

How Cognitive Science Impacts AI

And What We Can Learn From It

Authors: **Marcus Westberg**, Amro Najjar, Amber Zelvelder

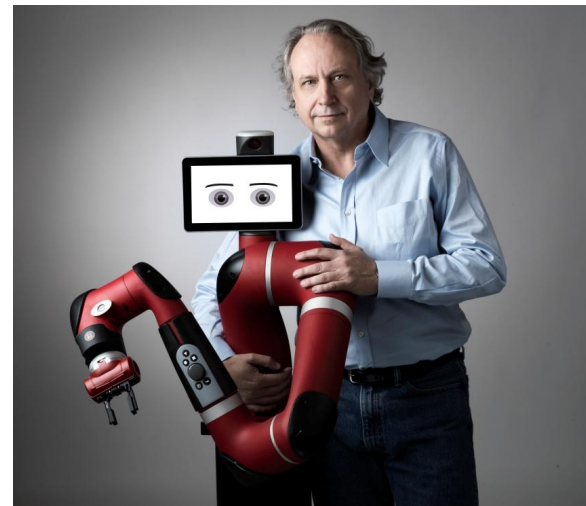
The rise of Computational Theory

- Turing Machine
 - If computation can solve problems and make decisions, can cognition be computational?
- Putnam & Fodor's Computational Theory
 - Manipulation of symbols - language of thought
 - Founding theory of cognitive science
- Marr's theory of vision
 - Neuroscience adaptation of CTM



The shift to world-oriented approaches

- Cognition integrated with motion and world
- Embodied, Embedded, Extended and Enactive cognition
- Focus on frugality and offloading



Embodiment in a nutshell

- Classical CTM traps the mind within the skull (input output sandwich)
- World-oriented approaches incorporate body and environment into cognition
- Body as not just tool, but determinator of cognitive aspects
- Embedded cognition: natural and social environment as offloading



Theory-Theory to Simulation Theory

Similar trend toward frugal action-integration in folk psychology

Theory-theory:

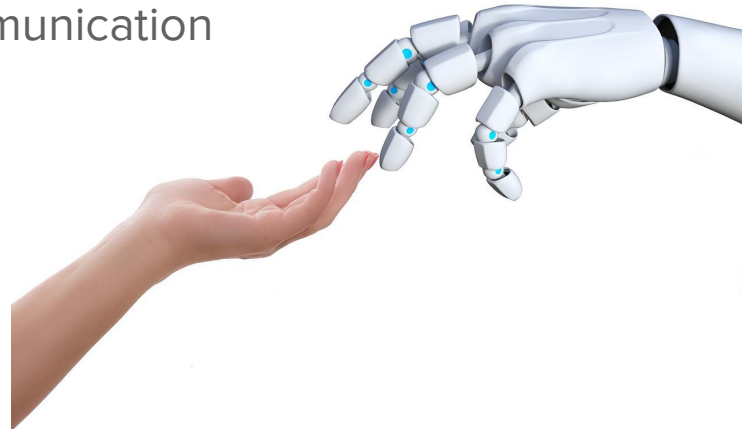
- Understanding through innate (nativist) or constructed (empiricist) knowledge base
- Information-rich

Simulation theory:

- Simulation through empathy or mirroring actions and behaviour
- Prediction through relatability

Folk Psychology in XAI

- Understanding an AI as an agent with mental states requires a mindreading relationship
- BDI agents display an application of theory-theory in XAI
- But theory-theory does not promote empathy or relatability
- This underutilises the human toolkit for communication
- Examples of simulation theory in XAI:
 - Kobayashi & Yamada's mobile robots
 - Usage of emotion-expressive avatars



Benefits of world-oriented XAI

- Simulation theory showcases importance of embodiment and world in cognition
- Embodied cognition expanding communicative toolkits
- Embodied cognition causes simulation (Boston Dynamics)
- More frugal pathways to communication
- Human nature does the heavy lifting
- Increased empathic bonding and trust

Prospects for the future

- XAI more naturally ingrained with human cognition and communication
- Embodied agents displaying relatable behaviour in communication and world interaction
- Bonding and understanding through simulation, refining through theory - adapting hybrid theories of folk psychology

Challenges

- How to design, develop, implement, test simulation theory inspired XAI
- Personalization: explain to lay users
- Agent to agent explanation
- Simulation of disembodied agents
- Uncanny valley - fear instead of trust