FUTURE DIRECTIONS

OUTLINE

- Agent Components

- Agent Architectures

- Are We Going in the Right Direction?

- What if AI Does Succeed?

AGENT COMPONENTS

- Interaction with the environment:
    - Sensors
    - Actuators
- Keeping track of the state of the world
    - Perception
    - Updating internal representations
- Projecting, evaluating, and selecting future courses of action
    - Hierarchical structure
    - Hierarchical reinforcement learning
- Utility as an expression of preferences
- Learning
    - Abstraction
    - Deep belief networks

AGENT ARCHITECTURES

- Hybrid architectures

  - Time-critical (real-time) vs. planning

    - Anytime algorithms

    - Decision theoretic metareasoning

      - Cost/benefit of computation (delay/better decision)

- Reflective architecture – sentience?

ARE WE GOING IN THE RIGHT DIRECTION?

- Perfect rationality?

- Calculative rationality?

    - Right answer at the wrong time has no value

- Bounded rationality?

    - Satisficing

- Bounded optimality?

    - Asymptotic bounded optimality

        - Optimal programs as opposed to optimal actions

WHAT IF AI DOES SUCCEED?

- Small scale success
  - We're there
- Medium scale success
  - On our way
    - AI used "behind the scenes" for the most part
    - Many things are called "AI" that aren't really
  - But what happens when AI systems are interacting with us?
  - What about built in bias?
- Large scale success
  - Human level and beyond intelligence

" We can see only a short distance ahead, but we can see that much remains to be done.

Alan Turing, 1950