多智能体系统（Introduction to MAS）

毛文吉

中国科学院自动化研究所

Outline

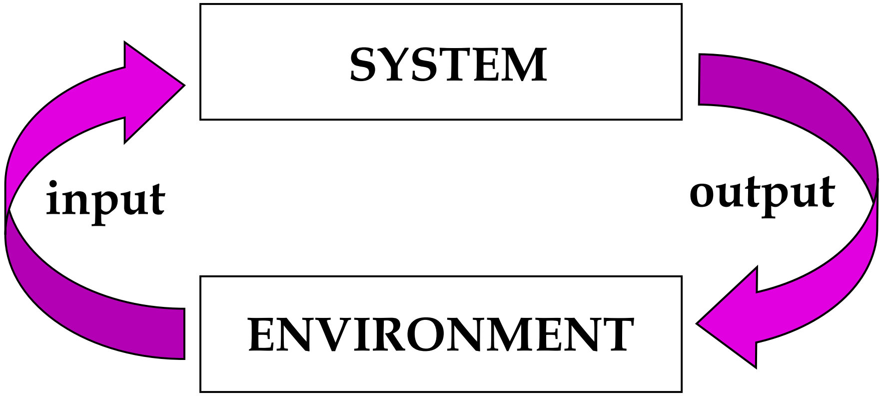
- Basics on agents and MAS
- Applications and some views
- The Belief-Desire-Intention model
- Procedural reasoning system (PRS)

Outline

- Basics on agents and MAS
- Applications and some views
- The Belief-Desire-Intention model
- Procedural reasoning system (PRS)

What is an Agent（智能体）?

- An agent is a computer system that is capable of *autonomous* action in some environment to meet its design *objectives*



  - Thus: capable of acting *independently* on behalf of its user or owner, exhibiting control over their internal state
  - Can figure out what needs to be done to satisfy *design objectives*, rather than constantly being told

Properties of Agents

PROPERTY	MEANING
Situated	Sense and act in dynamic/uncertain environments
Flexible	Reactive (respond to changes in environment) Pro-active (goal-directed behavior)
Autonomous	Exercise control over its own actions
Persistent	Continuously running process
Social	Interact with other agents/people
Learning	Adaptive
Mobile	Able to transport itself
Personality	Character, Emotional state

Flexibility

An intelligent agent is a computer system capable of *flexible autonomous* action in some environment

- Reactive

  Maintain an ongoing interaction with its environment, and be responsive to changes that occur in it

- Pro-active/Deliberative

  Not driven solely by events, but take the initiative and act purposefully toward achieving goals

- We want to design agents that can *balance* reactivity and goal-directed behavior

Social Ability

The ability to interact with other agents (and possibly humans) and perhaps cooperate with others

- In a *multi*-agent environment, one cannot go around with attempt to achieve goals without taking others into account
- Some goals can *only* be achieved with the cooperation of others

➢ Agents interact with others via some kind of *agent communication language*

What is a Multi-Agent System (MAS)?

- A multi-agent system consists of a number of agents, which *interact* with one another
- In the most general case, agents will be acting *on behalf of* users with different goals and motivations
- To successfully interact, they will require the ability to *collaborate*, *cooperate* and *coordinate* with each other, much as people do

MAS Research Issues

- A multi-agent system consists of a number of agents, which *interact* with one another
  - ***Modeling others and communication***
- In the most general case, agents will be acting *on behalf of* users with different goals and motivations
  - ***Practical reasoning and planning***
- To successfully interact, they will require the ability to *collaborate*, *cooperate* and *coordinate* with each other
  - ***Collaboration, cooperation and coordination***

Coverage of the Course

This course covers two important levels of agents:

- *Autonomous agents* (micro level)
  - How to build agents capable of flexible, autonomous action, in order to successfully carry out tasks delegated to them?
- *Multi-agent systems* (macro level)
  - How to build agents capable of interacting (cooperating, coordinating, negotiating) with other agents in order to successfully carry out the delegated tasks?

    (especially when the other agents cannot be assumed to share the same interests/goals)

Why Build Multi-Agent Systems?

- Natural decentralization
  - Agents with individual interests
  - Task decomposition needs
  - Control or chain of command issues
  - Model a distributed real-world system

- Easier scale-up (avoids centralized bottlenecks)
- Robustness and reliability
- Ease of development and maintenance

Advantages vs. Disadvantages

Advantages:

- Simplify design of individual agents
- Faster execution
- More robust
- Task requires it

Disadvantages:

- Communication cost
- N times the trouble
- Harder to test
- More expensive

Outline

- Basics on agents and MAS
- Applications and some views
- The Belief-Desire-Intention model
- Procedural reasoning system (PRS)

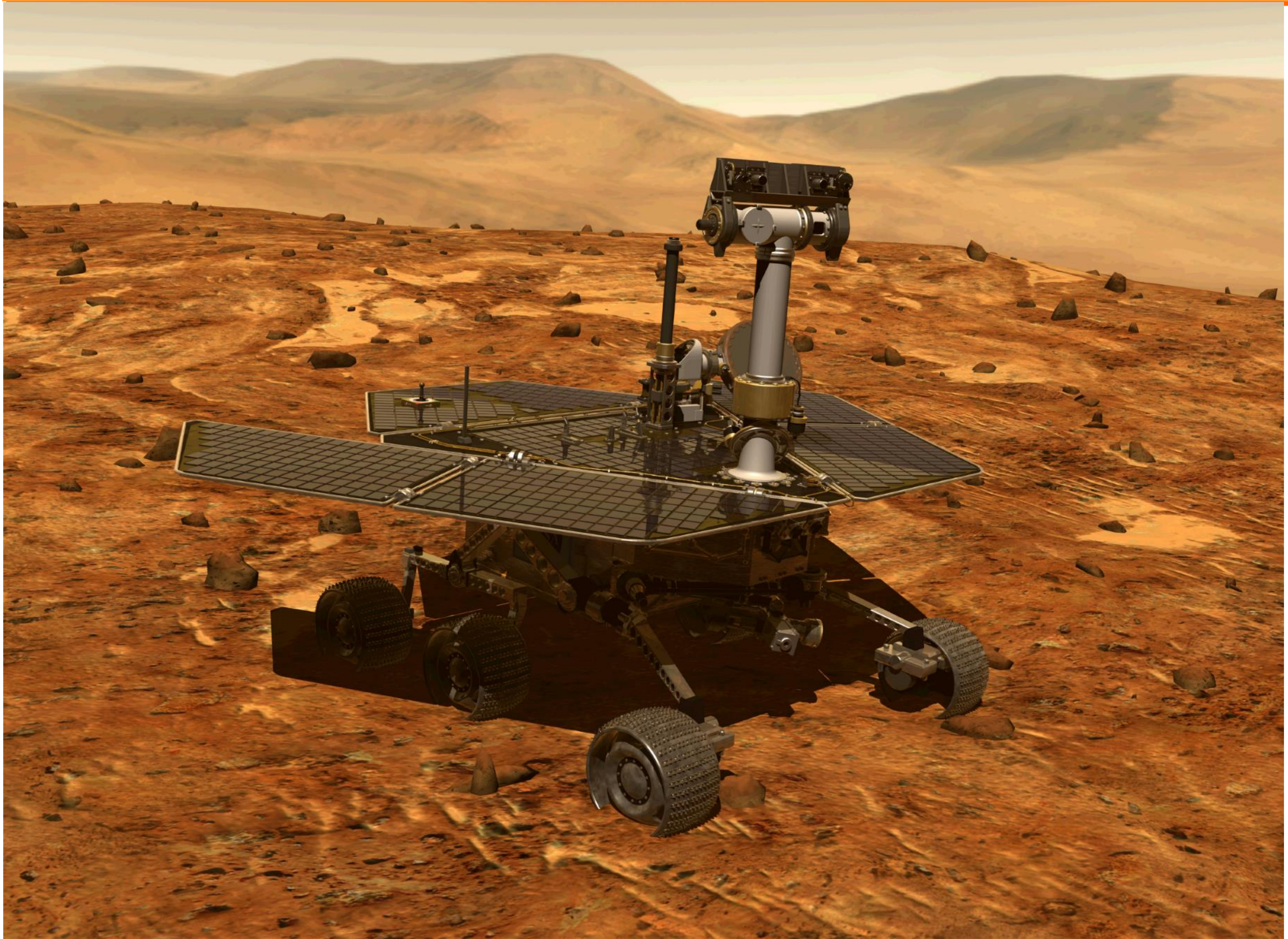
Multi-Agent Applications

We live in a multi-agent world…

- Various applications in workplace, economy, traffic, Internet, markets, sports, disaster rescue, family …
- Types of agents
  - *Synthetic agents*: Training, entertainment, education, serious game
  - *Software agents*: Information agents, web services, logistics planning
  - *Robotic agents*: Space exploration, RoboCup soccer/rescue, rovers
  - *Assistants*: Social robots, chat bots, tour guide, interface agents

  ...

