AAAI-11 Tutorial MP4

Philosophy as AI and AI as Philosophy

Those who are ignorant of philosophy are doomed to reinvent it – badly.

(Apologies to George Santayana)

Aaron Sloman

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

These slides will be available in in my 'talks' directory:

http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#aaai11 (talk 96)

WARNING

Note: August 22, 2011

These slides have been expanded considerably since the tutorial was presented on Monday 8th August.

They are likely to continue changing.

So please, if possible, save/communicate the link, not the pdf file.

`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#aaai11`

My slides have more material than can be covered in a presentation, partly because they are designed to be readable and (mostly) intelligble without someone presenting them.

Feel free to send me comments, questions, criticisms, suggestions for improvement.

Associated Web Site

The web site describing the tutorial for the benefit of people considering whether to attend is:

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

Contents of the web site (Also liable to change)

- Prerequisites for attendance
- Request to those thinking of attending
- Overview
- Provisional Schedule
- Example Topics (Expanded in these slides)
- Reading matter relevant to the tutorial
  (including several slide presentations related AI, Philosophy and Biology).
  - Relatively accessible
  - More technical
  - Other Sources
  - AI and Philosophy
- Speaker Bio
- Online papers, presentations, and teaching materials.

  Suggestions for additions are very welcome.

  See also the AITopics web site, especially the philosophy section.

  `http://www.aaai.org/AITopics`

PARTICIPANTS

Background and interests of participants.

Please see request for information from participants on the tutorial web site.

The presentation was intended to be suitable for people:

1. Doing research in or studying AI, as
   - students
   - academic faculty
   - researchers/industry

2. Philosophers (students, faculty, etc.)

3. Working in other disciplines, e.g.
   - Biology
   - Linguistics
   - Mathematics
   - Neuroscience
   - Physics/Chemistry
   - Psychology

A show of hands at the tutorial indicated that those present had varied backgrounds though most were in category (1).

I did not keep a record of the detailed break down. About 37 had registered for the tutorial, but a larger number (about 50?) attended.

Overview

Although most AI research has engineering objectives, some researchers are primarily interested in the scientific study of minds, both natural and artificial.

Some of the deep connections between both scientific and applied AI are linked to old problems in philosophy about the nature of mind and knowledge, what exists, how minds are related to matter, about causation and free will, about the nature of consciousness, about how language is possible, about creativity, and about whether non-biological machines can have minds.

Such questions linking AI and philosophy motivated AI pioneers such as Ada Lovelace, Alan Turing, Marvin Minsky, John McCarthy and Herbert Simon, and are also addressed in the writings of Margaret Boden, Andy Clark, David Chalmers, Daniel Dennett, John Searle, John Pollock, and others. Yet many questions remain unanswered.

Some philosophers and scientists think AI can contribute nothing except solutions to engineering problems. They are wrong!

The tutorial attempts to explain how some largely unnoticed relationships between AI, philosophy, biological evolution and individual development, along with some advances in computer systems engineering, provide the basis for major advances in several disciplines, including AI, Philosophy and the Biological sciences.

It also attempts to show how some philosophical confusions, e.g. about "symbol grounding", about relations between embodiment and cognition, and about how theories can be evaluated, can hold up progress. (More details are in other presentations and papers.)

High-level Plan

These topics are interleaved:

- What is Philosophy?
- What is Artificial Intelligence
- How are they related?
  (To each other and to Biology)
- Discussion and illustrations of relations
- References

What is Philosophy?

That's a philosophical question on which there are different views.

Here's one answer:

Professor Walter Sinnott-Armstrong (Dartmouth University)

`http://www.dartmouth.edu/~phil/whatis/wsa.html`

He writes:

> Philosophy's goal is nothing less than a systematic world view.
>
> Other fields study particular kinds of things. Philosophy asks how it all fits together. For example, if you want to learn about bodies, take a course in physics or biology. If you want to learn about minds, take a course in psychology. But if you want to learn about how minds are related to bodies, or how physics is related to psychology, then philosophy (of mind) is for you.

He also shows how sometimes trying to bring disparate fields of knowledge together can generate contradictions or paradoxes, which then add to the work of philosophy.

There are other views of philosophy taken by people who are interested in doing it or reading about it – but this is not the place for a survey.

What follows is consistent with, but expands on the above characterisation.

What is Philosophy?

If a question seems to need an answer, and cannot be answered by any of the available methods of investigating the contents of our world

e.g. in sciences like physics, astronomy, biology, geology, psychology, etc.,

or by doing mathematical reasoning

e.g. using arithmetic, or proving or using geometric theorems,

then it may be a philosophical question.

