CPE/CSC 580: Intelligent Agents

Franz J. Kurfess

Computer Science Department
California Polytechnic State University
San Luis Obispo, CA, U.S.A.

Course Overview

- **Introduction**
  - Intelligent Agent, Multi-Agent Systems
  - Agent Examples
- **Agent Architectures**
  - Agent Hierarchy, Agent Design Principles
- **Reasoning Agents**
  - Knowledge, Reasoning, Planning
- **Learning Agents**
  - Observation, Analysis, Performance Improvement
- **Multi-Agent Interactions**
  - Agent Encounters, Resource Sharing, Agreements
- **Communication**
  - Speech Acts, Agent Communication Languages
- **Collaboration**
  - Distributed Problem Solving, Task and Result Sharing
- **Agent Applications**
  - Information Gathering, Workflow, Human Interaction, E-Commerce, Embodied Agents, Virtual Environments
- **Conclusions and Outlook**

Overview Agent Architectures

- Motivation
- Objectives
- Agent Design Principles
- Agent Hierarchy
- Intentional Systems
- Abstract Agent Architecture
- Reactive Agents
- Important Concepts and Terms
- Chapter Summary

Motivation

Objectives

Agent Design Principles

Autonomy

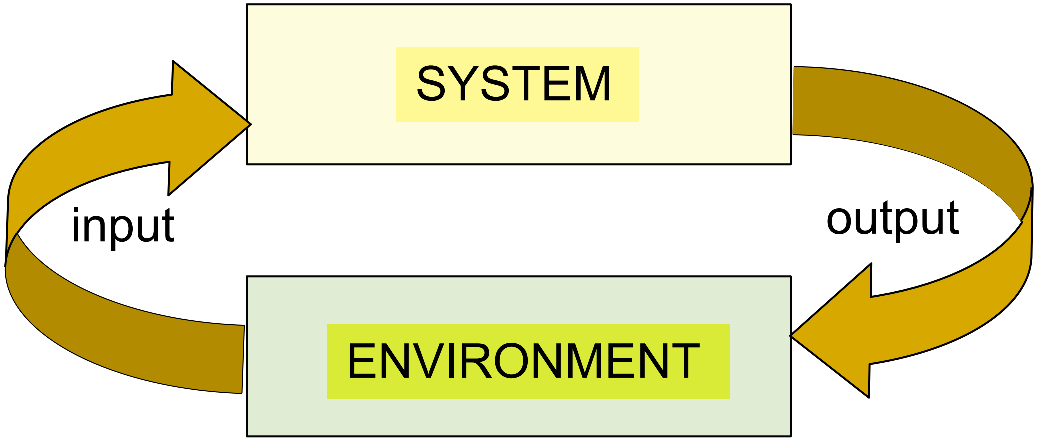
Embodiment

Belief, Desire, Intent

Social Behavior

Autonomous Agent

- An *agent* is
  - a computer system that is
  - capable of independent action on behalf of its user or owner



[Woolridge 2009]

Embodiment and Situatedness

- **An *embodied* agent has a physical manifestation**
  - often also called a robot
  - software agents typically are not embodied
- **Agents are situated in an environment**
  - often also referred to as context

Belief, Desire, Intention (BDI)

- software model developed for the design and programming of intelligent agents
- implements the principal aspects of Michael Bratman's theory of human practical reasoning
-

http://en.wikipedia.org/wiki/Belief%E2%80%93desire%E2%80%93intention_software_model

Beliefs

- **represent the informational state of the agent**
  - beliefs about the world (including itself and other agents)
- **beliefs can include inference rules**
  - for the generation of new beliefs
- **the term belief is used instead of knowledge**
  - expresses the subjective nature
  - may change over time

Desires

- **represent the motivational state of the agent**
  - situations that the agent would like to achieve
- **goals are desires adopted for active pursuit**
  - sets of multiple goals should be consistent
  - sets of desires can be inconsistent

Intentions

- **represent the deliberative state of the agent**
  - the agent has chosen to do something
- **intentions are desires to which the agent has committed**
  - to some extent
- **a plan is a sequences of actions to achieve an intention**
- **an event is a trigger for reactive activity by an agent**

