The Design-Based Approach to the Study of Mind

(in humans, other animals, and machines)

Including the Study of Behaviour
Involving Mental Processes

Aaron Sloman, School of Computer Science, UoB.

`http://www.cs.bham.ac.uk/~axs`

Slides will be here

　`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#aiib10`

(X) Apologies and Notes

Apologies/warnings:

- My work straddles several disciplines and I cannot keep up with most of what has been written that is relevant.
  I welcome pointers to things I should have read or known about.

- My slides are too cluttered for presentations: I write them so that they can be read by people who did not attend the presentation.
  So please ignore what's on the screen unless I draw attention to something.

Notes

- In what follows the word "information" is not restricted to what is **true**.
  Some philosophers mistakenly think the idea of **false information** is inconsistent.
  They ignore the fact that it is possible to have or acquire **false**, or **partly incorrect** information, e.g. government propaganda and bad philosophy. There is also control information, which is neither true nor false.

- I am grateful to Gill Harris for helpful comments regarding the diversity of developmental routes to similar end-points in both normal children and children with genetic or other disabilities.
  `http://psychology-people.bham.ac.uk/people-pages/detail.php?identity=harrisg`

- There's lots more here: `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/`

Table of Contents of Proceedings Paper

1 How to describe humans and other animals

2 What is the design-based approach ("designer-stance")?

3 What sorts of mechanism do we need?

4 Conceptual tools for explanation designers

5 Whole organism explanations

6 A designer's explanatory requirements

7 Developing a better explanatory ontology

8 Limitations and common errors in current AI

9 Different requirements and different designs

10 Kinds of requirements

11 Kinds of Observations Needed

12 A consequence of adopting this approach

13 An example, from AI

14 Relevance to ontology/ontologies

15 Logical geography vs logical topography

16 Conclusion?

Abstract – from paper in proceedings

There is much work in AI that is inspired by natural intelligence, whether in humans, other animals or evolutionary processes. In most of that work the main aim is to solve some practical problem, whether the design of useful robots, planning/scheduling systems, natural language interfaces, medical diagnosis systems or others.

Since the beginning of AI there has also been an interest in the scientific study of intelligence, including general principles relevant to the design of machines with various sorts of intelligence, whether biologically inspired or not. The first explicit champion of that approach to AI was John McCarthy, though many others have contributed, explicitly or implicitly, including Alan Turing, Herbert Simon, Marvin Minsky, Ada Lovelace a century earlier, and others.

A third kind of interest in AI, which is at least as old, and arguably older, is concerned with attempting to search for explanations of how biological systems work, including humans, where the explanations are sufficiently deep and detailed to be capable of inspiring working designs. That design-based attempt to understand natural intelligence, in part by analysing requirements for replicating it, is partly like and partly unlike the older mathematics-based attempt to understand physical phenomena, insofar as there is no requirement for an adequate mathematical model to be capable of replicating the phenomena to be explained: Newton's equations did not produce a new solar system, though they helped to explain and predict observed behaviours in the old one.

This paper attempts to explain some of the main features of the design-based approach to understanding natural intelligence, many of them already well known, though not all.

The design based approach makes heavy use of what we have learnt about computation since Ada Lovelace. But it should not be restricted to forms of computation that we already understand and which can be implemented on modern computers. We need an open mind as to what sorts of information-processing systems can exist and which varieties were produced by biological evolution.

Summary

AI as Science and Philosophy vs AI as engineering.

How to think about an AI/Cognitive science research roadmap.

All living things are information processors: informed control systems

We need a better understanding the problems evolution solved – in order to understand the solutions.

   That's hard: many problems are invisible. So people focus on their favourite problem.

   Vision is the hardest problem – to see and to solve.

Tradeoffs-between evolution and development
   precocial vs altricial species,
   preconfigured vs meta-configured competences (Chappell&Sloman 2005, 2007)

Some ways to think about evolution and development.

Embodiment is important but not in the way most embodiment theorists think.

The physical environment is a deeper driver of the evolution and development of animal intelligence than most people have noticed.

   (Ulric Neisser and John McCarthy noticed. "Artificial General Intelligence" theorists mostly don't.)

Social drivers came much later.
(And built on the results)

Human language has evolutionary and developmental precursors required by other species and pre-verbal children: languages for thought (including perceiving, deciding, acting, ...) rather than for communication.
(Generalised languages: GLS)

Can AI become more collaborative and multi-disciplinary?

AI as Science and Philosophy vs AI as engineering.
We need to understand wholes as well as parts.

- It was obvious to the main founders of AI that AI is as relevant to science (e.g. attempting to model aspects of human or animal mental functioning) as to engineering.

- Not all who are involved in AI teaching and research are interested in that aspect.

- The scientific goals do not necessarily involve building some kind of complete intelligent machine, since the work can focus on a subset of what is required,
  as often happens in AI research on vision, language processing, learning, planning, reasoning, game playing, etc.

- Building models of some subsystem of a human or animal without testing the model in the context of a multi-component architecture may be misleading.

- We may fail to notice that the model does not account for aspects of the subsystem involving interacting with, cooperating with, supporting, or interfering with other subsystems.
  E.g. what you see can influence how you understand a sentence, and vice versa.

Understanding wholes requires understanding architectures combining parts of different kinds of functionality.

Not all the functionality may be visible in behaviours, or in brain scanners.

The subtle and unsubtle roles of AI

The obvious way in which AI can be relevant to biology is by providing working models that are potential explanations for capabilities of organisms.

E.g. models of mechanisms required for various forms of perception, learning, motive generation, planning, plan execution, hypothesis generation, hypothesis testing, and many varieties of action.

But there is a more subtle role for AI in AIIB, namely generating research questions and providing new concepts that can be used in formulating such questions and testing answers.

For example, the questions can be concerned with how many subsystems are concurrently active, what sorts of things they do, how they are controlled, how they develop, what kinds of information they acquire, use manipulate, which species have them, how they evolved,

In order to formulate the deep questions we must abandon the categories of common sense, e.g. what does the organism want, believe, intend, think, know, etc, and talk about the functions of subsystems, including possibly unobservable subsystems, that are described using technical concepts linked to explanatory theories.

Living things are informed control systems

Multiple designs are found in living things.

Living things are information processors.

The world contains: matter, energy, information

Organisms are control systems:

They acquire and use information (external and internal),
in order to control how they use matter, energy and information

(in order to acquire more matter, energy and information,

and also reproduce, repair, defend against intruders, dispose of waste products...).
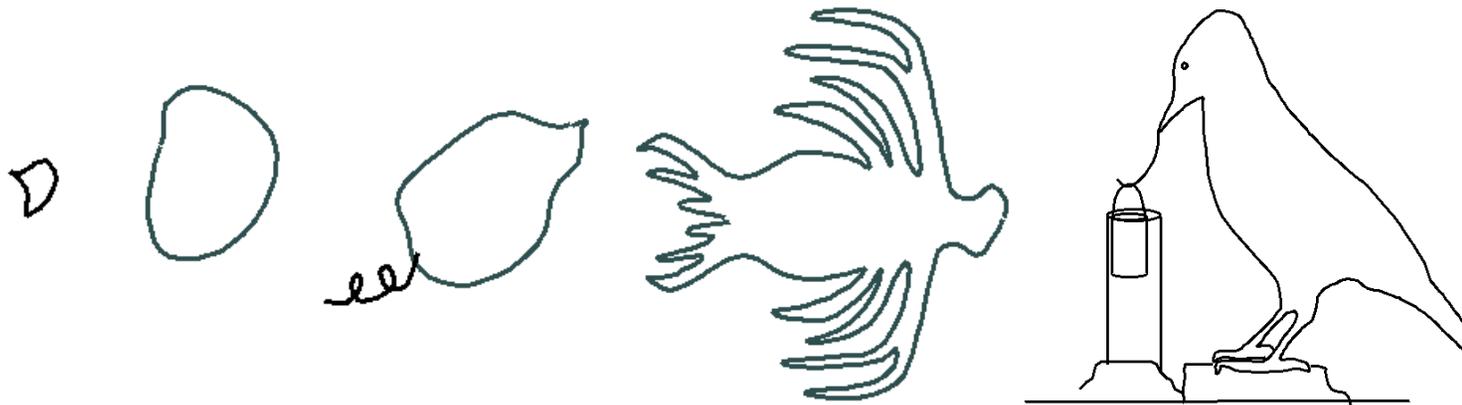
Evolution produced more and more sophisticated information processors,
driven partly by changes in environments, partly by results of prior evolutionary history.

Because of the way organisms use information they are informed control systems.