Such a question could be of various sorts (explained later)

- It could be a **practical** question:
- It could be an **epistemological** question:

  About the nature or limits of knowledge
- It could be a **metaphysical** question

  concerned with the nature (or "ultimate" nature) of reality.
- It could be a **conceptual** question
- It could turn out be a **nonsensical** question!

  Nonsense is often disguised as sense.

This options are illustrated below.

It could be a practical question:

- What should I do?
- How should I choose?
- what should I aim for?
- which option is better?

Such questions are studied in moral philosophy (ethics and meta-ethics), in social philosophy, in aesthetics.

I shall say nothing about most of those questions, though there are links with AI, e.g.

- How should a machine forming part of a social system relate to other things including humans and other biological organisms

  I don't think Asimov's laws of robotics give an acceptable answer

  See: Why Asimov's laws of robotics are immoral

  `http://www.cs.bham.ac.uk/research/projects/cogaff/misc/asimov-three-laws.html`
- What is required for a machine to enjoy music, great art, beautiful scenery, etc?
- What would enable a machine to create great art, music, poetry, etc.?

  See Harold Cohen's painting program AARON, and
  Margaret Boden's book
  The Creative Mind: Myths and Mechanisms, Weidenfeld & Nicolson, 1990

Some practical questions about what to do and how to decide also appear in problems investigated in AI, e.g. planning problems:

What can I infer from incomplete and/or noisy sensory data?
What would be the safest way to cross the canyon? etc.

It could be an epistemological question

- How is it possible to know anything exists outside our minds?
- How is it possible to know that there are past and future times?
- How is it possible to know that other minds exist?
- How can we tell that one scientific theory provides knowledge and another is false?
- How is it possible to know that something is a necessary truth, e.g. that
  - 3 + 5 = 8
  - Counting the same set in different orders must give the same restul
  - there are infinitely many prime numbers
  - angles of a triangle add up to a straight line (180 degrees – half a rotation.)
- How is it possible to draw sensible conclusions from noisy and unreliable data?

.... and many more

It could be a metaphysical question

- What kinds of things can exist?
- Why does anything exist at all?
- How can mind and matter interact?
- What is causation?
- Is everything that happens determined by previous states of the universe?
  Physical states? Other states? (Discusssed later)
- What are virtual machines, and how are they related to physical machines?
- How can events in VMs cause changes in physical machinery?
  In part that's a metaphysical question, and in part an engineering question. see
  `http://www.cs.bham.ac.uk/research/projects/cogaff/talks#sps11`

Often what appear to be metaphysical questions about the nature (or "ultimate nature") of reality, turn out to be closely related conceptual questions.

E.g. what does "exists" mean?

Or in some cases nonsensical questions.

Who created the universe?

Which bit of a brain gives a human free will?

It could be a conceptual question

- What does “true” or “good” or “know” mean?
- Under what conditions is it possible for knowledge to exist?
- What are concepts and what is their role in knowledge?
- Where do concepts come from?

It could be a nonsensical question

- Where is the universe?
- When did 3+5 first become 8 ?
- Where is nowhere?
- What is the self? How can we put one in a robot?
  `http://www.cs.bham.ac.uk/research/projects/cogaff/misc/the-self.html`
- Which bit of the brain contains your self?
- What happened before the beginning of time?
- What time is it now on the sun? (Wittgenstein.)

  Why is that nonsensical?

  How many sorts of "what time is it at?" questions can you fiind that are nonsensical? (See below.)

Philosophy of humour

There is also philosophy of humour (humor?) though it does not normally receive the same status as the other topics in a philosophy curriculum — if it is mentioned at all.

For online information search for both

Philosophy of humour
Philosophy of humor (spelling in lazy cultures?)

A quotation from this web page on humour `http://www.iep.utm.edu/humor/`

> Almost every major figure in the history of philosophy has proposed a theory, but after 2500 years of discussion there has been little consensus about what constitutes humor.

So don't believe anything you read about humour: it is probably only a minority opinion!

Humour and AI

Connection with AI

- There have been AI programs that generate jokes (especially puns).
  Kim Binstead 1994
- There have been AI programs that recognize jokes (harder to do)
- There have been no AI programs that enjoyed a joke because they found it funny.

Compare enjoying these:

music, dancing, art, movies, doing mathematics, playing soccer. winning at chess.

What would it mean to describe a robot or animal as enjoying a joke?

Some things to think about:

- Have you heard of a (non-biological) machine that can enjoy things?
- Finding funny is not the same as laughing or smiling.
- Enjoying is not the same as smiling or saying “I like this”
- What else is needed?