Social Ability

- The real world is a *multi*-agent environment: we cannot go around attempting to achieve goals without taking others into account
- Some goals can only be achieved with the cooperation of others
- Similarly for many computer environments: witness the Internet
- *Social ability* in agents is the ability to interact with other agents (and possibly humans) via some kind of *agent-communication language*, and perhaps cooperate with others

[Woolridge 2009]

Agent Hierarchy

Reflex Agent

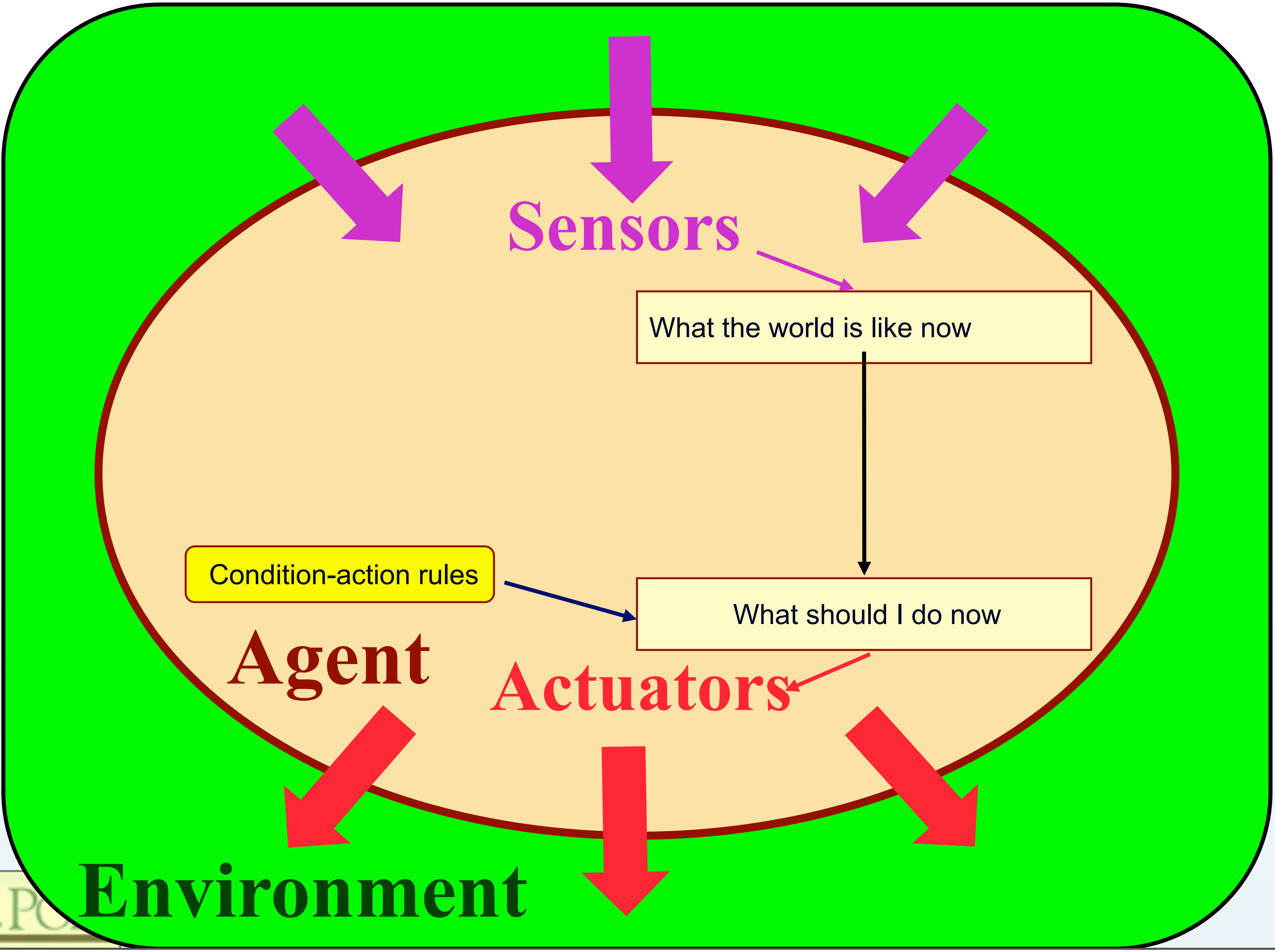
Model-Based Agent