All organisms are information-processors but the information to be processed has changed and so have the means
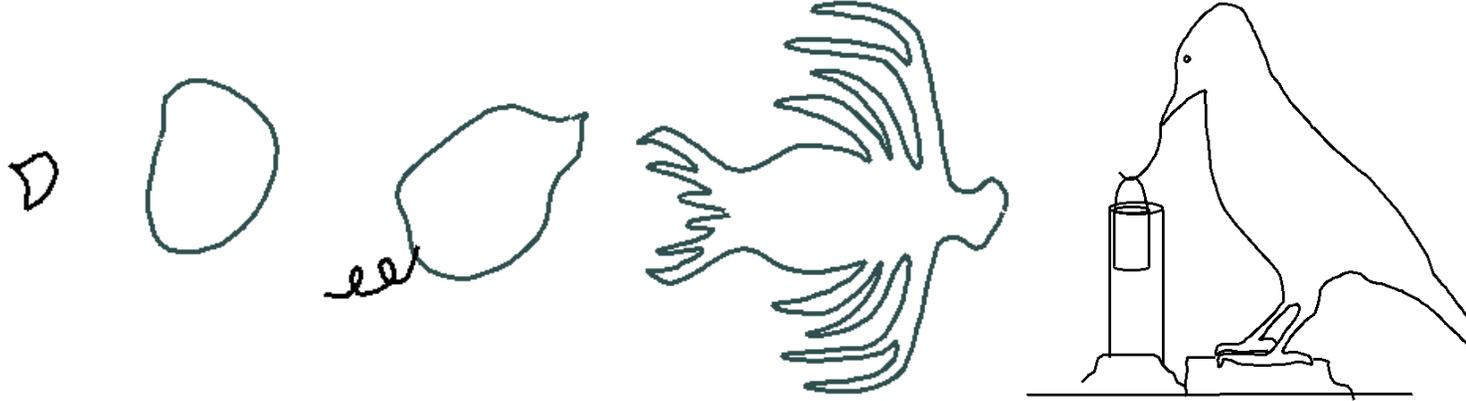
Types of environment with different information-processing requirements

- Chemical soup
- Soup with detectable gradients
- Soup plus some stable structures (places with good stuff, bad stuff, obstacles, supports, shelters)
- Things that have to be manipulated to be eaten (e.g. disassembled)
- Controllable manipulators
- Things that try to eat you
- Food that tries to escape
- Mates with preferences
- Competitors for food and mates
- Collaborators that need, or can supply, information.

The role of the environment

Ulric Neisser:

"We may have been lavishing too much effort on hypothetical models of the mind and **not enough on analyzing the environment** that the mind has been shaped to meet."

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

Compare: **John McCarthy**: "The well-designed child"

"Evolution solved a different problem than that of starting a baby with no a priori assumptions.

....

"Instead of building babies as Cartesian philosophers taking nothing but their sensations for granted, evolution produced babies with innate prejudices that correspond to facts about the world and babies' positions in it. Learning starts from these prejudices. What is the world like, and what are these instinctive prejudices?"

`http://www-formal.stanford.edu/jmc/child.html` Also in AI Journal, December 2008

All biological organisms are solutions to design problems that cannot be specified without specifying in detail the relevant features of the environment.

Turing, surprisingly got this wrong: he thought human-like learning was possible from a "clean slate".

How to look at the environments of organisms

All biological organisms are solutions to design problems that cannot be specified without specifying in detail the relevant features of the environment. (This does not imply that there is a designer.)

In order to understand which features of the environment are capable of influencing designs (or more precisely producing "pressures" to alter designs) we have to understand what the problems are that a team of engineers would have to solve – including hardware and software engineers.

That means understanding things like

- what information the organism needs in different parts of the environment while in different states (hungry, thirsty, escaping, competing, playing, exploring, etc.)
- what forms of representation of that information can be useful for the purposes of influencing internal and external processes including physical behaviours and information-processing (at the time or in future). in future).
- what information processing mechanisms can make use of the information
- what sort of architecture can combine a variety of forms of information processing possibly running concurrently.

Turing, surprisingly got this wrong: he thought human-like learning was possible from a "clean slate".

J.J. Gibson understood the general point, but missed many important details.
Other relevant authors: Piaget, Fodor, Chomsky, Mandler, Keil, Gopnik, Tenenbaum, Thomasello, Karmiloff-Smith, Spelke, ... (Not all seem to understand how to think about requirements and designs.)

To gain deep understanding, don't study just one species

Trying to do AI as science by developing only one type of system
  (e.g. "human level AI")

is like trying to do physics by investigating only how things behave near the leaning tower of Pisa.

Or studying only the motion of planets.

  we need to understand spaces of possibilities
  and the tradeoffs between alternative designs: so look at different species.

Don't assume all effective information-processing mechanisms have to be rational
  (as in Dennett's "intentional stance", Newell's "knowledge level".)

Engineers need to build reflexes into complex machinery to cope with the unexpected.

Likewise, evolution provided useful reflexes: reflexes are neither rational nor irrational.

Some useful reflexes are cognitive. Some even meta-cognitive.

Including reflexes that generate goals/motives warnings, things to remember, etc.

Not all goals are chosen because the individual knows associated rewards/costs/benefits:

See
  `http://www.cs.bham.ac.uk/research/projects/cogaff/misc/architecture-based-motivation.html`
  Architecture-based motivation vs reward-based motivation.

Example: Two ways to study architectures

1. Try to come up with a specification for a powerful, general (possibly human-like) architecture.

   vs

2. Try to understand ways in which architectures can differ –
   (The space of possibilities, and tradeoffs.)

   • what types of functionality (or combinations of types) they can support,
   • what design options there are for a given sort of functionality
   • what the tradeoffs are between the options
   • how different types could have evolved
   • how instances can develop or grow
   • what sorts of mechanism may or may not be available for particular sub-functions
   • what forms of representation are available and how they relate to particular sub-functions
   • tradeoffs between more or less parallel implementations.
   • tradeoffs between prior "hardware" commitments and run-time resource allocation.
     e.g. special hardware for meta-cognition or sharing?

Doing only (1) will produce limited understanding of the system studied, simulated, etc.

Biological evolution has already "explored" many such issues – though not completely

It will be useful to identify and explain some of the limits in what evolution has already done.

Some ways to think about architectures. 1

The CogAff schema: Possible components of architectures – differences in how functions relate to the environment (indicated by vertical divisions below), and differences in evolutionary (or developmental) stages (indicated by different horizontal divisions – lowest layers are oldest and most common).

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

This architecture schema allows different mechanisms (or none) to exist in each of the "boxes" and also allows different kinds of connectivity between them – different routes for factual and control information to flow between different mechanisms.

The CogAff schema covers very many very different architectures.

Architectures, evolution and development.

The CogAff Schema provides

A crude taxonomy of sub-functions
of possible architectures

divided (vertically) between (overlapping)
- sensory/perceptual functions
- more general (and more central) functions
- action/effector functions,
  (Compare Nilsson's "three towers".)

and divided (horizontally) between
- meta-semantic functionality
  (able to represent, reason about, and possibly control things that represent, including self and others).
- more flexible deliberative functionality
  (representing, reasoning about merely possible or unperceivable situations and processes)
- evolutionarily oldest
  (completely reactive – e.g. microbe-like or insect-like)

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

Architectures can differ according to which sorts of sub-functions they have (in the various boxes), how they are connected, whether the components are all fixed, or grown during development.

The latter is ruled out by some definitions of "architecture" as unchangeable.

Some ways to think about architectures. 2

H-Cogaff: A conjectured (human-like) subset of the CogAff Schema

See the web site: `http://www.cs.bham.ac.uk/research/cogaff/`

A special subset of the CogAff schema, containing many kinds of concurrently active mutually interacting components.

The papers and presentations on the Cognition & Affect web site give more information about the functional subdivisions in the proposed (but still very sketchy) H-Cogaff architecture, and show how many different kinds of familiar states (e.g. several varieties of emotions – primary, secondary, tertiary,....) could arise in such an architecture.

This is merely as an indication of the kind of complexity we can expect to find in some virtual machine architectures both naturally occurring (e.g. in humans and perhaps some other animals) and artificial (e.g. in future intelligent robots).

Compare Minsky's **Emotion Machine** architecture – with different subdivisions highlighted.
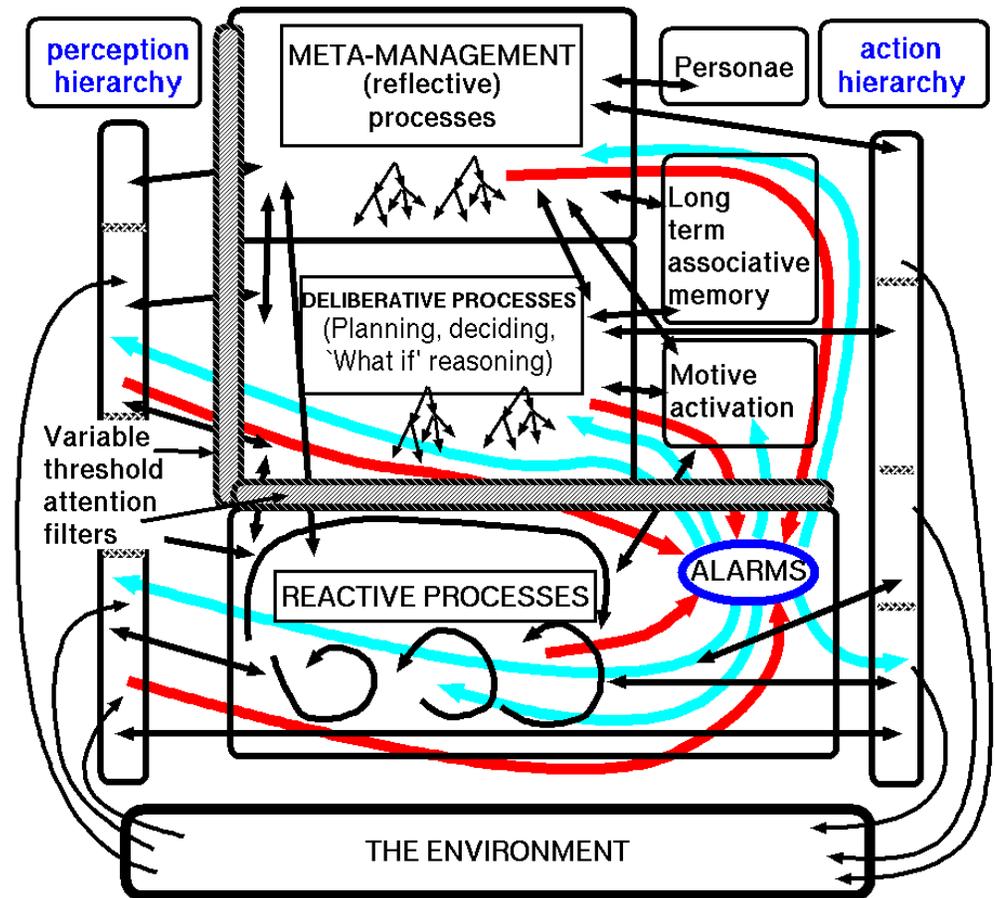
Virtual machinery vs brain machinery

Do not expect the conjectured sub-divisions to be detectable in the physical structure of the brain or in physically measureable boundaries between activities.

