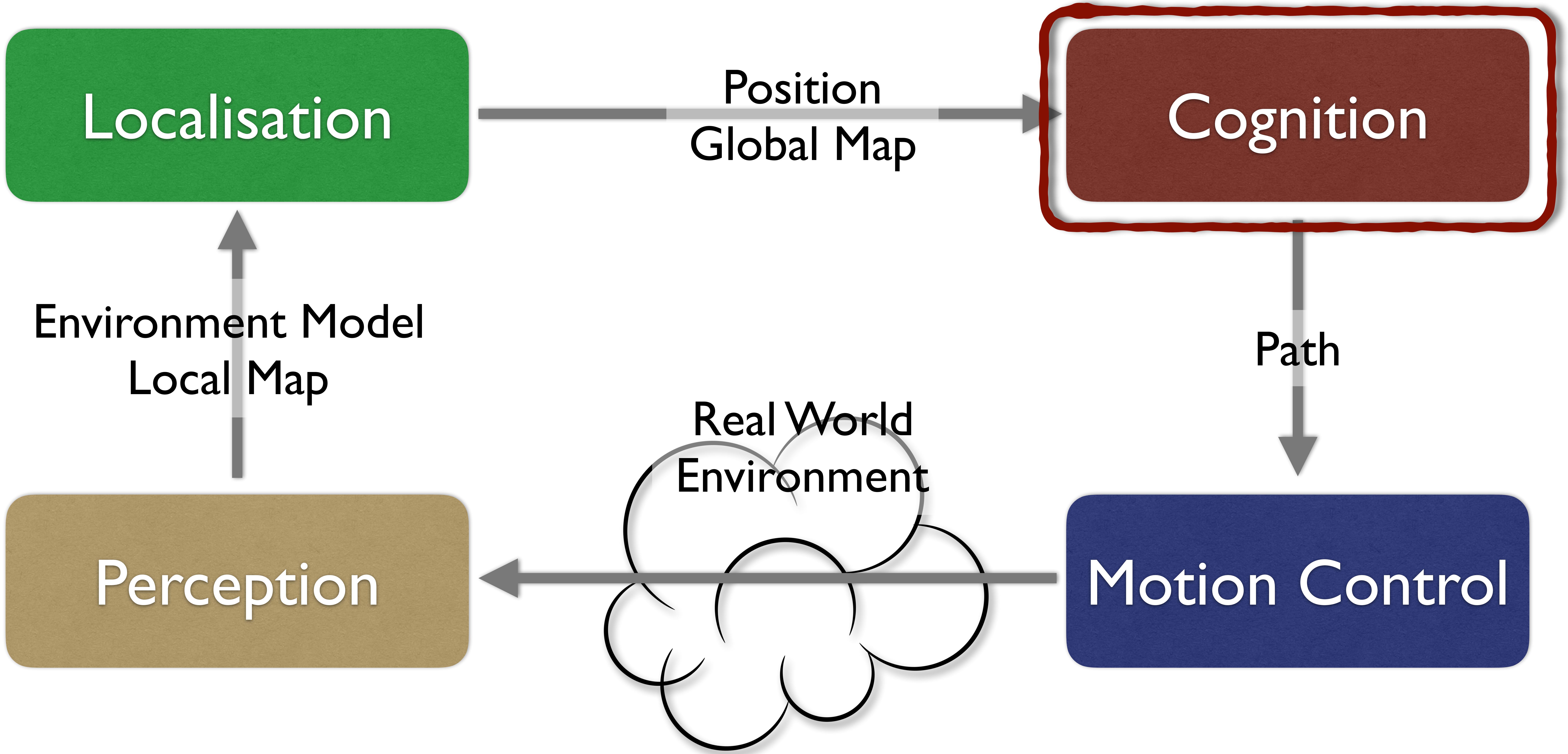
COMP329 Robotics and Autonomous Systems

Lecture 15: Agents and Intentions

Dr Terry R. Payne
Department of Computer Science
UNIVERSITY OF LIVERPOOL

General control architecture



Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

Aims of these slides

- We will start to focus on the second part of the module:
  - Autonomous agents (also check out COMP310)
  - Things you will need for the second assignment.

- We will recap some of the basic ideas about agents from earlier in the module.
  - Look at some aspects in more detail.
  - Introduce the idea of the intentional stance

What is an Agent?

