Causal Competences Of Many Kinds

Aaron Sloman
School of Computer Science, University of Birmingham
http://www.cs.bham.ac.uk/~axs/

Jackie Chappell
School of Biosciences, University of Birmingham
http://www.biosciences.bham.ac.uk/staff/staff.htm?ID=90

Location of these slides

These slides are available here:

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

They were originally part of this presentation at WONAC 2007

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

Jackie Chappell's WONAC slides are at

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

See also our papers at

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

For details of WONAC 2007

International Workshop on Natural and Artificial Cognition – Oxford June 2007

See

`http://tecolote.isi.edu/~wkerr/wonac/`

`http://tecolote.isi.edu/~wkerr/wonac/program.html`

Note on these slides

These slides were originally part of the Sloman/Chappell presentation for the NSF/euCognition Workshop on Natural and Artificial Cognition (WONAC) – Oxford June 24-26 2007, namely

Evolution of two ways of understanding causation: Humean and Kantian

`http://www.cs.bham.ac.uk/research/cogaff/talks/sloman-wonac-slides.pdf`

During the workshop a lot of time was spent discussing what it meant for an animal or machine to understand causation, and various suggestions were made regarding the sorts of tests that might reveal such understanding.

We had already prepared a first draft of the points made below, but there was not time for us to include the analysis during the workshop presentation, so we are making this available separately.

The summary of our presentation in the original set of slides is provided below.

The workshop slides include a brief explanation of the differences between Hume's and Kant's views on causation. Hume's view is much closer to current ideas about causation expressed in Bayesian causal nets. Kant's views suggest that understanding causation is very closely related to being able to do geometrical, or more generally mathematical or logical reasoning regarding conclusions that follow necessarily from premisses, especially premisses and conclusions involving spatial and temporal relations. We take that for granted in what follows.

NOTE: these slides were produced using LaTex.

Summary of our presentation at WONAC 2007

- Animals and robots need to grasp and use two types of causation:
  1. Humean causation: evidence-based, correlational, often statistical
  2. Kantian causation: structure-based, deterministic
- Most current theorising about causation in philosophy, psychology and AI is Humean in a modern form: e.g. theories about Bayesian causal nets.
- This ignores deep features in Kantian causation connected with reasoning about spatial and temporal structures and the role of properties of different kinds of stuff.
- Kantian understanding of causation, when available, also allows more creativity, and recombination of different kinds of knowledge to deal with new problems

  because of the way different structures and processes are embedded in the same spatial region.
- The growth of understanding of Kantian causation is linked to forms of learning and development found only in animals usually classified as altricial, for reasons that are only beginning to become clear.
- We need to revise and update some biological and computational ways of thinking about animals and machines and their evolution and development.
- Close observation of play and exploration in children and animals, including their failures as well as their successes, provide clues as to what is going on: including development of ontologies and forms of representation, requiring abduction.
- Systematic biological and psychological research, along with design and implementation of working models can add more clues and help to test the theories.

Our main point here

There are (at least) three reasons why it is pointless trying to define "understanding causation" in terms of some behavioural or experimental test that will decide whether an animal does or does not understand.

- We do not have a precisely defined widely shared concept of such understanding that is clear enough to be useful in specifying scientific research questions.
- Most of what is referred to by 'understanding' goes on in a virtual machine whose operations are only loosely connected with behaviour, as explained in this presentation:

  Virtual machine functionalism: `http://www.cs.bham.ac.uk/research/cogaff/talks/#inf`
- In any case the kind of competence we are investigating is not an all-or-nothing competence:

  instead there are several different sub-competences and different animals or machines may have different subsets of those competences.

  Arguing about which subset should be used to define the label "understands causation" is completely pointless (like many debates about definitions).

If instead we try to identify all those subcompetences and treat them as all worthy of scientific investigation we shall learn more than if we look for some dichotomy in nature.

That does not imply that there are only differences of degree.

SO...

Don’t ask which animals or machines understand causation

There is no fixed, sharp, distinction between understanding and not understanding causation.

Instead we can distinguish a fairly long list of types of competence related to causation (below), and then ask:

– which animals have which subset,

– how they evolved

– at what age or in what order they typically develop,

– under what conditions they typically develop,

– how those competences are acquired or extended,

– which are necessary precursors for others,

– what forms of representation, mechanisms and architectures support them

– what the trade-offs are between alternative sets of competences and alternative implementations,

That way we can replace futile debates, about how to label phenomena, with productive research, on what sorts of competences different animals have, what the implications of those competences are, and what mechanisms can explain them.

We can do that for many debates about cognition and development in humans, animals and machines.

I.e. replace ill-defined and poorly motivated dichotomies with analysis of spaces of possible designs and corresponding niches to be analysed, compared, explained, modelled, ...

Varieties of Humean causal sophistication

Animals and robots may have more or less sophisticated Humean causal knowledge used in the production of behaviour.

1. Abilities based on fixed, hard-wired predictive mechanisms and action-generating mechanisms.

   Essential for precocial organisms that don't live long enough to learn much.
   Even altricial organisms will need some hard-wiring for bootstrapping learning and for feeding, etc. (eg. suckling). (That applies also to cognitive bootstrapping: a tabula rasa will achieve nothing.)

2. Hard-wired but 'soft' predictive mechanisms with a fixed structure but adjustable parameters that can be modified by the statistics of experienced reality.