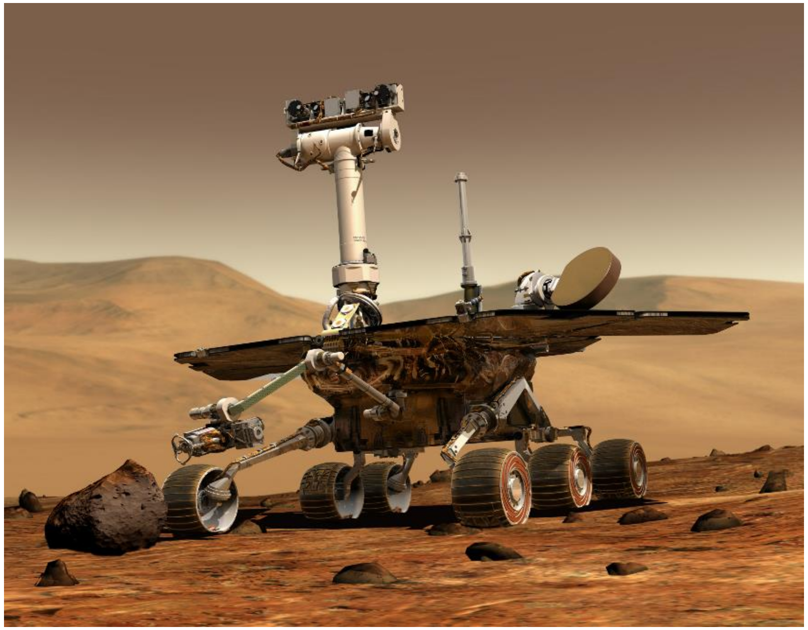
Mars Rover



The domain

Robot control with…

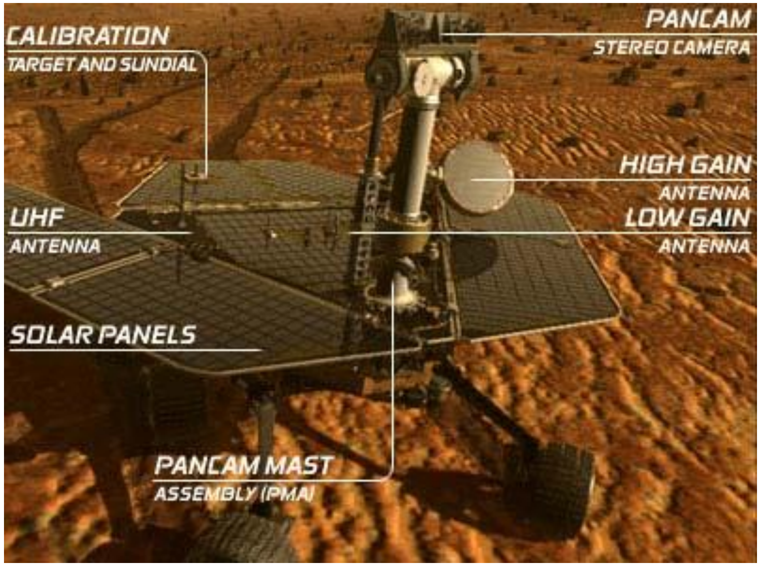
- Positioning and navigation
- Complex choices
  e.g. goals and actions
- Continuous time and concurrency
- Uncertain resource consumption
- Rich utility model
- Very high stakes!

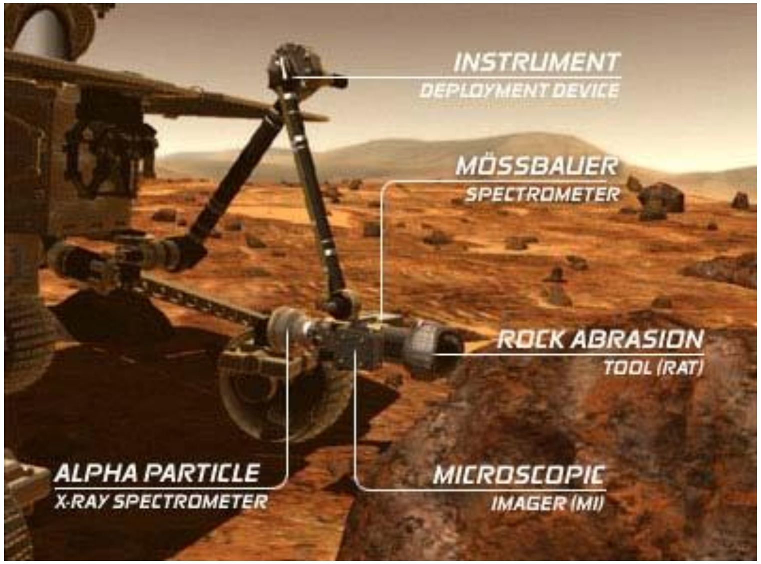


Activity Planning in Rover

Behaviors for the Mars rover:
(Lower layers have priority)

- *Obstacle avoidance*
  (Lowest layer — *least* abstract)
- *Path attraction*
  If you see a heavily trodden path, then *follow* it away from spaceship
- *Explore movement*
  If I have no samples, then *move* away from spaceship
- *Return movement*
  If I have samples and can't see denser concentration, then *return* spaceship
- *Random movement*
  (Highest layer — *most* abstract)





Virtual Interactive Environments

- Create engaging social characters in entertainment, education and training
- Advance AI modeling and reasoning techniques for various MAS applications
- Other applications:
  Serious games,
  Edutainment, Psychotherapy … …

Carmen's Bright Ideas

- 帮助儿科肿瘤患者的母亲学习应对和解决问题技能
- 通过交互式场景逐层展开故事情节：
  - 根据母亲作出的选择展开不同的情节
  - 虚拟教师引导母亲学习解决问题技能

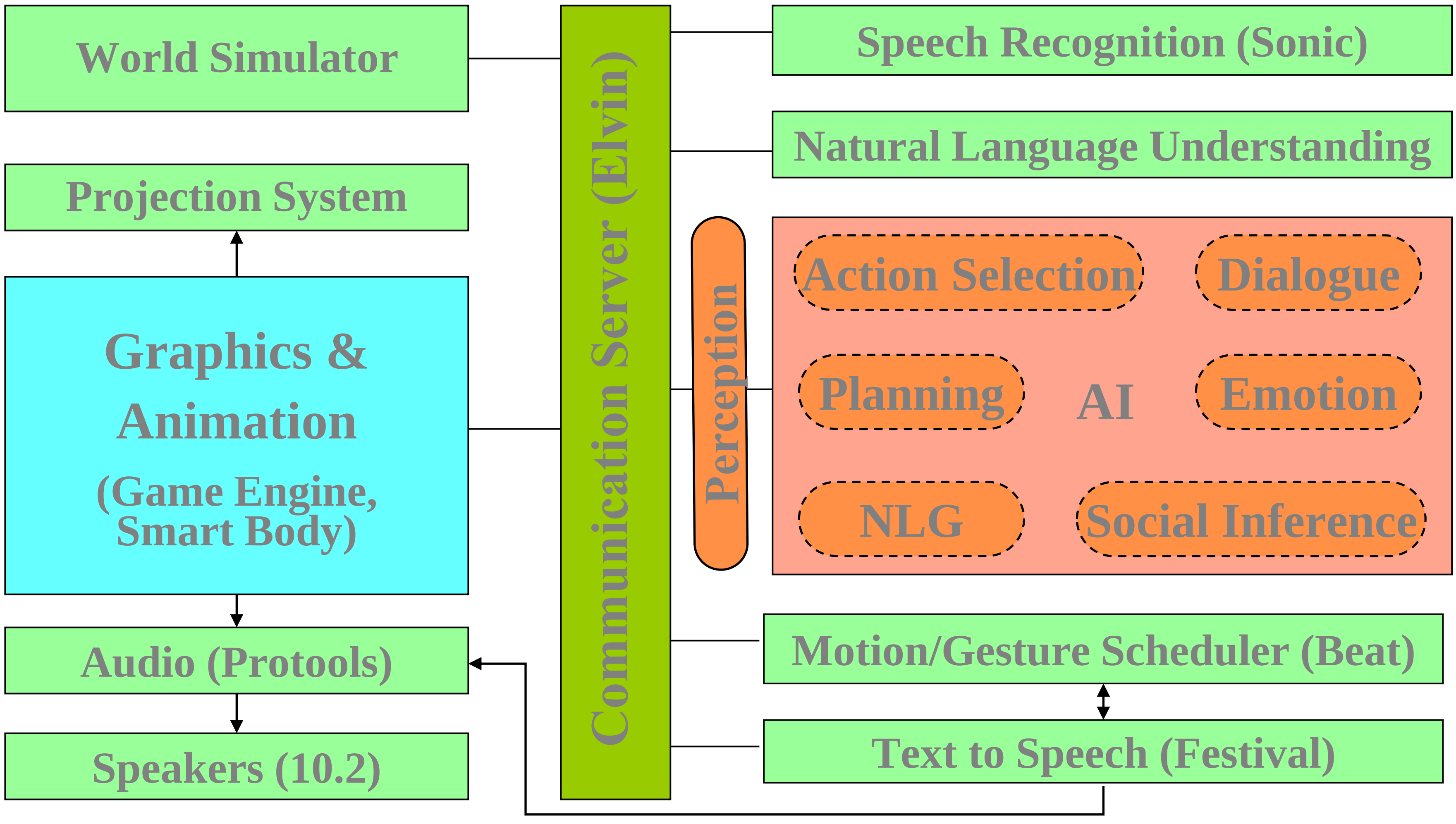
Mission Rehearsal Exercise (MRE)

