Varieties of Meta-cognition in Natural and Artificial Systems

Some pressures on design-space from niche-space

Aaron Sloman

University of Birmingham, UK

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

These slides are available in my 'talks' directory:
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#aaai08

The position paper for the meeting is available here
http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0803

Videos shown are available here
http://www.cs.bham.ac.uk/research/projects/cosy/conferences/mofm-paris-07/sloman/vid

These slides still need to be tidied up and compressed)

Apologies

I apologise

- For slides that are too cluttered: I write my slides 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.

NO APOLOGIES
For using Linux and LaTeX

Different Goals for AI Research

For some, the primary aim is to produce intelligent machines: not for me.
(Except as experiments testing the theories and models).

This is a presentation from the viewpoint of a researcher trying primarily to understand and model aspects of **both**

- natural intelligence in its many and diverse forms, and
- future possible forms of artificial intelligence.

... but not only human intelligence: also microbes, insects, birds, deer, hunting mammals, primates, landmark-using ants, portia spiders, Betty the hook-making crow, ...

However, if we ever do understand the products of biological evolution well enough to produce convincing working models, that may help us to become more effective at producing intelligent artifacts to meet many kinds of practical needs.

But producing something that works is not enough: we need to understand:

- What features of the design make it suited to the requirements (environment, niche);
- How different requirements might have required different designs;
- How different designs might have met the requirements in different ways;
- What the tradeoffs are.

Understanding goes beyond modelling.

The main feature common to all forms of natural intelligence

Organisms behave in an environment.

The nature of the environment defines many of the problems their designs have to solve.

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`

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

In contrast, much AI research attempts to formulate some very general problem and solve it, e.g. the problem of optimising a particular type of search.

The emphasis on generality/universality diverts attention away from the specific environment in which various organisms or robots have to function.

What is the environment?

The notion of “environment” required here needs to be specified with some care:

When a mouse is where a gorilla sat a few minutes earlier, is its environment the same?

Many of the requirements for which evolution has produced design solutions relate to the environment – but the environment in question for each species is not defined just by the location.

Even the (objective) geometrical and physical properties of things a type of organism or robot perceives and interacts with do not suffice to identify the relevant aspects of the environment – it also depends on the **type** of interaction, which depends on many details of the type of agent.

The environment for an organism includes the many ways in which things outside that organism relate to design features of the organism

Examples:

what the organism can perceive,
who likes to eat the organism,
what the organism needs for food,
whether some parts of the environment can support the weight of the organism.
the knowledge the organism has to use in dealing with prey, predators, competitors, etc.
(Comparable remarks can be made about a robot’s environment.)

Examples of diverse information-processing design problems faced by evolution

An arboreal environment can pose many problems for orangutans that don't arise for smaller primates.

E.g. the smaller ones can run along long relatively thin and weak branches to get from one tree to another, whereas the orangs are much bigger and heavier and many of the branches will not support them.

So orangs have deveveloped different modes of locomotion, e.g. making use of compliance of supports.

The orang mod of arporeal locomotion requires far greater cognitive sophistication.

I thank Susannah Thorpe and Jackie Chappell for drawing my attention to this.

The different sorts of nests that different birds build also require very different cognitive competences.

Could you build a nest using only one hand bringing one twig at a time to a branch of a tree?

Compare what crows, rooks, magpies and other corvids do.

Contrast bringing a piece of mud at a time.

The "environment" that determines design requirements includes the existing design on which new features need to build

Example: many of the problems discussed in the workshop papers assume that agents use a single CPU shared between the different control levels.

Contrast many biological organisms, including humans: they have many dedicated processing subsystems that are not interchangeable.

If there are already many portions of brains with different sorts of mechanisms for performing different sorts of tasks, there is no point trying to solve the problem of which task should be done next, as arises in many AI systems where one general-purpose CPU, or a pool of CPUs, has to be shared between different kinds of functions, including meta-cognitive functions.

Evolution "learns" that some brains need portions **dedicated** to kinds of metacognition, including monitoring, modulating, and learning about what other portions are doing.

Corollary:

Problems about how to optimise the distribution of the same set of tasks between the processors are replaced by problems of how to get each subsystem to perform its tasks well,

and perhaps how to give subsystems new kinds of tasks, suited to their capabilities: one of the effects of education and cultural influences in humans.

Towards an overview of design tradeoffs

I am trying:

- to understand how the tasks of modelling different sorts of organisms (and also possible machines) are related;
- analysing how demands and opportunities in the environment vary, and how they relate to possible design changes.

  But evolutionary changes and developmental changes differ in important ways.
- hoping to produce a unifying theoretical framework,

  relating different sorts of models, of different sorts of animals and machines,

  Don't think of **dimensions** of variation – it's not that kind of space:
  rather think of kinds of parts, subsystems and functional relationships that can be added or removed.

  Requirements and designs need a grammar, rather than a set of dimensions.
- focusing on capabilities of and constraints on whole organisms/machines

  rather than on component mechanisms (vision, motor control, language, learning, etc.)
- inspired by demands that changing environments place on evolutionary and developmental trajectories.
- looking at "inner logics" of some of the trajectories,

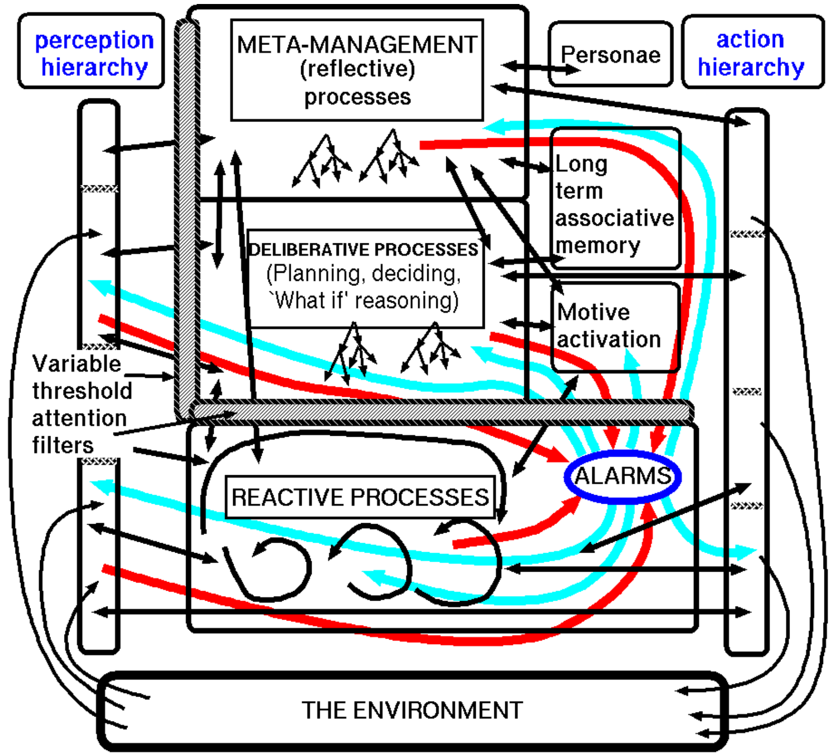
  e.g. trying to understand why design X should precede design Y, and why new pressures (changes in climate, predators, food supply, etc.) are not coped with as well by design X as by a new design Y.
  NB: the constraints probably produce partial not total orderings: different routes to similar ends.
- Understanding nature/nurture tradeoffs (Chappell & Sloman 2005,2007)

An architecture schema and an instance

Over many years, the Birmingham CogAff (cognition and affect) project has been investigating requirements for architectures for humans animals and robots. `http://www.cs.bham.ac.uk/research/projects/cogaff/`

The diagram on the left represents a generic schema for architectures with (at least) nine types of components, not all of which exist in all instances of the schema (e.g. insects).

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



On the right is a summary of a conjectured type of architecture, H-CogAff (Human-CogAff), which is a special case of CogAff, specifying a minimal, but complex set of requirements for human-like systems.

Notice that all levels have perceptual and action links with the environment.

For more details: `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#talk3`

Metacognitive subsystems also interact with the environment

Example: show video of child failing to join up his train.

He has to learn that there is something about the environment that he has not yet grasped.

Exactly how that bug/gap is addressed and what changes in the child's internal information processing is not known.