Goal/Utility-Based Agent

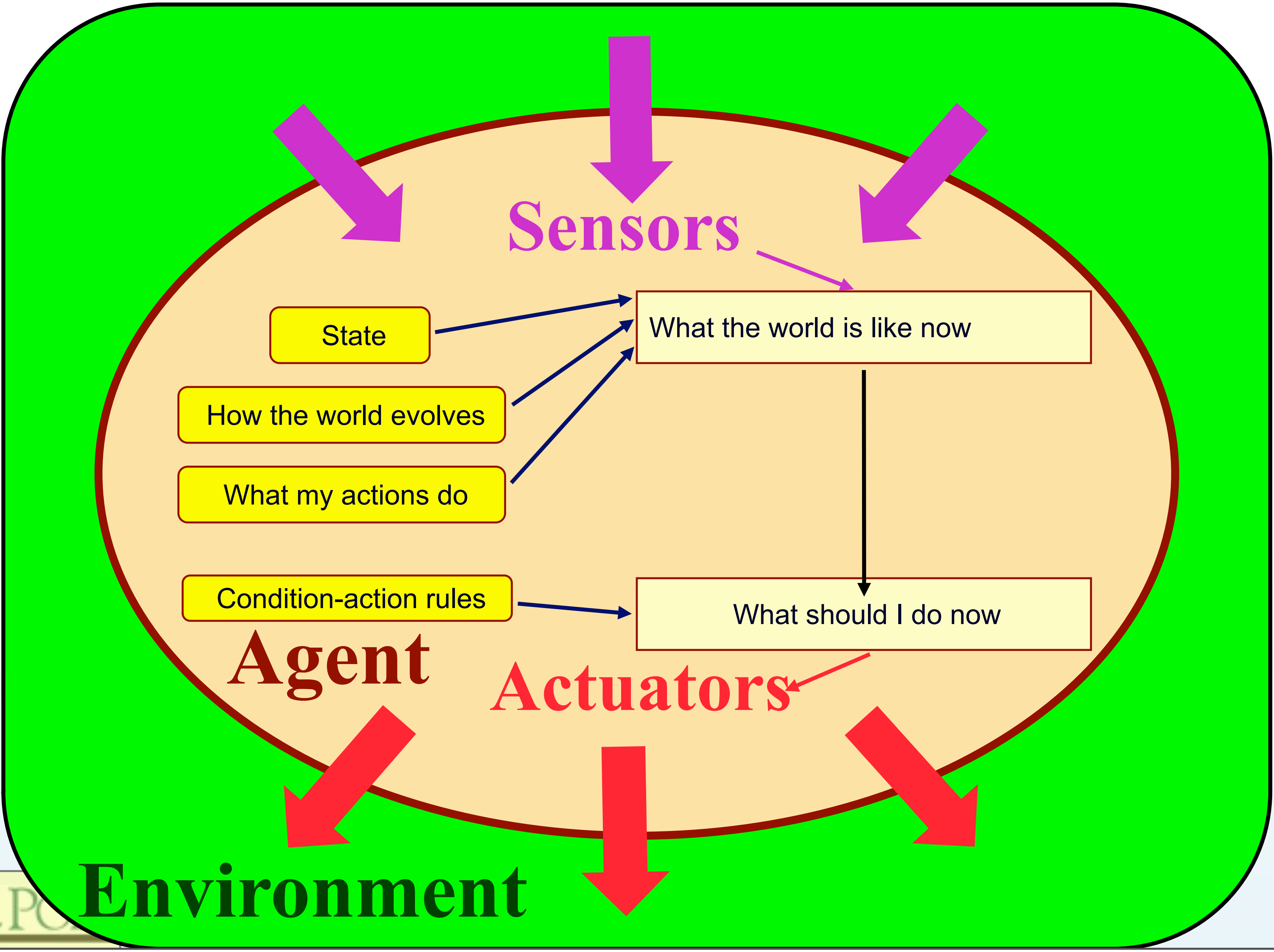
Learning Agent

Reasoning Agent

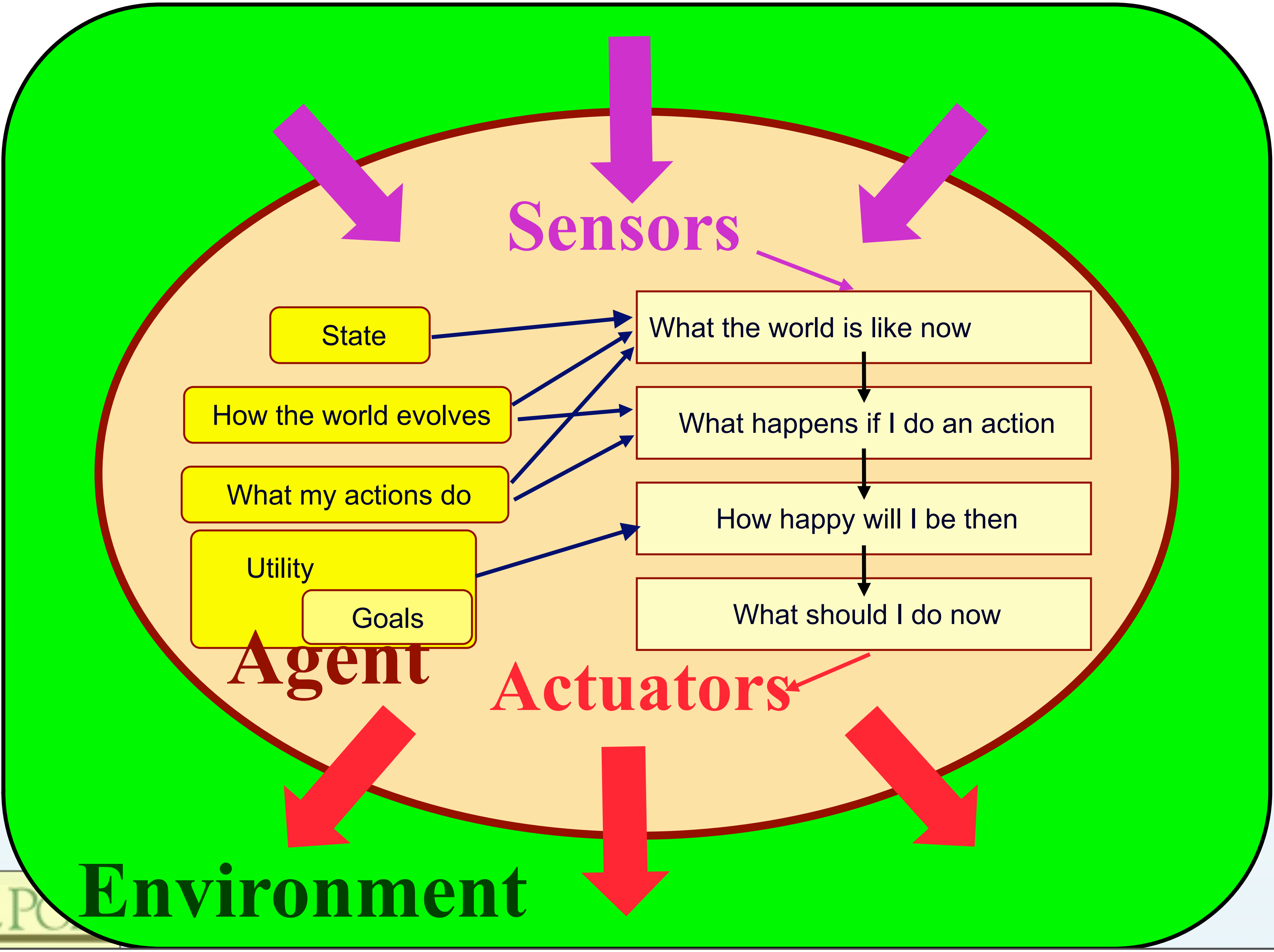
Reflex Agent Diagram 2



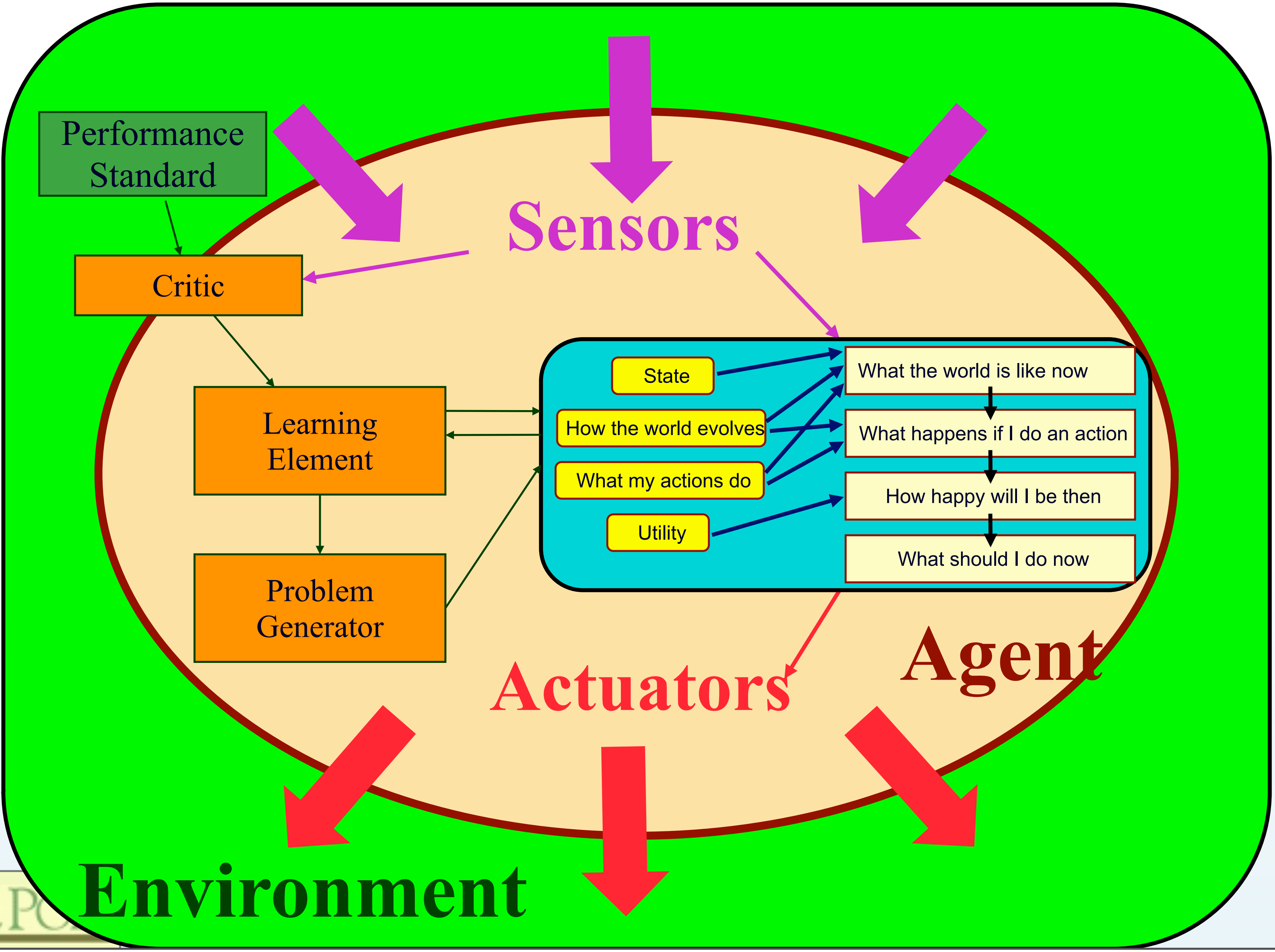
Model-Based Reflex Agent Diagram



Utility-Based Agent Diagram



Learning Agent Diagram



Intentional Systems

Agents as Intentional Systems
The Need for Abstraction
Representational Flexibility
Post-Declarative Systems

Agents as Intentional Systems

- 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.
- Human behavior is predicted and explained through the attribution of attitudes,
  - such as believing and wanting, hoping, fearing, ...
- The attitudes employed in such folk psychological descriptions are called the intentional notions

[Woolridge 2009]

Agents as Intentional Systems

- The philosopher Daniel Dennett coined the term intentional system
  - describes entities ‘whose behavior can be predicted by the method of attributing belief, desires and rational acumen’
- different ‘grades’ of intentional systems:
  - first-order intentional system has beliefs and desires but no beliefs and desires about beliefs and desires.
  - second-order intentional system is more sophisticated;
    - it has beliefs and desires about beliefs and desires
    - also has other intentional states
      - together with beliefs and desires about those pther intentional states
    - refers to states of others and its own

[Woolridge 2009]

Agents as Intentional Systems

- The answer seems to be that while the intentional stance description is consistent,

  > . . . it does not *buy us anything*, since 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*