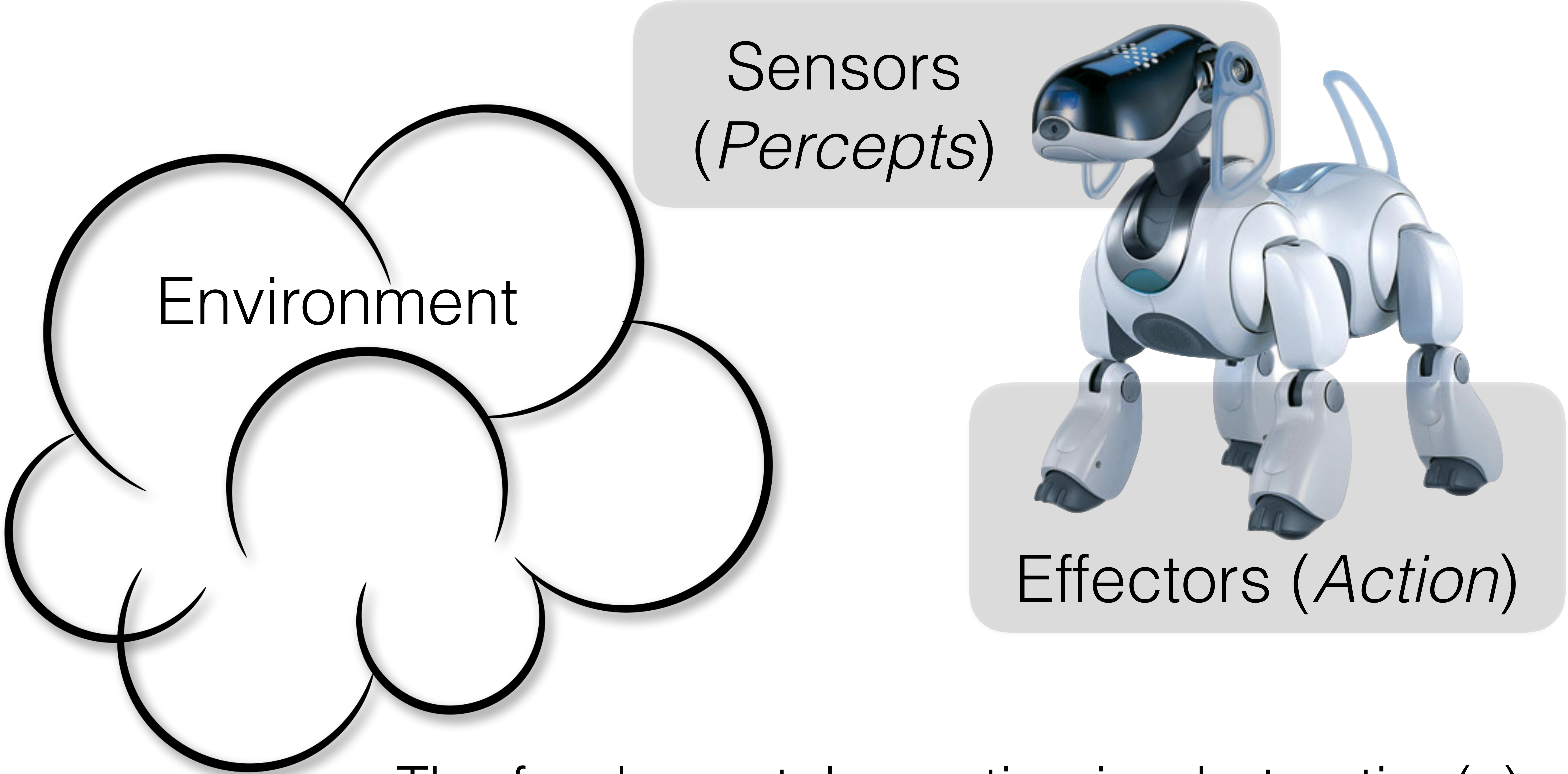
- As we have said before:
  - The main point about agents is they are **autonomous**: capable independent action.

> "...An agent is a computer system that is *situated* in some *environment*, and that is capable of *autonomous action* in that environment in order to meet its delegated objectives..."

- It is all about decisions
  - An agent has to choose ***what*** action to perform.
  - An agent has to decide ***when*** to perform an action.

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

Agent and Environment



The fundamental question is what action(s) to take for a given state of the environment

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

Intelligent Agents

- Making good decisions requires the agent to be intelligent.
  - Agent has to do the right thing.
- An intelligent agent is a computer system capable of flexible autonomous action in some environment.
- By flexible, we mean:
  - Reactive
  - Pro-active
  - Social
- All these properties make it able to respond to what is around it. (More on the next few slides).

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

Reactivity

- If a program's environment is guaranteed to be fixed, the program need never worry about its own success or failure

- Program just executes blindly.
  - Example of fixed environment: compiler.

- The real world is not like that: most environments are ***dynamic*** and information is ***incomplete***.

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

Reactivity

- Software is hard to build for dynamic domains: program must take into account possibility of failure
  - ask itself whether it is worth executing!

- A ***reactive*** system is one that maintains an ongoing interaction with its environment, and responds to changes that occur in it (in time for the response to be useful).

Proactiveness

- Reacting to an environment is easy
  - e.g., stimulus → response rules
- But we generally want agents to **do things for us**.
  - Hence **goal directed behaviour**.
- ***Pro-activeness*** = generating and attempting to achieve goals; not driven solely by events; taking the initiative.
  - Also: recognising opportunities.