In some cases, debugging a failure requires **substantive ontology extension**, i.e. learning about some feature of the environment that cannot be defined in terms of pre-existing concepts.

E.g. a toddler learning not just to put puzzle pieces back in the right location, but also learning about outer and inner boundaries and about alignment of boundaries.

Or learning that there are different kinds of stuff out of which things can be made, with different kinds of properties that cannot be directly sensed.

How substantive (non-definitional) ontology extension can work is explained here:

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

Varieties of affordance

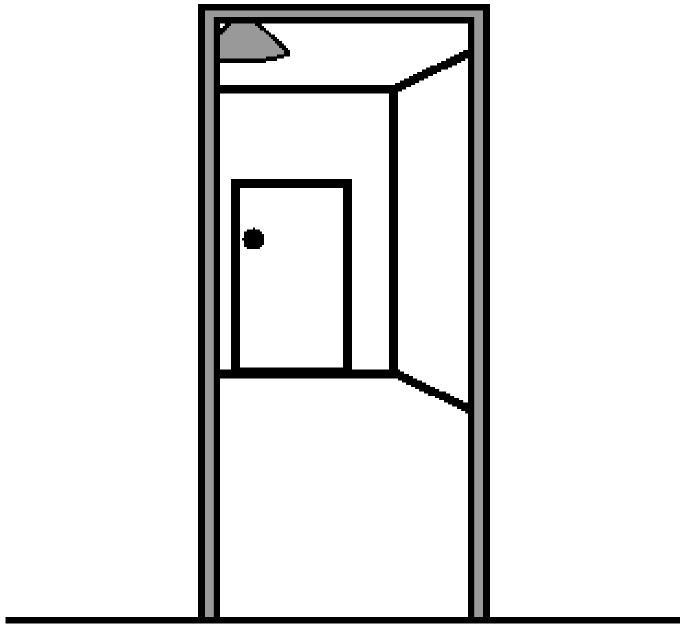
In the early years a child needs to learn hundreds or perhaps thousands of facts about what processes can and cannot occur in the environment.

A subset of those are processes it can produce: J.J. Gibson called learning what it can and cannot do in various situations learning about affordances.

As explained in the workshop paper, “proto-affordances” (possibilities for and constraints no processes) is a more general class of which affordances for an individual form a subset.

The affordances for an individual can be divided into

- action affordances (positive and negative)
- epistemic affordances (concerned with what information is and is not available)



An important kind of metacognitive learning is learning which actions alter or create which epistemic affordances. Examples:

If you move towards the door more information about the far room becomes available to you, and less if you move backwards.

If you move sideways the information about the room changes in a different way.

Uncertainty-removing affordances

The paper explains how in some cases where perceptual information is uncertain or incomplete, or predictions of consequencs of actions are uncertain, that is because the agent is in phase boundary between two regions of space, or direction, or intended motion, where information is certain.

By moving out of that phase boundary (out of the region of uncertainty) or by moving to a location with different epistemic affordances, it is sometimes possible to avoid having to compute with probabilities because information is definite.

Often that also requires using a more abstract level of description.

See `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0803`

(To be expanded)

Composition of affordances

Affordances and proto-affordances are concerned with what processes can and cannot occur in a situation.

Processes can be combined in different ways, including serial composition and concurrent composition.

Concurrent composition of **spatial** processes can produce interactions between parts of moving things (e.g. the teeth of meshed gears).

Learning about the interactions and effects of various kinds of composition of affordances is a major feature of human development, and would be required for human-like robots.

This is one of many cases where metacognitive subsystems need access not just to other parts of the agent's information processing architecture, but also to the environment.

This is indicated in some of the architecture diagrams (CogAff and HCogaff) later.

E.g.: If a partly open door has no handle you can open it more by pushing.
If it is wide open how can you shut it, if there is no handle?
A child may learn that the "obvious" thing to do is buggy.
(Compare Sussman's HACKER in *A computational model of skill acquisition* 1975)

In some cases, reasoning about such interactions seems to be much easier for humans if they use spatial (e.g. diagrammatic) forms of representation than if they use Fregean (e.g. logical) forms. (Sloman 1971)

For more on pre-linguistic forms of representation, including implications for evolution of language, see `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#glang`

An important transition in learning

A child playing with various things in its environment can discover regularities and impossibilities.

These are all initially empirical discoveries, represented as liable to be falsified by new experiments, changing circumstances, etc.

In a very important subset of cases the child comes to grasp that some of these are necessary truths, e.g.

- if you rotate one end of a rigid rod the other end moves in a big circular arc
- if you try to shut an open door by grasping its edge and pulling you will fail
- if you move further away from an open door you'll see less of the room beyond
- counting a set of fingers from left to right and from right to left (without making mistakes) produces the same result.

These transitions from understanding something as empirical and contingent to seeing its necessity, are not to be confused with increasing probabilities: they require completely different architectural support (not yet properly understood).

This is what makes it possible for humans to do mathematics.

It is also connected with Kantian causal reasoning (as opposed to Humean causal reasoning based on correlations and probabilities).

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

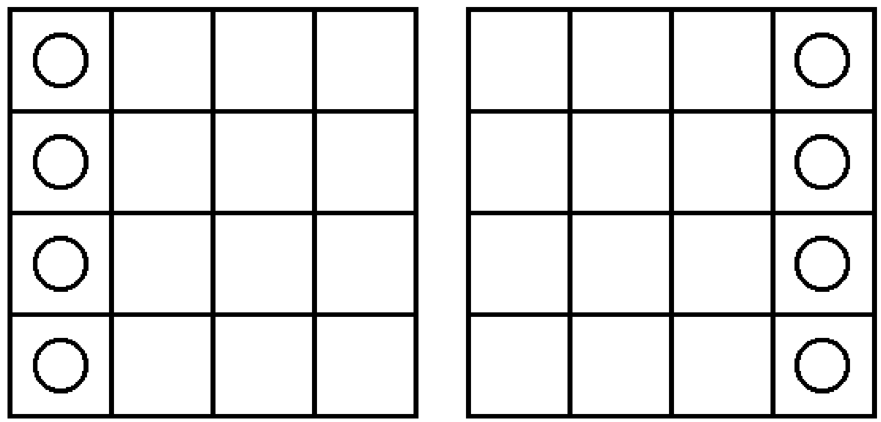
Compare Karmiloff-Smith on "representational redescription" in *Beyond Modularity*.

Seeing motion possibilities and impossibilities

Seeing affordances and proto-affordances, and how they can be combined.

Can you slide the 4 coins from column 1 to column 4 using only diagonal moves?

Using only diagonal slides, can you get from this to this?



What is the minumum number of moves required

For any configuration that happens to occupy some part of space, there are always variants that are possible: learning to see which changes are and which are not possible is not just a puzzle-solving activity: it is a crucially important aspect of learning to see, plan, predict and explain – for an animal or robot that can act in the world.

An unnoticed but important consequence of embodiment, for humans!

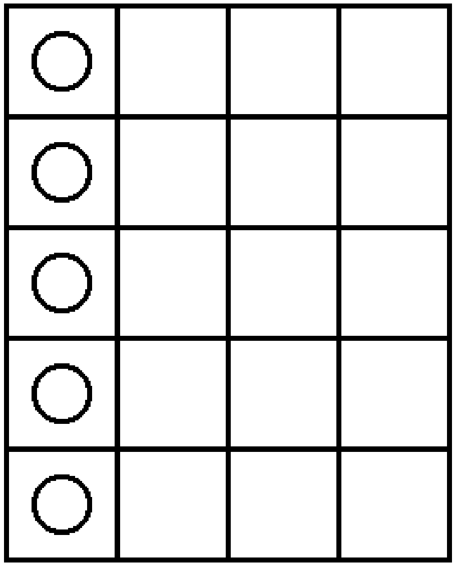
Seeing motion possibilities

Some possibilities and impossibilities are obvious, others not so obvious.

