Intro: AI

• Nilsson defines AI as "Intelligent behavior in artifacts"

• Definition depends on meaning of *intelligence*

• What is intelligence?
  – Is it a single faculty, or a collection of distinct and unrelated abilities?
  – How much is learned, and how much is hard-wired?
  – How does learning occur?
  – What is creativity?
  – What is intuition?
  – Can intelligence be inferred from behavior, or does it require a particular internal mechanism?
  – What are the biological representations of intelligence, and what are their implications for computational intelligence?
  – What is self-awareness, and what is its implications for computational intelligence?
  – Must intelligence be defined in human terms?
  – Does intelligence depend on biologic existence, requiring interaction with the environment?
Intro: AI Definitions

1. **Haugeland**: The effort to make computers think... machines with minds, in the full literal sense.

2. **Billman**: The automation of activities that we associate with human thinking, activities such as decision making, problem solving, learning, ...

3. **Charniak and McDermott**: The study of mental faculties through the use of computational models.

4. **Winston**: The study of the computations that make it possible to perceive, reason, act.

5. **Kurzweil**: The art of creating machines that perform functions that require intelligence when performed by people.

6. **Rich and Knight**: The study of how to make computers do things at which, at the moment, people do better.

7. **Schalkoff**: The field of study that seeks to explain and emulate intelligent behavior in terms of computational processes.

8. **Luger and Stubblefield**: The branch of CS concerned with automation of intelligent behavior.

Above can be categorized along 2 dimensions:

1. Human v ideal performance, and

2. Reasoning v behavior

<table>
<thead>
<tr>
<th></th>
<th>Human</th>
<th>Ideal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thought</td>
<td>Haugeland</td>
<td>Charniak/McDermott</td>
</tr>
<tr>
<td></td>
<td>Billman</td>
<td>Winston</td>
</tr>
<tr>
<td>Action</td>
<td>Kurzweil</td>
<td>Schalkoff</td>
</tr>
<tr>
<td></td>
<td>Rich/Knight</td>
<td>Luger/Stubblefield</td>
</tr>
</tbody>
</table>
• Cognitive Model Approach: Thinking human
  – Requires understanding how people think
  – 2 ways to study this:
    * Introspection
    * Human experiments
  – Strong v weak AI
    * Strong AI: The embodiment of human intellectual capabilities with a computer.
    * Weak AI: The study of mental facilities through the use of mental models implemented on a computer.
  – Cognitive science: Study of human thought processes
    Has utilized AI to model thought processes - weak AI

• Laws of Thought Approach: Thinking rationally
  – Mathematical approach to specifying correct thought
  – Uses formal logic to represent knowledge
  – Formal methods (inference, deduction) to prove validity of knowledge
  – Performance is provably correct -
    Is _ideal_ performance
  – Criticisms:
    1. Difficult language to use
    2. Limited computational resources
Intro: AI Definitions (3)

• Turing Test Approach: Acting human
  – Given
    1. An AI
    2. A human
    3. An interrogator
  – The AI and human are hidden from the interrogator
  – The interrogator can communicate with both via a given means (teletype)
  – The AI is successful if it can convince the interrogator it is human
  – Advantages:
    1. Empirical
    2. Does not try to define intelligence
    3. Eliminates bias
  – Criticisms:
    1. Only addresses symbolic tasks
      Doesn’t deal with perceptual, manual tasks
    2. Is human-based
    3. Lady Lovelace’s objection: Since computers are programmed, can only do as told and so cannot have original thought
    4. Argument from Informality of Behavior: Cannot create a set of rules that handle every circumstance
Intro: AI Definitions (4)

- Rational Agent Approach: Acting intelligently
  - Acting intelligently: Given a set of beliefs and goals, performing appropriate actions to achieve those goals
  - Agent-based approach
    Agent: Any entity that performs and acts
  - Advantages
    1. More general than Laws of Thought
      Sometimes need to act even when no provably correct action
    2. More applicable to scientific analysis (than humanistic approaches)
      Is independent of evolutionary processes
Intro: Goals of AI

• Basic goals of AI:

  1. Development of machines that can do things as well as - if not better than -
     humans
  2. Specify what intelligent behavior is and determine whether it is possible in non-
     human entities