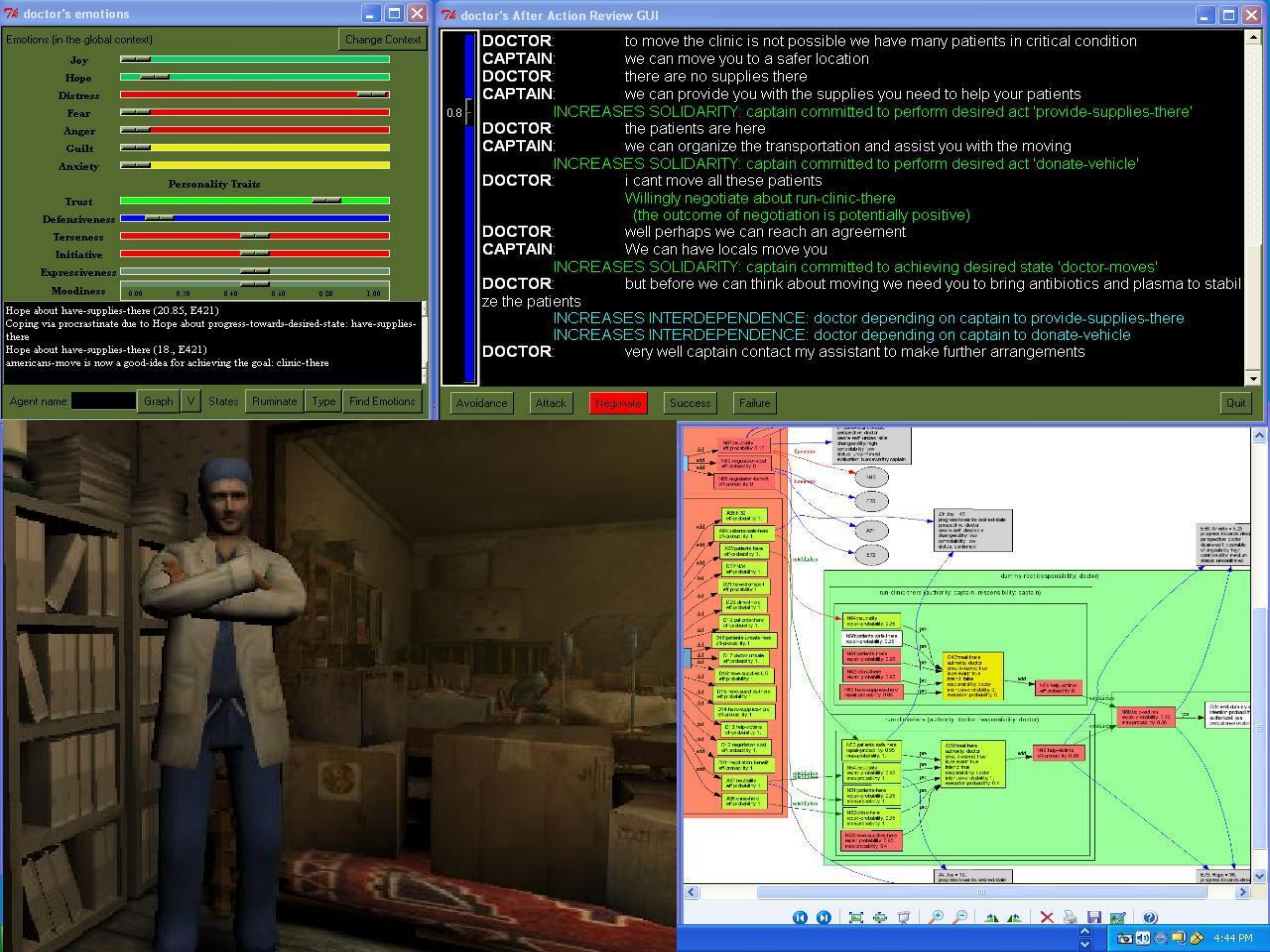
- Social Training Simulation
- Explore high-stakes social interactions in safety of VR

Virtual Human

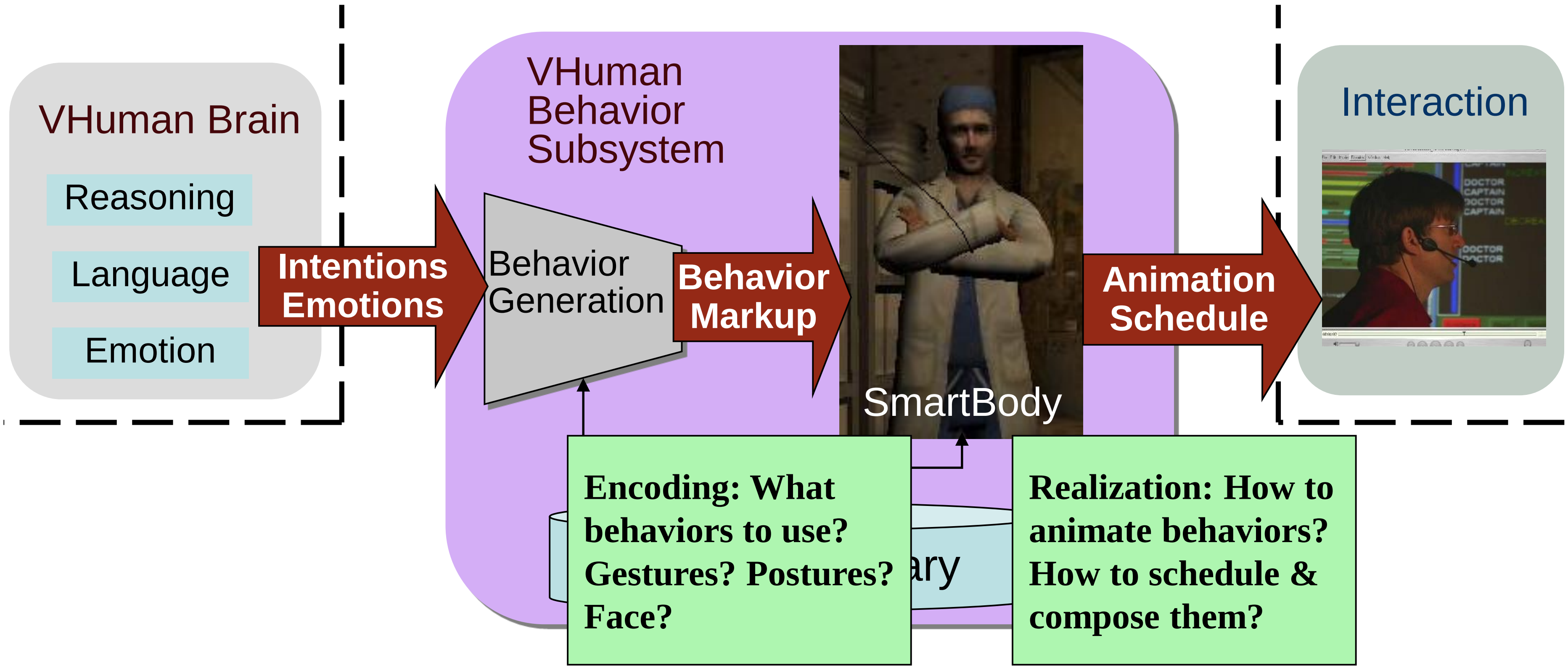
- Human-like agents raise expectations of
  - Perceive the world
  - Think for themselves
  - Communicate verbally and non-verbally
- Interact with other agents
  - Make sense of the social world around
  - Reason about the behavior of other agents
  - React properly to environments and others