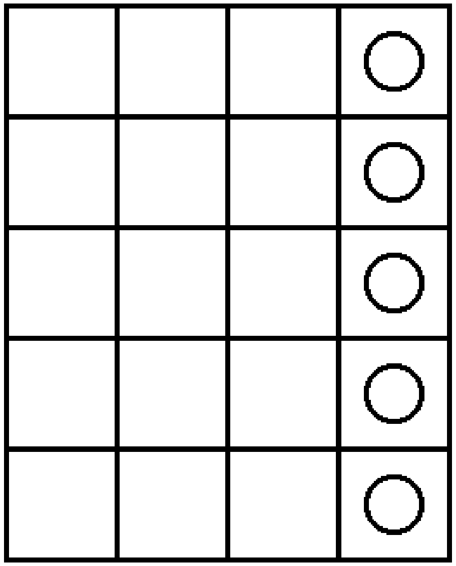
For example if you have coins placed on a board divided into squares, it is obvious that you can slide them around into different places.

Suppose you consider only moves that are diagonal: no coin can go straight up or down or horizontally.

Play around with various combinations of diagonal moves, looking for patterns and constraints in the space of possiblities.



(a)



(b)

Question:

Using only diagonal moves, can you transform configuration (a) to configuration (b)?
What is the minimum number of diagonal moves?

Some people will find the answer obvious, whereas others will have to experiment.

Some will solve the problem by using moderately advanced mathematics.

Others will use a very different kind of insight into the constraints.

Seeing impossibilities

Question:

How can you understand why it is impossible?

You could try all possible collections of diagonal moves: the board is finite and there is only a finite number of configurations using that number of coins.

But an exhaustive analysis is very tedious: is there a better way than exhaustive search?

(And what if the board were infinite, so that arbitrarily long routes can be attempted.)

Mathematical intelligence essentially involves laziness, i.e. productive laziness.

In this case we want a way to see why the transformation is impossible, in a much simpler and cleaner way than by trying all possible moves.

One way is to see some possibilities that are quite different from the possibility of moving coins around:

There is a new way to represent the problem, which makes the answer much easier to establish.

What sort of metacognitive architecture could discover that?

Discovering parity

I watched someone working on the five coin problem: at first she thought it was going to be easy. Then she tried, and found a problem. After trying a few different ways, she began to suspect it was impossible. After tracing routes she saw a pattern relating locations in the left column to relations in the right column. That pattern made it easy for her co conclude that the task was impossible. She had seen the link with chess boards.

Someone once discovered that a grid of squares has an interesting property: they can be divided into two colours in such a way that squares of the same colour are never adjacent: they meet only at corners.

That fact is used in chess boards.

On the basis of that clue you may be able to work out why it is impossible to perform the transformation from (a) on the previous slide to (b) on the previous slide, by moving coins only diagonally.

That requires you to notice a pattern that is never changed by such moves.

I don't know at what age young children are capable of discovering and using the information about diagonal colouring, but more importantly I don't know what has to change in their information-processing architectures to enable them to make the discoveries discussed here.

Perhaps we can come up with good hypotheses by trying to design robots that are capable of such discoveries.

Another route to solving the coin-sliding problem

The coin-sliding puzzle involves several different kinds of structures and processes superimposed, including spatial and numerical structures and processes: Consequently there are different routes to a solution, using different forms of representation, and different cognitive competences.

A mathematical expert saw this as a problem of translating a vertical column of coins horizontally to a new location, which he expected to be easy, so he initially tried moves that he expected would demonstrate a solution, and was surprised when they failed.

He suspected that parity was involved, and realised that diagonal moves preserve the parity of dx+dy, concluding (after a few more steps) that it was possible to shift the column of coins 2k steps horizontally, but not 2k+1 steps horizontally (e.g. column 1 to column 4).

Another person experimented with diagonal moves and soon discovered that three coins on the left were competing for two slots on the right, making the task impossible.

He then noticed the connection with a bishop's moves in chess, and suggested that the task would be much easier if the squares were coloured.

Some people might notice, after playing with the puzzle, that there are many superficially different variants of the same puzzle.

E.g. on an infinite board can you get a 2x2 square of coins to shift one column right using only diagonal moves for each coin? What about a 3x3 square?

Playing with the arithmetisation of geometry

Descarte's arithmetisation of geometry was one of the greatest and most important intellectual achievements in human history; that discovery required geometrical insight as well as arithmetical expertise.

Without it, Newton's mechanics would have been impossible
(Stephen Muggleton drew my attention to this.)

A child might find it convenient for some game (e.g. "battleships") to label rows and columns of a rectangular grid with numbers: then each square in the grid can be identified using two numbers.

After that the step to arithmetisation of continuous space is smaller but still non-trivial.

That invites various kinds of playing: e.g. what happens if you write into each box the sum of the two numbers that identify it?

Suddenly a deep link between colouring possibilities, diagonal moves and the difference between even and odd numbers becomes evident.

What enables a learner to realise that it does not matter how many squares there are, and that it works even if the grid has holes, like the odd slab of chocolate illustrated earlier.

You might suspect that if the colouring process went round a hole in the grid it might come round and be inconsistent with the starting layout. Why is this impossible?

What happens if you write the **difference** of the two numbers, into each box?

Try the **product** of the two numbers: is there anything interesting to be found in the resulting pattern?

Moving to the third dimension

Embedding the coin-sliding problem in a 3-D space allows a new version of the problem to become solvable.

A person who starts thinking about variants of the coin-sliding problem may notice partial analogies with other structures and processes.

If the problem is posed using a grid on a sheet of paper or other flexible material, then you might wonder what happens if you make a tube by joining two edges, e.g. top and bottom.

Two cases can occur, depending on the number of rows on the grid.

In some cases, diagonal moves across the join no longer preserve square-colour (parity).

As a result any configuration of coins could then be transformed into any other configuration using diagonal moves, but only because the 2-D grid is embedded in a 3-D space.

A logician could avoid using spatial intuition and map the tubular grid of squares into a purely logical structure, about which appropriate logical theorems could be proved, in the manner suggested by Bertrand Russell.

However, seeing the equivalence of the logical and the geometrical structures and processes would require use of geometrical insight.

Another option is to mark the grid on both sides of the sheet, and make a möbius strip.
See `http://en.wikipedia.org/wiki/M%C3%B6bius_strip`
`http://www.metacafe.com/watch/331665/no_magic_at_all_mobius_strip/`

Beyond toy examples

The examples given above of moving from empirical understanding of a generalisation or constraint to grasping its necessity mostly involve reasoning about discrete transitions: also the focus of much early AI.

For a child in the first few years of life there are many more examples, involving motion of rigid objects, strings, liquids, etc. where transitions occur that are continuous at one level (e.g. one object moving towards another) and discontinuous at a higher level of abstraction

(e.g. a finger-tip enters the convex hull of a cup).

In many cases there are consequences and constraints that we can learn to see as mathematically necessary rather than empirical, e.g.

if the opening of a rigidly fixed hollow cup (or cylinder) faces upwards and an object moves down until it is within the convex hull of the cup, that will impose new constraints on its possibilities for horizontal motion.

I suspect a typical child learns hundreds if not thousands of such non-empirical facts about geometry, topology, epistemic affordances, and properties of synchronised processes (e.g. counting and stepping) during the first few years of life.

These processes have not, as far as I know, been documented.

Some other animals that play with things that can be manipulated (e.g. hunting mammals, nest building birds) may also do this kind of learning, though probably not so much of it. (Why?)

A human-like embodied robot will need similar capabilities.

See `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#math-robot`

A pattern in the preceding examples

We can see some common patterns in the preceding examples, which may help us design more human-like machines, help us understand better how humans and some other animals work, and perhaps even help us design far better educational strategies.

The pattern seems to be something like this:

- Competences are acquired that allow actions to be performed on objects in the environment.
- Mechanisms involving those competences require use of structured, recombinable, representations of the structures and processes (with compositional semantics).
- Those representations of possible occurrences can to some extent be created and manipulated independently of what is actually going on in the environment.
- Consequences of the manipulations can be used for predicting or explaining actual occurrences, or for planning new ones to achieve goals.
- The forms of representation can themselves become objects of play and exploration, sometimes with the aid of externalisations (e.g. diagrams).
- This can allow the representations acquired for different competences to be combined playfully and the consequences explored (with or without external aids).
- A meta-management architectural layer observing things that happen in play and in use can notice and store patterns that have some interesting feature.
- Often those patterns allow new problems to be solved.
- Sometimes trying to solve a specific problem also leads to discovery of a new and powerful pattern or constraint. (Though such discoveries can be buggy.)

