
  - **triggering "alarm" mechanisms, ...**

Extending JJ Gibson's theory



Different perceptual sub-systems use different affordances, and different ontologies. LIKE DIFFERENT SUB-ORGANISMS

Different levels of perceptual abstraction required for different purposes. When a necker cube flips only geometrical properties and relationships change. When the others flip, the changes are more subtle and go beyond geometric and physical properties.

(Evidence from brain damage: selectively disabled sub-competences.)

See also: Sloman 1989 (In Journal of Theoretical and Experimental AI)

Compare ACTION **layers: low level motor control vs plan schema activation vs social interaction.**

THE THIRD LAYER ENABLES SELF-MONITORING, SELF-EVALUATION and SELF-CONTROL (and qualia!)

What kind of machine can have emotions?

PROBLEM:

MANY different definitions of "emotion", in psychology, philosophy, neuroscience . . . with many variants within each discipline

DIAGNOSIS:

Different theorists concentrate on different phenomena. We need a theory that encompasses all of them.

REPHRASE:

What are the architectural requirements for human-like mental states and processes? (Definitions can come later)

Metamanagement and emotions

The third layer makes possible "tertiary" emotions, involving loss of control of thoughts and attention:

- **Feeling overwhelmed with shame**
- **Feeling humiliated**
- **Aspects of grief, anger, excited anticipation, pride,**
- **Being infatuated, besotted and many more**
  typically HUMAN **emotions. (Contrast attitudes.)**

NOTES:

1. Different aspects of love, hate, jealousy, pride, ambition, embarrassment, grief, infatuation can be found in all three categories: primary, secondary and tertiary emotions.

2. Remember that these are not STATIC **states but** DEVELOPING **processes, with very varied aetiology.**

The meta-management layer need not have constant contents

Different 'personalities' (personae) in different contexts

- **At home with the family**
- **Driving on a motorway**
- **Interacting with subordinates at work**
- **Being interviewed by superiors**
- **In the pub with chums**

...and many more ...

WHERE CONTROL BY A PERSONALITY INVOLVES TURNING ON A LARGE COLLECTION OF:

- **skills,**
- **styles of thought and action,**
- **types of evaluations,**
- **decision-making strategies,**
- **reactive dispositions,**
- **....**

COMPARE THE MUCH FASTER GLOBAL CHANGES PRODUCED BY ALARM MECHANISMS: PERHAPS AN EVOLUTIONARY PRE-CURSOR OF METAMANAGEMENT?.

Meta-management and language

As sophistication of self-monitoring, self-evaluation, self-control develops,

the linguistic and conceptual requirements for performing that task also develop.

Some of the linguistic development is stimulated through social and cultural interaction.

However, as the linguistic competence grows, so does the ability to absorb more information relevant to meta-management from other individuals.

I.e. there's a positive feedback loop.

The meta-management system is a framework which can be occupied by different 'control regimes' at different times?

THIS REQUIRES

- **A store of 'personalities'**
- **Mechanism for acquiring and storing new ones and modifying extending old ones**
- **Mechanisms for 'switching control' between personalities.**

WHAT FOR?:

Different contexts have different requirements.

Global switching triggered by context may be more effective than always having to select individual rules, strategies, information items etc. on the basis of

TASK + LOCAL CONTEXT + GLOBAL CONTEXT

In people switching personality is often involuntary and even unconscious (i.e. unnoticed).

WHY?

Can we learn to be more self-aware?
What needs to change?

META-MANAGEMENT AND SOCIAL CONTROL

A SOCIETY OR CULTURE CAN INFLUENCE INDIVIDUALS

E.G. by

- **Training reactive mechanisms**
  e.g. using reinforcement learning.
- **Enabling successful plans, strategies, etc. to be transferred without having to be rediscovered.**
- **Training modes of coordination in collaborative activities,**
- **Transferring powerful formalisms**
- **Transferring useful modes of categorisation, ontologies** (including ontologies of mental phenomena)
- **Influencing evaluation mechanisms**
  including evaluating internal events, actions
  (e.g. I was selfish, selfless, brave, stupid, wise, lucky)

THIS CAN BE USEFUL OR HARMFUL:

E.G. RELIGIOUS INDOCTRINATION WHICH PRODUCES GUILT ABOUT NATURAL HEALTHY DESIRES, ETC.

