CS&AI Theory Seminar Friday 6 Sep 2001

Also presented at SSAISB/SGAI Joint Meeting City University March 2006

When is seeing (possibly in your minds eye) better than deducing, for reasoning?

Aaron Sloman

http://www.cs.bham.ac.uk/˜axs

School of Computer Science

The University of Birmingham

These slides are available online as

http://www.cs.bham.ac.uk/research/cogaff/talks/#talk7

THANKS

To the developers of Linux
and other free, portable, reliable, software systems,
e.g. Latex, Tgif, xdvi, ghostscript, Poplog/Pop-11, etc.

The problem

In Philosophy, Psychology and AI there is a very long history of discussion about different forms of reasoning, including

- **What they are**
- **how they work**
- **whether they are valid**

In particular, in AI there are many who emphasise the role of logic, and others who emphasise the need to understand and replicate non-logical human reasoning capabilities, including visual reasoning.

The latter is obviously a requirement for AI researchers mainly interested in understanding how human minds work.

But perhaps AI researchers who are primarily engineers trying to design effective reasoning systems could overcome some of the current difficulties if they were better able to model natural reasoning systems.

AI researchers should investigate varieties of forms of representation, and the various modes of reasoning associated with them, instead of simply assuming one mode, whatever it is, will do for all purposes.

What is logic?

That's actually a hard question to answer adequately.

Logic as we understand it is a combination of

- **form of representation**
- **modes of reasoning or inference**

with the following features:

- **Topic neutrality: logical concepts and methods are equally applicable to any subject matter whatsoever – physical, mental, theological, poetic, etc.**
- **Use of precisely (mathematically) specifiable syntax for terms, sentences, derivations.**
- **The main syntactic operation is application of a function to some arguments (where the arguments may themselves be functions), e.g.**

  P(a), R(a,b,c), S1 and S2, $\neg$ S1, S1 or S2, S1 $\iff$ S2, S1 $\Rightarrow$ S2
  $\forall$ y ($\exists$ x P(x,y) or $\exists$ x P(y, x))

- **Ability to embed non-logical terms within the syntax, e.g. 'red', 'bigger', 'Fred'.**
- **Compositional semantics: the meaning of a complex expression is completely determined by the meanings of the atomic components and the syntactic structure of the expression.**

Pop11 has a propositional logic tutor

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

Run pop11

```
: lib logic
: logic();
```

Type h to get a list of commands.

Exit with: bye
To find out more, in VED/XVed do

```
help logic
```

Examples:

```
P    print current formula (using symbols)
TB   print truth-table for current formula.
HT   get help with truth-table for current formula.
EV   evaluate current formula for different truth-values
NF   generate a new formula at random.  (These get increasingly complex.)
G    play the "guessing" game.
EQ   Test your ability to construct equivalent formulas
VAL Test your ability to detect valid inferences
H    Help -- prints out this file
HH    Prints out 'beginners' help file
STOP Stop current command
BYE Finish.
```

One logic, or many logics?

Logic was once thought to be *unique* though we now know that there are different logics which can differ in their primitive concepts, their syntactic forms, and their rules for derivations. **E.g.**

- **Propositional logic**
- **Predicate logic**
  - **First order**
  - **Higher order**
- **Many kinds of modal logics – using operators: □ ◊**
  - **Alethic logic (necessarily, possibly)**
  - **Deontic logic (obligatory, permitted)**
  - **Epistemic logic (knows, believes, ...)**
  - **Temporal logic**

    it will (sometimes, always) be the case that P

    it was (once, always) the case that P, etc.
  - **Parametrised modal logics (Modal operators have parameters, e.g. actions.)**
- **Intuitionistic logics**

  In “classical” logics from not(not(p)) you can infer p

  but not in “intuitionistic” logics

If there are many varieties of logic, what’s to stop us using even more forms of representation and inference – e.g. using pictures?

Why has logic been important in AI?

Logics, especially propositional and predicate logic, have been useful in AI mainly because their syntax can be mechanised.

