

CSCI 400 Artificial Intelligence
David Keil, Framingham State University

Topic 8: Philosophical considerations and future prospects

1. Theories of mind
2. Objections to weak and strong AI
3. Future prospects and goals

Reading: Ch. 26-27

Inquiry

- Could clarity in research in cognitive science be served by relaxing the association of intelligence with humanness and by reframing AI and cognition as *rational adaptive behavior*?
- Does the notion of *bounded optimality* (a property of the best *program* achievable to solve a problem that entails adaptation) offer a sound theoretical foundation for AI research?

Objectives

- 8a. Discuss philosophical issues raised by the notion of artificial intelligence
- 8b. Discuss the weak and strong AI theses, and future prospects for AI
- 8c. Explain the notion of bounded optimality

1. Theories of mind

- *Dualism* (Descartes): Soul and body are distinct
- *Monism* (materialism): Material objects exist, not immaterial souls
- *Problems*: consciousness, understanding, self-awareness
- *Intentional states* (propositional attitudes): beliefs, knowledge, wishes, feelings about external world

Epistemology

- The study of how we know what we know
- *Theories:*
 - *Rationalism* (Descartes, Plato):
We remember or figure it out
 - *Empiricism* (Locke, Hume): We obtain
knowledge via our senses
- *Ontology*, in contrast, is the theory of what *is*

Philosophical considerations re AI

Questions:

- How does a mind work?
- Can machines act as if intelligent?
(Weak AI hypothesis)
- Can a machine think or have a mind?
(Strong AI)
- “Can machines think?” is a poorly defined
question, because “think” has different usages

2. Objections to weak and strong AI

- *Weak AI thesis*: the claim that an artificial system can *simulate* intelligence
- *Strong AI thesis*: the claim that an artificial system can *be* intelligent

Objections to strong AI

- *Argument based on phenomenology* (study of experience): machines are said to lack the experience of thought
- *Argument based on intentionality*: machines are said not to be referencing actual things in the world
- Turing's reply to possible objections: lacking evidence that humans think, we politely say that they do
- *Comparisons*: artificial urea, legs, sweeteners, insemination, flowers
- Q: Is computer simulation of mental process an actual mental process?

Objections to AI based on Gödel's theorem

- Lucas and Penrose challenged weak AI based on Gödel's theorem (1931) that showed no formal system in mathematics could be both complete (enable proof of all truths) and consistent (enable proof only of true assertions)
- The claim of human superiority is based on the assumption that humans are capable of what no formal system can do and are not subject to the same limitations as formal systems

Chinese room argument (J. Searle)

- An argument against strong AI
- Imagine an non-Chinese-speaking person in a room with a large rule book about how to reply in Chinese to Chinese-language utterances
- This system simulates speaking and understanding Chinese
- But if the person doesn't, and the rule book doesn't, then where is the understanding of Chinese?

Is discussion of AI too human-centered?

- *Anti-AI view* is based on assumptions about humanness of intelligence
- *Turing test* is based on assumption that intelligence is imitation of human behavior
- Are we not really interested in “AI” as *the formalization of generalized rational, adaptive behavior?*

The Turing test

Other definitions of intelligence

3. Future prospects and goals

- Ethical issues
- Future of agent computing
- Goals of AI
- The possible goal of bounded optimality

Ethical issues

Concerns:

- Job loss via automation
- Loss or excess of leisure time
- Loss of sense of uniqueness
- Loss of privacy rights
- Loss of accountability
- End of human race (technological singularity when machines invent machines)
- Civil-rights for robots?

Agent components and functions

- Interaction with environment
- Tracking state of environment
- Developing and selecting courses of action
- Expression of preferences as utility
- Learning

Agent architectures

- Hybrid architectures are used
- Real-time AI requires *anytime algorithms* that always have a reasonable decision when stopped
- *Decision-theoretic metareasoning* uses theory of value of information
- *Reflective architecture* supports deliberation about computation within the architecture, where state space consists of environment plus agent state

Possible goals of AI

Alternatives:

- *Perfect rationality*: maximum utility
- *Calculative rationality*: eventual arrival at optimal action at time of start of computation
[CLARIFY]
- *Bounded rationality* using satisficing
- *Bounded optimality*: maximum utility, given computational resources

Bounded optimality

- *Definition*: a property of the best *program* achievable to solve a problem
- “Seems to offer the best hope for a strong theoretical foundation for AI”
- BO entails adaptation to environment
- *Asymptotic bounded optimality*: ability to compete with a BO program running on a machine weaker by not more than a constant factor

Concepts

anytime algorithm	meaning
artificial intelligence	metareasoning
bounded optimality	mind
bounded rationality	monism
consciousness	perfect rationality
constructivism	phenomenology
dualism	rationalism
empiricism	reflective architecture
epistemology	satisficing
experience	strong AI
intentional state	thinking
	weak AI

References

George Luger. *Artificial Intelligence*. Addison Wesley, 2005.

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach, 2nd ed.* Prentice Hall, 2003.