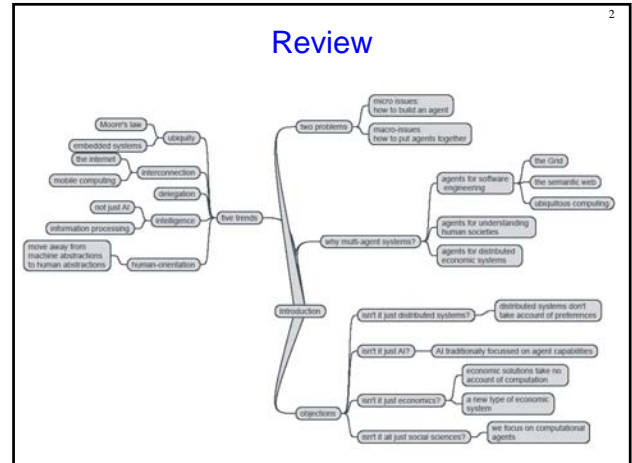


**Intelligent Autonomous Agents**  
**ICS 606 / EE 606**  
**Fall 2011**

**Nancy E. Reed**  
**nreed@hawaii.edu**



### Lecture #2A – Agent Environments and Intentional Systems

- Agent Environments
- Agents as Intentional Systems
- References
  - Wooldridge MAS Ch. 1 & 2
  - Russell & Norvig AIMA Ch. 2
  - Weiss Ch. 1 & 2

Table Page 4. Weiss

	attribute	range
agents	number	from two upward
	uniformity	homogeneous ... heterogeneous
	goals	contradicting ... complementary
	architecture	reactive ... deliberative
	abilities (sensors, effectors, cognition)	simple ... advanced
interaction	frequency	low ... high
	persistence	short-term ... long-term
	level	signal-passing ... knowledge-intensive
	pattern (flow of data and control)	decentralized ... hierarchical
	variability	fixed ... changeable
environment	purpose	competitive ... cooperative
	predictability	foreseeable ... unforeseeable
	accessibility and knowability	unlimited ... limited
	dynamics	fixed ... variable
	diversity	poor ... rich
availability of resources	restricted ... ample	

### Characteristics of Environments

- Accessible vs. inaccessible
- Deterministic vs. non-deterministic
- Episodic vs. non-episodic
- Static vs. dynamic
- Discrete vs. continuous

### Accessible vs. Inaccessible Environments

- An **accessible** environment is one in which the agent can obtain **complete, accurate, up-to-date information** about the environment's state.
- Most moderately complex environments (including, for example, the everyday physical world and the Internet) are **inaccessible**.
- The more accessible an environment is, the more straightforward it is to build agents to operate in it.

### Deterministic vs. Non-deterministic Environments <sup>7</sup>

- A **deterministic** environment is one in which every action has a **single guaranteed effect** — there is no uncertainty about the state that will result from performing an action.
- The physical world can, for all intents and purposes, be regarded as **non-deterministic**.
- Non-deterministic environments present greater problems for the agent designer. Why?

### Episodic vs. Non-episodic Environments <sup>8</sup>

- In an **episodic** environment, the performance of an agent is **dependent on a number of discrete episodes**, with no link between the performance of an agent in different scenarios.
- Episodic environments are simpler from the agent developer's perspective because the agent can decide what action to perform based only on the current episode — it need not reason about the interactions between this and future episodes.

### Static vs. Dynamic Environments <sup>9</sup>

- A **static** environment is one in which everything can be assumed to **remain unchanged**, except by actions of the agent.
- A **dynamic** environment is one that has other processes operating on it, and which hence **changes in ways beyond the agent's control**.
- The physical world is a highly dynamic environment.

### Discrete vs. Continuous Environments <sup>10</sup>

- An environment is **discrete** if there are a **fixed, finite number** of actions and percepts in it.
- Russell and Norvig give
  - a chess game as an example of a discrete environment, and
  - taxi driving as an example of a continuous one.
- Continuous environments may be modeled as discrete ones

### Which Environmental Characteristics Make Constructing Agents More Difficult? <sup>11</sup>

- Accessible vs. inaccessible
- Deterministic vs. non-deterministic
- Episodic vs. non-episodic
- Static vs. dynamic
- Discrete vs. continuous

### Agents as Intentional Systems <sup>12</sup>

- When explaining human activity, it is often useful to make statements such as the following:
  - Janine took her umbrella **because she believed** it was going to rain.
  - Michael worked hard **because he wanted** to possess a PhD.
- These statements make use of a *folk psychology*, by which human behavior is predicted and explained through the attribution of *attitudes*, such as believing and wanting (as in the above examples), hoping, fearing, and so on.
- The attitudes employed in such folk psychological descriptions are called the *intentional* notions.

## Intentional Systems (2)

13

The philosopher Daniel Dennett coined the term *intentional system* to describe entities “whose behavior can be predicted by the method of attributing belief, desires and rational acumen”.

