Artificial Intelligence II

CSC 720—Spring 2010

Jon Doyle

Department of Computer Science
North Carolina State University

January 12, 2011

Course introduction

What is this course about again?

- Logical, economic, psychological, social, environmental, and mechanical conceptions of
  - Rationality and irrationality
  - Knowledge, belief, and uncertainty
  - Desire, preference, intention, and motivation
  - Planning, deliberation, action, and self-management
  - Learning and knowledge acquisition
  - Consciousness and personhood
  - Representation, embodiment, and self-government
- Understanding the structure and principles of artificial agents
  - Theoretical understanding and foundations
  - Illuminating applications

Big problems

AI as automation of all human knowledge and skills

- Big problems are the main problems of each field of knowledge
- Politics, economics, psychology, biology, philosophy, etc.
- For example, economics studies rational action (resource allocation), at both individual and group level

Big questions concerning

- Types, origins, and characteristics of intelligence
- Effects of artificial intelligence

Types of intelligence

- Human
- Animal
- Plant?
- Inanimate?
- Machine?
  - Equal to any of these?
  - Possibly superhuman?
- Different kinds of human minds?
- Social and organizational

Origins of intelligence

- Increasing size and complexity
- Increasing breadth and depth
- Nature versus nurture

Increasing size and complexity

- Size alone?
  - A bigger pile of sand is just a pile of sand
- Complexity alone?
  - Erdös and Spencer random graph theorem

Increasing breadth and depth

- Expert systems “smart” about one thing, dumb otherwise
- Add depth to narrow intelligence?
  - Still dumb?
- Add breadth (more narrow things)?
  - Ever get intelligence?
- Is intelligence approximable?
  - If so, by what sorts of limiting processes?
  - depth = closer approximation to particular human skills
  - breadth = closer approximation to variety of skills

Nature versus nurture

- Locke's blank slate: initial structure knows nothing, learns all
  - Structure = knowledge?
  - Universal Turing Machines?
  - Minsky's 2 symbol, 7 state UTM
- Chomsky's innate knowledge of language structure
  - Infant language learning of $\approx$ 16 bits
- Reverse time; chisel away at an intelligent person
  - Remove eyes; still intelligent?
  - Remove all senses; still intelligent?
    - Even if from conception?
  - Paralyze all voluntary muscles; still intelligent?
  - Remove particular neurons or bits of brain
    - Old-age memory loss, stroke victims; still intelligent?

Characteristics of intelligence

- Rationality
- Knowledge and skills
- Consciousness?
  - Intelligence without consciousness?
  - Consciousness without intelligence?
- Comprehensibility and content
  - Is there simple underlying structure we can uncover?
  - Or is intelligence a "kluge"?
  - Chaitin/Kolmogorov information content
    - Logic $\neq$ knowledge
    - Knowledge is nonlogical

Characteristics of intelligence

- Personhood and humanity
  - Is an intelligent thing a person?
  - Do we need to care about it?
- Personality and character
- Adaptiveness
  - Is something that stops learning still intelligent?
  - That learns some things but not others?
- Emergence and reducibility
- Computability?
  - With ordinary computers?
  - With hypercomputation?

Effects of intelligence

- How do people change in the brave new world?
  - All technology changes how humans live and think
- Does it change human nature?
  - What if no one has to reason (plan, choose, etc.) any more?

General research questions

- What should we aim for?
  - What is desirable, computability aside?
- What can we aim for?
  - What is feasible?
- How do we get there?
  - What are the subproblems and approximations?

Always ask

- What does it mean?
- Is it true?
- Why should we be interested in that?
- Does it make a difference?
- What difference does it make?
- Do the details matter?
- Would infinite instantaneous computation solve the problem?