- **I.e. it is not difficult, using computers, to create programs that construct, recognise, analyse and transform expressions that conform to the precisely defined syntax of a particular logic.**
- **That is in part because computers were designed to perform operations onto which such tasks could be mapped.**

In contrast, little is known about the other varieties of forms of representations used by human and animal brains, and attempts to model them have made far less progress.
Modelling human visual (spatial) reasoning abilities has proved extremely difficult, except for very simple and restricted cases.

Often this has been based on the assumption that mental visual operations use something like physical images, which can apparently be represented in 2-D arrays.

Many people have developed algorithms for manipulating 2-D images arrays in the hope that a sufficiently large collection of such algorithms will model or replicate human visual/spatial reasoning.

This may turn out to have been a huge red herring, if we actually use quite different forms of representation.

Visual/Spatial reasoning

Humans seem to have different sorts of reasoning capabilities.

Logical reasoning is one of them.

- **Premisses**
  - **All fleas are insects**
  - **Horace is a flea**
- **Conclusion**
  - **Therefore Horace is an insect**

Another special case is our ability to use diagrams and visual images, even in reasoning about very abstract mathematical problems, e.g. thinking about the complexity of a search strategy.

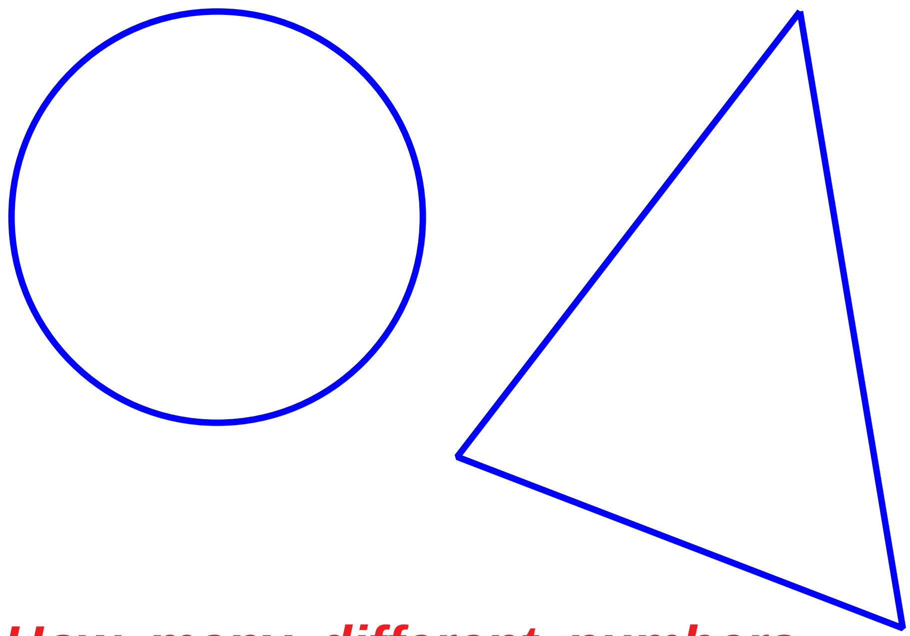
These do not depend on the syntactic forms and rules that characterise logical reasoning, for they involve other modes of representation, with different structures and different types of transformations.

Visual reasoning in humans

E.g. No points are common to the triangle and the circle. Suppose the circle and triangle change their size and shape and move around in the surface. They could come into contact.

If a vertex touches the circle, or one side becomes a tangent to the circle, there will be one point common to both figures. If one vertex moves into the circle and the rest of the triangle is outside the circle how many points are common to the circle and the triangle?

How do humans answer the question on the right?



How many different numbers of contact points can there be?

This requires the ability to see empty space as containing possible paths of motion, and fixed objects as things that it is possible to move, rotate and deform. Does it require *continuous* change?

Perhaps: but only in a virtual machine! **To be discussed another time.**

Some people (e.g. Penrose) have argued that computers cannot possibly do human-like visual reasoning e.g. to find the answer 'seven' to the question.