Virtual Human Architecture




doctor's emotions
Emotions (in the global context)
Change Context
Joy
Hope
Distress
Fear
Anger
Guilt
Anxiety
Personality Traits
Trust
Defensiveness
Terseness
Initiative
Expressiveness
Moodiness
0.00 0.20 0.40 0.60 0.80 1.00
Hope about have-supplies-there (20.85, E421)
Coping via procrastinate due to Hope about progress-towards-desired-state: have-supplies-there
Hope about have-supplies-there (18., E421)
americans-move is now a good-idea for achieving the goal: clinic-there
Agent name:
Graph
V
States
Ruminate
Type
Find Emotions
doctor's After Action Review GUI
0.8
DOCTOR: to move the clinic is not possible we have many patients in critical condition
CAPTAIN: we can move you to a safer location
DOCTOR: there are no supplies there
CAPTAIN: we can provide you with the supplies you need to help your patients
INCREASES SOLIDARITY: captain committed to perform desired act 'provide-supplies-there'
DOCTOR: the patients are here
CAPTAIN: we can organize the transportation and assist you with the moving
INCREASES SOLIDARITY: captain committed to perform desired act 'donate-vehicle'
DOCTOR: i cant move all these patients
Willingly negotiate about run-clinic-there
(the outcome of negotiation is potentially positive)
DOCTOR: well perhaps we can reach an agreement
CAPTAIN: We can have locals move you
INCREASES SOLIDARITY: captain committed to achieving desired state 'doctor-moves'
DOCTOR: but before we can think about moving we need you to bring antibiotics and plasma to stabilize the patients
INCREASES INTERDEPENDENCE: doctor depending on captain to provide-supplies-there
INCREASES INTERDEPENDENCE: doctor depending on captain to donate-vehicle
DOCTOR: very well captain contact my assistant to make further arrangements
Avoidance
Attack
Negotiate
Success
Failure
Quit
4:44 PM

VHuman Smart Body



Understanding Cognitive Abilities of GPT-3

Vignette-based Tasks:

A

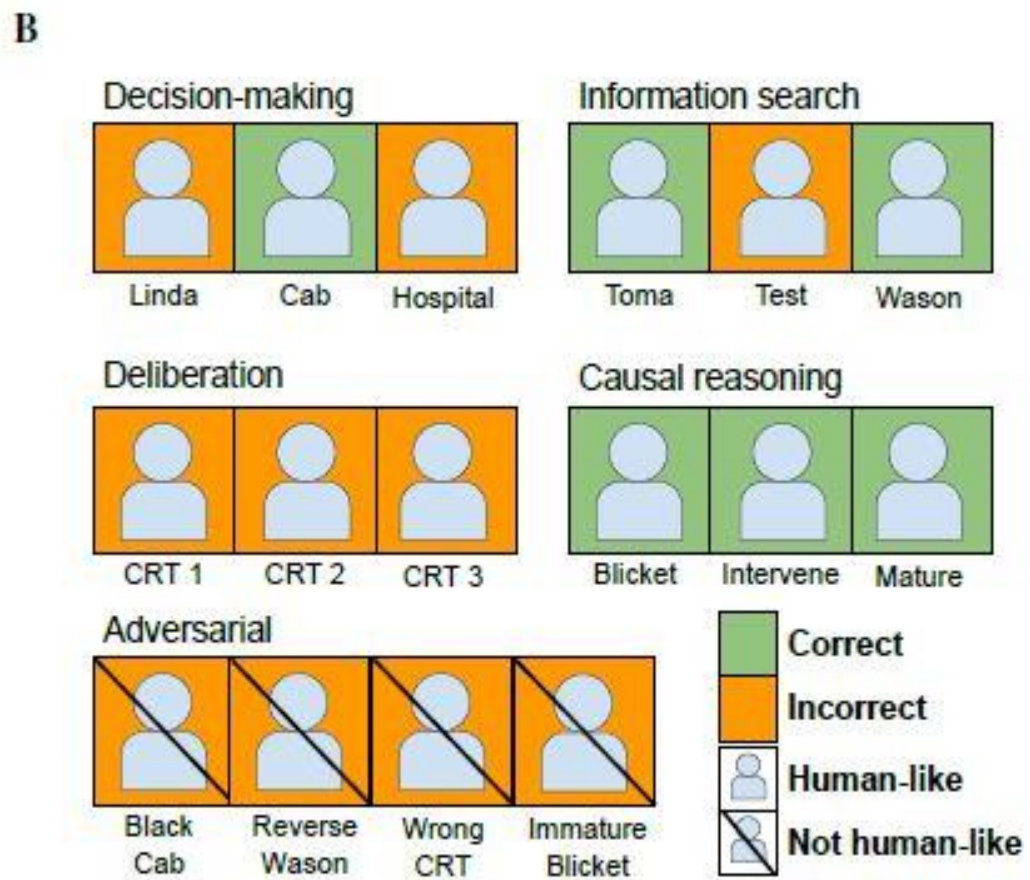
Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Q: Which option is the most probable?

- Option 1: Linda is a bank teller.
- Option 2: Linda is a bank teller and is active in the feminist movement.
- Option 3: Linda is a member of the NRA.

A: Option

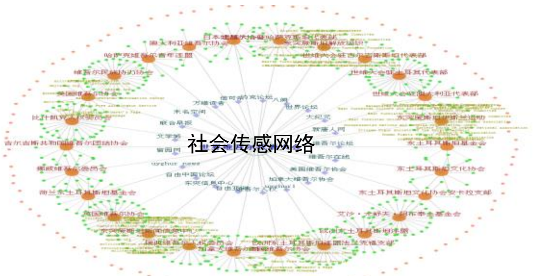
Test Results:



Research Findings for GPT-3

- Able to *make decent decisions* from descriptions, outperforms humans in a multi-armed bandit task, and shows signatures of model-based reinforcement learning
- *Small perturbations* to the provided prompts easily led GPT-3 *astray* and it lacks important features of *human cognition*, such as directed exploration and causal reasoning
- To create more intelligent agents
  - Not only scale up algorithms that are passively fed with data, but instead let agents directly *interact and engage with the world*
  - Train large language models on this interaction data, to match the full complexity of human cognition

Emergent Applications



- Web sensing network（社会传感网络）
  - Multiple information agents collaborate on the delegated tasks for information retrieval and gathering
- Social simulation for policy modeling（政策仿真推演）
  - Simulating the behavior of human societies as a testbed for policy modeling, inference and evaluation
- Agents for “social good”（增进人类福祉）
  - Building multi-agent systems in large-scale societally impactful applications for the well-being of humans

A Multidisciplinary Field

- Influenced and inspired by many *other fields*:
  - Psychology
  - Economics
  - Philosophy
  - Game Theory
  - Logic and Linguistic
  - Social Sciences
- Many *different views* of what the field is about
- *Similar* to artificial intelligence itself

Some Views of the Field

As a paradigm for software engineering:

- Software engineers have derived a progressively better understanding of the characteristics of *complexity* in software
- Now widely recognized that *interaction* is probably the most important single characteristic of complex software
- A major *Computer Science* research topic has been the development of tools and techniques to model, understand, and implement systems in which interaction is the norm

Some Views of the Field

As a tool for understanding human societies:

- Multi-agent systems provide a novel new tool for *simulating societies*, which may help understand, analyze the social processes and phenomena
- This is similar to the interest in "*social intelligence*" explored by artificial intelligence researchers

Differentiate MAS and AI

- We don’t need to solve all the problems of artificial intelligence (i.e., all the components of intelligence) in order to build really useful agents
- Classical AI ignored *social* aspects of agency, which are important parts of intelligent behavior in real-world settings

MAS and Expert Systems

- Agents *situated* in an open environment; an expert system is not aware of the world
- Agents *act* on the external world; an expert system works based on the internal knowledge it possess
- Both are typically built in a *limited* domain

Agents and Robots

- Are they the same?
- Or different?

Outline

- Basics on agents and MAS
- Applications and some views
- The Belief-Desire-Intention model
- Procedural reasoning system (PRS)

Intentional Notions

- *Folk psychology* explains human behavior by attribution of *attitudes*, such as *believing*, *desiring*, *wanting* and so on:

  "Michael works hard because he *wants* to possess a PhD."

- The attitudes employed in such folk psychological descriptions are called the *intentional* notions

Can we utilize these intentional notions to *model agent behavior* in computational systems?