Logical patterns and spatial patterns

I have tried to bring out a deep similarity between spatial and logical forms of reasoning: in both cases relationships between patterns can be discovered that allow a single abstract representation to cover a wide range of phenomena in a way that allows predictions to be made in individual cases

But there is also a deep difference:

- Insofar as logical formulae are all Fregean (in the sense defined in Sloman 1971) the only way to alter a logical formula LF is to replace a function representation with another symbol, or an function-argument representation with another symbol, where the thing substituted may be more or less complex than the original, or to embed LF as an argument or function representation in a larger logical formula.

- In contrast, the ways of modifying and combining spatial representations (or spatio-temporal representations, when simulations are used) are far less constrained: many kinds of superposition, entanglement, adjoining, partial removal, extrusion, etc. can transform one structure to another, with many relationships, including causal relationships altered simultaneously.

- Moreover, the changes that are possible when a Fregean complex formula is modified are all discrete, whereas alterations of properties and relationships in a spatial or spatio-temporal representation can be continuous.

  There are techniques for accommodating an approximation to continuous change into a Fregean formula, namely by associating a numerical value with a property or relationship, and allowing the numerical value to undergo small discrete changes: this is how many graphical simulation programs work: but getting effects of these changes to propagate requires more work than in an analogical spatial representation (the “frame problem”).

Patterns versus Analogies and Metaphors

The discovery of reusable patterns – new abstractions that have wide applicability – is subtly and importantly different from discovery of analogies/metaphors.

A metaphor (or analogy) uses concepts from two domains and the notion of a mapping.

For example: the sequence of numbers 1 to 7 can be used as a metaphor for the days of a week, using the mapping

```
Sunday Monday Tuesday Wednesday Thursday Friday Saturday
One    Two    Three   Four      Five     Six    Seven
```

However instead of learning about and remembering all the many mappings between similar structures, you can create and store a new representation for the pattern that is common to all the “ends” of those mappings, something like this:

```
[ ( )   ( )   ( )   ( )   ( )   ( )   ( ) ]
```

This abstract specification (of an ordered collection of seven items), is applicable to very many ordered sets, and is also an entity that can be manipulated in its own right, along with other such entities, and relationships between them can be explored.

For example, you can discover different ways in which this structure can be constructed from smaller structures, and what you learn will be applicable to all the concrete ordered sets that the patterns can be applied to.

In short: abstraction is part of the power of mathematical thinking:
It is more powerful than metaphor/analogy, but often confused with it.

Tasks for the future

The ideas sketched above have been explored in restricted ways by some AI researchers, e.g. Doug Lenat (AM, Eurisko), Simon Colton (HR), Alison Pease (Modelling Lakatos).

This work has, as far as I know, been completely ignored by roboticists.

Likewise those researchers used forms of representation and reasoning that were not suitable for use by robots: e.g. they did not use mechanisms that could be part of a 2-D or 3-D visual perception system or action control system.

It remains an open question whether a geometric intuition mechanism (of the sort indicated in Sloman 1971) could be implemented in a virtual machine running on a von Neumann machine: I see no reason why not, but Penrose claims (using flawed philosophy) that it is impossible:

See *The Emperor's New Mind: Concerning Computers Minds and the Laws of Physics*, OUP, 1989.

My review is in AIJ 1992 and in the Cogaff web site.

One of the many tasks remaining is to specify a kind of information-processing system that could discover new abstractions and represent them independently of the instances from which they are derived, as suggested in the previous slide; this well lead to a much better theory of concepts than the partly similar "prototype" theory of Rosch.

I am suggesting that a new synthesis, guided by investigation of tasks that arise naturally when a child or animal learns about the environment, may lead to major advances relevant to several disciplines.

However we also need to look at ways in which things can go wrong: they may be clues to the types of mechanisms and forms of representation that play a key role, and indicate some unavoidable flawed intermediate stages in the development of new competences, including new mathematical competences.

Compare Imre Lakatos *Proofs and Refutations* 1970.
He sometimes seems to draw a mistaken conclusion: that mathematics is empirical. But discovering and fixing buggy proofs differs from refuting empirical generalisations and consequently modifying them.

Towards a ‘logic’ of evolutionary and developmental change

We aim to treat evolutionary and developmental trajectories not just as brute empirical facts,

but as partly explainable by the logical relationships between:

- sets of requirements
- existing designs.
- possible modifications of those designs.
  (Evolution also has affordances)

How to carve up the research space

One possibility

- Everyone chooses some particular practical goal (or modelling goal) and tries to design a machine to achieve that goal, or works on a particular type of mechanism trying to improve its performance on an agreed set of tests.

Another possibility

- Some people try to work on understanding the space of possible goals, and requirements, and the space of possible designs for meeting those requirements.

How can we carve up such spaces?

- study different sorts of competences and behaviours, and types of mechanisms capable of supporting them; and use competence-based metrics for comparing designs (This is how a great deal of AI research now works.)
- study different sorts of complete systems and how to model or implement them
- identify different kinds of environment and study requirements for coping within them
- identify and compare different sorts of evolutionary and developmental trajectories and
  - the mechanisms involved in supporting the trajectories
  - the varying relationships between designs linked by trajectories
  - perhaps focusing on particular neighbourhoods in design space or niche space

Some Observations, Conjectures, Speculations

1. Evolution produced far more design solutions for far more problems than we have so far discovered. Identifying those problems is not easy.
2. You can't understand designs without knowing what problems they solved – and a complex design may have solved a very large number of problems.
3. You cannot understand any design without understand how and why it differs from other designs in its neighbourhood in design space (alternative designs) and how it relates to neighbourhoods in niche spaces (alternative sets of requirements).
4. Solutions to problems may have important and useful side-effects for which they were not selected, and not every important feature has a mechanism producing it.

   (E.g. Be very suspicious of a box labelled "emotion" in an architecture diagram.)
5. Human competences DO NOT scale up (Humans are defeated by combinatorics.)
6. Human competences DO scale out – they are combinable with others in novel ways.
7. Some high performance competences may not scale out.
8. Many novel combinations of human competences arise in 3-D spatial processes:
   Spatial coincidence and proximity (with and without temporal proximity) supports novel combinations of initially unrelated structures and processes:

   some animals acquire orthogonal recombinable competences
9. We do not yet understand what forms of representation of 3-D structures and processes (and related affordances) humans and other animals use.
10. A human without a normal body nevertheless inherits mechanisms that evolved in normal human bodies:
    human cognitive development, requires embodied ancestors more than it requires embodiment.

Main 'meta-' themes

Themes regarding the role of meta-X for various values of X:

- Meta-lavel content
- Meta-level formalisms and forms of representation
- Meta-level processes
- Meta-level architecture

  Architecture as implicit in the workings of a complex algorithm that has different phases, is not the same as a design which has processing resources "permanently" allocated to different functions.
  Example: **goal generation, prioritising, alarm mechanism, bug detection, ontology extension**

- Focus on requirements and designs arising out of internal processes and structures **vs** focus on requirements arising from the environment (with or without other agents)
- Issues about performance or efficiencey vs issues about what is possible at all.
- The nature of mathematical discovery and knowledge and its relation to meta-reasoning.
- Tools to support various kinds of meta-research (and meta-tools)

Warnings

We should not assume that the problem is simply a collection of inadequate tools, and immediately start trying to develop new tools, which has happened repeatedly in the history of AI.

The most obstructive gap in our knowledge is inadequate understanding of the problems, or, in engineering terms, the different sets of requirements that need to be satisfied by working systems of different sorts.

That includes, but is not restricted to, understanding the problems of learning and acting in a richly structured, changing 3-D environment.

Without doing the requirements analysis, building new tools, even biologically inspired tools, can lead us up blind alleys.

Contrast requirements based on the detailed structures, processes, and causal properties in the environment, vs requirements formulated in a general, domain-indepdent fashion.

This relates to, but reformulates, old debates between “neats” and “scruffies” in AI.
(Schank? Abselson?)