SOCIALLY IMPORTANT HUMAN EMOTIONS INVOLVE RICH CONCEPTS AND KNOWLEDGE and RICH CONTROL MECHANISMS

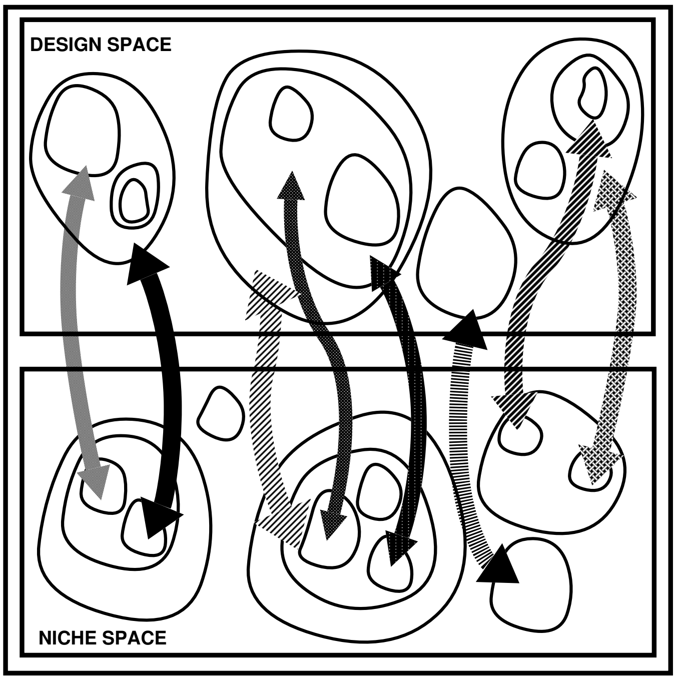
(architectures)

- **Our everyday attributions of emotions, moods, attitudes, desires, and other affective states implicitly presuppose that people are information processors.**
- **To long for something you need to know of its existence, its remoteness, and the possibility of being together again.**
- **Besides these *semantic* information states, longing also involves *control* states.**

  ONE WHO HAS DEEP LONGING FOR X DOES NOT MERELY OCCASIONALLY THINK IT WOULD BE WONDERFUL TO BE WITH X. IN DEEP LONGING THOUGHTS ARE OFTEN *uncontrollably* DRAWN TO X.

- **Physiological processes (outside the brain) may or may not be involved. Their importance is normally over-stressed by experimental psychologists under the malign influence of the James-Lange theory of emotions. (Contrast Oatley, and poets.)**

Design space and niche space



A design can be related to many possible niches and *vice versa*.

We need to understand the structure of both design space and niche-space, and the variety of mappings between them (NOT **numerical-valued 'fitness functions' – perhaps vector valued?)**

Then we can understand the variety of types of evolutionary pressures, and the feedback loops involved in co-evolution.

What sorts of evolutionary and developmental trajectories in design space and niche space are possible and how.

(Remember: Biological evolution is DISCONTINUOUS**)**

Different sorts of trajectories in design space

There are different evolutionary trajectories pursued in parallel: different solutions with different tradeoffs.

i-trajectories: **possible for an individual (development, learning)**

e-trajectories: **possible for a species, gene pool (evolution)**

s-trajectories: **possible for a society, culture.**

r-trajectories: **possible for an external repairer, designer to produce**

Others?

CONCLUSION: THE SCIENCE

• Much of this is conjectural – many details still have to be filled in and consequences developed (both of which can come partly from building working models, partly from multi-disciplinary empirical investigations).
• An architecture-based ontology can bring some order into the morass of studies of affect (e.g. myriad definitions of "emotion"). ***Towards a periodic table for the mind*.**
• This can lead to a better approach to comparative psychology, developmental psychology (the architecture develops after birth), and effects of brain damage and disease.
• It will provide an improved conceptual framework for the study of language, its functions, its mechanisms, its development, its evolution.

CONCLUSION: ENGINEERING

Designers need to understand these issues:
(a) to model human affective processes,
(b) to design systems which engage fruitfully with human affective processes (e.g. 'believable' softbots)
(c) to produce teaching/training packages for would-be counsellors, psychotherapists, psychologists.
(d) for convincing synthetic characters in computer entertainments
(e) For successful natural language interaction understood by machines

FOR SCIENCE AND ENGINEERING:

Consider an ‘Ecosystem of mind’ rather than just a ‘society of mind’.