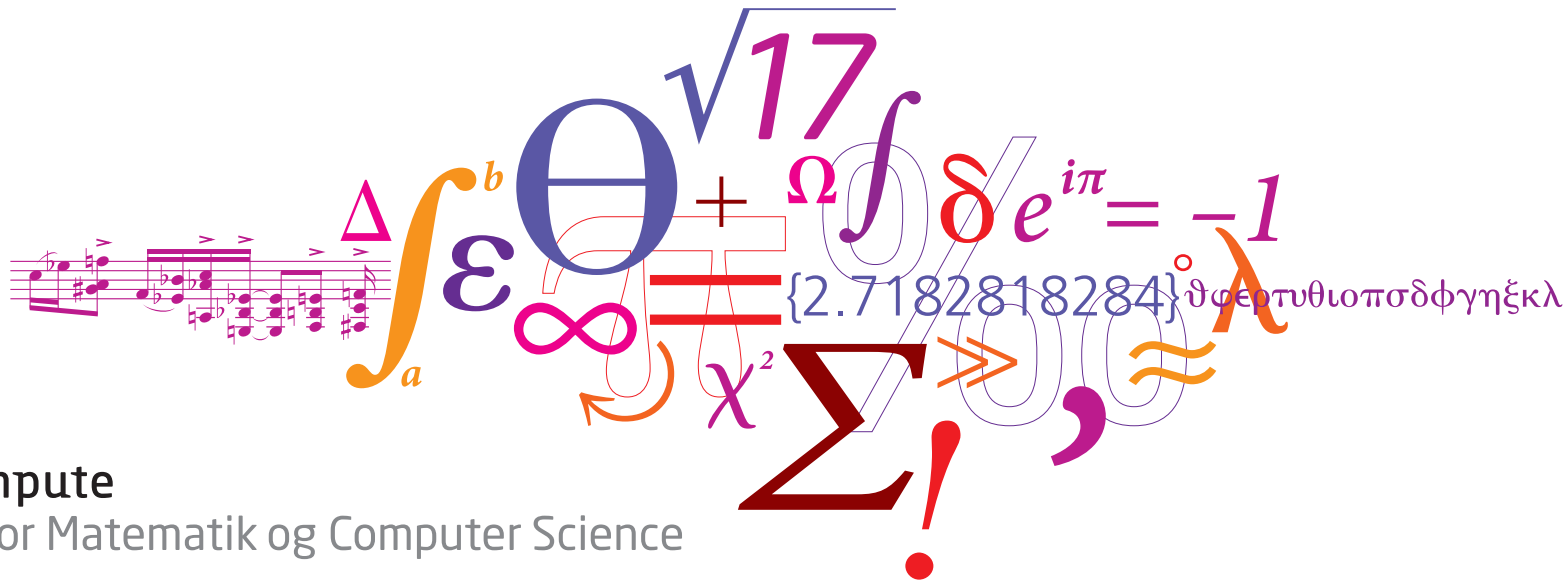
Synergies Between Symbolic and Sub-symbolic Artificial Intelligence