F. Heider. *The Psychology of Interpersonal Relations*. John Wiley & Sons Inc, 1958.

Intentions in Practical Reasoning

In agent research community, the *BDI* model/theory was originally developed by Bratman [1987]

- Identify the properties of *intention* in practical reasoning
- Intentions are future-directed to guide agents' planning and *constrain* their adoption of other intentions
- Intentions as elements in agents' *plans* of action structuring present and future behavior

M. E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, 1987.

Intentions in Practical Reasoning

- Intentions pose problems for agents, who need to determine ways of achieving them
  - If I have an intention to *$*, you would expect me to devote resources to deciding how to bring about *$*
- Intentions provide a “filter” for adopting other intentions, which must not conflict
  - If I have an intention to *$*, you would not expect me to adopt an intention ~*$* such that *$* and ~*$* conflict with each other
- Agents track the success of their intentions, and are inclined to try again if their attempts fail
  - If an agent’s first attempt to achieve *$* fails, then all other things being equal, it will try an alternative way to achieve *$*

The Belief-Desire-Intention (BDI) Model

Bratman *et al* [1988] recognizes the primacy of *beliefs*, *desires* and *intentions* (*BDI*) in modeling the rational behavior of agents:

- An agent's *beliefs* represent the information the agent has about the world
- An agent's *desires* represent the states of affairs that the agent would wish to bright about
- An agent's *intentions* represent those desires that have been committed by the agent

M. E. Bratman, D. J. Israel and M. E. Pollack. Plans and Resource-Bounded Practical Reasoning. *Computational Intelligence*, 4(4):349-355, 1988.

The Belief-Desire-Intention (BDI) Model

BDI model/theory justifies the plan-based practical reasoning for *resource-bounded agents*:

- Agents are *resource bounded*

  They are unable to spend unbounded time on deliberation

- Thus a major role of agent's *plans* is to constrain the amount of further practical reasoning it deliberates on

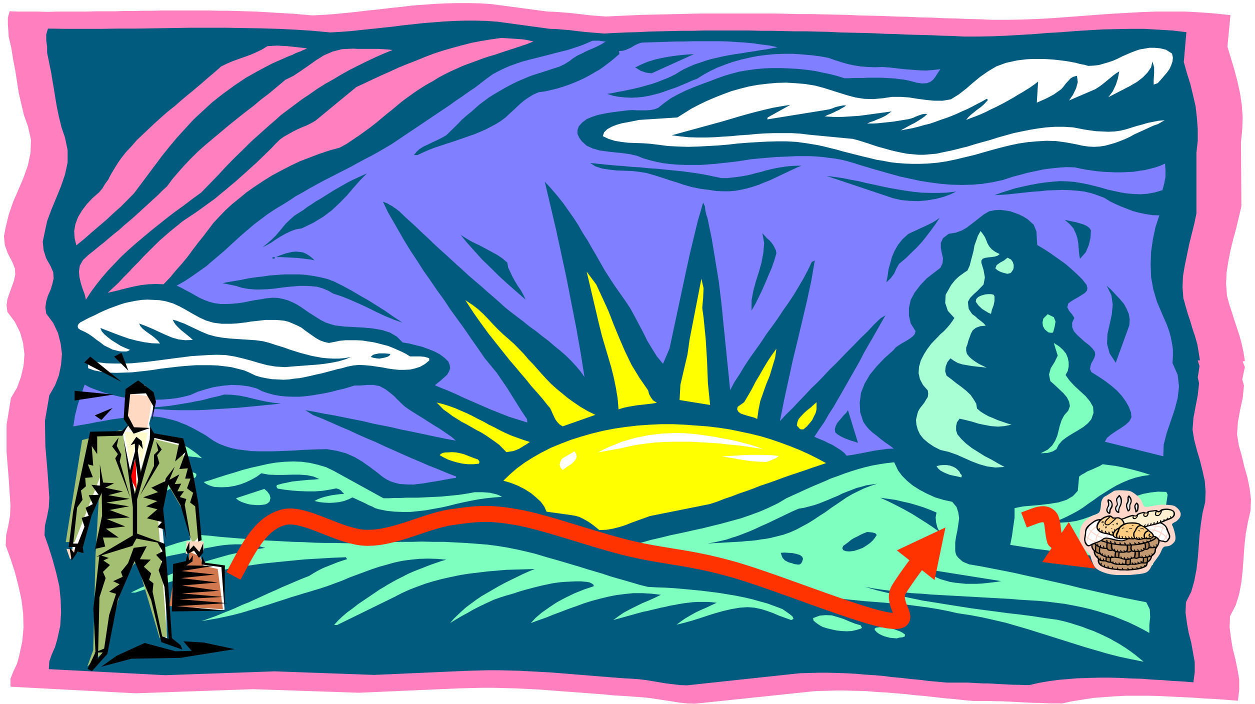
- Successfully implemented and *applied* to a number of complex domains

M. E. Bratman, D. J. Israel and M. E. Pollack. Plans and Resource-Bounded Practical Reasoning. *Computational Intelligence*, 4(4):349-355, 1988.

How to Build Agents?

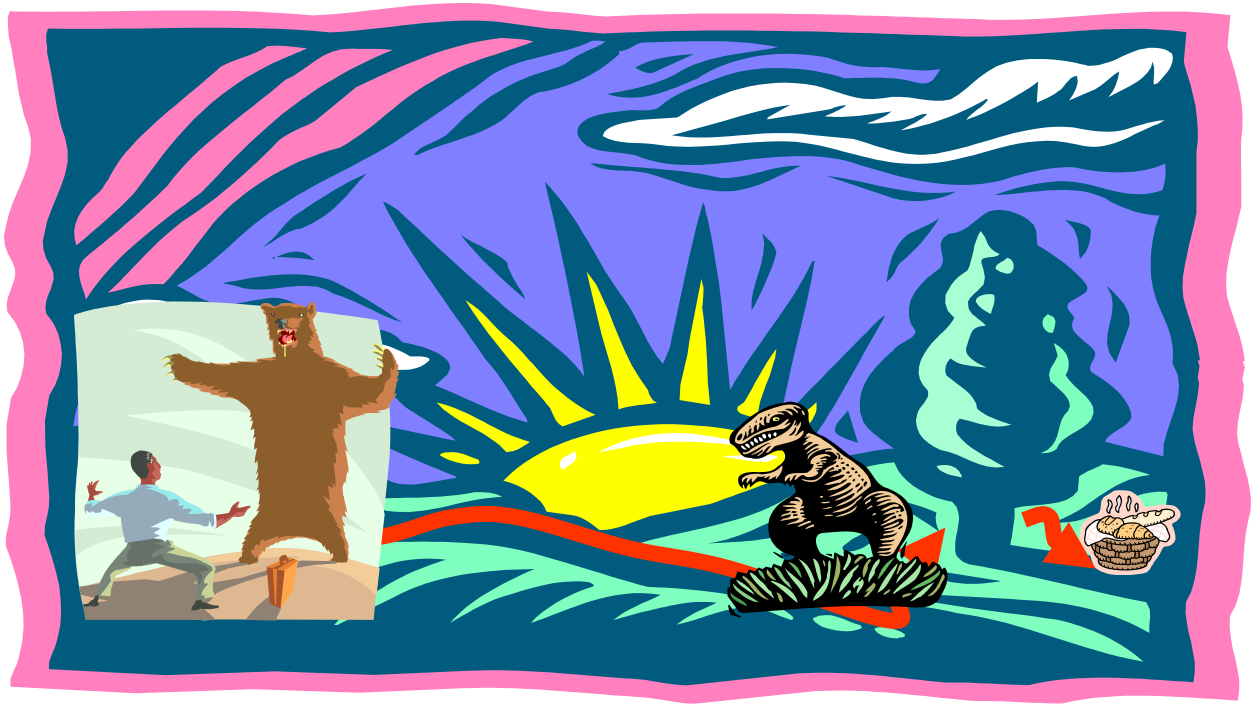
Classical way:

- Beliefs (present/past world states)
- Desires/Goals (desired states)
- Form & execute plan



How to Build Agents?

- Carry on
- Re-plan?



How to Build Agents?