Machines vary in what they use, or manipulate:

1. Matter
2. Energy
3. Information (of many kinds)

- Scientists and engineers have built and studied the first two types of machines for centuries. Newton provided the first deep systematisation of this knowledge.
- Until recently we have designed and built only very primitive machines of the third type, and our understanding of those machines is still limited.
- Evolution 'designed' and built a fantastic variety of machines of all three types – with amazing versatility and power long before human scientists and engineers ever began to think about them.
  - Biological organisms are information-processing machines, but vary enormously in their information-processing capabilities.
  - There are myriad biological niches supporting enormously varied designs, with many trade-offs that we do not yet understand (e.g. trade-offs between cheapness and sophistication of individuals).
  - The vast majority of organisms have special-purpose information-processing mechanisms with nothing remotely like the abstractness and generality of TMs
  - A tiny subset of species (including humans) developed more abstract, more general, more powerful systems. Turing's ideas about TMs were derived from his intuitions about this aspect of human minds. But at best that's a small part of a human mind.

  Understanding all this requires us to think more about architectures than about algorithms.
  See: http://www.cs.bham.ac.uk/research/cogaff/talks/

We don't yet know how many ways there are to represent and manipulate information

Even if logic can be used to represent anything at all (which is doubtful) it should not be assumed that it is always best form of representation for every task.

E.g. why do we use maps, musical notation, flow-charts?

Note that not all maps and pictures need to preserve metrical properties of what they represent. Example: London tube map.

Exercises

- Consider some of the tasks for which you use maps.
- Investigate how those tasks would change if, instead, you had logical descriptions of everything represented in the map.
- For which tasks is logic better?
- Why do maps have to have symbols on them,
  and why do they need an index of places (gazetteer)?

A little philosophical history

Hume thought that there are only two forms of knowledge:

- analytic (based only on definitions and logical deduction)
- empirical (requiring observation and experiment)

  (Hume thought everything else was nonsense,
  e.g. theology, and the books should be burnt!)

Immanuel Kant, in his *Critique of Pure Reason* argued against Hume.

He claimed

- there is also non-analytic, non-empirical (synthetic apriori) knowledge,
- e.g. in mathematics: we can gain new insights by examining structures and their properties, including the use of diagrams (on paper or in the mind) in doing Euclidean geometry.

Even Frege agreed, partly

Gottlob Frege was one of the greatest logicians of all time. He invented predicate calculus and was the first to clarify the notion of higher order functions (functions of functions) e.g. quantifiers.

Frege disagreed with Kant about arithmetic, since he thought arithmetic could be reduced to logic. I.e.

- He thought the concepts of arithmetic could be defined in terms of purely logical concepts
- He thought that all the truths of arithmetic could be derived purely from the axioms of logic plus the aforementioned definitions.

  (In modern jargon, that would prove that the truths of arithmetic are all ‘analytic’, not synthetic.)

Frege tried to prove that all of arithmetic could be derived from logic

He tried to demonstrate this as follows:

- He tried to show that all arithmetical concepts could be defined in terms of purely logical concepts.
- He tried to show that all arithmetical truths could be proved on the basis only of logical axioms, rules and the definitions.

But the attempt fell foul of Russell's paradox.

Let S be the set of all sets that are not members of themselves.

Then

S is a member of S $\Longleftrightarrow$ S is NOT a member of S

Therefore:

S is and is not a member of S

Subsequent attempts to fix this this remain controversial.

But Frege thought Kant was right about truths of geometry being non-empirical, and non-analytic, i.e. he too thought they were *synthetic apriori* truths.

(Show pythagoras demo.)

Focus on modes of reasoning not kinds of truths

Disagreement over who was right still continues.

- **One problem with all this is that there may be different ways of arriving at the same (or very closely related) results.**

  E.g. you may be able to prove using something like Frege's method or a later variant, that this is a truth of logic

  $$7 + 5 = 12$$

- **But there may be another proof of a very similar result, where the concepts are defined not in terms of logic, but in terms of visually detectable structures.**

- **So perhaps we should not ask about the status of the propositions themselves, i.e. attempting to distinguish different kinds of truths, but instead talk primarily about different means of proof.**
- **The things proved by different means may look the same, but actually have subtly different contents.**

KANT'S EXAMPLE: 7 + 5 = 12

It is obvious that this equivalence is preserved if you spatially rearrange the blobs within their groups:

```
ooo          o      oooo
ooo   +      o   =  oooo
o          ooo      oooo
```

Or is it?
How can it be obvious?
Can you *see* such a general fact?
How?