Why did the robot cross the road?

Many of the questions lead to conceptual questions

E.g. to show that this question is nonsensical

"What time is it now at the centre of the earth?"

you may have to analyse what it means to say

"The time at/in X is now HH:MM"

e.g.

"The time in London is now is now 01:25"

A question that is non-sensical cannot have an answer (though it can have a rebuttal).

A statement that is nonsensical cannot be true or false.

The number 99 is more intelligent than the colour yellow.

There are different sources of nonsense.

Is this question nonsensical?

What time is it now at the equator?

If so, why?

What can you say about these?

At what time was the father of the subject of this sentence born?

Colourless green ideas sleep furiously (Chomsky)

Green furiously sleep ideas colourless

Back to What is AI?

For now I'll leave the high level question "What is philosophy?" and turn to "What is AI?"

Later we'll need to address both, and their relationships.

What is AI?

AI has always had at least three strands

- AI as engineering

  This has had the most attention and the most funding.

  There has been a lot of progress in a collection of disparate application domains involving

  Data mining, Image processing, Robot control, Mathematical tools, Language processing, Intelligent teaching aids and many more

  Recently there has been much work on Biologically Inspired AI (BI-AI)

- AI as science

  The general science of intelligent systems – natural and artificial.

  Minsky, McCarthy, Simon, Newell, Boden and many more.

  Almost all have focused entirely on human intelligence, though there are some who focus almost entirely on animal intelligence, e.g.

  Barbara Webb and others interested in insect intelligence

  We can call that AI-Inspired Biology (AIIB)
  See the 2010 Symposium `http://www.cs.bham.ac.uk/research/projects/cogaff/aiib/`

- AI as philosophy

  (Minsky, 1968), (McCarthy & Hayes, 1969) (Sloman, 1971, 1978), (Boden, 1990).

  And many more.

AI and Philosophy: Key Idea

The key ideas linking philosophy and AI are:

- Many philosophical questions arise out of the nature of human capabilities and activities, and how these relate to the world:

  perceiving, learning, thinking, wanting, deciding, acquiring concepts, finding explanations, making aesthetic and moral judgements, using language, doing mathematics, making scientific discoveries, finding scientific explanations and theories.

- If we can learn more about what human minds are and how they work, and how they are similar to and different from other sorts of minds (e.g. many animals, current robots, future robots) then we'll be in a better position to think about those philosophical problems.

- Since it is very difficult to find out out how human minds work, we can gain new insights by designing, implementing, testing, debugging, and extending various fragments of minds

- One consequence of this is that the current set of concepts used for thinking and talking about minds may need to be revised, as has happened in other areas where advances in science suggested new and better ways of carving up the world than the older pre-scientific ontology.

- **Some new concepts allow entirely new philosophical theories to be formulated.**

  As we'll see.

Top Level Idea I

How philosophy is often practised

- Many philosophers, especially so-called “analytical philosophers” have been taught that philosophy is a special non-empirical discipline, investigating eternal conceptual truths using methods that are distinct from and cannot be affected by results of methods in other disciplines – especially the sciences.

  However those other disciplines can be the object of philosophical investigation, e.g. philosophy of history, philosophy of art, philosophy of physics, philosophy of biology, philosophy of mathematics etc.

A common objection

- It has often been said that philosophy done in ignorance of the other disciplines, especially the sciences, ends up being arid and disconnected from the original problems that sparked philosophical investigations.

  To avoid that, some philosophy books and journals include rich and detailed discussions of quantum mechanics, biology, neuroscience, linguistics, etc.

Top Level Idea II

Something new has happened in the last six or seven decades.

- What has not been widely appreciated by philosophers is that the science and technology of information offers something radically different, providing new ways of addressing some very old problems, e.g. about the nature of consciousness, free will, the mind body problem, the nature of emotions, the nature of language, and metaphysical questions about what is possible.

The situation is changing, but very slowly – in part because very few people understand what we have been learning in the last 60 years or so as a result of the science and technology of computing, even though they use computers every day.

For more on this see paper and presentation for SPS conference Nancy, France, July 2011

> Evolution of mind as a feat of computer systems engineering Lessons from decades of development of virtual machinery, including self-monitoring virtual machinery. Varieties of Self-Awareness and Their Uses in Natural and Artificial Systems
>
> Paper: `http://www.cs.bham.ac.uk/research/projects/cogaff/11.html#1103`
>
> Talk: `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#sps11`

I'll say more below.

Conceptual analysis and science – including AI

Studying logical geography vs studying logical topography.

