Artificial Intelligence II
CSC 720—Spring 2010

Jon Doyle

Department of Computer Science
North Carolina State University

January 12, 2011

Rationality

Pervasive rationality

Notions of rationality and irrationality pervade the study of

- Knowledge and belief
- Decision and action
- Planning and resource allocation
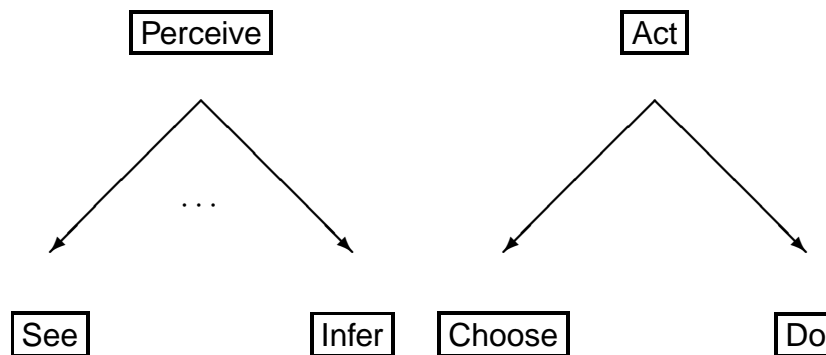- Representation and learning
- Personhood

Competing conceptions of rationality

- Psychological rationality
  - Focus on reasoning and reasonability
- Logical rationality
  - Focus on entailment and consistency
- Economic rationality
  - Focus on preference and optimality
- Social or cultural rationality
  - Focus on judgment by neighbors

Some history

- Greeks: rhetoric and logic
- Enlightenment philosophers: probability and utility
- Mathematics: mathematical logic and economics
- Knight, Simon: psychological economics
- AI: economical (nonmonotonic) logic and representation

The *True* and the *Good* and the *Reasonable*

Perceive		Act	

. . .

See	Infer	Choose	Do

Roles for rationality

Descriptive uses

- Concepts for formalization
- Mathematical tools for analysis

Normative uses

- Correctness standards for judging reasoning and action
- Targets for mechanization

Artificial Intelligence II
© 2010 by Jon Doyle

What is the character of rational thought?

Epistemology and economics study idealized rationality

- Perfect epistemic omniscience and consistency
- Instantaneous and complete assimilation of new information
- Perfect decisions on demand

Artificial Intelligence II
© 2010 by Jon Doyle

What is the character of rational thought?

Realistic rationality exhibits limits

- Reasoning and deliberation require effort and concentration
- Action taken despite inconsistency and ignorance of indefinite duration
- Learning takes time, and slows as habits accumulate

NC State University

34 / 400

Impediments to rationality

Ideal rationality is infeasible

- *Informational* limitations
  Incompleteness, inertia, inconsistency, indeterminacy
- *Resource* limitations
  Time, memory, human efforts, etc.
- *Organizational* limitations
  Operation sets, costs, reliabilities, adaptability
- *Physical* limitations
  c, $\hbar$, finiteness
- *Metaphysical* limitations
  Consistency, truth, pluralism

NC State University

35 / 400

Rationality in limited agents

Result and process rationality (Good, Simon)

- Type 1 or substantive rationality
- Type 2 or procedural rationality

Balancing control and reasoning computation

- Allocate these efforts rationally?
- Default choices vs. infinite regress
- Reflective equilibrium and ratified decisions

NC State University

36 / 400

Artificial Intelligence II
© 2010 by Jon Doyle