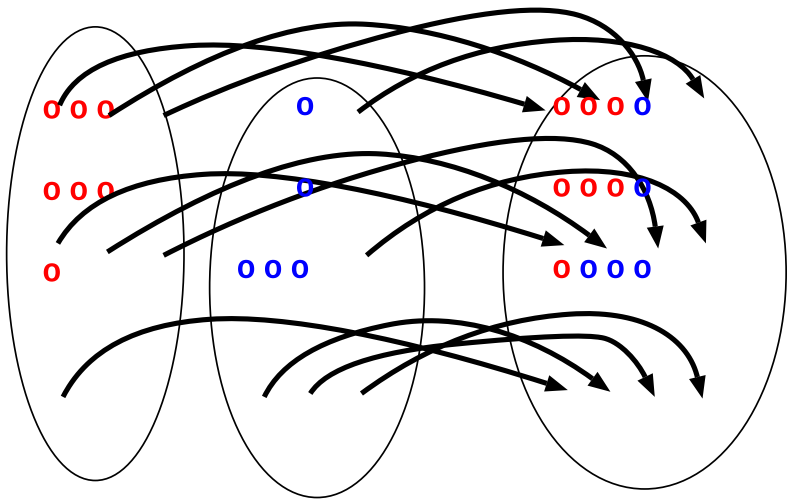
What sort of equivalence are we talking about?

I.e. what does "=" mean here?

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

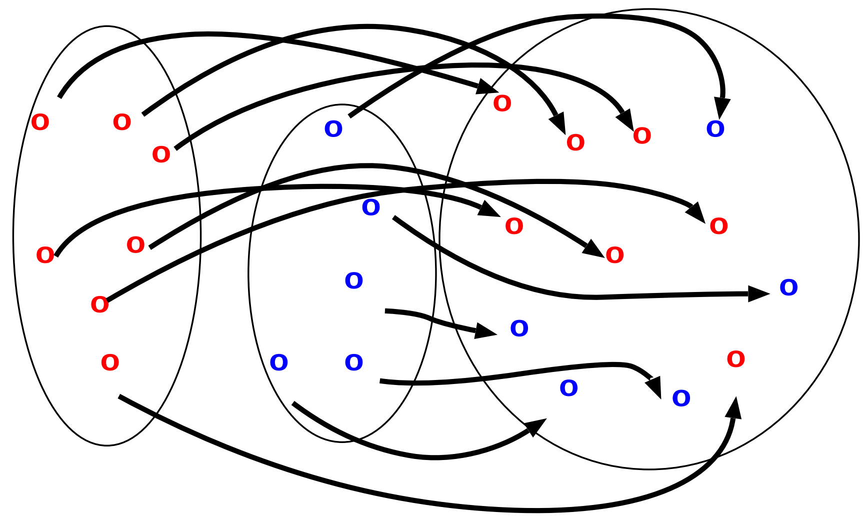
That can be defined logically, but the idea can also be understood by people who do not yet grasp the logical apparatus required to define the notion of a bijection.

SEEING that 7 + 5 = 12



Join up corresponding items with imaginary strings.

Then rearrange the items, leaving the strings attached.



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

Is it 'obvious' that the same mode of reasoning will also work for other additions, e.g.

777 + 555 = 1332

What does a diagrammatic proof prove?

Even if this is a way of discovering a truth,

is it the same truth as is proved starting from purely logical definitions and axioms?

Or do we have several *different* concepts of number?

And different arithmetical truths expressed using the same forms?

How many different meanings are there for:

777 + 555 = 1332

Consider

- **numbers of oranges,**
- **numbers of numbers,**
- **adding areas,**
- **adding operations on numbers,**

etc.

Both right?

Perhaps Kant and Frege were both right:

- **there are some analytic truths of arithmetic and some non-analytic ones.**
- **they differ in that they use concepts that are understood (defined) in different ways, using different conceptual resources (and different cognitive mechanisms)**
- **but there is a very strong structural correspondence between them.**
- **Understanding that relationship would be part of the task of philosophy of mathematics, and of AI.**
- **Maybe educators also need to understand it, in order to teach mathematics effectively.**

Who can do it?