- Many philosophers believe that a major function for philosophy is to clarify the nature of some of the most general and hard to define concepts we use, e.g.

  truth and falsity, knowledge, explanation, evidence, theory, goodness, beauty, obligation, duty, mind, consciousness, freedom, ...
- Gilbert Ryle described this as doing "logical geography" e.g. in (Ryle, 1949).
- But that would not include criticising ordinary concepts as inadequate or mistaken, or offering improvements. (Wittgenstein also seemed to have this view.)
- Science often shows that our ordinary concepts are based on confusion, ignorance, or false theories: when we discover that we often revise our concepts.

  E.g. we no longer think of whales and dolphins as fish since we now have deeper classification criteria for living things than appearance, behaviour and habitat.

  Likewise we now think of (pure) water as by definition a compound of hydrogen and oxygen.
- Just as the discovery of the architecture of matter led us to revise our ontology for thinking about the physical world, so can discoveries about the architectures of working minds can lead us to revise our ontologies for thinking about mental states, processes events, interactions, etc. (Sloman, 2002)
- Scientific knowledge gives us deeper understanding of the (fixed) topography, whereas common sense concepts are like (changeable) social/political geography.
- If we improve our understanding of the logical topography that may help us develop better logical geography. `http://www.cs.bham.ac.uk/research/projects/cogaff/misc/logical-geography.html`

What's in Philosophy?

Philosophy can be divided in different ways

- Problems investigated
  many subdivisions labelled "Philosophy of X"
- Approach used
- Whether
  (a) primarily theoretical (trying to understand), or
  (b) practical (trying to make the world better).
- Whether grand
  (a) trying to understand something fundamental about the universe, or
  (b) modest, e.g. trying to analyse concepts e.g. analysing notions like

  truth
  knowledge
  explanation
  cause
  consciousness
  language
  intention
  theory
  etc.

Relations to AI

Epistemology

What are concepts?

Where do they come from?

concept empiricism = (roughly) symbol grounding theory, states:
all concepts must be "grounded", i.e. based either on abstraction from experience of instances, or definition in terms of pre-existing concepts.

Refuted by Immanuel Kant in (Kant, 1781)
Finally demolished by 20th Century philosophy of science.
`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#models`
Talk 49: Why symbol-grounding is both impossible and unnecessary, and why theory-tethering is more powerful anyway. (Introduction to key ideas of semantic models, implicit definitions and symbol tethering through theory tethering.)

What is knowledge?

What are the sources of knowledge?

Experience? Reasoning? What about innate knowledge?

Can there be any other form of non-empirical knowledge?
(Karmiloff-Smith, 1992) (Chappell & Sloman, 2007)

Can anything really be known?

What can we know about the past?

About other minds?

About unobservable entities, e.g. sub-atomic particles, genes, gravitational fields, electrons.

Metaphysics/ontology

To be added

Philosophy of mind

Including the nature of qualia and the multifarious forms of consciousness)

The standpoint of this tutorial is that these are all to be understood as forms of biological information processing.

- Think of a mind as an information-using control system for the body.
- What does that imply about kinds of informaino required?
- many of the philosophical puzzles arise because of the architecture of virtual machinery

Varieties of functionalism

- Atomic state functionalism (ASF) vs virtual machine functionalism (VMF)

Philosophy of science

Including the question – do standard views of philosophy of science do justice to AI as science?

(See Chapter 2 of Sloman 1978

`http://www.cs.bham.ac.uk/research/projects/cogaff/crp/chap6.html`

Topics

- induction
- Popper
- Lakatos

Philosophy of causation and virtual machinery

- Can standard philosophical theories of causation do justice to the causal interactions within virtual machinery in complex information processing systems?

  This is a question linking philosophy and general computer science, software engineering.

  But the role of virtual machinery is crucial to the design of intelligent machines. Newell and Simon were seriously misleading in their talk of "Physical Symbol Systems" since the symbols are not physical but contents of virtual machines. They should have referred to "Physically-implemented Symbol Systems".

  What are the implications for mind/brain relations? (Including causal relations)

- Many philosophers, psychologists, neuroscientists, biologists, and AI/Robotics researchers think that causal learning processes are captured by various forms of associative/statistical learning, e.g. using Bayesian nets.

  This is essentially a Humean view of causation: causation is just reliable association. Our feeling that we understand something more, e.g. some kind of necessitation, is just illusory, and ruled out by concept empiricism for we cannot experience necessitation, only correlations, regularities, statistical relations, etc. (Hume, 1748)

  Immanuel Kant argued that Hume must be wrong and that since concept empiricism is wrong (since concepts are needed for experience and therefore cannot all come from experience) we can have a non-empirical concept of causal necessitation.