That's because we are talking about what computer scientists and software engineers call "virtual machinery": non-physical machinery that is implemented in physical machinery.

The relationships between virtual and physical machinery are far more complex and tangled than most people realise, as a result of half a century of design and development of many hardware and software mechanisms supporting the relationships, including compilers, interpreters, schedulers, memory mananagement, paging systems, device drivers, interrupt handlers, networking protocols, and many more.

As a result there are webs of causal relations between states and processes in portions of virtual machinery that do not map in any simple way onto the states and processes in the underlying (mostly digital electronic) physical machinery.

For example, the virtual mechanisms and states and causal interactions are much less fine-grained than the physical ones.

Conjecture: biological evolution created virtual machines long before we understood the need for them.

If mind brain relationships are like this, mind-brain identity theories are false. Implementation is not identity.

Moreover, not all virtual machine processes can be detected via input-output relationships.

See `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#mos09`
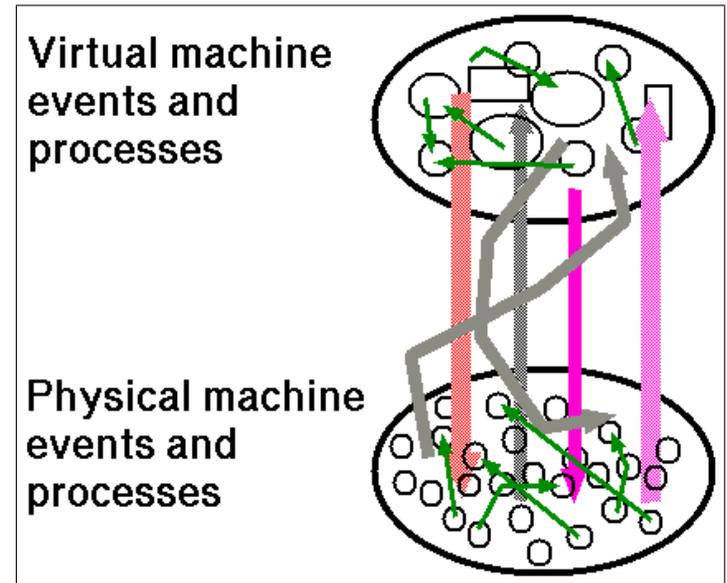
Two spaces needed to understand evolution

A design is a specification for a class of solutions to a collection of problems.

A niche is a set of requirements against which designs can be evaluated.

Niche space is a space of possible sets of requirements.

Some design solutions are meta-solutions: they specify classes of designs, and their instances grow particular solutions (particular designs) driven partly by conditions at "run-time".

Compare altricial vs precocial species.
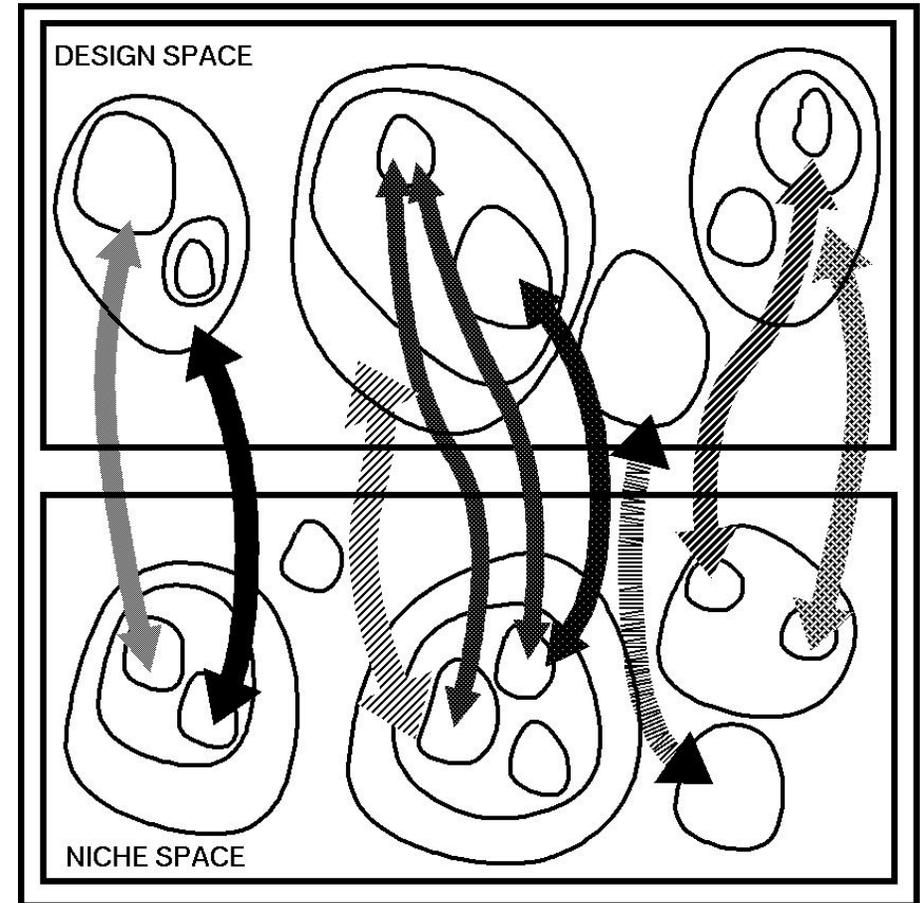
Too many people (in other disciplines) think only of dichotomies and continuous spaces (differences of degree).

We need spaces with different levels of abstraction allowing both discontinuities (differences of kind) and some continuous regions and trajectories (differences of degree).

Evolution and development: Trajectories in two spaces

Various interacting trajectories are possible in design space and niche space: dynamics of biological virtual machines in an ecosystem.

- i-trajectories: individuals develop and learn
  (altricial – precocial spectrum)

- e-trajectories: species evolve across generations

- r-trajectories: a 'repairer' takes things apart and alters them (hence discontinuities).

- s-trajectories: societies and cultures develop (Not shown)

- c-trajectories: e-trajectories where the Cognitive mechanisms and processes in the individuals influence the trajectory, as in mate selection, or adults choosing which offspring to foster in times of shortage. (Also not shown)

DESIGN SPACE

NICHE SPACE

→ i-trajectories

→ e-trajectories

⟶ r-trajectories
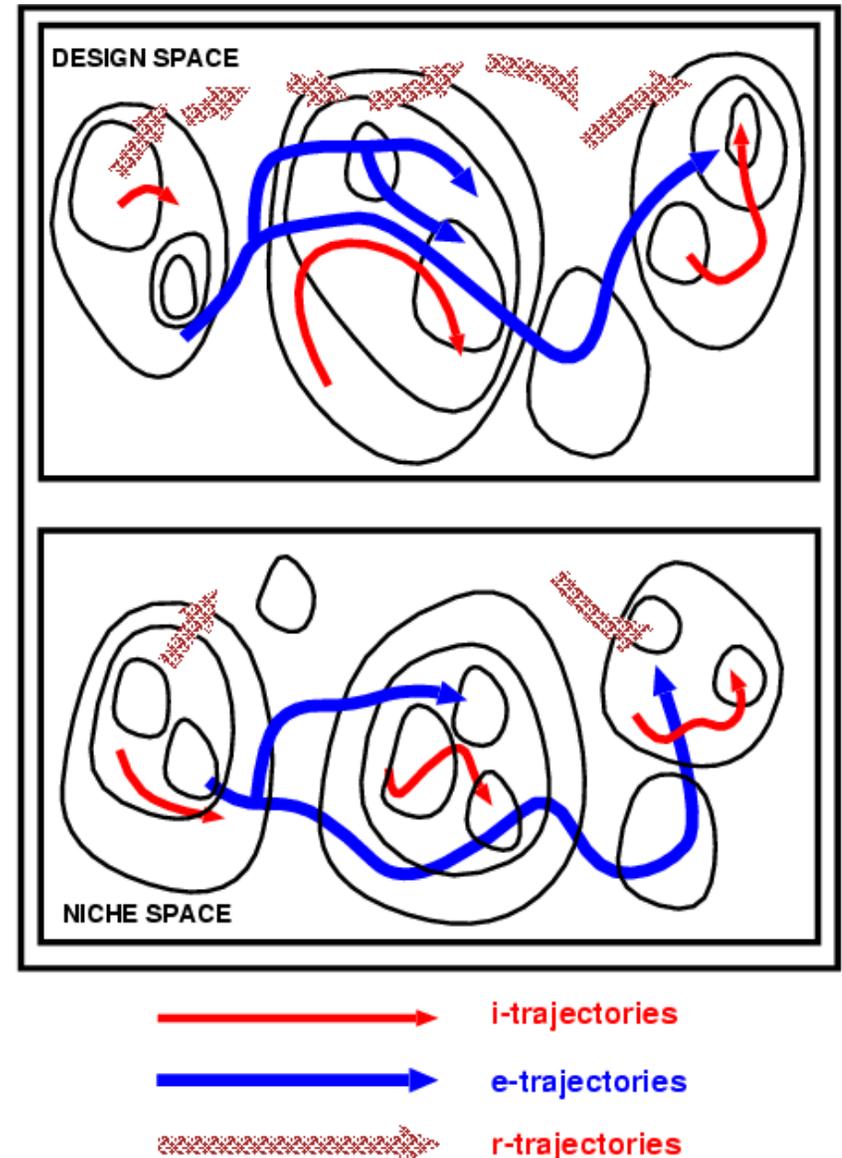
Tradeoffs-between evolution and development

Biologists distinguish

Precocial species: individuals born physiologically advanced and competent
e.g. chicks follow hen, and peck for food; new born deer run with the herd

Altricial species: individuals born under-developed and incompetent – very dependent on parents.