• Basic question (ala second goal above) is

  *Can machines think?*

  1. **What is a machine?**
     - Can it be biological?
       - If so, humans are existence proof of machines’ ability to think
     - Proposed requirements for intelligence:
       * Massively parallel processing
       * Fuzzy logic
       * Connectionist models (holographic memories)

  2. **What does it mean for an entity to think?**
     (What is intelligent behavior?)
     See earlier definitions of AI

  3. **What does *can* mean?**
     - Main concern here is whether machines will eventually have the capability of
       thought
     - Arguments against:
       (a) Thinking is too complex a process to ever implement on a machine
       (b) Embodiment
         Human intelligence is result of species’ interaction with environment over
         long time
         Machines are not *situated*
         Note: Not so true now
       (c) Human spark
         Humans are special
       (d) Machines cannot think like humans
         Processing power too limited
         Note: Does intelligence require thinking like a human?
Intro: Goals of AI (2)

- Physical Symbol System
  - Proposed by Newell and Simon
  - Considered to be at core of AI
  - "A symbol system consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (or symbol structure). Thus, a symbol structure is composed of a number of instances (or tokens) of symbols related in some physical way (such as one token being next to another). At any instant of time the system will contain a collection of processes that operate on expressions to produce other expressions: processes of creation, modification, reproduction and destruction. A Physical Symbol System is a machine that produces through time an evolving collection of symbol structures. Such a system exists in a world of objects wider than just these symbolic expressions themselves.
  
  - Such a system is capable of manipulating abstractions (symbols and expressions) via a set of procedures
  - The symbols can be combined into more complex structures (expressions)

- Physical Symbol System Hypothesis:
  A Physical Symbol System has the necessary and sufficient means for general intelligent action
1. Symbol-processing approaches

- Top-down
- ”Classic” AI
- Consist of
  - Knowledge base of declarative facts
  - Procedures for reasoning with the knowledge
- General levels of such systems:
  - Knowledge level
    - Specification of desired knowledge
  - Symbol level
    - Representation of knowledge via symbols
  - Implementational level
    - Low-level (implementational) representation
2. Sub-symbolic approaches

- Bottom-up
- Behavior-based systems (Brooks):

  Behavior-based System Architecture

  - General levels of such systems:
    (a) Signal level
    Entity processes signals from environment
    (b) Behavior modules
    * Generally one module per behavior
    * Each receives processed signals
    * If trigger conditions met, behavior is executed
    * Referred to as physical grounding
    * Modules independent
    * No centralized control module
    (c) Emergent behavior
    Interaction of multiple independent behaviors results in overall complex, intelligent action
    (d) Based on ethology - study of animal behavior
Intro: Implementational Approaches (3)

• Connectionist systems:
  
  ![Connectionist Architecture]

  – Simulate brain in software/hardware
  – Models systems of interconnected neurons

• Based on biologic entities

3. Situated automata

• Hybrid systems
Intro: AI Foundations: Philosophy

• Concerned with nature of intelligence

1. Rationalist approach
   – Aristotle
     * Proposed basic components for rational mind
       (a) Algorithm (for "correct" thinking)
       (b) Physical system (the "brain")
       (c) Knowledge (data for thought)
       Acquired via senses and by reasoning
       (d) Actions
       Results of correct thinking
     * Postulated a duality between matter and its form, which can be changed
   – Descartes
     * Proposed that reflection can reveal mind’s inner workings
     * "Cogito ergo sum": I think; therefore I am.
     * This epitomized the separation of the mind from the physical world
     The 2 are separate and independent
   – Mind-body problem
     * Descarte’s philosophy proposes complete separation of mind from physical world
     * However, there must be a connection to the physical world in order for entities to interact with it
     * This is the mind-body problem
     * Several philosophies relating the 2:
       · Materialism
         Mind functions in obeyance of physical laws of the universe
       · Dualism
         Mind obeys physical laws, but there is also a metaphysical component outside of physical world
       · Hybrid
         Mind obeys physical laws, but these are ultimately unknowable
2. Emergent Intelligence