  Several of his examples can be shown to be similar to mathematical relationships (Kant, 1781).

  My examples; changing the height of a triangle causes the area to change, and adding three coins to a jar with five coins causes the number of coins in the jar to become eight.

- Question for AI Researchers/Roboticists: what forms of causal understanding are possible for a baby robot, and how can that understanding grow? How might this related to animal cognition?

Philosophy of mathematics

Initially part of epistemology (philosophy of knowledge).

David Hume: there are two kinds of knowledge

- empirical – based on observation, experiment, measurement, etc.
- "relations between ideas" e.g. matters of definition and logical consequences of definitions (later called "analytic knowledge"

Immanuel Kant (who would have loved AI) (Kant, 1781), objected that rhere is something in-between, knowledge that is

- non-empirical (a priori), in the sense that it can be lnown to be true independently of observations, experiments, personal experiences, etc.
- synthetic (non-analytic) insofar as when you acquire it you have learnt something new and significant, unlike definitional truths (e.g. "All bachelors are unmarried")

  Examples included truths of arithmetic (3+5 = 8, Thee are infinitely many prime numbers) and geometry, e.g. angles of a triangle sum to a straight line = 180 degrees, pythagoras' theorem and many more.

My DPhil (Oxford 1962) was an attempt to show why Kant was righy and Hume wrong (like most analytical phylosophers when I was a student). Links with requirements for future robots and products of biological evolution..

I did that mainly by attempting to analyse a variety of exampless to illustrate whaat I thought Kant was getting at and why the examples supported his claims against Hume.

Kant thought Euclidean geometry was necessariliy true and synthetic. It later turned out that the parallel axiom was false as regards physical space - following Einstein's work on General Relativity.

But that left a great deal of Euclidean geometry intact, the parts common to various non-euclidean geometries.

A few years later (around 1969) I started learning about AI, including learning to program, and became that doing AI was the best way to do various parts of philosohy, including philosohy of matheatics, by building various fragments of human minds to demonstrate in a deeper way than either philosopohical analysis or empirical psychology, or neuroscience, what sorts of mechanisms could underly mathematical, linguistic, and other competences.

For an example of how AI thinking led to a new way of understanding what is involved in acquiring basic knowledge of concepts of positive integers, see chapter 8 of (Sloman, 1978):
`http://www.cs.bham.ac.uk/research/projects/cogaff/crp/chap8.html`

On toddler theorems:

More recently I have been trying to show that there are many domains of knwledge ("exploration domains") in which a young learner can develop new concepts and learn empirical generalisations and then later reorganise that knowledge so that what was previoiusly learnt empirically (e.g. that order of counting of a set of objects does not affect the outcome, if the set remains the same and the counting process includes no errors.

itry to work out for yourself why that must be so.

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

Philosophy of information

Consequences of the view of the universe as made up of matter, energy, information, and processes involving them, in space-time.

What does information add?

What's Information:

(Sloman, 2011)

`http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#905`

Links with biology and neuroscience

- Biologically inspired AI/Robotics (BIAI)
- AI-Inspired Biology (AIIB)
  See `http://www.cs.bham.ac.uk/research/projects/cogaff/aiib/`

Evolutionary transitions/Design distinctions

How to think about evolution of information processing: What were the major transitions?

Relations with major transitions during individual development and relationships with philosophy of mathematics.

Relationships to design options.

A framework for developing these ideas

John Maynard Smith and Eörs Szathmáry (1995) proposed that there are eight major transitions in evolution, summarised here:

`http://en.wikipedia.org/wiki/The_Major_Transitions_in_Evolution`

Transition from:	Transition to:
1 Replicating molecules	"Populations" of molecules in compartments
2 Independent replicators (probably RNA)	Chromosomes
3 RNA as both genes and enzymes	DNA as genes; proteins as enzymes
4 Prokaryotes	Eukaryotes
5 Asexual clones	Sexual populations
6 Protists	Multicellular organisms - animals, plants, fungi
7 Solitary individuals	Colonies with non-reproductive castes
8 Primate societies	Human societies with language, enabling memes

They identified features common to the eight transitions:

1. Smaller entities come together to form larger entities.
2. Smaller entities become differentiated as part of a larger entity.
3. Smaller entities become unable to replicate in the absence of the larger entity.
4. Smaller entities become able to disrupt the development of a larger entity.
5. **New ways arise of transmitting information.**

   Is that the only kind of information-processing transition?