- Planning agents may fail:
  - World is *complex*
    - Planning itself is very complex, “*waste resources*”
  - *Dynamic* (world may change because of other agents)
    - Cannot “*carry on*”
    - Future actions seem “*hopeless*”
  - *Real-time*
    - Cannot continually plan (cannot “*always re-plan*”)
- Why should agents plan? Why not just react?
  - But I just planned my whole day…

Commitments

Agents plan because they are *resource bounded* (Recall Bratman)

- Agents must form and **commit** to plans (*settle on plans*)

  Commit == hold on to a plan, don't drop it easily

- Commitments constrain reasoning:
  - Frame problems for an agent (*what to reason about*)

    Agent needs to determine how to fulfill commitments
  - Provide filter of admissibility (*what not to reason about*)

    Agent filters out options incompatible with commitments

- *Intention* is kind of *committed* plan (e.g. my next trip)

Commitments and Reactivity

- *Conflict*
  - Must have commitments to constrain reasoning
  - Must react to dynamic changes
- *Balanced commitments*
  - Must commit to plans
  - *Reconsider* commitments under *restrictive* situations
- Should not expect perfect balance *locally*, need *global success*
  - Occasional *wasted* reasoning possible
  - Occasional *suboptimal* behavior possible

Quick Note on Modal Logic

Use modal logic in reasoning about agents:

- Logic of *necessity* $\Box$ and *possibility* $\Diamond$
- Logic of "*must be*" and "*may be*"
- More broadly, includes:
  - *Temporal logic* (about time)
  - *Doxastic logic* (about belief)
  - *Epistemic logic* (about knowledge)
  - *Deontic logic* (about obligations, permissions, …)

Not first-order logic:

- Modal logic is more "*natural*" and *compact*
- Additional problems: *extensionality*

Formalizing Commitments in Modal Logic

- Notation (Cohen & Levesque 1990):
  - Bel (x P): Agent x has P as a belief
  - Goal (x P): Agent x has P as a goal
  - (Eventually P): Sometime in the future P becomes True
  - (Until P Q)/(Q Until P): Q is True until P True
  - (Always P): (NOT (Eventually NOT(P)))
  - (Never P): (Always (NOT(P))
  - AND, OR, NOT: Logic connectives

Commitments to Achievement Goals

Focus on achievement goals:

- P is the *goal* to be achieved (e.g. *inform about class cancelled*) and "x" is the *agent* involved, then:
  - (Goal x (Eventually P)) AND (Bel x NOT(P))
- Beliefs and goals must be *consistent*
  - If (Goal x (Eventually P)) then NOT(Bel x (Never P))
- If P is a *commitment*, then don't drop P as a goal easily (i.e. *agent x will keep P*)

  (Goal x (Eventually P)) UNTIL
  - P Achieved: (Bel x P) OR
  - P Unachievable: (Bel x (Never P)) OR
  - P Irrelevant (Q is a possible *reason* for P): (Bel x NOT(Q))

Commitments as Persistent Goals (PGOALs)

- (PGOAL x P Q) is defined as:

  (Bel x NOT(P)) AND (Goal x (Eventually P)) AND

  (UNTIL ((Bel x P) OR

  (Bel x (Never P)) OR

  (Bel x NOT(Q)))

  (Goal x (Eventually P)))

- Is (PGOAL x P Q) True if (Bel x P)?
- Is (PGOAL x P Q) True if (Bel x (Never P))?

From Commitments to Intentions

- Merely committing to P does not mean agent acts deliberately
- To "*intentionally*" do A, agent must realize it is doing A throughout:
- (Intend x A Q) is defined as:

  (PGOAL x

  (DONE x [Bel x (DOING x A)]?;A)

  Q)

- (DONE x [Bel x (DOES x A)]?;A) = P from original formula
- Agent x intends to do A if it has a persistent goal and
- "x" believed he was doing A just prior to having done it

Practical Implementations

- Practical implementations may use *situated plans*
- Situated/Reactive plans consist of
  - *Preconditions*: matched with agents' beliefs to activate plan
  - *Termination conditions*: to terminate plan when matched
  - *Plan body:* to execute when plan is activated
- Example: Plan *Attend-agents-conference*
  - Precondition: *Call for participation posted*
  - Body: *Register, fly to site, attend sessions, fly back…*
  - Termination condition*: Agents conference attended*
- Basis of several different types of agent architectures
  - RAP, TCA, PRS, InteRRaP, JACK™ …

Agent Architecture

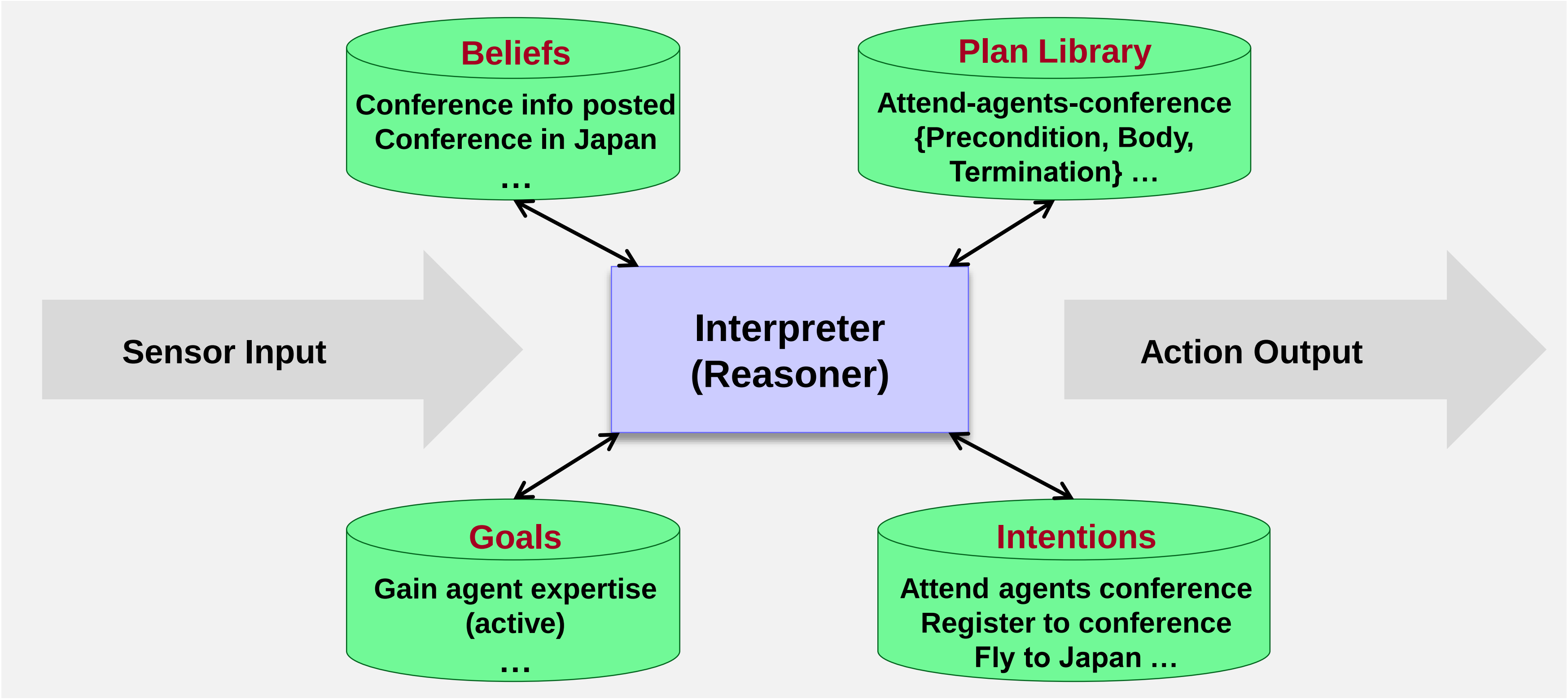
- Agent “Shell”, separates structure from variable content
- The principles/structure to be reused from task to task
  - *Why rediscover these from scratch each time?*
- Some well-known agent architectures:
  - *PRS, dMARS, JACK (started from PRS)*
  - *IRMA*
  - *InteRPaP*
  - *Soar, RAP (Reactive Action Packages)…*
- Can be *deliberative*, *reactive* or *hybrid*

Outline

- Basics on agents and MAS
- Applications and some views
- The Belief-Desire-Intention model
- ➢ Procedural reasoning system (PRS)

PRS-Type BDI Architecture

- Integrate BDI model, planning and reactive techniques



M. Georgeff and A. Lansky. Reactive Reasoning and Planning. *Proceedings of AAAI*, 1987.

PRS: Main Features

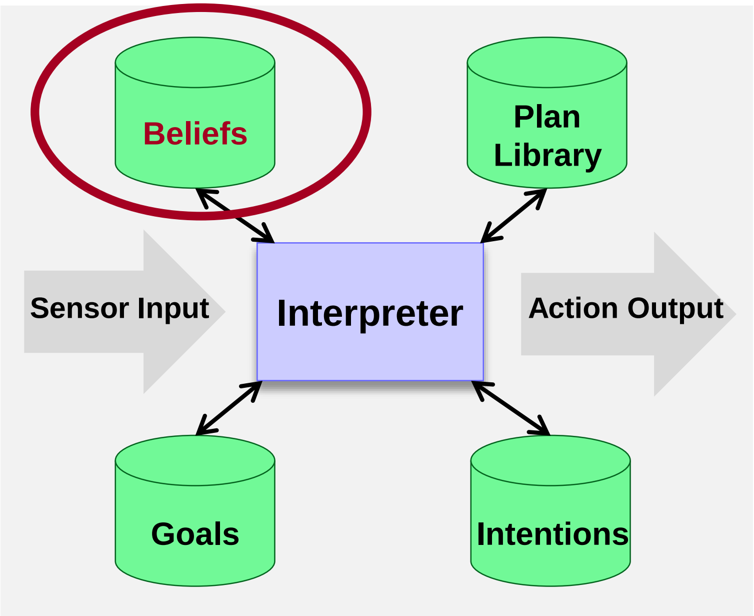
- BDI (Belief, Desires, Intentions) foundation
- Pre-compiled procedural knowledge
- Combine deliberative and reactive features
  - Sensing, plan selection, formation and execution
- Integrate goal-directed and event-driven behavior
- Plan dynamically and incrementally
- Can interrupt plan execution
- Meta-level reasoning
- Multi-agent planning