- Dennett identifies different ‘grades’ of intentional system:
  - ‘A *first-order* intentional system has beliefs and desires (etc.) but no beliefs and desires *about* beliefs and desires’.
  - ‘A *second-order* intentional system is more sophisticated; it has beliefs and desires (and no doubt other intentional states) about beliefs and desires (and other intentional states) — both those of others and its own’.

## Intentional Stance – Why?

14

- Is it legitimate or useful to attribute beliefs, desires, and so on, to computer systems?
- McCarthy argues that there are occasions when the intentional stance is appropriate:
  - ‘To ascribe beliefs, free will, intentions, consciousness, abilities, or wants to a machine is *legitimate* when such an ascription *expresses the same information* about the machine that it expresses about a person.
  - It is *useful* when the ascription *helps us understand the structure* of the machine, its past or future behavior, or how to repair or improve it.

## Intentional Stance (2)

15

- It is perhaps never *logically required* even for humans, but expressing reasonably briefly what is actually known about the state of a machine in a particular situation may require mental qualities or qualities isomorphic to them.
- Theories of belief, knowledge and wanting can be constructed for machines in a simpler setting than for humans, and later applied to humans.
- Ascription of mental qualities is *most straightforward* for machines of known structure such as thermostats and computer operating systems, but is *most useful when applied to entities whose structure is incompletely known*’.

## What Can be Described by the Intentional Stance?

- As it turns out, *more or less anything can*. . . consider a light switch:

‘It is perfectly coherent to treat a light switch as a (very cooperative) agent with the capability of transmitting current at will, who invariably transmits current when it believes that we want it transmitted and not otherwise; flicking the switch is simply our way of communicating our desires’. (Yoav Shoham)

- But most adults would find such a description absurd! Why is this?

16

## Does the Intentional Stance Buy us Anything?

17

- The answer seems to be that while the intentional stance description is consistent, it does not *buy us anything*, if we essentially understand the mechanism sufficiently to have a *simpler, mechanistic description* of its behavior (Yoav Shoham).
- Put crudely, the more we know about a system, the less we need to rely on animistic, intentional explanations of its behavior.
- But with *very complex systems*, a mechanistic, explanation of its behavior may not be practicable.
- *As computer systems become ever more complex, we need more powerful abstractions and metaphors to explain their operation — low level explanations become impractical. The intentional stance is such an abstraction.*

## Abstractions

18

- The *intentional notions* are thus *abstraction tools*, which provide us with a convenient and familiar way of describing, explaining, and predicting the behavior of complex systems.
- Remember: most important developments in computing are based on *new abstractions*:
  - procedural abstraction
  - abstract data types
  - objects
- Agents, and agents as intentional systems, represent a further, and increasingly powerful abstraction.
- So agent theorists start from the (strong) view of *agents as intentional systems*: whose *simplest consistent description requires the intentional stance*.

## Abstractions (2)

19

- Thus **the intentional stance is an abstraction tool** — a convenient way of talking about complex systems, which allows us to **predict and explain their behavior** without having to understand how the mechanism actually works.
- Now, much of computer science is concerned with looking for abstraction mechanisms (witness procedural abstraction, ADTs, objects, . . .)
- So **why not use the intentional stance as an abstraction tool in computing** — *to explain, understand, and, crucially, program computer systems?*
- This is an important argument in favor of agents.

## Connections to Other Areas of Computer Science

- We find that researchers from a more mainstream computing discipline have adopted a similar set of ideas...
- In distributed systems theory, *logics of knowledge* are used in the development of *knowledge based protocols*
- The rationale is that when constructing protocols, one often encounters reasoning such as the following:
 

```
IF    process i knows process j has
      received message m1
      THEN process i should send process j
           the message m2
```
- In DS theory, knowledge is *grounded* — given a precise interpretation in terms of the states of a process; we'll examine this point in detail later

20

## Nested Representations

21

- The intentional stance gives us the potential to specify systems that *include representations of other systems*.
- It is widely accepted that such nested representations are essential for agents that must **cooperate** with other agents.

## Post-Declarative Systems

22

- This view of agents leads to a kind of post-declarative programming:
  - in **procedural programming**, we say exactly what a system should do;
  - in **declarative programming**, we state something that we want to achieve, give the system general info about the relationships between objects, and let a built-in control mechanism (e.g., goal-directed theorem proving) figure out what to do;
  - with agents, we give a **very abstract specification of the system**, and let the control mechanism figure out what to do, knowing that it will act in accordance with some built-in theory of agency (e.g., the well-known Cohen-Levesque model of intention).

## Summary

23

- Characteristics of environments agents inhabit
- The *intentional stance* provides us with a familiar, non-technical way of *understanding and explaining* agents.
- Reference:
  - Wooldridge, Ch. 2
  - Reference AIMA, Ch. 2
  - Weiss, Ch. 1, 2

## Questions

24

