Overview of Commonsense Knowledge and Explanation: ubiquitous in everyday thinking, speaking and perceiving in our ordinary interaction with the world.

“Computers will end up with the psychology that is convenient to their designers” —John McCarthy

"The little thoughts of thinking machines."


Ontology Summit 2019 pre-Session
Co-champions: Gary Berg-Cross & Torsten Hahmann (U of Maine)
Outline

1. Why/how is Commonsense Knowledge & Reasoning relevant to Explanation?

2. Complementing Ram's AI Evolution Slide with a short historical AI perspective: McCarthy, production systems, naive physics to ontologies.

3. Connections to earlier Summits & recent CommonSense efforts

4. Examples of research and Issues

Overview of Commonsense Knowledge and Explanation
Why a commonsense thrust?

- For fluidity, explanations require knowledge about the world
  - Explanations are intimately connected to both commonsense (AI) and ontologies.
- A major long-term goal of AI has been to endow computers with standard commonsense reasoning capabilities.
- Another, related goal, is to endow AI systems with NL understanding (and production).
- These goals support one another since any general NLP must possess the commonsense that is assumed in the text.
- A conclusion is that together (in the Symbiosis phase) commonsense & NL understanding are key to a robust AI explanation system.
AI, Commonsense & Explanations

- To achieve human-level performance in domains such as natural language processing, vision, and robotics, basic knowledge of the commonsense world—time, space, physical interactions, people, and so on—will be necessary.

- Although a few forms of commonsense reasoning, such as taxonomic reasoning and temporal reasoning are well understood, progress has been slow.

- Extant techniques for implementing commonsense include logical analysis, handcrafting large knowledge bases, Web mining, and crowdsourcing. Each of these is valuable, but none by itself is a full solution.

- Intelligent machines need not replicate human cognition directly, but a better understanding of human commonsense might be a good place to start.

From, *Commonsense Reasoning and Commonsense Knowledge in Artificial Intelligence* Ernest Davis, Gary Marcus  *Communications of the ACM*, September 2015, Vol. 58 No. 9, Pages 92-103
A Major Part of Early AI was about Formalizing Commonsense

• John McCarthy described 3 ways for AI to proceed – (1) imitate the human CNS, (2) study human cognition or
  
  − (3) “understand the common sense world in which people achieve their goals.”

• “One will be able to assume that the advice taker will have available to it a fairly wide class of immediate logical consequences of anything it is told and its previous knowledge.

• This property is expected to have much in common with what makes us describe certain humans as having common sense.

• We shall therefore say that a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows” from “Programs with Common Sense” (McCarthy 1959)
Production Systems have simple ideas about K & Explanation (why was my loan denied?)

Basic structure of a rule-based expert system

Brittleness of handcrafted rules encoding useful info about diseases, but this knowledge is isolated and opaque & breaks in the face of obvious errors due in part to a lack of common sense. (CYC)

Vonzuben diagram - lecture on Rule based Systems
The How & Not Why of Proof in Early Symbolic Systems

- Proofs found by Automated Theorem Provers provide a map from inputs to outputs.
  - But do these make something clear?
  - They may know the “how” but not the “why.”
- And scripts & rule-based systems got so complex that the trace was hard to follow with conditional paths.
- As Derek pointed out, people asked: Can you explain this in simple terms that is understandable via my mental model???
- CS reasoning is needed to deal with incomplete and uncertain information in dynamic environments responsively & appropriately.
- So we support the thesis that explanation systems need to understand the context of the user and need to be able to communicate to a person and be trusted.
Problems with Rule Based Explanations

Clancey found that Mycin's individual rules play different roles, have different kinds of justifications, & are constructed using different rationales for the ordering & choice of premise clauses.

There are structural and strategic concepts which lie outside the rule representation, and they can only be supported by appealing to this deeper level of (background) knowledge.