Jackie Chappell convinced me in 2005 that:

It's not species but competences that should be distinguished.

Even humans have some "precocial" competences (e.g. sucking is highly developed very early).

We argued in a paper at IJCAI 2005 that individuals need a "spectrum" of competences:

Some largely genetically determined, others developed through interactions between products of genome and environment.

A similar spread will be required for advanced robots of the future.

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

A later paper (invited for a special issue of IJUC, 2007) enlarged the theory and introduced new terminology

There are preconfigured vs meta-configured competences

The same distinction can be applied to meta-competences and meta-meta-competences.

The idea is that an architecture is grown in stages, or layers, where new layers learn by observing not only the environment but also contents of prior layers, and their interactions with the environment.

Staggered brain development is a prerequisite for this.

It is very likely that humans are not unique in this. (We are looking at some other species.)

Embodiment is important but not in the way most embodiment theorists think.

The physical environment is a deeper driver of the evolution and development of animal intelligence than most people have noticed

Evolution seems to have produced some overlapping cognitive functions in animals with very different morphologies, e.g. primates, elephants, corvids, octopuses: convergent evolution of cognition.

The deep explanatory similarities in these competences will be closely related to features of a complex 3-D environment with various mixtures of persistence and change, including possibilities, opportunities, threats, dangers, routes, obstructions, food, usable objects, etc. (proto-affordances and affordances).

McCarthy:
"Instead of building babies as Cartesian philosophers taking nothing but their sensations for granted, evolution produced babies with innate prejudices that correspond to facts about the world and babies' positions in it. Learning starts from these prejudices. What is the world like, and what are these instinctive prejudices?" (op. cit.)

There will be some species differences, and possibly also differences within a species (e.g. humans).

Social drivers came much later, building on the results

Individual trajectories – more or less environment-driven

Routes from genome to behaviour
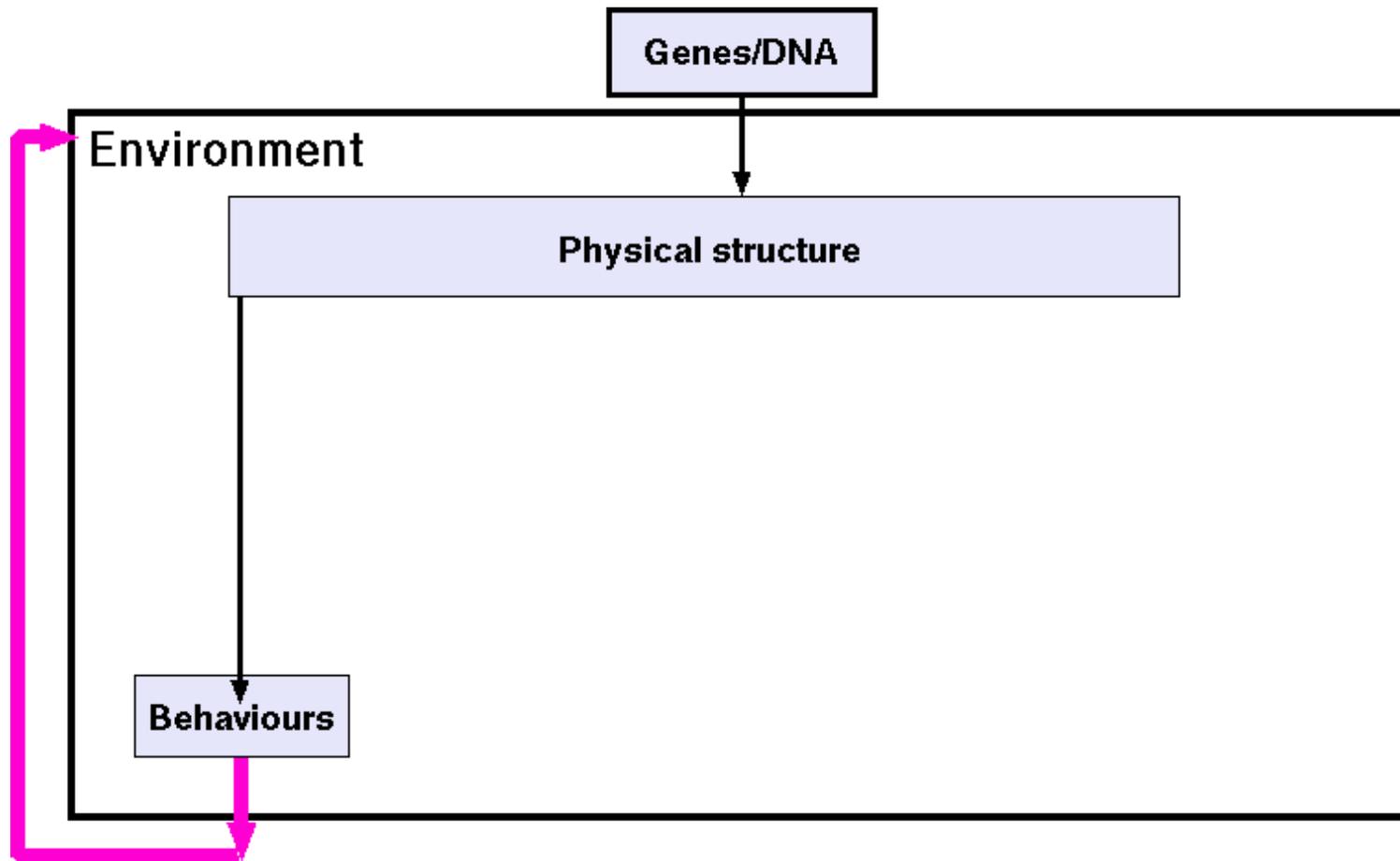
Here behavioural repertoires are mostly fixed.
The vast majority of organisms (including micro-organisms) are like this.
Many don't live long enough to learn much.   But ....

Individual trajectories

Routes from genome to behaviour
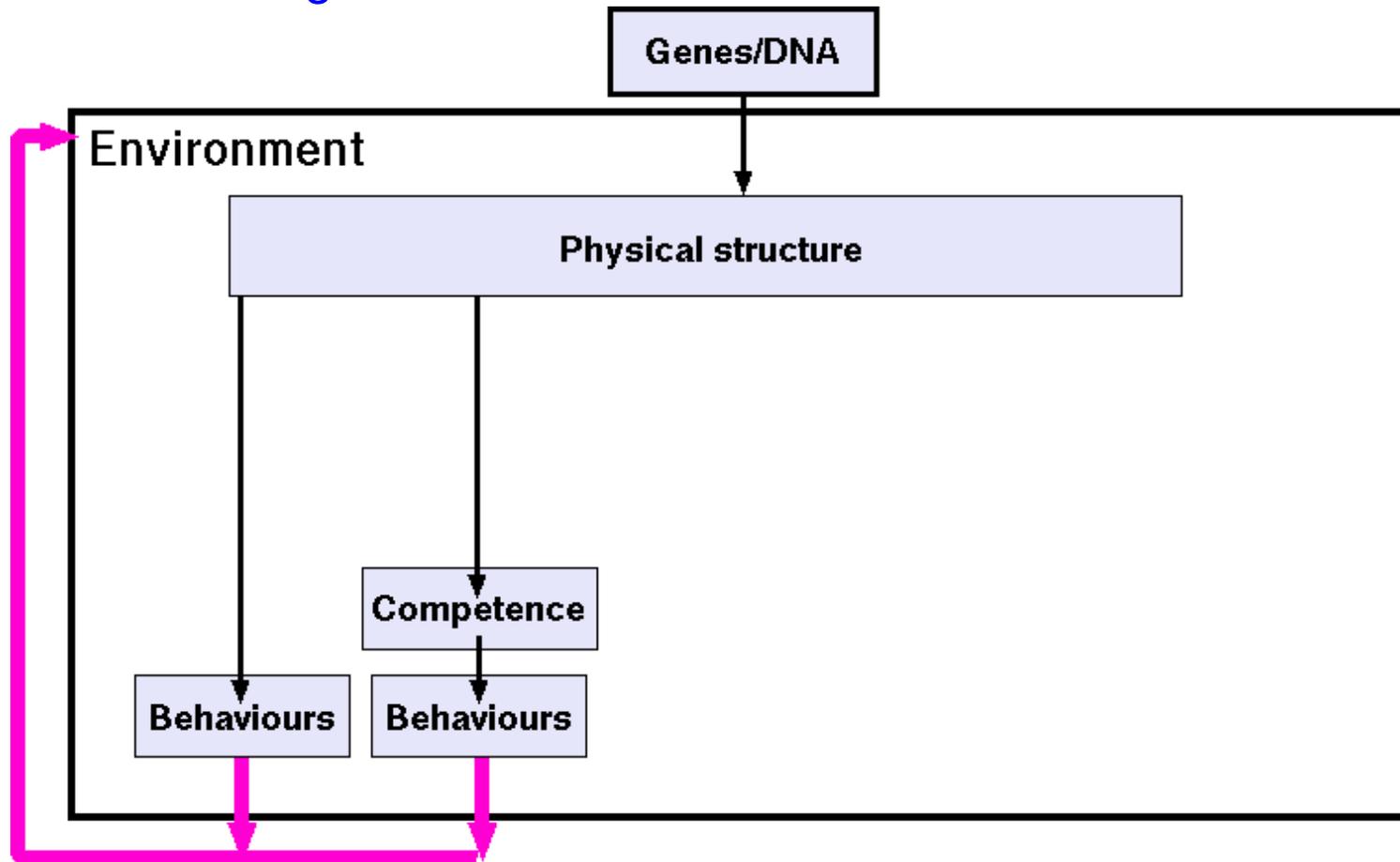
Some organisms instead of having only fixed behaviours triggered by events, have competences: packages of behaviours with conditional tests that allow variations in the environment to be accommodated. (There are nearly always intermediate cases in Biology!)