What are the information-processing requirements for

- **being able to grasp structural relationships**
- **being able to visualise transformations**
  (e.g. seeing that dots can be rearranged),
- **being able to grasp higher level generalisations that are preserved by such transformations**
  (e.g. seeing that the one-to-one correspondence is preserved)?

Can a dog or a monkey see such truths?

Can a two year old child?

If not, then why not?

What changes when a child becomes able to see them?

What do other animals (and some humans) lack that prevents them learning to see such things?

(Kant thought that sort of ability was innate.
I don't think he ever considered other animals, or dreamed of the possibility of intelligent machines.
Had he been born 200 years later he would been an AI enthusiast.)

A partial analysis

The ability to see arithmetical truths using a grasp of spatial structures requires at least:

- **The ability see the spatial structures involved in the proof.**
- **The ability to see possibilities for variation in those structures (e.g. rearrangements of components, as in logical reasoning)**
- **The ability to grasp features that are invariant under those rearrangements.**
- **The ability to grasp a collection of structures, possibilities for change, and invariants, involved in a sequence of configurations or maybe even a continuous transformation covering a range of configurations.**

Compare logic:

Everything is discrete and all syntactic composition involves function application

Information processing architectures that are able to support human visual reasoning capabilities will be far more complex than those required for logical reasoning.

Reasoning is part of what we call seeing

That specification of requirements for visual reasoning is very vague, and would not be easy to mechanise in a general way in an AI program.

Those features are involved in much of our ordinary use of seeing, e.g. when we think about possible ways of rearranging furniture in a room, or possible routes from here to there.

This requires grasping structures, seeing possibilities for change, seeing "affordances" etc.

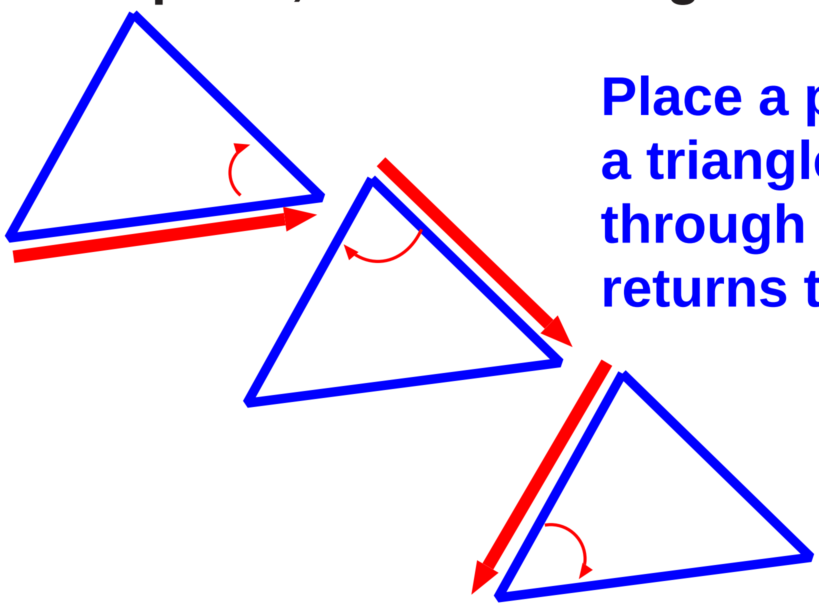
(See the talks on vision here
http://www.cs.bham.ac.uk/research/cogaff/talks/ and
my 'Actual Possibilities' paper in http://www.cs.bham.ac.uk/research/cogaff/)

When we have understood how human (and animal) visual capabilities evolved, what their functions are, what sorts of architectures support them, and what sorts of mechanisms and representations are used in those architectures, then perhaps we shall be in a far better position to understand what our reasoning capabilities are.

Example 2: The angles of a triangle

Why do the angles of a triangle add up to a straight line?