Suggestion: We should try to produce a taxonomy of types of evolutionary or developmental transition (gradual or discontinuous) connected with information processing – in animals, and in future animats.

Besides information transmission there is also manipulation.

(A century or two should suffice.)

Another set of transitions

We can generalise the last feature to include

- new kinds of information (new information contents)
- new forms of representation (new physical and virtual media)

  For a discussion of some of the language-like forms of representation that must have evolved in other animals for internal use, and must be present in human children before they begin to speak

  (e.g. it is required for visual and other forms of perception, for wanting, for trying, for planning and executing actions, and in order to have a need or a desire to communicate)

  see `http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#glang`
  Evolution of minds and languages. What evolved first and develops first in children: Languages for communicating, or languages for thinking (Generalised Languages: GLs)
- new ways of acquiring, processing, or using information.
- new information-processing architectures

These changes can be related to different sorts of trajectories in evolution and in development.

Trajectories in different spaces

- Trajectories in the space of possible sets of requirements (niches)
- Trajectories in the space of designs for systems satisfying requirements
- Trajectories in The space of implementations for each design

Is language essentially for communication? NO

GLs: Generalised Languages: some used internally for perceiving, thinking, wanting, planning, executing plans, reasoning, learning ...

Implications for nature of human languages used internally (for perceiving, having desires, planning, control of actions etc.)

(Also philosophy of language.)

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

Talk 52: Evolution of minds and languages. What evolved first and develops first in children: Languages for communicating, or languages for thinking (Generalised Languages: GLs)

Implications for nature of language learning: collaborative design rather than data-mining in a corpus.

See this paper on the need for a variety of forms of representation: (Sloman, 1971) – criticising (McCarthy & Hayes, 1969).

Also chapter 7 of (Sloman, 1978)

Confusions about embodiment

Most thinkers who are concerned about embodiment consider only the real-time online interactions between animal or robot and immediate environment.

They ignore abilities to think about, learn about, reason about, and make use of information about the past, distant places, unobservable entities, the future, unactualised possibilities and mathematical abstractions.

The result is seriously blinkered research.

See Some Requirements for Human-like Robots: Why the recent over-emphasis on embodiment has held up progress,
(Sloman, 2009b)

Also this discussion of kinds of dynamical system:
`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/kinds-of-dynamical-system.html`

Reward-based vs Architecture-based motivation

Must all motives be adopted because of their links with rewards?

NO!

Architecture-Based Motivation vs Reward-Based Motivation (Sloman, 2009a)

http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#907

Confusions about emotions

Are emotions required for intelligence?

NO!

http://www.cs.bham.ac.uk/research/projects/cogaff/talks/#cafe04

Confusions about free will

What implications does AI have for debates about free will? What implications do philosophical discussions of free will have for AI?

See

`http://www.cs.bham.ac.uk/research/projects/cogaff/misc/four-kinds-freewill.html`
Four Concepts of Freewill: Two of them incoherent

Also: `http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#8`
How to dispose of the free will issue (1988)

(Expanded as Chapter 2 of Stan Franklin's 1995 book: **Artificial Minds**)

See also Daniel Dennett's Elbow Room (Dennett, 1984)

Where next?

Where to go next in AI and long term philosophical implications.

What's special about computation?

What exactly is special about computers, and computation, that makes a computational approach to understanding mind not just the latest a series of fashions for thinking about minds and brains in terms of new technology?

When I was a child it was fashionable to say, and write, and think that the brain was a sort of telephone exchange. Telephones were still relatively new then, and they were all connected by wires.

Example topic:

The philosophical significance of virtual machines composed of interacting coexisting virtual machines, some of them connected to input and/or output devices.

Could biological evolution have "discovered" the need for virtual, as opposed to physical, machinery long before engineers did?

Could self-monitoring virtual machines, including perceptual sub-systems linked to sensory devices, provide the key to the nature of perceptual qualia, including explaining all their philosophically puzzling features? (E.g. their privacy.)

See `http://www.cs.bham.ac.uk/research/projects/cogaff/08.html#803`
Title: Virtual Machines in Philosophy, Engineering & Biology

(And other items referred to there, including talks on virtual machinery.)

Causation in virtual machinery

Example topic:

What implications do causes and effects within complex virtual machinery have for philosophical theories of causation?

Or for metaphysics more generally?

A paper on this topic, written for a philosophy of science conference held in July 2011 is online here:

`http://www.cs.bham.ac.uk/research/projects/cogaff/11.html#1103`

Evolution of mind as a feat of computer systems engineering: Lessons from decades of development of self-monitoring virtual machinery.