3. The inductive ability to generate new predictive rules as a result of learning:
   - rules that are rigidly fixed once formulated.
     E.g. things that are coloured red and black are noxious.
   - rules that are modifiable though further experience (e.g. altering a bayesian net)
     quantitatively, e.g. changing (conditional or prior) probabilities, or
     qualitatively e.g. adding or removing nodes or arrows, (altering the structure).

4. Ability to deal with conflicts between predictive/causal rules
   - Implicit conflict resolution, e.g. using competing activation weights
   - Explicit representation of which rules are available and the ability to learn which to trust in which situations, using conflict-resolution rules.
     e.g. depending on current goals and constraints, such as time constraints.

5. Ability to notice evidence that is inconsistent with, or casts doubt on, stored rules and to reduce trust in those rules as a result of making inferences from the evidence.

Varieties of Humean causal sophistication (2)

6. Ability to notice an information gap when attempting to deploy a known rule.

   E.g. noticing a need to look in a new place or a new direction to get some information to plug into a condition that will allow a rule to be used to take a decision or make a prediction.

7. Ability to conduct experiments or observations to test or modify a doubtful generalisation or theory.

   We can distinguish
   - experiments conducted automatically when failed or conflicting predictions directly trigger behaviour modifications that test which variations produce which results
   - experiments selected as a result of reasoning about consequences of alternative explanatory theories and searching for good tests
   - evidence seeking actions: without having specific expectations inspect things in an appropriate spatial region, looking for previously unnoticed details that might give clues to distinguish one situation from another.
   - memory seeking actions: as before, but searching episodic memory for remembered details that might be relevant to the problem.

8. Ability to use representations with hypothetical, or counter-factual content:

   e.g. possible futures, possible histories, possible theories, questions to be answered, possible strategies, etc.
   I.e. there's a distinction between what is represented and committed to, and what is represented as part of a deliberative or other process (story telling, daydreaming, wondering about, doing thought experiments, etc.)

   Special architectural features are required to support this distinction in the processing of information, whether factual information or control information

Varieties of Humean causal sophistication (3)

9. Ability to generate new explanatory theories non-inductively (using abduction) and then test them, including hypotheses explaining observations:
   - single step explanations
   - multi step explanations
   - structured explanatory models (see Kantian causation)

   For an incomplete discussion of the role of abduction in science and philosophy see
   `http://www.cs.bham.ac.uk/research/projects/cogaff/misc/logical-geography.html#abduction`

10. Ability to extend the ontology available for expressing causal hypotheses.
    - definitional extensions (e.g. putting more tests into network nodes)
    - substantive extensions – hypothesising new kinds of entities, events, processes, etc. not definable in terms of currently known ones (which is how deep science proceeds).

      See `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0604`
      'Ontology extension' in evolution and in development, in animals and machines.

11. Ability to record externally, or communicate to others, what causes what.
12. Ability to seek information from others about what causes what.
13. Philosophical ability to think about causation

    what it is, whether it is rational to assume there are causal connections, whether determinism is true, worries about free will, etc.

NB: that is not a complete list!
Some of these alternatives also apply to Kantian causation – but there are great differences in the details.

Things still to be added

The above list is incomplete in various ways, especially insofar as it focuses only on the abstract structures common to all sorts of causal models, as expressible in Bayesian nets.

We need to list ways in which the ability to do Kantian causal reasoning extends those general causal competences.

The most basic cases will involve being able to represent complex structures and processes embedded in a common region of space-time so that different parts of objects and different ongoing sub-processes can interact with one another. (As in rotating meshed gear wheels, or a steam engine with speed governor.)

We also need to list kinds of causal reasoning about things with semantic competences, e.g. to reason about beliefs, desires, preferences, intentions, decisions, planning processes, plan execution processes, etc. in oneself and others.

This requires meta-semantic competences combined with meta-management competences for using meta-semantic information (information about information and information-processing).

Examples of Kantian causal competences

New Placeholder 28Jun

– ability to do various kinds of Kantian causal reasoning

o E.g. about different sorts of shapes, about distances and velocities, about kinds of stuff, about mental states and processes, about effects of different sorts of actions ...

– ability to build a causal model of what's going on in the current environment

(e.g. why only part of the toy train is visible: the rest is in the tunnel, and still moving)

– ability to work out what needs changing before some goal can be achieved.

– ability to plan actions using causal features of objects or situations.

o Can you lift a cup full of liquid using only one finger?

– ability to work out what to do to get new information

o where to look, what to try manipulating, what to test, etc.

See the original workshop slides from which these were extracted (and expanded)
`http://www.cs.bham.ac.uk/research/projects/cogaff/talks/wonac/#sloman`

Examples of meta-semantic causal competences

New Placeholder 28Jun

– ability to reason causally about mental processes in others

lots of literature to refer to in dev psych and animal behaviour studies

Also John Barnden's ATT-META system (which supports recursively nested simulative reasoning)

`http://www.cs.bham.ac.uk/~jab/ATT-Meta/`

– ability to do mathematics

– ability to do philosophical thinking

Some limitations of current AI

Alas, current AI systems, including robots, have only a small set of the previous types of causal competence:

e.g. they may be able to do things,
but, when doing something, don't know
what they do, why they do it, how they do it,
what they did not do but could have done
why they did not do it differently,
what would have happened
if they had done it differently,
etc.

They also cannot tell whether someone else
is doing it the same way or a different way,
and if something goes wrong explain why it went wrong.

All this is related to lacking Humean and Kantian causal competences.