Thomas Bolander, Associate Professor, DTU Compute

Current Trends in AI, 23 November 2016

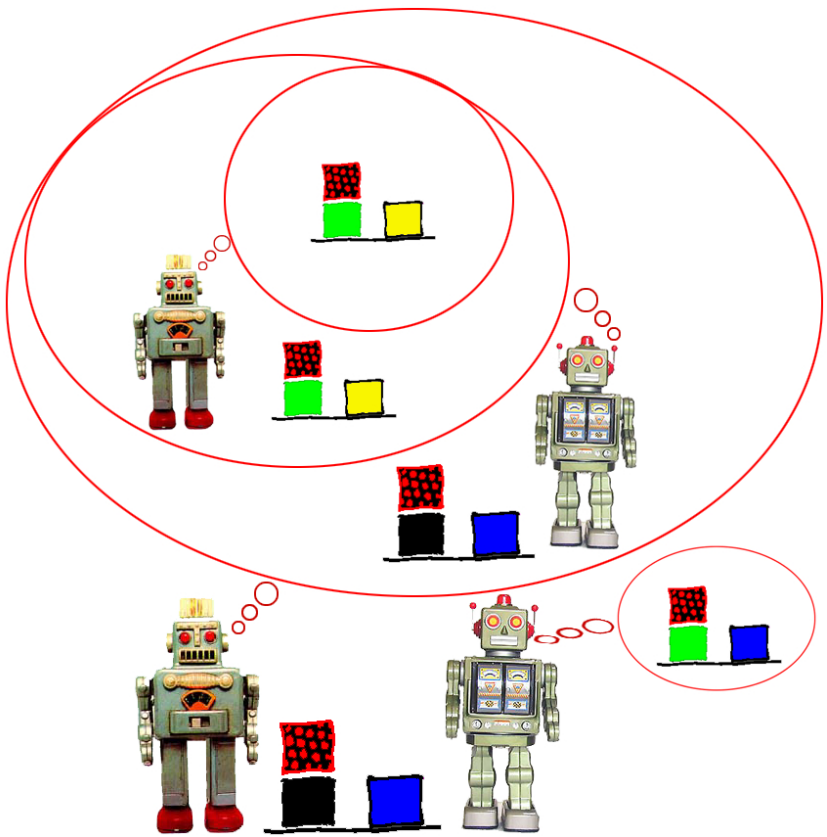


DTU Compute
Institut for Matematik og Computer Science

A bit about myself

Thomas Bolander

- Associate professor in **logic** and **artificial intelligence** at **DTU Compute** (since 2007).
- Member of the newly established **Siri Commission** (future of AI in DK. Est. by Ida Auken and IDA).
- **Current research**: How to equip AI systems with a **Theory of Mind** (ToM)?



Content

Symbolic vs sub-symbolic AI

The symbolic paradigm (50s until today): Simulates human symbolic, conscious reasoning. Explicit/symbolic world models. Search, planning, logical reasoning.

robust, predictable

flexible, learning

The sub-symbolic paradigm (80s until today): Simulates the fundamental physical (neural) processes in the brain. Artificial neural networks.

Symbolic or sub-symbolic: both!

- **Combined symbolic and sub-symbolic methods** (00s until today): Integration of the (so far) competing paradigms. E.g. Google Deepmind (learning arcade game) and AlphaGo.

Aspects of human cognition... and of AI

- Perception (sub-symbolic)
- (Knowledge representation and) reasoning (symbolic)
- Learning (sub-symbolic)
- Planning (symbolic)
- Interaction (multi-agent systems) (symbolic)

Combining approaches and paradigms

Last sentence of the 2016 report of the **One Hundred Year Study on Artificial Intelligence** (Stanford University, September 2016):

> *Although the separation of AI into subfields has enabled deep technical progress along several different fronts, synthesizing intelligence at any reasonable scale invariably requires many different ideas to be integrated. For example, the AlphaGo program that recently defeated the current human champion at the game of Go used multiple machine learning algorithms for training itself, and also used a sophisticated search procedure while playing the game.*

Another recent succesful example of integrating symbolic AI (*reinforcement learning*) and sub-symbolic AI (*deep neural networks*): **Google DeepMind** learning to play Artari arcade games (Nature, 2015)...