– Situated models
  * Intelligence is emergent as a result of the encompassing culture
  * Intelligence reflected by collective behavior of large number of simple, interacting agents
– Agent is element of society that is aware of environment and can affect the environment either directly or via cooperation with other agents
– Agent characteristics:
  * Autonomous or semiautonomous
  * Situated in own environment
  * Interactive
  * Exist in structured society
– Requirements for such a society:
  * Structures for representing info
  * Search strategies for alternate solutions
  * Architectures that support agent interaction
Intro: AI Foundations: Mathematics

• Algorithms

• Logic
  – Aristotle: Logic
  – Boole: Boolean logic
  – Frege: First order predicate logic
  – Tarski: Theory of reference
    Relates formal logic to physical world (i.e., provides semantics)

• Computability
  – Core question to machine intelligence: What are limitations on computation?
  – Undecidability
    * Is there a limit to power of proof procedures?
    * **Goedel’s Incompleteness Theorem**: In any language expressive enough to describe properties of natural numbers, there exist statements that are undecidable.
      Ramifications: Truth of some statements cannot be algorithmically proved
    * **Church-Turing Thesis**: A Turing machine is capable of computing any computable function.

  – Intractability
    * Problem is *intractable* if \( f(n) \in \mathcal{O}(n^m) \)
    * Such problems called *NP-complete*
    * Take unreasonable amounts of time to execute
    * Intelligent behavior must be based on *tractable* algorithms

• Probability
  Intelligent systems must deal with uncertain worlds
Intro: AI Foundations Psychology and Linguistics

• Psychology
  – Concept of intelligence evolved over time
    * Introspection
      · Reflection can reveal mind’s inner workings (ala Descartes)
      · Problematic - Biased
        When self-reflecting, thinker influences results
    * Behaviorism
      · Everything based on stimulus-response
      · Knowledge, beliefs, goals, etc. not important
    * Cognitive Science
      · Reintroduced Knowledge, beliefs, goals
      · Craik’s hypothesis:
        1. Stimulus → internal rep
        2. Internal rep → new structure
        3. Internal rep → action
      · Enables planning and prediction of results as result of proposed action

• Linguistics
  – Study of language
  – Chomsky responsible for modern approach
  – Overthrew behaviorist approach to language
Intro: Agents

- Agent is entity that perceives and acts
  - Perception occurs via sensors
  - Action accomplished via effectors
- Rational agent is agent that "does the right thing"
- Rationality depends on
  - **Performance measure**: a criterion for success
    * Should be objective
    * Should be defined by outside agency
    * Must be careful *what* is being measured
    * Must be careful *when* it is measured
  - **Percept sequence**: series of inputs from environment
  - Knowledge about its environment
  - Set of available actions
  - Concept of "what is right"
    \[\text{rationality} \neq \text{omniscience}\]
- Ideal rational agent: One that performs whatever action is expected to maximize its performance measure for each possible percept sequence on the basis of the evidence provided by its percept sequence and whatever built-in knowledge it has available.
- Agent is a function that maps perception sequences to actions
- Actions must be based on agent’s experiences
  - If pre-programmed, behavior not intelligent
  - Cannot adapt
- Want *autonomous* agent
  - Autonomy does not preclude *a priori* knowledge
• Goal of recent AI is to construct *agent programs*
  
  To design such a program, need to establish
  1. Types of percepts needed (or available)
  2. Actions needed (or available)
  3. Goals and performance measures
  4. Environment
  
  These implemented in terms of
  1. Sensors (physical or soft)
  2. Effectors (physical or soft)
  3. Computing architecture
  4. The program itself
  
  Basic agent structure (function):

```python
function BasicAgent (percept) returns action {
    static memory

    memory <- UpdateMemory(memory, percept)
    action <- ChooseBestAction(memory)
    memory <- UpdateMemory(memory, action)
    return action
}
```
Intro: AI Disciplines

• Core Areas/techniques:
  – Search
  – Knowledge representation and reasoning
  – Planning
• Game playing
• Automated reasoning and theorem proving
• Expert systems
• Natural language processing
• Computer vision
• Robotics
• Machine learning
• Cognitive modeling
• Philosophical issues
• AI languages