Most people cannot remember the proof they were taught as children. Mary Ensor (former Sussex student) invented this highly memorable proof, while teaching mathematics:



Place a pencil along an edge of a triangle and rotate it in turn through the three angles until it returns to the original edge.

How much has it rotated?

Of course, like the standard proof, this gives wrong results on a sphere. (Visualise a triangle with three right angles.)

But such a proof still teaches one something.

Example 3: List processing

Consider the list-process operation rotate, which when given an integer N and a list L, creates a new list by moving N items in turn from the beginning to the end of the original list. Most programmers could easily define such a procedure. So:

rotate(2, [a b c d e]) = [c d e a b])

What can you say about the following general expression, where L is any list:

rotate(length(L), L) = ???

Does anyone reason about this by starting from logical axioms and definitions and using logical deduction?

That’s what many AI theorem provers would do, e.g. using structural induction over binary trees, etc.

But do they prove the same thing as we *see* to be obviously true?

We abstract away from the tree-structured implementation of lists and reason about them as spatially manipulable linear structures.

This is important for teaching list processing.

... continued

If you can see the operations in your mind's eye the result is obvious.

But seeing a particular case is one thing, e.g.

rotate(5, [a b c d e]) = [a b c d e])

Seeing the general principle is quite another.

Can you see in your mind's eye the core properties of a spatio-temporal process that is common to a large (infinite) set of cases, covering all possible lists?

Example 4: Transfinite ordinals

We can visualise, and reason about, things that are impossible to see – including the infinitely thin and infinitely long lines of Euclidean geometry. More recently, mathematicians have discovered transfinite ordinals: infinite discrete structures.

The simplest one is the familiar sequence of natural numbers

TO1: $1, 2, 3, 4, ...$

From that we can easily derive new ones by rearranging them. E.g. take all the even numbers followed by all the odd ones:

TO2: $2, 4, 6, ..., 1, 3, 5, ...$

Questions:

1. In TO1, is it possible to move between the numbers 398 and 300002 without passing the number 1057?
2. Same question for TO2
3. In TO2, is it possible to move between the numbers 398 and 300003 without passing the number 1057?
4. In TO2, going from 999999 to 222, which will you meet first, 777 or 888 ?

How do you think about such questions?

More on transfinite ordinals

Consider the transfinite ordinal obtained by taking each prime number in turn, and for each of them forming a sequence of all their powers, followed by the sequence of the powers of the next prime.

$$2^1, 2^2, 2^3, \ldots\ldots \quad 3^1, 3^2, 3^3 \ldots\ldots \quad 5^1, 5^2, 5^3, \ldots\ldots \quad 7^1, 7^2, 7^3, \ldots\ldots$$

Is the sequence well-ordered? I.e. does every arbitrary sub-sequence of the sequence have an earliest member?

What if we reversed the order of the powers of every second prime?

$$2^1, 2^2, 2^3, \ldots\ldots \quad \ldots\ldots 3^3, 3^2, 3^1, \quad 5^1, 5^2, 5^3, \ldots\ldots \quad \ldots\ldots 7^3, 7^2, 7^1, \ldots\ldots$$

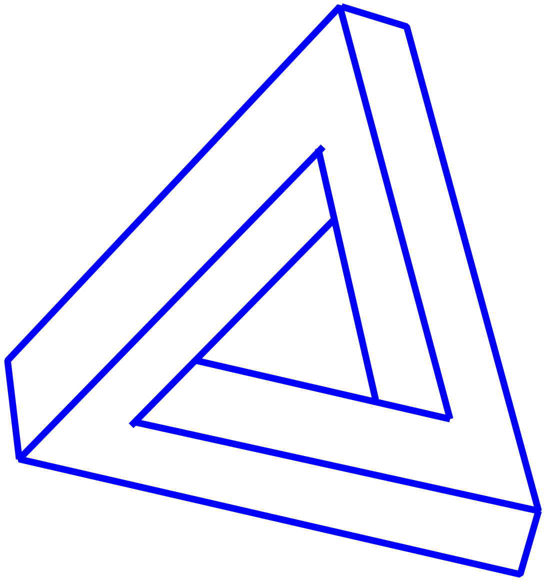
It is possible to produce logical axioms characterising these infinite ordinals, and derive answers to questions using those axioms.