DeepMind playing Breakout

http://www2.compute.dtu.dk/~tobo/DeepMind.MP4

DeepMind playing Pac-Man

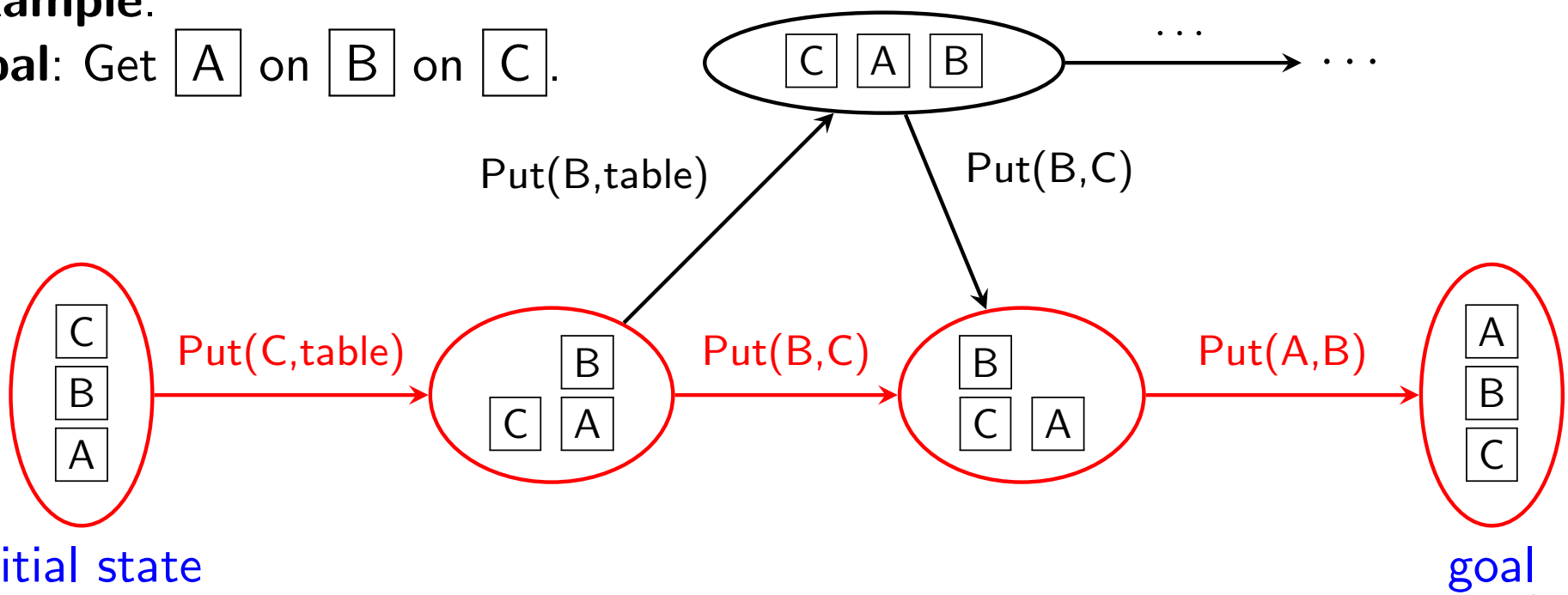
http://www2.compute.dtu.dk/~tobo/DeepMindPacman.mp4

Automated planning

Automated planning:

- Computing **plans** (**seq. of actions**) leading to desired outcomes.
- More precisely: Given a **goal formula**, an **initial state** and a set of **possible actions**, an **automated planner** outputs a plan that leads from the initial state to a state satisfying the goal formula.
- Actions are represented compactly via **action models** (**action schemas**).

Example.
Goal: Get A on B on C.



Learning action models

Action models are traditionally manually crafted (no learning).

Learning approaches to learning action models:

- Machine learning approaches (e.g. kernel perceptrons). (Mourão et al., 2012)
- Logical/qualitative approaches. (Walsh & Littman, 2008; Amir & Chang, 2008; Bolander & Gierasimczuk, 2015)

Learning the right action model for Pac-Man generalises to all levels of the game, opposite the reinforcement learning of DeepMind!