One Solution Approach? – Ontology KBs to formalize K

Ontologies are needed to make explicit structural, strategic & support knowledge which enhanced the ability to understand & modify the system as well as support suitable explanations.

Naive Physics - close to Commonsense

Pat Hayes “Naive Physics Manifesto” (1978) proposed to develop a formal theory encompassing the large body of knowledge of the physical world:

1. in an axiomatically dense form and using a unified conceptualization (he used symbolic logic to specify the axioms), and

2. supported by commonsense reasoners.

Emphasis was on formalizing foundational “everyday” physical concepts, including: measurements and scales; spatial concepts (shape, orientation, direction, containment); substances and states (solids, liquids), physical forces, energy and movement; manufactured objects and assemblies.
An alternative is to focus on justified inferences using an abstraction called a microworld (MT), where each MT covers a small range of physical phenomena in which valid inferences can be made.

Each microworld has enough competence in a unified model (using some basic concepts - space, time, rigid objects, roads..) to justify/explain commonsense physical inferences and answer Qs.

A strength is the ability to consider MTs for limited purposes - it makes the analysis much easier; we can focus on getting some particular class of inferences to work without worrying how these inferences will fit with all the rest of the naive physics.

• Similar to competency questions’ role in ontological engineering
• And it allows much closer, model unified ties to practical Apps.
• But, integrated reasoning over MTs is a problem & picking them may be arbitrary.
Explanations from CYC Knowledge

Arguments consist of justification chains showing which knowledge (ground facts, rules, and inference methods) was used to arrive at a conclusion.

CYC can provide a partial inference chain constructed in response to queries such as: “Can the Earth run a marathon?”

No, Earth is not animate and the role capability of running a marathon is detailed in a sports MT.
Structure of related MTs, integrated through metarules to answer Qs (explanation)

At the top of this structure is a single Hayes-like theory to which all Qs can ultimately be referred (after E/ Davis: *The Naive Physics Perplex*, 98).

Quasi-statics, limiting case where dissipative forces are always >> compared to momentum. Objects only move if forced.

Overview of Commonsense Knowledge and Explanation
80s-90s Associate System Vision

By the mid-80s there was a view that intelligent systems must be able to interpret user questions, formulate appropriate replies, provide answers to follow-up questions, make backward references and “intelligently” index and retrieve relevant information.


User Models: CYC was arguing “systems should not only handle tasks automatically, but also actively anticipate the need to perform them....agents are actively trying to classify the user’s activities, predict useful subtasks and expected future tasks, and, proactively, perform those tasks or at least the sub-tasks that can be performed automatically.”
A Common Sense Psychology includes Wants and Beliefs as part of (Associate) Explanation

- A child learns to ascribe wants and beliefs to others in a complex way that she/he never learns to encapsulate in definitions. (McCarthy on CS)

- Approximate criteria for some specific properties relating them to the more implicit properties of believing and wanting.
  - **Intends** — We say that a machine intends to do something if we can regard it as believing that it will attempt to do it. We may know something that will deter it from making the attempt. Like most mental concepts, intention is an intermediate in the causal chain; an intention may be caused by a variety of stimuli and predispositions and may result in action or be frustrated in a variety of ways.
  - **Tries** — This is important in understanding machines that have a variety of ways of achieving a goal including possibly ways that we don’t know about.

If the machine may do something we don’t know about but that can later be explained in relation to a goal, we have no choice but to use ‘is trying’ or some synonym to explain the behavior.
Examples of AI-system “Beliefs” Illustrate Natural/Common/Naive Reasoning

Belief is a slippery concept. Consider Derek Doran's Bank or Factory examples:

Explaining its actions - ‘There’s enough money in the account,’ or ‘I don’t give out that much money.’

There is some natural reasoning done in your head to come up with a category for an image you see as “factory.”

What was used to arrive at this?

We have common background knowledge (we believe in) that objects/lights, actions & relations are associated with factory.
Explanation and Background K

Famous example of knowledge needed to understand:

“The city council refused the demonstrators a permit because they feared violence,” vs. “… because they advocated violence.”