Design space(s) Niche space(s) and their relationships

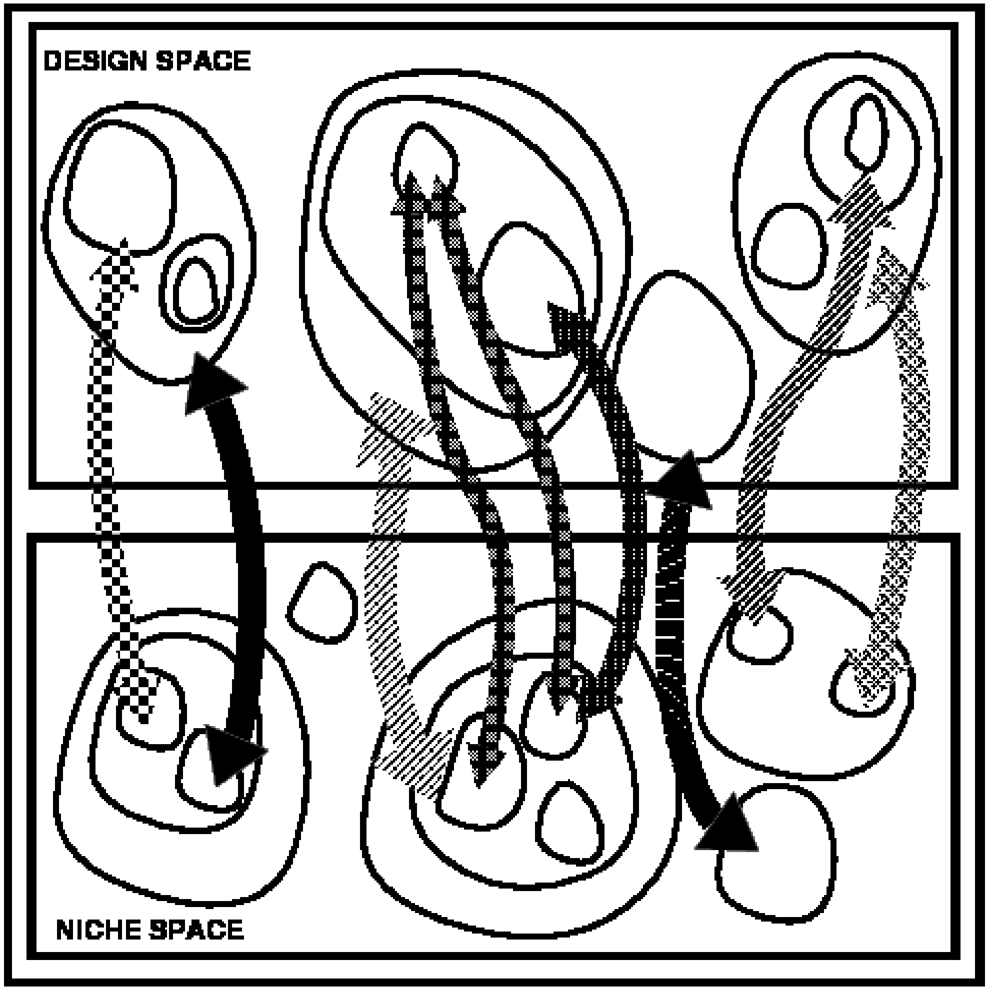
Design space: a space of possible architectures (including mechanisms, formalisms, etc.)

Niche space: a space of possible sets of requirements for whole animals, robots...

There are discontinuities in both design spaces and niche spaces: not all changes are continuous.

Do not expect one fitness function. Instead expect diverse structured fitness relations between designs and niches.

We can also talk about designs and niches for parts of an existing system: e.g. the niche for a digestive subsystem, or a motor control subsystem, or a perceptual subsystem depends in part on what is already in the rest of the machine or animal.



Thus not only species and whole organisms but also subsystems and their designs can co-evolve and co-develop.

Trajectories in both spaces

There are different sorts of trajectories through the two spaces.

i-trajectory: possible for an individual organism or machine, via development, adaptation and learning processes (of many types): egg to chicken, acorn to oak tree, etc.

e-trajectory: possible for a sequence of designs evolving through natural or artificial evolution. Requires multiple re-starts in slightly different locations.

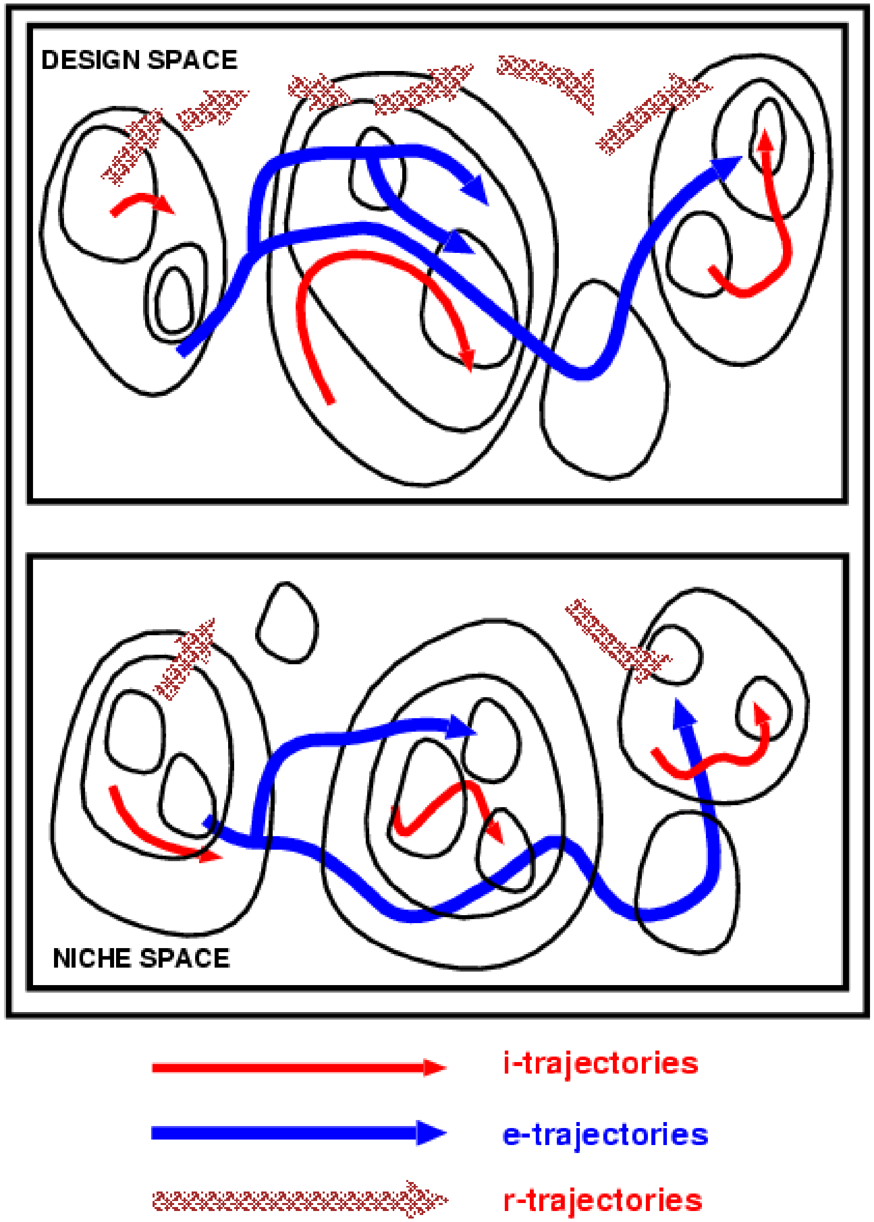
r-trajectory: possible for a system being repaired or built by an external designer whose actions turn non-functioning part-built systems into functioning wholes, or add a new feature: can produce discontinuous trajectories.

---

s-trajectory: possible for social systems with multiple communicating individuals. (Can be viewed as a type of i-trajectory.)

c-trajectory: trajectory made possible by the use of cognitive capabilities of individuals, e.g. mate selection or differential parental caring for young of different capabilities.

NOTE: All but r-trajectories are constrained by the requirement for “viable” systems at every stage.



Dynamics of Linked trajectories

Movements along trajectories in design space cause, and are caused by motion along trajectories in niche space:

Change in design of species A changes requirements for species B, which causes changes in design of species B, which changes requirements for A, etc.: there are networks of such interactions.

This obviously applies to e-trajectories, and less obviously to i-trajectories

Niches for an unborn foetus, for a newborn infant, a schoolchild, a parent, a professor, etc. are different – though the same human can be in them all at different times.