Humans can use both logical and visual modes of thinking about such structures. Something about how our visual systems evolved produced, as a side-effect, the ability to visualise not only naturally occurring but also physically impossible objects.

Two physically impossible objects

Just as logic does not prevent us forming contradictory assertions, the syntax of spatial representations does not prevent us depicting incoherent spatial structures. E.g.

- **The Penrose triangle:**



- **A round square:**

(viewed from the edge)

There are many more examples

There are many uses of human spatial reasoning.

- Knowing where to look for an object thrown over a wall
- Route planning
- Engineering design
- Seeing how to assemble a toy crane from components in a meccano set: including seeing that a particular assembly will NOT work (e.g. because a quadrilateral is not rigid, or because a triangle is).
- The use of spatial concepts in many programming designs, e.g. the notion of a search space.
- Understanding the relationships between spatially defined search strategies and syntactically specified programs, e.g. depth first search uses a stack and breadth first search a queue.
- Many uses in physics, e.g. the notion of a phase space, a trajectory in phase space, an attractor in phase space
- Many examples in control engineering, e.g. the notion of a feedback "loop"

Can we replicate, or even explain, these human capabilities?

Conjecture:

Our visual reasoning capabilities in realms like mathematics are in part a side-effect of the interactions between different components in an animal architecture.
Those components evolved at different times, to serve different sorts of purposes.

For more on these ideas see these papers:

- http://www.cs.bham.ac.uk/research/cogaff/sloman.ppsn00.pdf
- http://www.cs.bham.ac.uk/research/cogaff/sloman.bmvc01.pdf
- http://www.cs.bham.ac.uk/research/cogaff/Sloman.actual.possibilities.pdf
- http://www.cs.bham.ac.uk/ axs/misc/draft/esslli.pdf
- http://www.cs.bham.ac.uk/research/cogaff/sloman.diagbook.pdf
- (or Postscript versions)

There are also many books and journal articles on diagrammatic or spatial reasoning.

E.g. The latest issue of the AI Journal (Vol 145, Nos 1–2, April 2003) has an article by M. Anderson and R. McCartney

Towards an explanatory theory

The Birmingham Cognition and Affect project has been investigation architectures for “complete” intelligent agents.

It is conjectured that the various architectural components can be located in the nine main regions of the COGAFF architecture schema:

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

...continued

- There are reactive mechanisms involved both in triggering a variety of responses, including internal and external reflexes, and in managing tight feedback control loops (e.g. posture control, grasping accurately)
- There are deliberative mechanisms that require the ability to reason in a discrete fashion about possible futures, possible unobserved entities (the back of that tree), possible explanations (possible pasts) possible sequences of actions (plans for oneself, predictions regarding others).
- There are self-monitoring (meta-management, reflective) capabilities that involve the ability to monitor, categorise, evaluate, and perhaps modify internal states, processes and strategies.

These evolved at different times, are present to different degrees in different organisms, develop at different stages in human beings.

They involve multiple forms of representation, with multiple mechanisms for manipulating those representations, and multiple varieties of semantics.

Interactions between concurrent modules

If there are so many concurrently active components

- doing different tasks,
- using different representations (possibly expressing information derived from the same source, such as a retina)
- including tasks where one module monitors another

then powerful new capabilities (and bugs) can emerge from the interactions.

Investigating all this requires a long hard slog, including attempting to characterise with far greater precision than before the types of visual competences found in humans, and the varieties of forms of representation and information manipulation that can occur.

Maybe logic will have some useful role as part of this.

Apologies

In fact I believe there are practically no adequate theories yet about how any of this works

There are fragments of theories, but it is not clear that the fragments can be put together to form a coherent whole with the right explanatory properties.

Looking into brains won't necessarily help any more than looking into computers will tell you about the virtual machines running in them.

SO THIS IS REALLY JUST AN INTRODUCTION TO A LONG TERM RESEARCH PROGRAMME. VOLUNTEERS WELCOME.

There is of course some useful work that has been already done by others.

For additional presentations relevant to this talk see

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

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