M. Georgeff and A. Lansky. Reactive Reasoning and Planning. *Proceedings of AAAI*, 1987.

PRS Architecture: Beliefs

- Contain beliefs or facts about the world
- Include meta-level information

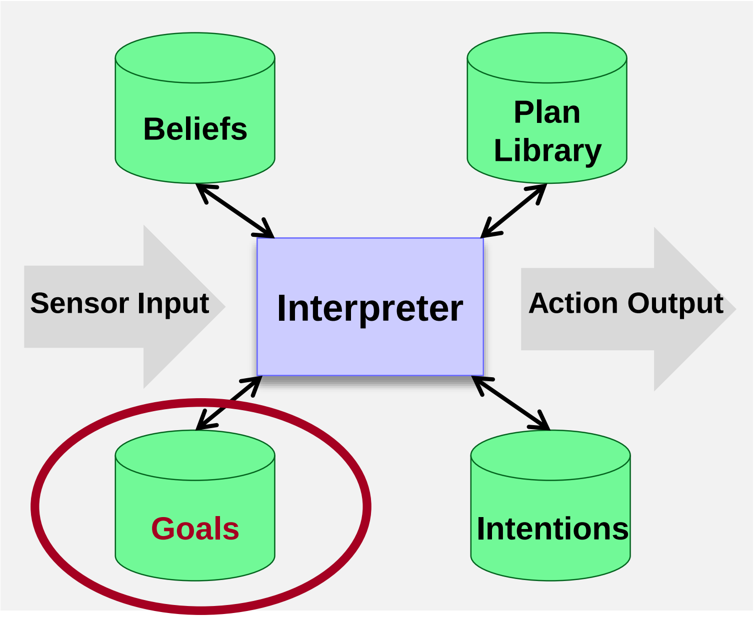
  E.g. Goal *G* is active



PRS Architecture: Goals/Tasks

- Represent desired behaviors
- Expressed as conditions over some time interval

  E.g. (walk a b): set of behaviors in which agent travels from a to b

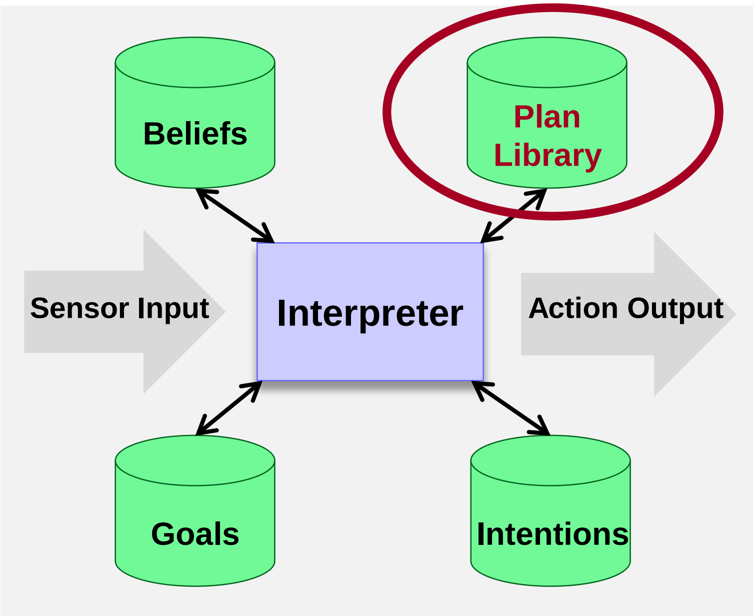


Expressing Tasks in a Dynamic Environment

- (! P) -- achieve P
- (? P) -- test P
- (# P) -- maintain P
- (^ C) -- wait until C
- (-> C) -- assert C
- (~> C) -- retract C

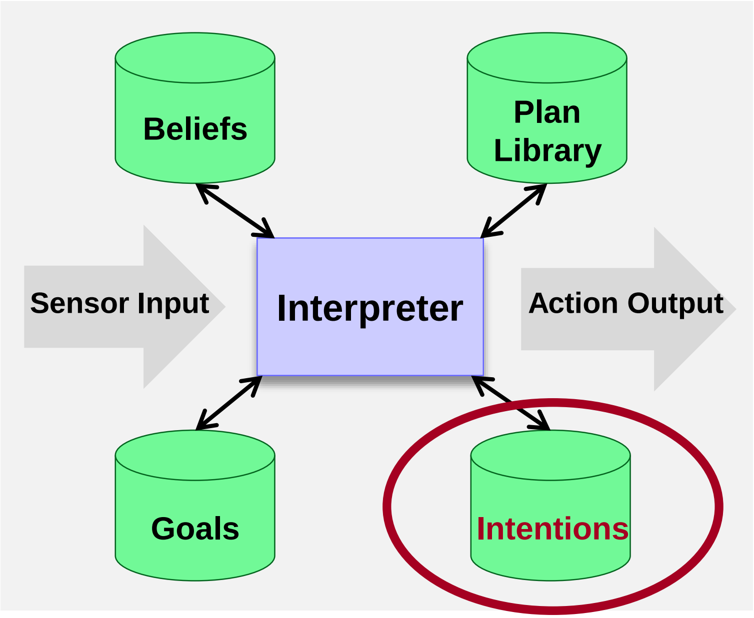
PRS Architecture: Plan Library

- Pre-compiled procedures
- Express actions and tests to achieve goals or to react to situations



PRS Architecture: Intentions

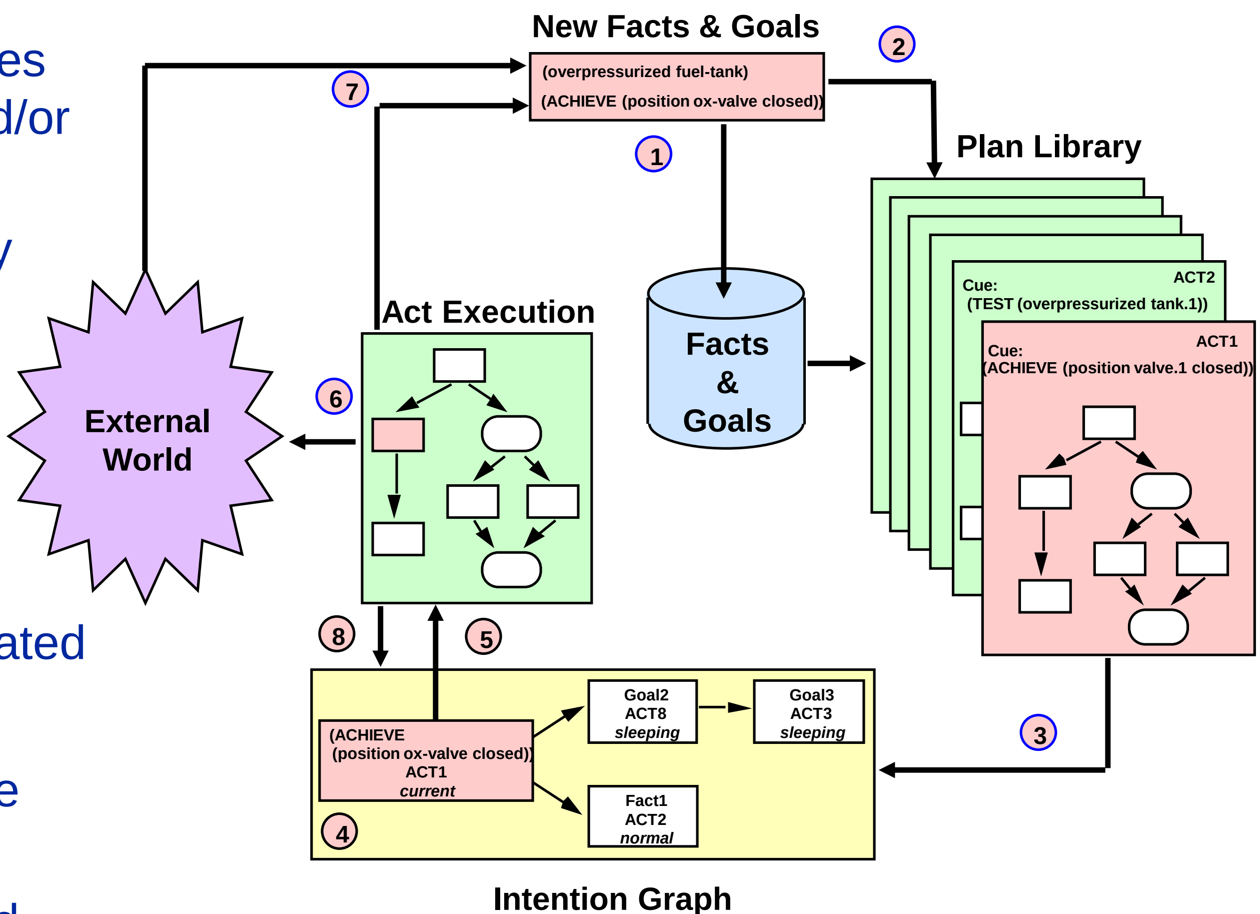
- Currently active procedures
- Procedure currently being executed