Moreover, an individual can instantiate more than one design, satisfying more than one niche: e.g. switching between being

- protector and provider,

  or
- parent and professor

To cope with development of multi-functional designs we can include *composite niches* in niche space, just as there are composite designs in design space.

Composite niches lead to composite designs and vice versa.

New mathematics needed?

The history of the biosphere involves multiple interacting e-trajectories for designs and niches, with many interacting feedback loops.

Previous slides suggest a need for new mathematics linking all the many complex feedback loops involved at different levels of abstraction between niches and interacting design instances.

Biological evolution:

As more and more complex organisms evolved, their i-trajectories became longer and more diverse.

Much later came s-trajectories and c-trajectories,

Maybe there will soon be r-trajectories (genetic engineering?)

Many questions: e.g. why are there so few "intelligent" species or individuals?
(Count species, individuals or biomass.)
Answer: the most sophisticated brains have to be near the tip of food pyramid.

Under what conditions does the (expensive) transition to deliberative capabilities pay off, compared with other design options?

`http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0604`
Requirements for a Fully Deliberative Architecture

Are those conditions very rare?

How many intermediate cases are there between purely reactive and fully deliberative mechanisms?

How many different forms of representation did evolution produce, and why?

What led to development of a-modal non-sensorimotor ontologies?

Why do we get some apparently intelligent species (mammals, birds) that are precocial while others are altricial?

Are there some constraints that require more sophisticated adults to start from less sophisticated infants?

How i-trajectories start: tabula rasa or ...?

The fact that some species, the precocial species, start so competent provokes the question:

> Why do other species, such as primates, hunting mammals and nest-building birds, start so helpless and incompetent?

Such species are labelled ‘altricial’.

Do they start with no knowledge and only a very general learning mechanism, or do they have sophisticated but ‘generic’ knowledge or ‘meta-knowledge’?

Evolution did not solve one problem with one solution:

(See papers and presentations by Chappell and Sloman)

John McCarthy on The Well Designed Child

Most AI researchers try to find a small number (preferably only one) of powerful, general, learning mechanisms that can learn from arbitrary data.

Compare John McCarthy: The Well-Designed Child

http://www-formal.stanford.edu/jmc/child.html

"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?

"Animal behavior, including human intelligence, evolved to survive and succeed in this complex, partially observable and very slightly controllable world. The main features of this world have existed for several billion years and should not have to be learned anew by each person or animal."

http://www-formal.stanford.edu/jmc/child.html

Biological facts support McCarthy

Most animals start life with most of the competences they need apart from some fine tuning – e.g. deer that run with the herd soon after birth, chickens(?), invertebrates, etc.

For them, there's no blooming, buzzing confusion (William James)

So why not humans and other primates, hunting mammals, nest building birds? ... and some future robots

Perhaps we have not been asking the right questions about learning.

We need to understand the nature/nurture tradeoffs, much better than we currently do, and that includes understanding what resources, opportunities and selection pressures existed during the evolution of our precursors, and how evolution responded to them.

This requires us to understand the environments involved, as well as mechanisms, architectures, etc..

Blooming buzzing confusion????

If John McCarthy is right, William James got it badly wrong.

Of course he did identify a problem.

But it is not the problem solved by baby mammals or birds.

It may be a problem solved by some other organisms, possibly long extinct organisms.

Compare the book by Eleanor Gibson and Anne Pick
An Ecological Approach to Perceptual Learning and Development,
2000

Varieties Of Worlds To Learn In

One way to explore the issues is to consider different sorts of worlds in which organisms might

- behave
- develop
- evolve

We can see how different environments produce both

- Different constraints
- Different opportunities

  (Gibson's notion of "affordances" for an individual can be generalised to affordances for a species, or gene pool.)

It's not only the environment:

As a species evolves, acquiring more complex body parts and more complex information processing capabilities, its constraints and opportunities change.

Changes in designs produce new requirements: new niches for subsystems

E.g. having an articulated body provides major new opportunities, and far more demanding requirements beyond mere mobility.

From chemical soups to skyscrapers

The AAAI'07 Symposium paper presented a few snapshots from a large space of possible environments, illustrating some of the ways in which the affordances in those environments vary.

`http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0704`

The examples include:

- amorphous chemical soups
- soups with some spatial structure
- soups with additional impervious structures, e.g. walls, forming corridors and doorways
- objects that need to be assembled from parts, or disassembled to provide parts, etc.
- Structured animals with independently movable parts
- Animals that need to enhance themselves with tools as the things they build get larger, more complex or more sophisticated.
- Animals with complex virtual machines attempting to understand themselves
- Animals with social and competitive relationships with other animals.

This is not an exhaustive survey, just an illustration of a type of methodology not widely recognized as appropriate to AI

Soup dwellers:

A chemical soup, even with different chemicals needed at different times, does not give much scope for cognition

- at most information about local gradients can be used.
- If there is some other field, e.g. gravitation or illumination field-strengths may correlate with nutrients and 'noxients': that may allow controlled traversal based on gradients to optimise feeding.

If things are changing there may or may not be learnable patterns in the changes.

Varying soup-challenges:

- If there are permanent obstacles separating regions with different chemicals, use of gradients may still suffice but more flexibility is possible if the geography can be discovered and used
  - E.g. maps of walls, corridors, rooms, with regions where different chemicals have greatest concentrations, etc.
  - NB: new architectures, forms of representation, mechanisms are needed to acquire, store and use such information.
- Other possibilities, distribution of nutrients and noxients, competitors for the same nutrients, predators, mobile food (prey organisms)

NB: Learning can be done either by evolution or by individuals – with different costs and benefits

Noxious and walled soups:

Fixed, impenetrable barriers to motion can also add opportunities to benefit from the ability to learn and remember their locations.

The barriers impose rigid constraints on motion whereas the noxients impose soft constraints.

Atomic sensors:

If new types of substance (nutrients and noxients) are introduced or removed from time to time, then the problem for evolution is much harder.

It needs to allow the information-processing architecture and the contents of representations to be take different forms in different generations.

There are two main ways of doing this.

If the changes are slow enough then, as the soup-world changes, the fixed architecture of the organisms can slowly vary from generation to generation, and over time some sequences of variation may produce surviving sequences of organisms in the new environments.

If environmental changes occur faster, then individual organisms need to be able to learn about the new substances.

That requires sensors to have more generality:

> e.g. by being able to detect 'lower-level' features that are common to different chemicals and finding out which combinations are good and bad, etc.

NB. It makes a big difference whether different types can be characterised simply by different vectors or whether structural descriptions are needed (e.g. for same atoms arranged differently).

From Moving to Manipulating:

Some environments make it possible for smaller fragments to be combined to form useful larger objects, and larger objects to be disassembled and fragments re-used.

This provides yet more challenges for evolution and learning.

This is also a requirement for digestion, metabolism, growth, repair of damage, immune responses – all of which use chemical mechanisms produced mostly by evolution, which operate in purely reactive ways in various parts of an organism's body)

Manipulation of external complex objects introduces new demands.

The effectors need to be more complex, so that objects or their parts can be moved together or apart or rearranged.

This requires the ability either to move one object to or away from another, including rotational movements when constructing 3-D objects (unless made from something like mud).

Some complex concurrent movements may be required.

These opportunities also substantially extend the requirements for information-processing.

Whereas previously the only kind of future to be considered in formulating goals or predicting consequences of actions was a future in which the organism's location (and possibly orientation) changed, manipulation involves far more varied changes than transformation of a fixed-size vector.

If planned objects can be more or less complex, that will require the ability to construct more or less complex representations of objects, instead of using only fixed-size, fixed-complexity vectors of measurements.

Exosomatic ontologies:

Exactly how these new structures should be represented is debatable.

An organism that can represent only relations between signal-patterns within its body (sensorimotor patterns and contingencies) uses only a 'somatic' ontology.
The ontology can be uni-modal or multi-modal.

If an animal or machine uses an ontology that refers to things that can exist or occur independently of how they are sensed or acted on by that agent, it uses an 'exosomatic' ontology.

Components of an exosomatic ontology require a type of semantic content that is not definable in terms of sensori-motor (somatic) concepts.

Such concepts cannot be accommodated within symbol-grounding theory.

However they can be represented by undefined symbols in a theory about the environment.