Different kinds of learning, and causation

Many animals (e.g. corvids, elephants, primates, squirrels) seem to be able to **work out** solutions to new problems, instead of having to use trial and error, or imitation, or explicit instruction, or genetically programmed solutions. What mechanisms enable them to do solve such problems, and how are they related to the abilities of humans to do mathematics?

Can designing robots with similar capabilities be a contribution to philosophy of mathematics?

E.g.: can a computational theory of development of mathematical competences in young humans, or young robots, shed light on philosophical questions about the nature of mathematical knowledge?

Is research in philosophy of mathematics relevant to the task of designing machines capable of doing or learning to do mathematics?

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

The work of Annette Karmiloff-Smith is very relevant (Karmiloff-Smith, 1992)

Baby robot mathematicians

Example topic:

Under what conditions could a young robot begin to make mathematical discoveries, e.g. about geometry, topology, sets, orderings, numbers, ...

XXX

Example topic:

It is often assumed that all motivation must be based on (positive or negative) rewards. I'll argue that that's a false assumption and there are good reasons why evolution should have produced mechanisms for "architecture-based motivation", which have consequences that the individual concerned cannot possibly anticipate.

If this is correct, what are the implications for robotics? For philosophical theories of motivation and affect?

See: `http://www.cs.bham.ac.uk/research/projects/cogaff/09.html#907`

Robot philosophers

Example topic:

Under what conditions could a young robot discover for itself some old philosophical problems, e.g. about the nature of qualia, about relations between mind and matter, about what knowledge is, about whether free will is possible, about what words like “good”, “right” and “ought” mean, and whether there are objective moral values?

Could the same initial philosophical potential in young robots lead to different “adult” philosophical theories in different robots with the same initial design? E.g. could some end up thinking like John Searle, others like Daniel Dennett, others like David Hume, others like Immanuel Kant, ...etc...?

Meaning and "Symbol Grounding"

Example topic:

Symbol-grounding theory is often taken as axiomatic by researchers in AI and robotics. Yet it is just a new version of an old philosophical theory "concept empiricism", first refuted by Immanuel Kant (around 1781) and more thoroughly demolished by philosophers of science in the 10th Century. What alternative is there? Can "theory tethering" provide the answer?

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

How can a machine or animal acquire, create, derive and use semantic contents? Why do so many people (and not just John Searle) regard it as obvious that computers cannot understand the symbols they manipulate? Contrast: `http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#4` Aaron Sloman, What enables a machine to understand?, in Proc 9th IJCAI, Los Angeles, pp. 995–1001, 1985,

XXX

Example topic:

In the past decade or two, there has been great enthusiasm among philosophers, cognitive scientists and roboticists for the claim that cognition must be embodied, and that acknowledging the role of embodiment revolutionises theories about mind and intelligence.

Is that claim correct, or does the emphasis on embodiment ignore some important features of biological evolution and important requirements for future robots?

Compare: `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0804`

Some Requirements for Human-like Robots: Why the recent over-emphasis on embodiment has held up progress.

(Might include discussion of claims for “mirror neurons”.)

XXX

Example topic:

How do **affective** states and processes (e.g. desires, attitudes, preferences, values, ideals, emotions, moods, interests, etc.) differ from things like perception, belief, reasoning, planning, explaining?

How can thinking about **architectures** for minds help us answer this question? Do information-processing theories provide a better alternative than older philosophical answers, e.g. dualist theories, logical behaviourism, the intentional stance?

XXX

Example topic:

Could consciousness have been produced by biological evolution? If not, why not? If so how?

If evolution can produce conscious animals does that have implications for whether human engineers can produce conscious machines?

Similar questions can be asked about having desires, preferences, ideals, moral values, etc. Can non-human animals have them, and if not why not, and if so how? Does this help us understand whether robots could?

XXX

Example topic:

Are Asimov's "Laws of robotics" unethical towards intelligent machines?

See `http://www.cs.bham.ac.uk/research/projects/cogaff/misc/asimov-three-laws.html`

XXX

Request for help with AITopics web site.

`http://www.aaai.org/AITopics`

XXX

See also the abstract for invited talk at AGI 2011 the week before AAAI 2011:

Conference: http://agi-conf.org/2011/

AGI = Artificial General Intelligence

I don't believe there is an such thing. There are many special kinds of intelligence, found in different organisms and different machines.