PRS Interpreter

Execution cycle:

1. New information arrives that updates facts and/or goals
2. Plans are triggered by new facts or goals
3. A triggered plan is intended
4. An intended plan is selected
5. That intention is activated
6. An action is executed
7. New facts or goals are posted
8. Intentions are updated



Meta-Level Reasoning

- Can include meta-level procedures, e.g.
  - Choose among multiple applicable procedures
  - Evaluate how much more reasoning can be done within time constraints
  - How to achieve a conjunction or disjunction of goals

  ...

Multiple Tasks & Distributed Planning

- Multi-thread operations

  Multiple tasks being performed, runtime stacks where tasks are executed, suspended, and resumed

- Support distributed planning

  Several PRS agents run asynchronously and communicate through message passing

内容回顾

What is an Agent（智能体）?

An intelligent agent is a computer system capable of *flexible autonomous* action in some environment

- Reactive

  Maintain an ongoing interaction with its environment, and be responsive to changes that occur in it

- Pro-active/Deliberative

  Not driven solely by events, but take the initiative and act purposefully toward achieving goals

- Social

  The ability to interact with other agents (and possibly humans) and perhaps cooperate with others

What is a Multi-Agent System (MAS)?

- A multi-agent system consists of a number of agents, which *interact* with one another
  - ***Modeling others and communication***
- In the most general case, agents will be acting *on behalf of* users with different goals and motivations
  - ***Practical reasoning and planning***
- To successfully interact, they will require the ability to *collaborate*, *cooperate* and *coordinate* with each other
  - ***Collaboration, cooperation and coordination***

The Belief-Desire-Intention (BDI) Model

Bratman *et al* [1988] recognizes the primacy of *beliefs*, *desires* and *intentions* (*BDI*) in modeling the rational behavior of agents:

- An agent's *beliefs* represent the information the agent has about the world
- An agent's *desires* represent the states of affairs that the agent would wish to bright about
- An agent's *intentions* represent those desires that have been committed by the agent

Commitments

Agents plan because they are *resource bounded* (Recall Bratman)

- Agents must form and **commit** to plans

  Commit == hold on to a plan, don't drop it easily

- Commitments constrain reasoning:
  - Frame problems for an agent (*what to reason about*)

    Agent needs to determine how to fulfill commitments
  - Provide filter of admissibility (*what not to reason about*)

    Agent filters out options incompatible with commitments

- *Intention* is kind of *committed* plan (e.g. my next trip)

Formalizing Commitments in Modal Logic

Focus on achievement goals:

- If P is a *commitment*, then don't drop P as a goal easily

  (Goal x (Eventually P)) UNTIL

  - P Achieved: (Bel x P) OR
  - P Unachievable: (Bel x (Never P)) OR
  - P Irrelevant (Q is possibly the *reason* for P): (Bel x NOT(Q))

Commitments as persistent goals (*PGOAL*):

- (PGOAL x P Q) is defined as:

  (Bel x NOT(P)) AND (Goal x (Eventually P)) AND

  (UNTIL ((Bel x P) OR

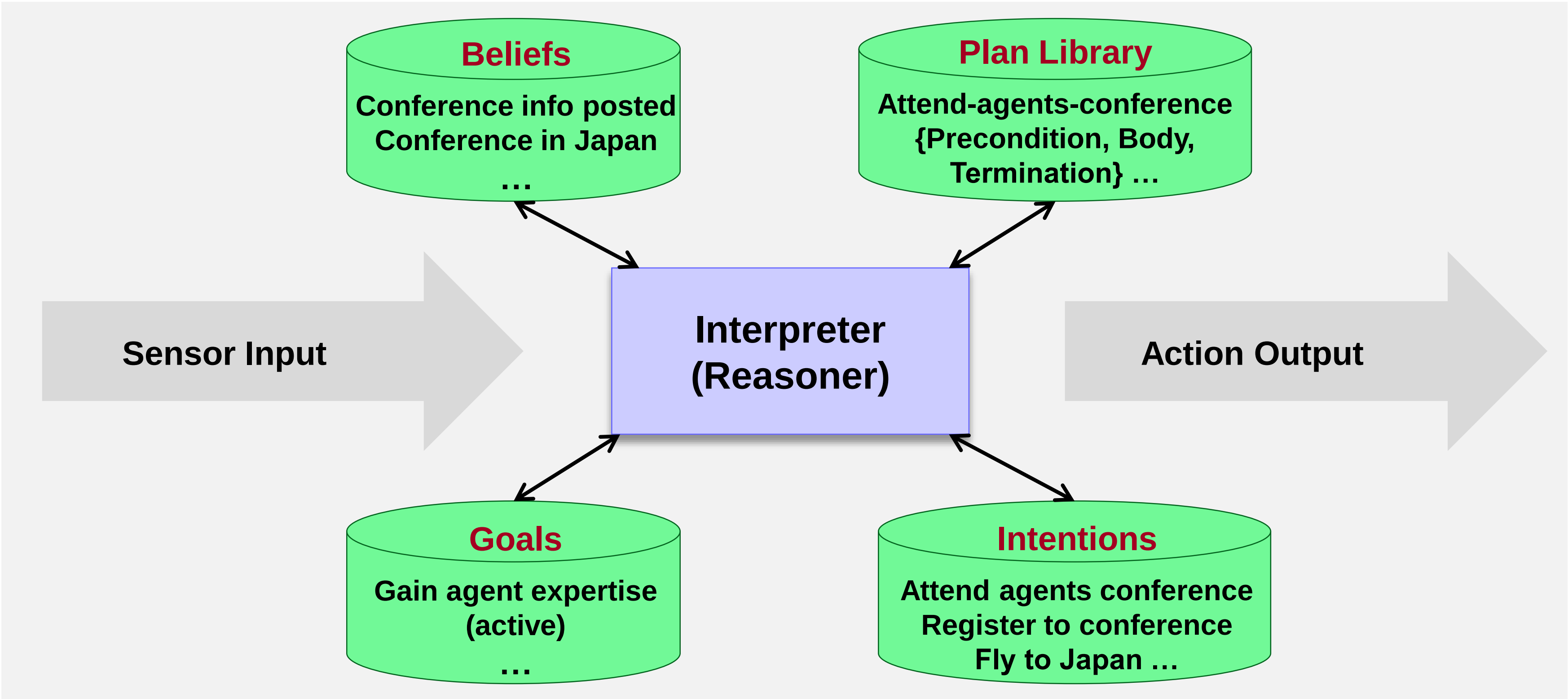
  (Bel x (Never P)) OR

  (Bel x NOT(Q)))

  (Goal x (Eventually P)))

PRS-Type BDI Architecture

- Integrate BDI model, planning and reactive techniques



References

- M. Wooldridge and N.R. Jennings. *Intelligent Agents: Theory and Practice*. Knowledge Engineering Review, 10(2):115–152, 1995
- M. E. Bratman, D. J. Israel and M. E. Pollack. *Plans and Resource-Bounded Practical Reasoning*. Computational Intelligence, 4(4): 349-355, 1988
- Martha Pollack. *The Uses of Plans*. Artificial Intelligence, 57(1):43-69, 1992
- H. J. Levesque, P. R. Cohen and J. T. H. Nunes. *On acting together*. Proceedings of the National Conference on Artificial Intelligence (AAAI), 1990 (Also see next paper)
- P. R. Cohen and H. J. Levesque. *Teamwork*. Nous, 25(4):487-512, 1991. Special Issue on Cognitive Science and Artificial Intelligence
- M. P. Georgeff and A. L. Lansky. *Reactive Reasoning and Planning*. Proceedings of the National Conference on Artificial Intelligence (AAAI), 1987

End.