[Woolridge 2009]

Intentional Systems as Abstraction

- the more we know about a system, the less we need to rely on animistic, intentional explanations of its behavior
- with very complex systems, a mechanistic, explanation of its behavior may not be practicable
- intentions can be used to describe complex systems at a higher level of abstraction
  - to express aspects like
    - autonomy
    - goals
    - self-preservation
    - social behavior

[Woolridge 2009]

Agents as Intentional Systems

- additional points in favor of this idea:
  - Characterizing Agents:
    - provides a familiar, non-technical way of *understanding & explaining* agents
  - Nested Representations:
    - offers the potential to specify systems that *include representations of other systems*
    - widely accepted that such nested representations are essential for agents that must cooperate with other agents

[Woolridge 2009]

Post-Declarative Systems

- 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,
    - let a built-in control mechanism figure out what to do
    - e.g., goal-directed theorem proving
- intentional agents
  - very abstract specification of the system
  - let the control mechanism figure out what to do
    - knowing that it will act in accordance with some built-in theory of agency

[Woolridge 2009]

Abstract Agent Architecture

Environment, States
Actions, Runs
State Transformations
Agent as Function
System

Abstract Architecture for Agents

- Assume the environment may be in any of a finite set E of discrete, instantaneous states:

$$E = \{e, e', \ldots\}.$$

- Agents are assumed to have a repertoire of possible actions available to them, which transform the state of the environment:

$$Ac = \{\alpha, \alpha', \ldots\}$$

- A *run*, r, of an agent in an environment is a sequence of interleaved environment states and actions:

$$r : e_0 \xrightarrow{\alpha_0} e_1 \xrightarrow{\alpha_1} e_2 \xrightarrow{\alpha_2} e_3 \xrightarrow{\alpha_3} \cdots \xrightarrow{\alpha_{u-1}} e_u$$

[Woolridge 2009]

Abstract Architecture for Agents

- Let:
  - R be the set of all such possible finite sequences (over E and Ac)
  - R^{Ac} be the subset of these that end with an action
  - R^{E} be the subset of these that end with an environment state

[Woolridge 2009]

State Transformer Functions

- A *state transformer* function represents behavior of the environment:

$$\tau : \mathcal{R}^{Ac} \rightarrow \wp(E)$$

- Note that environments are…
  - *history dependent*
  - *non-deterministic*
- If $\tau(r)=\emptyset$, then there are no possible successor states to r. In this case, we say that the system has *ended* its run
- Formally, we say an environment Env is a triple $Env = \langle E, e_0, \tau \rangle$ where: E is a set of environment states, $e_0 \in E$ is the initial state, and τ is a state transformer function

[Woolridge 2009]

Agents

- Agent is a function which maps runs to actions:

$$Ag : \mathcal{R}^E \to Ac$$

An agent makes a decision about what action to perform based on the history of the system that it has witnessed to date. Let AG be the set of all agents

[Woolridge 2009]

Systems

Franz J. Kurfess

Computer Science Department
California Polytechnic State University
San Luis Obispo, CA, U.S.A.