Tutorial http://tinyurl.com/aaronagi11

The biological bases of mathematical competences: a challenge for AGI

There's more here: http://www.cs.bham.ac.uk/research/projects/cogaff/misc/AREADME.html ¡/ul¿

¡a name="reading"¿¡/a¿ ¡h3¿**Reading matter relevant to the tutorial. (To be extended)** ¡/h3¿

Additional Reading

Please see the tutorial web site for pointers:

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

Further reading

References

Boden, M. A. (1990). *The creative mind: Myths and mechanisms*. London: Weidenfeld & Nicolson.

Chappell, J., & Sloman, A. (2007). Natural and artificial meta-configured altricial information-processing systems. *International Journal of Unconventional Computing*, *3*(3), 211–239. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609)

Dennett, D. (1984). *Elbow room: the varieties of free will worth wanting*. Oxford: The Clarendon Press.

Hume, D. (1748). *An Enquiry Concerning Human Understanding*.

Kant, I. (1781). *Critique of pure reason*. London: Macmillan. (Translated (1929) by Norman Kemp Smith)

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Maynard Smith, J., & Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford, England:: Oxford University Press.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of AI. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence 4* (pp. 463–502). Edinburgh, Scotland: Edinburgh University Press. (http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html)

Minsky, M. L. (1968). Matter Mind and Models. In M. L. Minsky (Ed.), *Semantic Information Processing.* Cambridge, MA: MIT Press.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson.

Sloman, A. (1971). Interactions between philosophy and AI: The role of intuition and non-logical reasoning in intelligence. In *Proc 2nd ijcai* (pp. 209–226). London: William Kaufmann. (http://www.cs.bham.ac.uk/research/cogaff/04.html#200407)

Sloman, A. (1978). *The computer revolution in philosophy*. Hassocks, Sussex: Harvester Press (and Humanities Press).

Sloman, A. (2002). Architecture-based conceptions of mind. In P. Gärdenfors, K. Kijania-Placek, & J. Woleński (Eds.), *In the Scope of Logic, Methodology, and Philosophy of Science (Vol II)* (pp. 403–427). Dordrecht: Kluwer. (http://www.cs.bham.ac.uk/research/projects/cogaff/00-02.html#57)

Sloman, A. (2009a). Architecture-Based Motivation vs Reward-Based Motivation. *Newsletter on Philosophy and Computers*, *09*(1), 10–13.

Sloman, A. (2009b). Some Requirements for Human-like Robots: Why the recent over-emphasis on embodiment has held up progress. In B. Sendhoff, E. Koerner, O. Sporns, H. Ritter, & K. Doya (Eds.), *Creating Brain-like Intelligence* (pp. 248–277). Berlin: Springer-Verlag.

Sloman, A. (2011). What's information, for an organism or intelligent machine? How can a machine or organism mean? In G. Dodig-Crnkovic & M. Burgin (Eds.), *Information and Computation* (pp. 393–438). New Jersey: World Scientific.

Important propositions in the philosophy of AI include: *Turing's "polite convention": "If a machine acts as intelligently as a human being, then it is as intelligent as a human being."This is a paraphrase of the essential point of the Turing Test. Harvnb|Turing|1950, Harvnb|Haugeland|1985|pp=6-9, Harvnb|Crevier|1993|p=24, Harvnb|Russell|Norvig|2003|pp=2-3 and 948] * The Dartmouth proposal: "Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it." Recent papers in Philosophy of Mind and Artificial Intelligence & AI. Papers. People. Review: Heuristics Intelligent Search Strategies for Computer Problem Solving (Judea Pearl, 1984). Book review, published 15 March, 1986. SUMMARY: The view of AI science offered by Judea Pearl is thoroughly traditional and standard, and therein lie both this book's strengths and its weaknesses as a monograph, a reference, or a textbook. Save to Library. Download. Artificial intelligence has close connections with philosophy because both use concepts that have the same names and these include intelligence, action, consciousness, epistemology, and even free will. Furthermore, the technology is concerned with the creation of artificial animals or artificial people (or, at least, artificial creatures; see Artificial life) so the discipline is of considerable interest to philosophers. These factors contributed to the emergence of the philosophy of artificial The philosophy of artificial intelligence is a collection of issues primarily concerned with whether or not AI is possible -- with whether or not it is possible to build an intelligent thinking machine. Also of concern is whether humans and other animals are best thought of as machines (computational robots, say) themselves. The most important of the "whether-possible" problems lie at the intersection of theories of the semantic contents of thought and the nature of computation.Â Organizations such as the EU High-Level Expert Group on AI and the IEEE have recently formulated ethical principles and values that should be adhered to in the design and deployment of artificial intelligence. These include respect for autonomy, non-maleficence, fairness, transparency, explainability, and accountability.