Individual trajectories

Routes from genome to behaviour
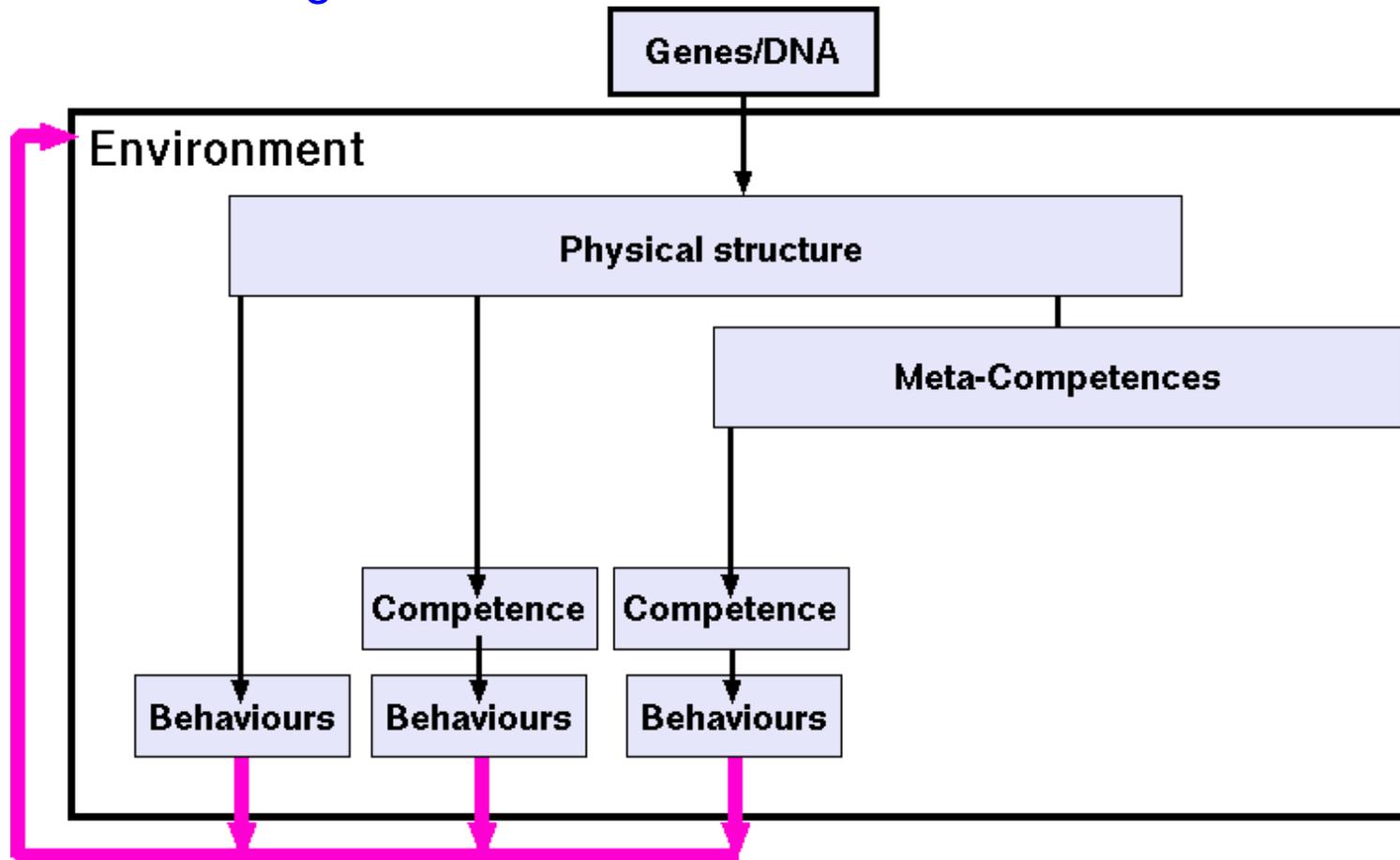
Competences may be meta-configured, i.e. developed by a learning process.

The learning processes may use a lot of genetically provided information about the environment, including the sorts of things that can be learnt and the kinds of things worth doing in order to learn them.

Individual trajectories

Routes from genome to behaviour
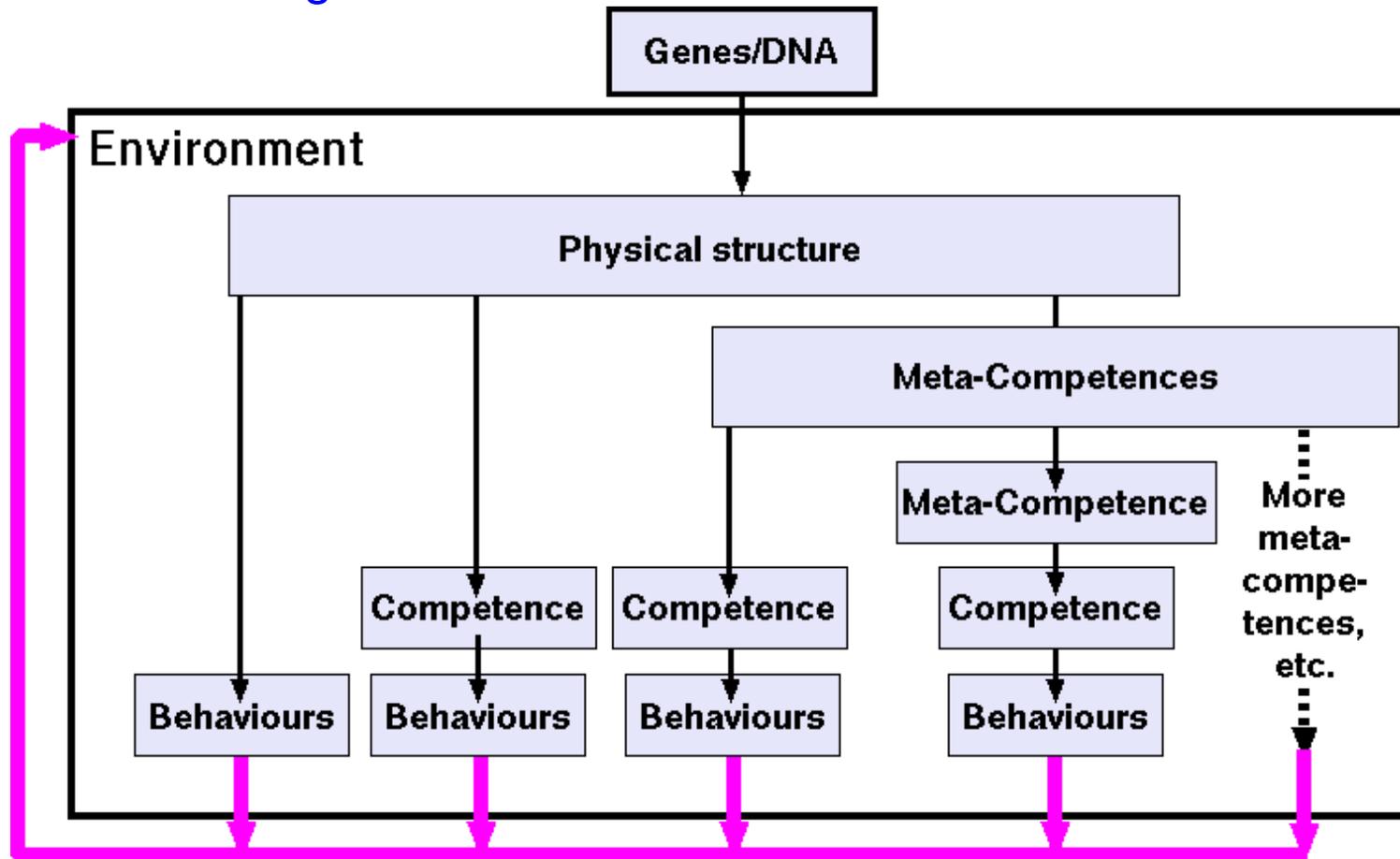
Meta-competences may also be meta-configured, i.e. developed by a learning process, using a meta-meta-competence, which itself may have been preconfigured or grown.