Systems

$$(e_0, \alpha_0, e_1, \alpha_1, e_2, \ldots)$$

Franz J. Kurfess

Co

Califo

S

$$e_u \in \tau((e_0, \alpha_0, \ldots, \alpha_{u-1})) \quad \text{where}$$

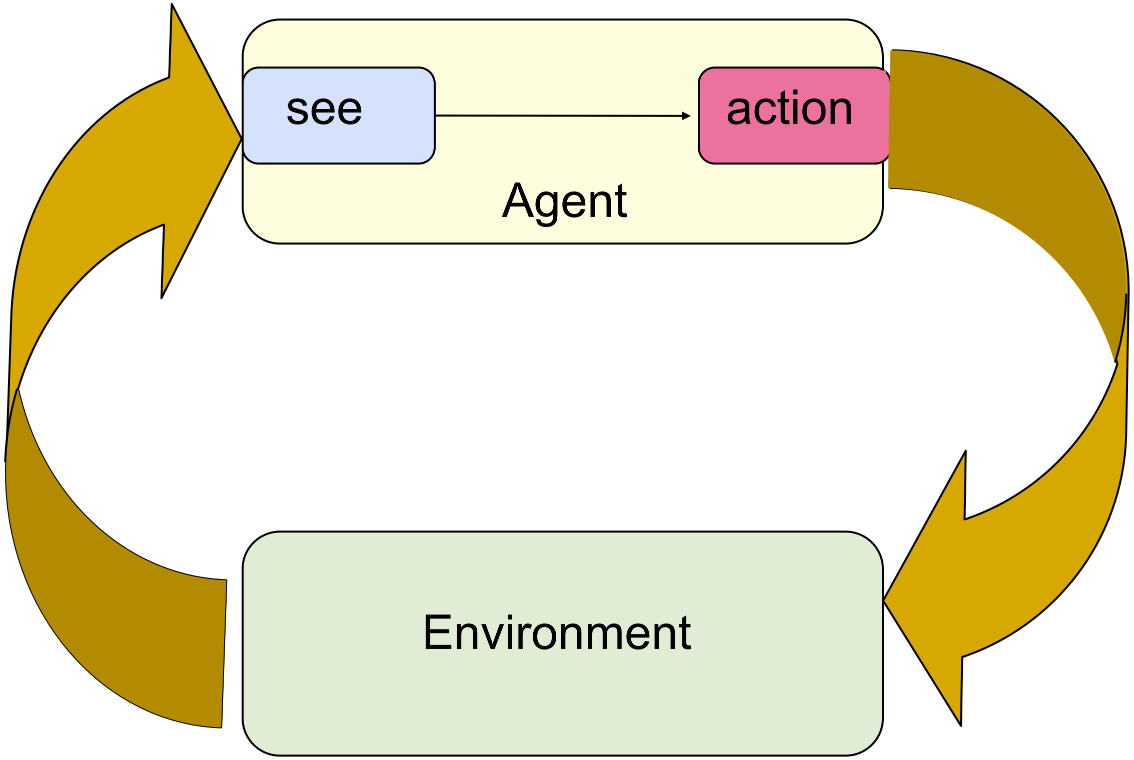
$$\alpha_u = Ag((e_0, \alpha_0, \ldots, e_u))$$

Reactive Agents

Perception
Agents with State
Tasks
Utility Functions

Perception

- Now introduce *perception* system:



[Woolridge 2009]

Perception

- the *see* function is the agent's ability to observe its environment,
- the *action* function represents the agent's decision making process
- *Output* of the *see* function is a *percept*:

$$see : E \to Per$$

  - maps environment states to percepts

- *action* is now a function

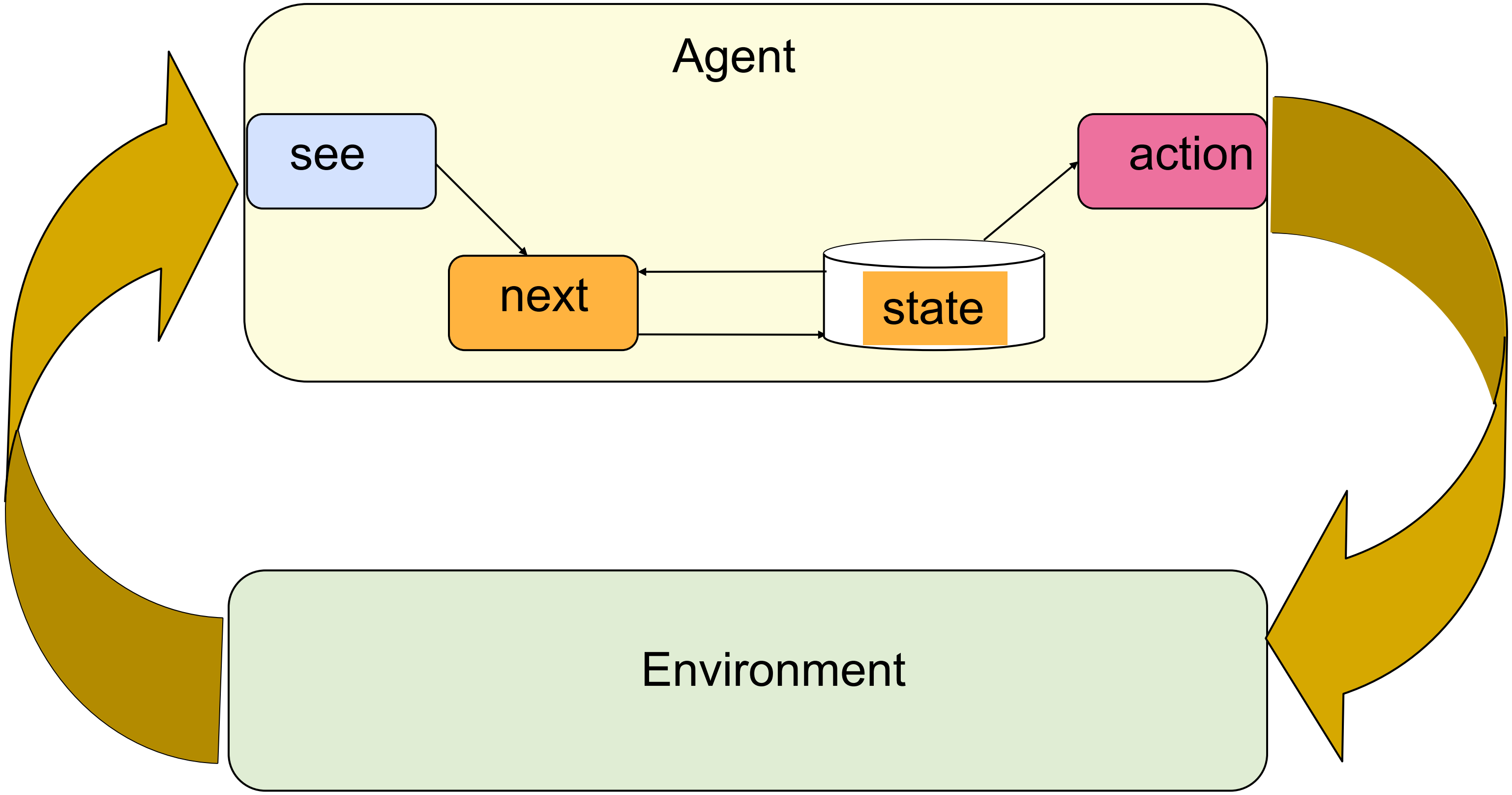
$$action : Per^{*} \to A$$

  - maps sequences of percepts to actions

[Woolridge 2009]

Agents with State

- We now consider agents that *maintain state*:



[Woolridge 2009]

Agents with State

- internal data structure
  - typically used to record information about the environment state and history.
- let I be the set of all internal states of the agent
- the perception function *see* for a state-based agent is unchanged:

$$see : E \rightarrow Per$$

- the action-selection function *action* is now defined as a mapping from internal states to actions:

$$action : I \rightarrow Ac$$

- An additional function *next* is introduced:

$$next : I \times Per \rightarrow I$$

  - maps an internal state and percept to an internal state

[Woolridge 2009]

Agent Control Loop

1. Agent starts in some initial internal state i_0
2. Observes its environment state e, and generates a percept $see(e)$
3. Internal state of the agent is then updated via $next$ function, becoming $next(i_0, see(e))$
4. The action selected by the agent is $action(next(i_0, see(e)))$
5. Goto 2

[Woolridge 2009]

Tasks for Agents

- agents carry out *tasks* for users
  - tasks must be *specified* by users
- tell agents what to do *without* telling them how to do it

[Woolridge 2009]

Utility Functions over States

- associate *utilities* with individual states
  - the task of the agent is then to bring about states that maximize utility
- a task specification is a function

$$u : E \rightarrow \text{ú}$$

  - associates a real number with every environment state

[Woolridge 2009]

Utility Functions over States

- value of a *run*
  - minimum utility of state on run?
  - maximum utility of state on run?
  - sum of utilities of states on run?
  - average?
- disadvantage:
  - difficult to specify a *long term* view when assigning utilities to individual states
    one possibility: a *discount* for states later on

[Woolridge 2009]

Utilities over Runs

- another possibility
  - assigns a utility not to individual states, but to runs themselves:

$$u : \mathrm{R} \to \text{ú}$$

  - inherently *long term* view
- other variations
  - incorporate probabilities of different states emerging
- difficulties with utility-based approaches:
  - where do the numbers come from?
  - humans don’t think in terms of utilities
  - hard to formulate tasks in these terms

[Woolridge 2009]

Summary Agent Architectures

Important Concepts and Terms

- agent
- agent society
- architecture
- deduction
- environment
- hybrid architecture
- intelligence
- intention
- multi-agent system
- reactivity
- subsumption
-