To determine that “they” in the first sentence refers to the council if the verb is “feared,” but refers to the demonstrators if the verb is “advocated” demands knowledge about the characteristic relations of city councils and demonstrators to violence.

Does a typical ontology account for the way people habitually describe the factory world or city councils actions?

Does it highlight the same/similar objects & relationships to events that we notice?

Lesson - no purely linguistic clue suffices.

CYC was geared to “learn” such background K to understand text.
Need to Connect to a Reasoner

Some areas are still hard. Is the answer to this Q encoded in a KB or reasoned out?:

"Can an elephant fit through a doorway?"

I probably didn't know this but reason this out (Sensemaking) using some size criteria.

Or Sensemaking for:

“I saw the Grand Canyon flying to NY”.

From DARPA’s Machine Common Sense (MCS) proposal:
DARPA wants to give AI common sense using child psychology:

• Common sense is “perhaps the most significant barrier” between the focus of AI applications today, and the human-like systems we dream of.  2 November 2018

AI Explanations Now

We seem close to AI systems that will do common tasks (drive or give advice on common tasks like eating) and they need to exhibit robust commonsense knowledge and reasoning to be trusted.

As intelligent agents become more autonomous, sophisticated, and prevalent, it becomes increasingly important that humans interact with them effectively – why did my self-driving vehicle take an unfamiliar turn?

Need what DARPA called Associate Systems in the 80s – which learned to cooperate
We have vast amounts of this not usually formalized in KBs: “you can’t be in two places at the same time” or “when drinking from a cup, hold the open end up.”
Contemporary AI Efforts Feature CS

“To make real progress in A.I., we have to overcome the big challenges in the area of common sense,”

Paul Allen, 2018

- A.I. “recognizes objects, but can’t explain what it sees.
- It can’t read a textbook and understand the questions in the back of the book. ….It is devoid of common sense.”

Oren Etzioni, formerly University of Washington professor who oversees the Allen Institute for Artificial Intelligence.

Overview of Commonsense Knowledge and Explanation
"We'll never produce NL understanding systems until we have systems that react to arbitrary sentences with "I knew that, or didn't know & accept or disagree because..."
Alessandro Oltramari (Research Scientist at Bosch)
"From machines that learn to machines that know: the role of ontologies in machine intelligence"

Ontologies play different and yet crucial roles for enabling

- **Semantic transparency** in Deep Learning Networks where the algorithmic processes are opaque, not the I/O

- The *Black Box problem* can reduced to a more “tractable” issue: how can we clarify the connection between data as input and data as output?

- **Quantitative methods**: measure the correlation E.g., Family of **metrics** that capture the degree of influence of inputs on outputs of the system*

- **Semantic transparency**: ontologies of input data (across layered NNs) can be ultimately interlinked to ontologies of output data

---

Some General Recurring Questions

1. How can we leverage the best of the two most common approaches to achieving commonsense?
   - formal representations of commonsense knowledge (e.g. encoded in an ontology's content as in Cyc or Pat Hayes’ Ontology of Liquids) vs.
   - strategies for commonsense reasoning (e.g. default reasoning, prototypes, uncertainty quantification, etc.)

2. How to best inject commonsense knowledge into machine learning approaches?
   - Some progress on learning using taxonomic labels, but just just scratches the surface

3. How to bridge formal knowledge representations (formal concepts and relations as axiomatized in logic) and representations of language use (e.g. Wordnet)
Sessions

16 January 2019 Introduction to the Tracks

23 January 2019 Commonsense Session 1
   Pascal Hitzler and Mike Gruninger

6 March 2019 Commonsense Session 2
   Speakers TBD
Additional References


- McCarthy, The Logic & Philosophy of ARTIFICIAL INTELLIGENCE, 1988


Overview of Commonsense Knowledge and Explanation