Humans go on learning how to learn through and beyond university.

The last few slides

The ideas in the last few slides were developed with Jackie Chappell.

Jackie Chappell and Aaron Sloman, 2007,
Natural and artificial meta-configured altricial information-processing systems,
International Journal of Unconventional Computing, 3, 3, pp. 211–239,
`http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609,`

Both Jackie and Chris Miall helped with the diagrams.

There is much more to be done: turning these ideas into designs for working systems, and in the process bugs in the ideas will surely show up and the ideas will change.

(Compare Lakatos on how scientific research programmes develop.

Some turn out to be progressive and some degenerative:
it is often impossible to decide in advance which.)

We need a better understanding of the problems evolution solved – in order to investigate possible solutions

Alas, it's hard: many problems are invisible.

So people focus on their favourite problem(s).

Then wonder why their machines seem to be so far behind what they are aiming for.

Sometimes part of the problem is assuming that their favourite forms of representation, or favourite programming paradigms are suitable for developing explanatory models.

I've been to many seminars and conferences where speakers have just assumed

- that a robot's visual system has to solve the problem of computing distances using some familiar unit of measurement,
- that locations and orientations of objects need to be represented in some global coordinate system using some familiar units of length and angle,
- that objects have to be recognized and labelled as being in some category in order to be seen
- that only what exists can be seen.

Human language has evolutionary and developmental precursors

CONJECTURE: Both pre-verbal humans and also many non-human species make use of internal forms of representation, used for:

- perception
- reasoning, planning
- goal formation
- motive selection
- question formation
- plan execution

Such forms of representation will need to provide features noted in human languages

- variable complexity
- structural variability
- compositional semantics
- manipulability

But they need not be restricted to linear forms or discrete elements. Some will use analogical rather than Fregean forms of representation.(Compare human sign languages.)

We can generalise the notion of "language" (Generalised Language, GL) to include these forms of representation used for internal purposes, rather than for communication.

```
http://www.cs.bham.ac.uk/research/cogaff/04.html#200407 (1971)
http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#43 (1979)
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#glang (2008)
```

Vision is the hardest problem – to see and to solve.

Vision and mathematics

politics – art of the possible

mathematics – science of possible and impossible.

vision (especially perception) also about what's possible.

Too many people think vision is about recognition.

Compare J.J.Gibson: perception involves getting information about affordances.

Generalise this: information about structures, including fragments of surfaces, structural relationships and possibilities for change + constraints on change among the relationships: i.e. "proto-affordances".

Despite the (very) low resolution and the noise you can see a great deal of structure in these images, and you can see many proto-affordances – things that could change, including available information.

What processes using two fingers of one hand could transform one of the configurations into the other?

See `http://www.cs.bham.ac.uk/research/projects/cogaff/challenge.pdf`

The Popeye programme could see several levels of structure in a picture like this. The perception at different levels was done collaboratively and incrementally.
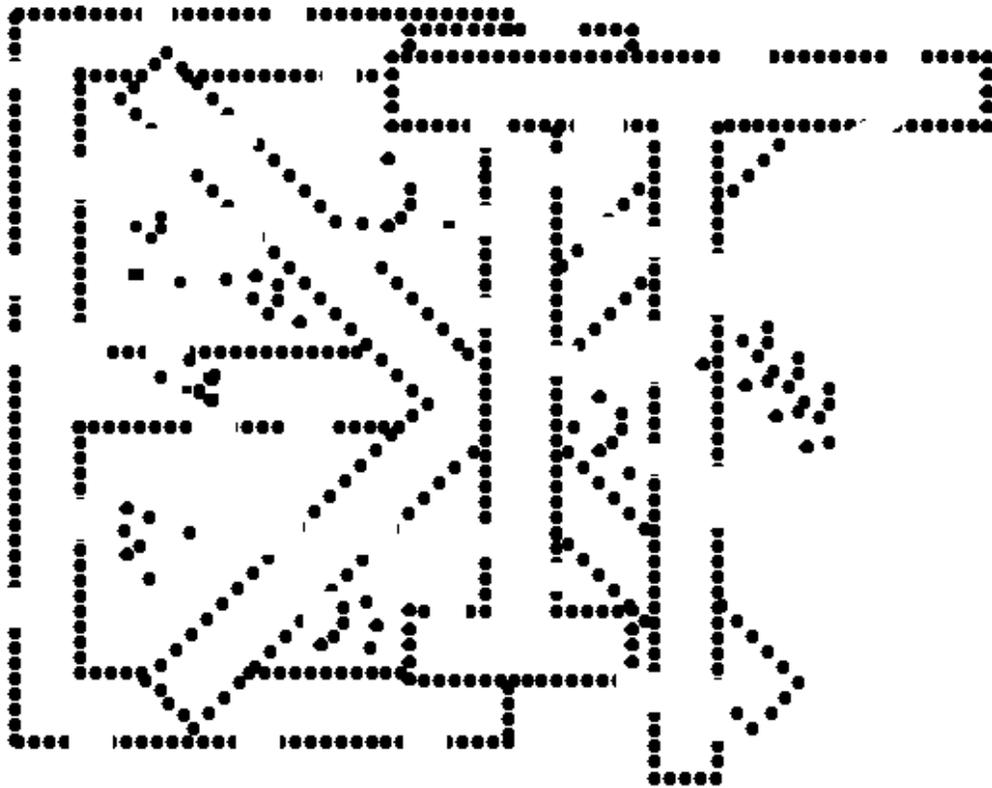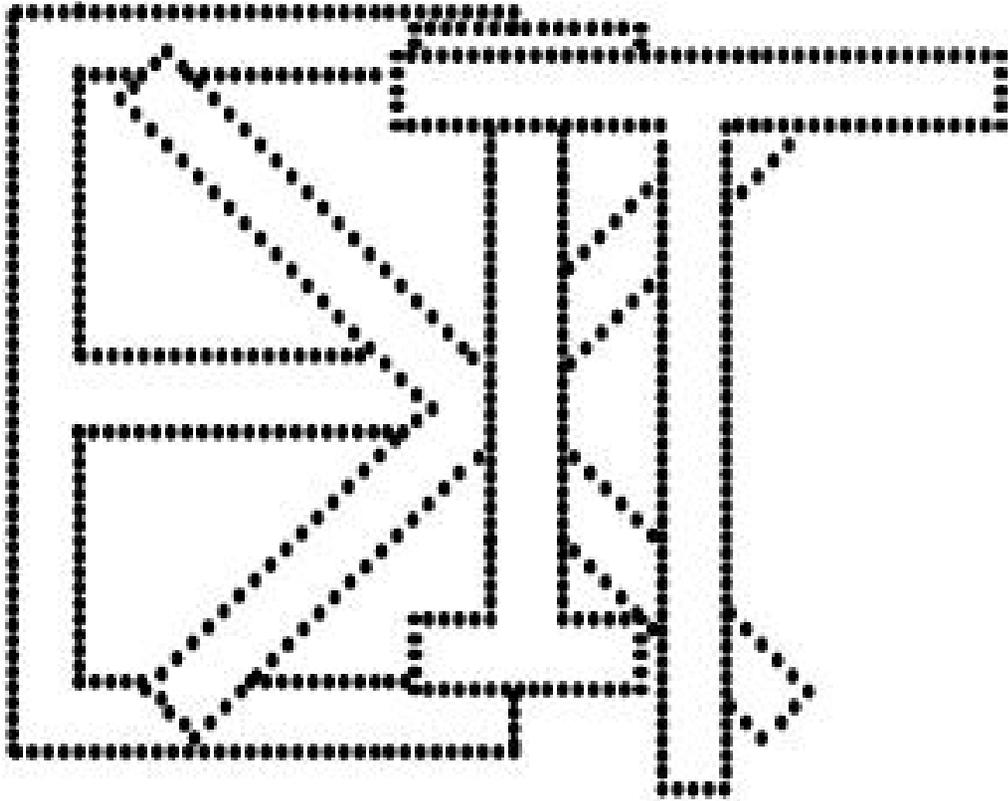
In case you did not see the structure in the previous image

This shows the various layers of interpretation making use of several different ontologies, in different "domains", as we called them, following Max Clowes.

Inferences (hypotheses) were propagated up and down the layers, across ontology boundaries.

(f) Words at the top level

(e) Letter sequences, where the letters were suggested by the stroke collections and by partly recognised words.

(d) collections of strokes (abstract linear entities)

(c) Collections of bars and bar-junctions made of opaque laminas, hallucinated onto the lines and line junctions

(b) Collections of lines and junctions hallucinated onto the dots

(a) Binary pictures (we called them "dotty pictures").
    Scanned initially looking for evidence of line-fragments grouped using the hough transform, suggesting which were good clues.