Their semantics may be partly implicitly defined by their role in the theory (model-based semantics), and partly by bridging rules (theory-tethering).
See `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#models`

There are many advantages to exosomatic ontologies including their economy in representing what is common to many different processes (or actions) and their use in making predictions, forming plans, forming explanations, including hypothesised histories.

From Inheriting to Discovering:

If the kinds of object and kinds of manipulation required in a particular environment do not change much then it is possible for evolution to produce combinations of sensors, effectors, and information-processing mechanisms, including ontologies and forms of representation required.

Compare precocial species.

Evolution needs to be replaced by learning if the environment changes faster than evolved designs can, for instance

- by providing new sorts of materials,
- producing new configurations of objects,
- producing new climatic conditions, e.g. temperature,
- new environments found by migration
- new behaviours in rival species

  (prey, or predators, or competitors for the same food or habitats)
- new arrivals from other locations

If these changes happen too fast, it will not be possible for evolution to hard-code all the required competences, even if the sensors and effectors provided are potentially very general.

Then some of the burden of accommodation needs to be shifted to learning.

Manipulation Changes What Can Be Learnt

In a 3-D world, as more and more complex objects are constructed by assembling available components, those complex objects themselves can be parts of still more complex objects.

So the fact that certain objects have been made makes it possible to discover that there are new kinds of objects that can be made in fairly small steps that could not previously be made in small steps.

As objects become larger the problems of manipulating them change, also requiring more complex combinations of actions to be created.

Likewise as actions are produced that assemble such objects, 'chunked' combinations of actions become available to be used as components in more complex actions

Sometimes the complexity involves doing longer sequences of things, and at other times it involves doing more things in parallel, using cooperative agents.

An important point on which much more research is needed is that spatial relationships between processes can cause complex interactions which may provide rich and complex positive and negative affordances.

This is relevant to understanding causation in 3-D structures and processes
`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/wonac`

The value of **sequential** chunking in reducing complex search spaces has been well known in AI for decades. I don't think chunking of **concurrent** spatially integrated actions has been studied much.

Ontology extension:

The ontology of the learner may also have to be extended to include new kinds of design, new kinds of tool, new kinds of construction process, new ways of collaborating with others, providing new contents for thinking, seeing, planning, and learning processes.

[Show video:
Yogurt can be food for the mind as well as for the body in an 11 month old scientist.]

Watching young children playing with toys of various sorts shows that things that seem obvious to older children may be completely incomprehensible to younger ones for a while, such as why putting a a puzzle piece in its recess and pressing does not make it return to its previous location.

[Show video of betty?]

Videos of the crow, Betty, spontaneously making hooks in several different ways, are available online at the Oxford zoology web site. Use a web search for “betty”, “crow” and “hook”.

Learning begets new needs:

By enabling the learner to produce novel structures, those learnt abilities themselves rapidly produce new opportunities (and sometimes threats – since new constructions can be dangerous as well as useful), possibly requiring another ‘layer’ of learning.

The learning capabilities that produced the early competences will not always suffice for producing the newer more complex ones.

It seems that somehow this was discovered by evolution.

What can be learnt by altricial individuals at various stages changes significantly.

Could evolution have produced this? (Chappell&Sloman, 2007, IJUC)

Cognitive epigenesis: Multiple routes from DNA to behaviour, some via the environment

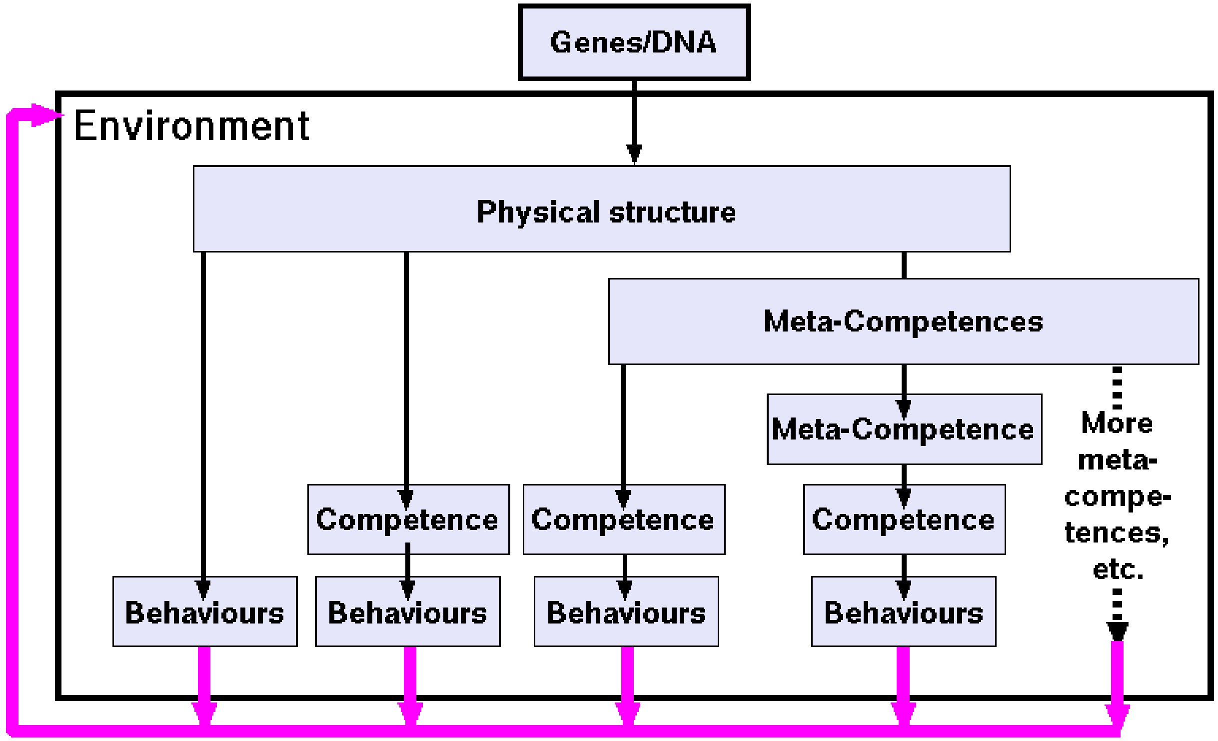
Pre-configured competences:

are genetically pre-determined, though they may be inactive till long after birth (e.g. sexual competences), and their growth may depend on standard, predictable, features of the environment, as well as on DNA.

They occur towards the left.

Meta-configured competences:

(towards the right of the diagram) are produced through interaction of pre-configured or previously produced meta-configured competences with the environment (internal or external).



(Chris Miall helped with the diagram.)

The environment changes the learning and development mechanisms.

Evolution 'discovered' that speed of learning is increased by active intervention: it produced some species that discover many facts about the environment, and themselves, through creative exploration and play, in which ontologies, theories and strategies are developed, tested and debugged.

PERHAPS INFANTS THAT STARE LONGER AT SOMETHING ARE TRYING TO DEBUG A THEORY?

Innate Context-Specific Meta-knowledge:

If the previous arguments are correct, then some meta-competences that enable or facilitate the acquisition of new competences (or new meta-competences), far from being general learning algorithms, are specifically tailored to finding things out about restricted environments.

Definitional vs Substantive Ontology-Extension

Patterns in sensory and motor signals define an immediately accessible ontology which is the only kind simplest organisms can use. (Compare Braitenberg's "vehicles".)

Exosomatic ontologies, e.g. referring to things in the environment that exist independently of being sensed or acted on require substantive ontology extension – a hard task, whether for e-trajectories or i-trajectories.

- This would be impossible if 'symbol-grounding' theory (concept empiricism) were true!
- Definitional (abbreviative) ontology extension uses new combinations of existing concepts (including relational concepts and logical connectives), to form new concepts that are found to be useful: they are stored for re-use and associated with a new label.

  A more sophisticated type of abbreviative ontology uses invariants found in patterns of change.

  E.g. if electric currents and voltages can be measured then a conductor may be found to exhibit an invariant ratio between current and voltage as current and voltage change: that is given a label "resistance".

  Many new concepts are introduced that way in science, and AI researchers and cognitive scientists have attempted to generalise this to more sophisticated sensorimotor invariants (e.g. O'Regan).
- Abbreviative labels are heuristically useful but do not extend the expressive competence of the learner.