Example of an action model (in ADL). (Pacman with one ghost and no magic pills)

ACTION: $Move(startpos, dir)$
PRECONDITION: $Neighbour(startpos, resultpos, dir) \wedge \neg Dead(pacman)$
EFFECT:
$At(pacman, resultpos) \wedge \neg At(pacman, startpos) \wedge \neg At(resultpos, pill) \wedge$
when $At(ghost, resultpos)$: $Dead(pacman)$

Social robot bartender

(Petrick & Foster, ICAPS 2013)

Robustness, trust, explanations, generalisability, predictability...

... are arguments in favour of symbolic and logical methods.

CoBot, Carnegie Mellon U.

Proposal: A multi-paradigmatic competition in AI

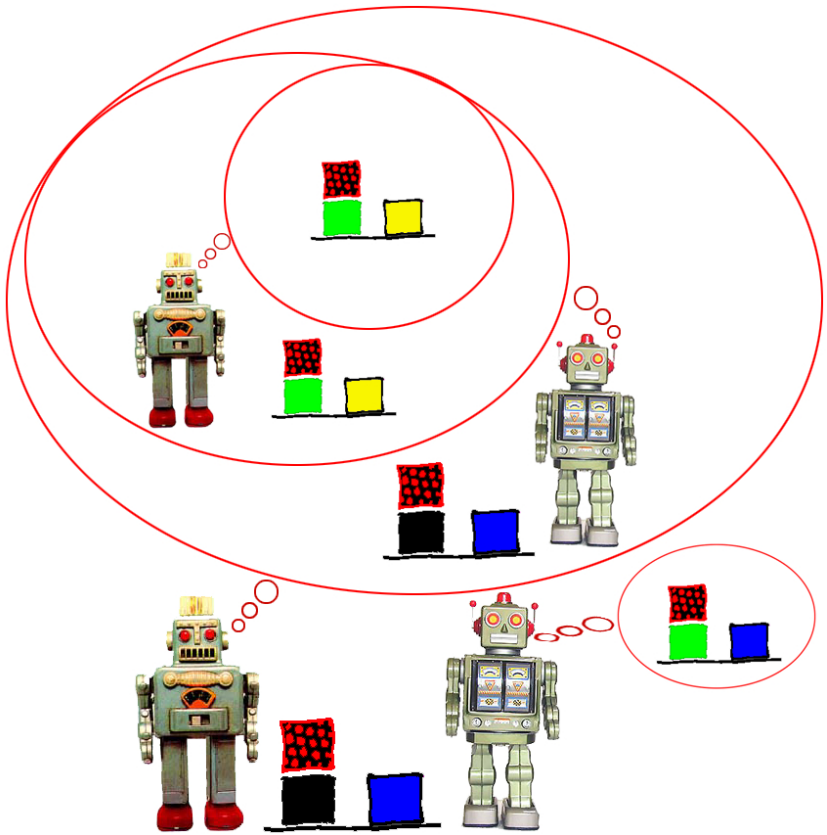
http://www2.compute.dtu.dk/~tobo/AImuffins_on_MAsojourner.mov

What does it take to build a socially intelligent robot?

Social intelligence: The ability to understand others and the social context effectively and thus to interact with other agents successfully.

Theory of Mind (**ToM**): The ability of attributing mental states—beliefs, intentions, desires, etc.—to other agents.

Theory of Mind (ToM) is essential to human social intelligence (Baron-Cohen, 1997).



Anti-social TUG hospital robots (2009)

Frustrated users of hospital robots in USA:

- *"TUG was a hospital worker, and its colleagues expected it to have some social smarts, the absence of which led to frustration—for example, when it always spoke in the same way in both quiet and busy situations."*
- *"I'm on the phone! If you say 'TUG has arrived' one more time I'm going to kick you in your camera."*
- *"It doesn't have the manners we teach our children. I find it insulting that I stand out of the way for patients... but it just barrels right on."*

TUG hospital robot

(Barras, 2009)

Helpful robots = Automated Planning + Theory of Mind