For more see Chapter 9 of *The Computer Revolution in Philosophy* (1978)

http://www.cs.bham.ac.uk/research/projects/cogaff/crp/chap9.html
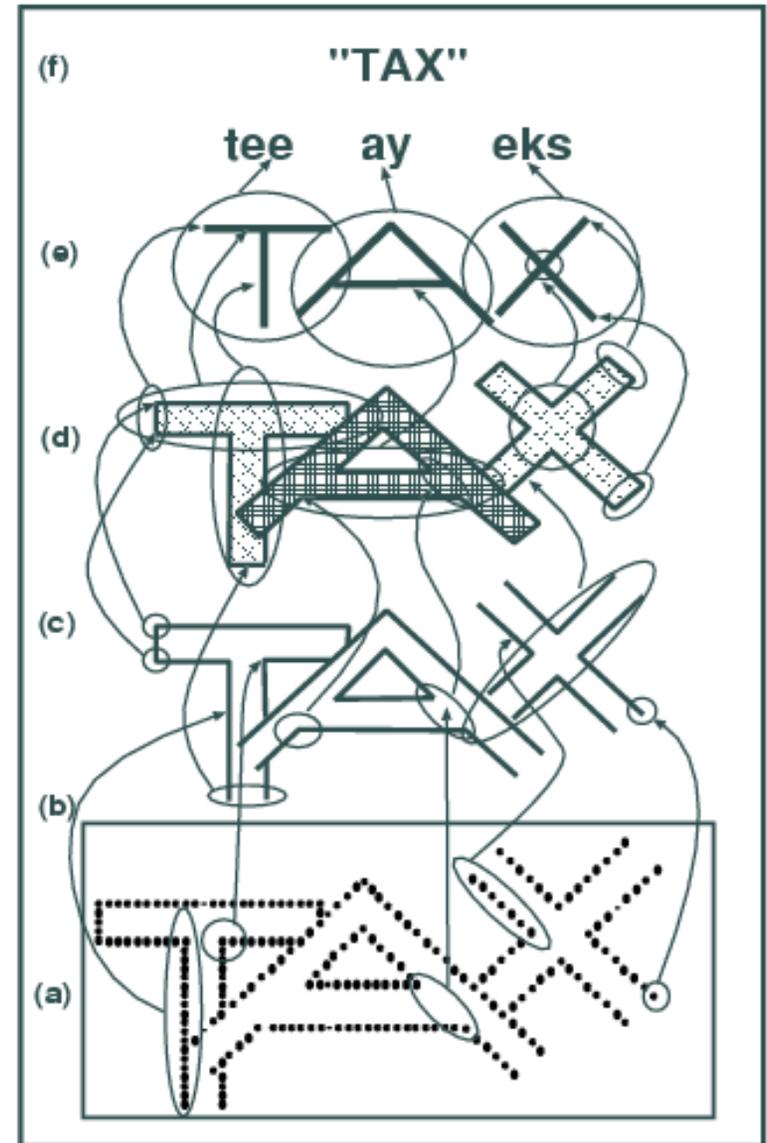
Structures vs Processes and Possibilities

It took me many years to realise that

- for biological organisms perception of static scenes and structures must be a special case of perception of processes, and

- just as perception of structures requires an understanding of relationships and constraints on relationships, perception of processes requires understanding of changing relationships and constraints on changing relationships.

This is deeply connected with the nature of mathematics: discovering what is and is not possible, and why some necessary connections are necessary.

That goes beyond J.J.Gibson's understanding of affordances as merely empirical discoveries.

If the constraints are not merely recorded and used but also understood they can be used to work out what is and is not possible in novel situations: as young children and many animals can do – but not yet robots.

NB:
this is more like proving theorems in geometry or topology than like working out probabilities from evidence.

Toddler Theorems – and Bugs

There are lots of "toddler theorems" discovered, which adults do not notice.

I suggest that that ability later grows into the basis of mathematical competence. I am trying to use that as a basis for defending a (slightly modified) version of Kant's philosophy of mathematics.

The main modification is to allow that the mechanisms involved in mathematical discovery and proof are not infallible: bugs can occur, but can also be found, diagnosed and sometimes removed.

For that we have to combine ideas from Sussman's Hacker, and Lakatos' *Proofs and Refutations*

For more on this see

```
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#toddler
```

What came first: mathematics teachers or pupils?

It is normally assumed that mathematical knowledge has to be acquired with the help of adults who already know mathematics.

If that were true, mathematical knowledge could never have developed on earth, for there were no mathematics teachers in our early evolution.

These slides develop (using a spiral of theory, experiments, challenges and evidence), a conjecture about how mathematical competences originally evolved under pressure from biological needs.

Such pressures led to mechanisms, forms of representation and architectures that supported two very different forms of knowledge acquisition

- **Empirical learning**, including use of statistical information, correlations, etc.
- **Working things out**, including designing novel structures and mechanisms, and also making predictions, forming explanations, using not statistical but structural information – e.g. geometry, topology, properties of matter... etc.

Later on, mathematical knowledge became formalised, ritualised, and transformed into an object of study for its own sake, and culturally transmitted: through mathematics teaching.

But that teaching works only because the mechanisms produced by evolution are there in every normal child, though there may be differences leading to different speeds of learning and different preferred directions of learning.

What is empirical information?

Empirical information (not just beliefs), e.g.

- about specific situations,
- about general truths,
- about associations,
- about how to do something,
- about what will happen in certain circumstances,

can be **acquired** in various ways, e.g. by:

- observing something that happens, possibly as a result of performing some observation, introspection, measurement or manipulation (in some cases repeatedly),
- being informed by another individual, newspapers, a book, a dream, ...
- derivation from other information, using logic, mathematics, statistics, analogy, etc.
- guess-work, hunches
- from the genome (Genetic inheritance is the main source of general information in most species.)

What makes many kinds of information empirical (as Popper noted) is not **the source**, but **the possibility of observations and experiments contradicting the information**:

If a perceived counter-example could turn up, a prediction could fail, a situation could be observed to have been previously misperceived, or a method could sometimes not work, etc. (as often happens in science) the information must then be empirical. (Finding exactly which portions are empirical may be tricky.)

I. Lakatos in *Proofs and Refutations*: showed that mathematics can be "quasi empirical".
The counter-examples in mathematics can occur in thought, not just perceived phenomena.

Theories about what is **possible** (e.g. formation of neutrinos) can only be **confirmed** empirically, not **refuted**. (This needs more explanation: See Chapter 2 of *The Computer Revolution in Philosophy*)

Theories of kinds of stuff, bits of stuff, bits of process

Experiments with pins and rubber bands involve interactions between bits of stuff of different kinds – the interactions are bits of process.

The bits of stuff include:

- portions of the elastic band – all flexible and stretchable

- rigid, pointed, graspable pins, with a pointed end and a blunt end, and other portions

- the surface on which the band and pins lie, and into which the pins can be pushed

- the materials composing the experimenter's hands and fingers.

The bits of process involve changes described in previous slides, including both local changes (e.g. pin pushes piece of rubber sideways), and more global changes

(e.g. a straight line becomes bent, or *vice versa*, the rubber becomes more taught and more resistant to further pushing, an asymmetric shape becomes symmetric, or *vice versa*).

Someone playing with pins and bands could acquire a moderately complex, but finite, "micro-theory" about many local structures and processes that can occur.

The theory can generate predictions about results of indefinitely many (but not all) different combinations of actions involving the rubber band and pins

- The predictions will use topological and geometrical reasoning, not just observed correlations.

- This is a form of intuitive child science, that might later be fully explained by adult official science (physics, materials science, plus mathematics).

- Compare the work by Pat Hayes on "The naive physics manifesto".

- The theory will not predict exactly when the band will break if stretched a long way. (Why not?)

Alternatives to empirical knowledge

Empirical beliefs normally cannot be proved to be correct.

Not even in science, contrary to popular misconceptions
See `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#talk18`
    "What is science?"

Contrast empirical information with:

- Necessary truths – some derivable using only logical or mathematical methods.

    Examples – nothing could make these false:
    P or (not P),   **not(P and (not P))**,   **Containment is transitive**, $3 + 5 = 8$
    If P is true, and (P logically implies Q) is true, then Q is true.
    If Q is false, and (P logically implies Q) is true, then P is false.

- Intermediate case (Semi-necessary/Relatively-necessary truths):

    Truths logically/mathematically derivable from the best available deep, general, well-established theory about how some important part of the world works. E.g.
    Any portion of pure water contains twice as many hydrogen atoms as oxygen atoms.
    Electric currents flowing through a metal produce heat.
    A rubber-band star needs an even number of pins. (Or is that necessary?)

- "Framework Theories" (in one sense of the term)

    Unavoidable/indispensable (implicit) theories required for perceiving, thinking, learning, acting, etc.

    Explaining precisely what such a theory is would require us to build something like a working robot using one: too hard, at present!
    So instead I'll merely offer sketchy examples, below.

Varieties of semi-necessary knowledge

Semi-necessary information includes deep, general, well-tested theories about how some important parts of the world work.

It also includes whatever is derivable from the theories, using derivations that depend on the form in which the information is encoded.

There are different types of semi-necessary knowledge, including the following:

- Official-adult science

  Whatever is in, or is derivable from, the best established general theories about the nature of space, time, and the things that can occupy space and time.

  That would include modern physics, chemistry, and biology, for example. (What else?)

  This is the most explicitly formalised type of theory, and many of the derivations use logical and mathematical forms of inference.

- Unofficial-adult science

  This includes a vast amount of information about kinds of things that can exist, kinds of process that can occur, and constraints on what can occur. (I.e. a rich ontology, much of it unarticulated.)

  There is a core **shared** among all adults, and also much that **varies** according to culture, professions, and individual interests and histories: e.g. rubber-band science must be relatively new.