  Like "macros" they sometimes help the search for new useful combinations of old concepts.
- Abbreviations pick out useful subsets from the very large space of possible concepts and possible laws already expressible in the learner's ontology.
- Substantive ontology extensions go further.

Substantive concept learning:

- In contrast, substantive concept learning produces new concepts that are not definable in terms of the initial set, allowing construction of new theories or hypotheses that were not previously expressible.
- Some previously known facts may later turn out to support or contradict those theories, and old puzzles may be explained by the new theories.
- Such cases are familiar from the history of science.
- Substantive ontology extension includes learning about different kinds of matter and properties of various materials that are not detectable using available sensors, for instance, solubility or atomic number of elements.
- What is learnt through the application of meta-competences includes what sort of ontology is useful in the environment, as well as which laws using that ontology work well for making predictions in the environment.
- But it is not a simple algorithmic process, for it involves controlling search in the vast space of ontology extensions.
- The search is controlled by introducing new concepts only when a current explanatory theory needs to be extended or modified because some plans, predictions and explanations turn out to be inconsistent with available evidence.
- Theory formation is called “abduction”, but AI abduction systems do not usually allow ontology extension (use of new undefined symbols in “axioms” added to a theory).

"Inward" ontology extension

An organism may need sophisticated forms of control that requires it to have a model of itself, as has often been proposed.

E.g. Owen Holland, and the meta-management architectural layer in H-CogAff, and many others

This inward looking theory about the world will often be most usefully a theory of high level virtual machinery rather than the detailed physical mechanisms and processes.

For this the system will have to beyond what internal physical sensors can sense – it will need sensors that perceive virtual machine states: another example of substantive ontology extension.

None of this implies that such a system has a special part which could be described as "the self".

Even if it has a self-model, that may be much more complex and fluid than many people suppose.

For a tutorial introduction to confusions about the notion of a "self" and some of the unobvious complexities of self-models see

`http://www.cs.bham.ac.uk/research/projects/cogaff/misc/the-self.html`
"The self" – a bogus concept.

Forms of representation for ‘inner languages’

Previous work with biologist Jackie Chappell suggests that in some species the kinds of perceptual, planning, problem-solving, and plan execution competences that develop require sophisticated internal forms of representation that we call ‘generalised-languages’ abbreviated as ‘g-languages’, or GLs.

A GL has three features often assumed to exist only in human languages,

- structural variability (including Fregean and analogical formalisms)
- (context sensitive) compositional semantics
- goal directed manipulation

GLs

(a) Structural variability in a GL allows the construction of complex information structures with varying numbers of parts in varying relations.

This includes the use of lists, trees, and networks containing different sorts of items of information. It also includes distributed as well as localised and geometrical as well as logical forms of composition.

(b) Compositional semantics in a GL allows any information structure to occur as parts of several different larger information structures,

the information (meaning) expressed in the larger structure will be determined by (i) the structures of which it is composed, (ii) how they are organised, along with (iii) relevant contextual information.

Different notions of part, whole and composition to form complex information structures are possible.

Would use of a GL be essential for 'fully deliberative' architectures?

(described in `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0604`)

The parts in an intended configuration of objects could include parts of the animal, e.g. hands, feet, or jaws, which might be required in intermediate phases of the production of some new complex object. Animals building nests, peeling bananas, or using two rocks to crack goal state (e.g. termites?).

Alternatively such actions may depend on using a GL to represent a new goal which is then used to control actions constructing, disassembling or rearranging objects, using do next.

Very few types of animal can do this.

Deciding which ones can is hard. (Chappell/wonac)

Representation is always for an information user and constrained by that user's information-processing capabilities.

GLs precede communicate languages

GLs must have evolved for internal use prior to the evolution of human language – though their use for external communication probably accelerated their development.

Likewise GLs must be available for internal use prior to the

This transforms the nature of the language learning task: for the learner already has rich, structured, semantic contents available to communicate,

including possibly questions, goals, and plans,

This contrasts with theories of language learning that assume the child has to learn both how to mean and what to mean at the same time as learning how to communicate meanings.

It also allows learning an external language to be a collaborative creative problem-solving task in which the children are normally in a minority and “politically weak”.

Contrast the Nicaraguan deaf children

More on inner languages

GLs are not restricted to linear strings of symbols or to Fregean languages using a syntactic form composed entirely of applications of functions to arguments.

(Sloman 1971) suggested long ago that analogical representations using other modes of composition are sometimes useful for representing and reasoning about spatial configurations among other things.

Analogical representations, including diagrams and maps are capable of supporting structural variability and (context sensitive) compositional semantics since parts of diagrams can be interchanged, new components added, etc.,

But they don't have to be isomorphic with what they represent, as should be obvious from the fact that 2-D pictures can represent 3-D objects (e.g. the Necker cube).

The relationship is more subtle and complex than isomorphism, and can be highly context sensitive.

Internal GLs may use analogical representations not yet known to science.

The use of such representations externally (e.g. on paper, in 3-D models) usually has to be learnt or developed – the representations only work for people, animals, or machines that have suitable information-processing mechanisms.

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

Comment from Karmiloff-Smith

Annette Karmiloff-Smith:

"Decades of developmental research were wasted, in my view, because the focus was entirely on lowering the age at which children could perform a task successfully, without concern for how they processed the information."

Karmiloff-Smith, A. (1994).
Precis of Beyond modularity: A developmental perspective on cognitive science.
Behavioral and Brain Sciences 17 (4): 693-745.
(Preprint: `http://www.bbsonline.org/documents/a/00/00/05/33/index.html`)

We could make a similar comment about people studying under what conditions various animals do interesting things, without explaining how they do any of those things.

Comment: We need both

- deep understanding of important PROPERTIES OF ENVIRONMENTS animals or children interact with, and
- deep theories about the INFORMATION PROCESSING MECHANISMS that make it possible to engage fruitfully with those environments

Should we teach students to study environments in greater depth?

We need to expand alternatives to be considered

- Evolutionary knowledge about the environment could be partly encoded in strategies for learning about 3-D structures and processes by performing experiments, and for debugging what is learnt.
- What appears to be random ‘motor babbling’ in an infant could be part of a controlled set of experiments. (Compare Gibson and Pick, 2000)
- Humans and animals are not unitary entities so that you can ask: what does it perceive, know, want? There are many different subsystems operating in parallel, and they need not communicate fully.
- It is often assumed that discovering causes is discovering correlations, or laws relating observed phenomena (Hume). I’ll try to show that some causal understanding goes deeper and is based on understanding of interacting structures – not necessarily all visible (Kant).

Learning About Causation

Discussion notes and presentations on the CoSy web site

Claim: the ability to manipulate (possibly analogical) representations of spatial structures can be the basis of a kind of causal competence that enables a reasoner to understand why a certain event or process must have certain effects.

This uses a Kantian conception of causation that involves more than mere reliable correlation: there is a geometrical necessity in the relation between cause and effect.

It is not easy to determine what forms of representation and inference are used in animals (or children) that cannot talk.

Some of the problems of investigating causal understanding in non-human animals, were discussed by Jackie Chappell in her WONAC presentation:

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

Also Gibson and Pick (2000).

THANK YOU!

For the importance of virtual machines and supervenience see

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

Why robot designers need to be philosophers (and vice versa)

For ideas about how machines or animals can use symbols to refer to unobservable (theoretical) entities see

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

Introduction to key ideas of semantic models, implicit definitions and symbol tethering

For an argument that internal generalised languages (GLs) preceded use of external languages for communication, both in evolution and in development, see

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

What evolved first: Languages for communicating, or languages for thinking (Generalised Languages: GLs) ?

For a challenge to learning theorists: how to do dimensionality expansion to invent 3-D explanations for 2-D experiences (rotating necker cube):

http://www.cs.bham.ac.uk/research/projects/cogaff/misc/nature-nurture-cube.html

Requirements for going beyond sensorimotor contingencies to representing what's out there
(Learning to see a set of moving lines as a rotating cube.)

Invited presentation to machine consciousness symposium AAAI 2007 fall symposium:

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

Additional papers and presentations, including presentations on causation

http://www.cs.bham.ac.uk/research/projects/cogaff/talks/
http://www.cs.bham.ac.uk/research/projects/cosy/papers/
http://www.cs.bham.ac.uk/research/projects/cogaff/talks/wonac

Especially papers and presentations by Chappell and Sloman