  The contents of this sort of science are not usually articulated explicitly: much is represented internally implicitly in a usable form, and often used, though rarely communicated – e.g. craft knowledge.

  Everything derivable (how?) from the contents of this science is also part of this unofficial science.

- Baby-toddler science

  This includes precursors of adult-unofficial science, developed in early life – like rubber-band science.

Baby-toddler semi-necessary knowledge

Official-adult science has received most academic attention.
Baby-toddler science has only been investigated in narrow fields, in a shallow way, depending on changing fashions in psychology.
Unofficial adult science is somewhere in between (some of it studied in anthropology and ethnography).

A deep study of Baby-toddler science requires investigation of environment-driven requirements for the design of **working** animals, potentially testable in robots that develop as many animals do, e.g. through play and exploration in the environment.

Baby-toddler science includes deep and widely applicable theories developed in infancy and childhood that make sense of the environment, including:

- theories about kinds of stuff that can exist

    e.g. various kinds of solids, fluids, sticky stuff, etc., with various kinds of flexibility, elasticity, etc. Some of this will be mentioned later.

- theories about kinds of relationship that can hold between bits of stuff of various kinds

    e.g. relationships between rigid fixed pins and elastic bands in various shapes.

- theories about kinds of process (changing properties and relationships) that can occur – e.g. what happens if a pin is removed in the capital "T"? What can happen if you pull a piece of string?

An incomplete presentation on the ontology developed in baby-toddler science is here:

`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#babystuff`
DRAFT: Assembling bits of stuff and bits of process, in a baby robot's world: A Kantian approach to robotics and developmental psychology. (Help needed extending this.)

On being a proto-mathematician

There are many things a child can learn **empirically**, then later come to understand as **mathematically necessary** not just highly probable.

Examples (there are many more):

- Counting the same things from left to right and from right to left gives the same result.

- Going round a house one way features are seen in a fixed order, which is reversed if you go round the opposite way (example due to Kant).

- If A contains B, and B contains C, then A contains C.

- For many types of process there are things that they cannot produce:

  e.g. moving a coin from bottom left to top left on a normal chess board, using only diagonal moves.
  (Is it obvious that that is impossible?)

- An object whose sideways motion is hindered by rigid obstacles can (sometimes) be liberated from the constraints by first moving in a different direction.

  (Show movie of toddler with broom.)
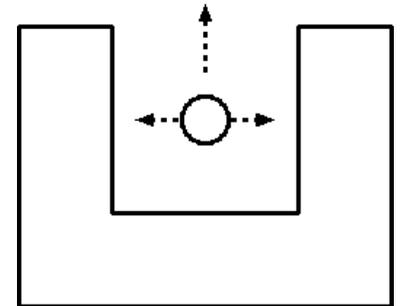
CONJECTURE The ability to transform some empirical discoveries into "theorems" arises out of solutions to biological problems "discovered" by evolution: problems posed by a complex and changing environment.

(This work is relevant to UKCRC "Grand Challenge 5: (GC5) Architecture of Brain and Mind".)
   `http://www.cs.bham.ac.uk/research/projects/cogaff/gc/`

Why Is Progress in AI/Robotics Slow?

Usual (wrong or incomplete) answers

- Wrong answer 1: It's not really slow – look at all the new results

  Visual recognition, visual tracking, 3-D reconstruction, DARPA challenges, robot soccer, impressive robots, e.g. BigDog .... but these are all very limited

  e.g. doing but not thinking about doing:
  real-time control but not reflection on what was and wasn't done or why, or what could have been done.

- Wrong answer 2: Progress is slow because the problems are really really hard, and in order to solve them we need the following new approaches ...

  – Biologically inspired mechanisms (neural nets, evolutionary computation....)

  – Biologically inspired morphology for robots (emphasis on "embodiment")

  – New architectures

  – Development of quantum computers, or some other new form of computation

  – .....

These answers miss the main point.

An alternative answer - **Progress is slow because:**

- **We have not understood the problems**

- Most of the problems come from the nature of the environment

  Which has changed in the course of evolution:

  There's not much scope for intelligence if you are a micro-organism living in an amorphous chemical soup

- We need to understand the cognitive problems and opportunities provided by the environment.

  – These arise from the fact that the environment is structured in many ways – some of which, at least, can be discovered and used.

  – This requires something deeper than using statistics to learn correlations and probabilities (e.g. Bayesian learning)

  – Some species, though not all, can do this.

  – Humans go further, and discover what they have learnt, formalise it, and transmit it across generations.

  – And they use it to do increasingly complex things with little intellectual effort: support for productive laziness is a key feature of mathematical competences

Our task for the next few hundred years: understanding requirements and possible designs.

One way to think about an AI/Cognitive science roadmap.

Things we would like human-like machines to be able to do one day

A partially ordered network of intermediate competences and scenarios

Tempting dead-ends

Ordered by dependency and difficulty

Things machines can do now
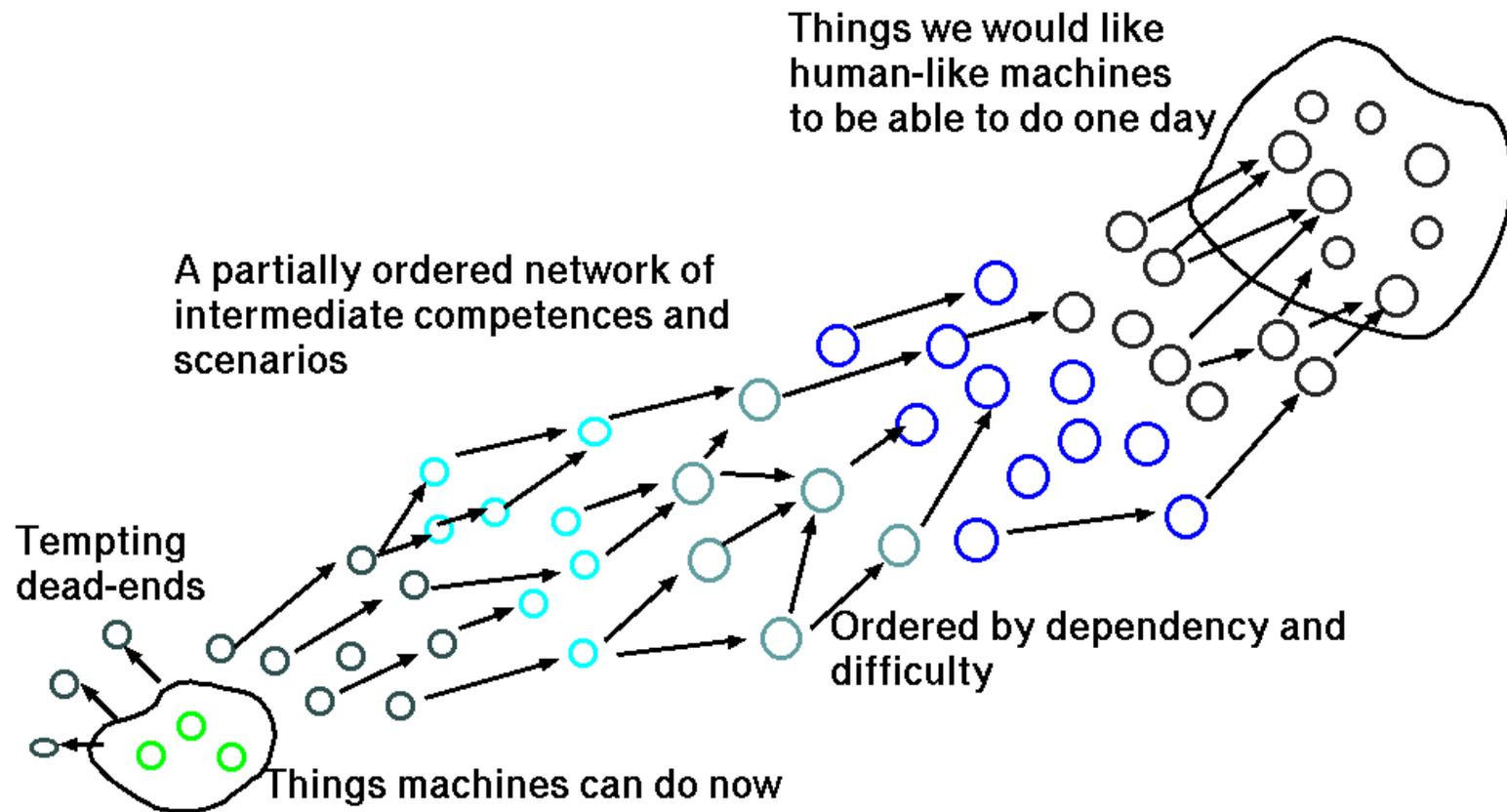
But we must not think only about designing one type of system, e.g. a system with human-like capabilities.

- Deep understanding of any design requires understanding how it is like and unlike other designs. That requires comparative analysis of multiple designs.
- Moreover, even "human-like" covers a multitude of cases: neonates, infants, toddlers, teen-agers, professors of psychology, poets, philosophers, philatelists, people with genetic or trauma-induced abnormality, people deprived of any systematic education, etc.

More things to look at

In my talks directory

`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/`

On the Cognition and Affect web site

`http://www.cs.bham.ac.uk/research/projects/cogaff/`

On the CoSy project web site

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

Messy overview of my doings:

`http://www.cs.bham.ac.uk/~axs/my-doings.html`

I have learnt from many colleagues, friends, students, and from thinkers and writers remote in time and space whom I wish I could meet face to face.