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

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 by interacting with 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 ***cooperation***, ***coordination***, and ***negotiation***.
  - At the very least, it means the ability to communicate. . .

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

Social Ability: Cooperation

- Cooperation is ***working together as a team to achieve a shared goal.***
- Often prompted either by the fact that no one agent can achieve the goal alone, or that cooperation will obtain a better result (e.g., get result faster).

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

Social Ability: Coordination

- Coordination is ***managing the interdependencies between activities.***

- For example, if there is a non-sharable resource that you want to use and I want to use, then we need to coordinate.

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

Social Ability: Negotiation

- Negotiation is the ability to reach agreements on matters of common interest.

- For example:
  - You have one TV in your house; you want to watch a movie, your housemate wants to watch football.
  - A possible deal: watch football tonight, and a movie tomorrow.

- Typically involves offer and counter-offer, with compromises made by participants.

Properties of Environments

- Since agents are in close contact with their environment, the properties of the environment affect agents.
  - Also have a big effect on those of us who build agents.

- Common to categorise environments along some different dimensions.
  - Fully observable vs partially observable
  - Deterministic vs non-deterministic
  - Episodic vs non-episodic
  - Static vs dynamic
  - Discrete vs continuous

Properties of Environments

- Fully observable vs partially observable.
  - An accessible or ***fully observable*** 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, or ***partially observable***.
  - The more accessible an environment is, the simpler it is to build agents to operate in it.

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

Properties of Environments

- Deterministic vs non-deterministic.
  - A deterministic environment is one in which any action has a single guaranteed effect — there is no uncertainty about the state that will result from performing an action.
  - The physical world can to all intents and purposes be regarded as non-deterministic.
  - We'll follow Russell and Norvig in calling environments ***stochastic*** if we quantify the non-determinism using probability theory.
  - Non-deterministic environments present greater problems for the agent designer.

Properties of Environments

- Episodic vs non-episodic.
  - 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.
    - An example of an episodic environment would be an assembly line where an agent had to spot defective parts.
  - 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.
    - Relations to the Markov property
  - Environments that are not episodic are called either ***non-episodic*** or ***sequential***. Here the current decision affects future decisions.
    - Driving a car is sequential.

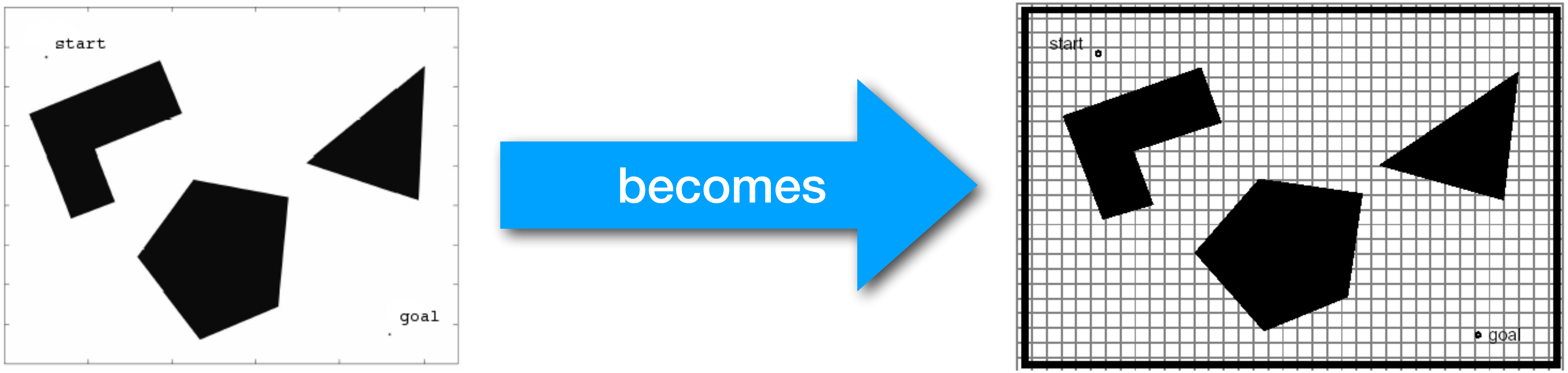
Properties of Environments

- Static vs dynamic.
  - A ***static*** environment is one that can be assumed to remain unchanged except by the performance of actions by 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.
  - One reason an environment may be dynamic is the presence of other agents.

Properties of Environments

- Discrete vs continuous.
  - An environment is discrete if there are a fixed, finite number of actions and percepts in it.
  - Otherwise it is continuous

- Often we treat a continuous environment as descrete for simplicity



Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

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.
- These statements make use of a ***folk psychology***, by which human behaviour is predicted and explained through the attribution of ***attitudes***
  - *e.g. believing, wanting, hoping, fearing ...*
- The attitudes employed in such folk psychological descriptions are called the ***intentional*** notions.

Dennett on Intentional Systems

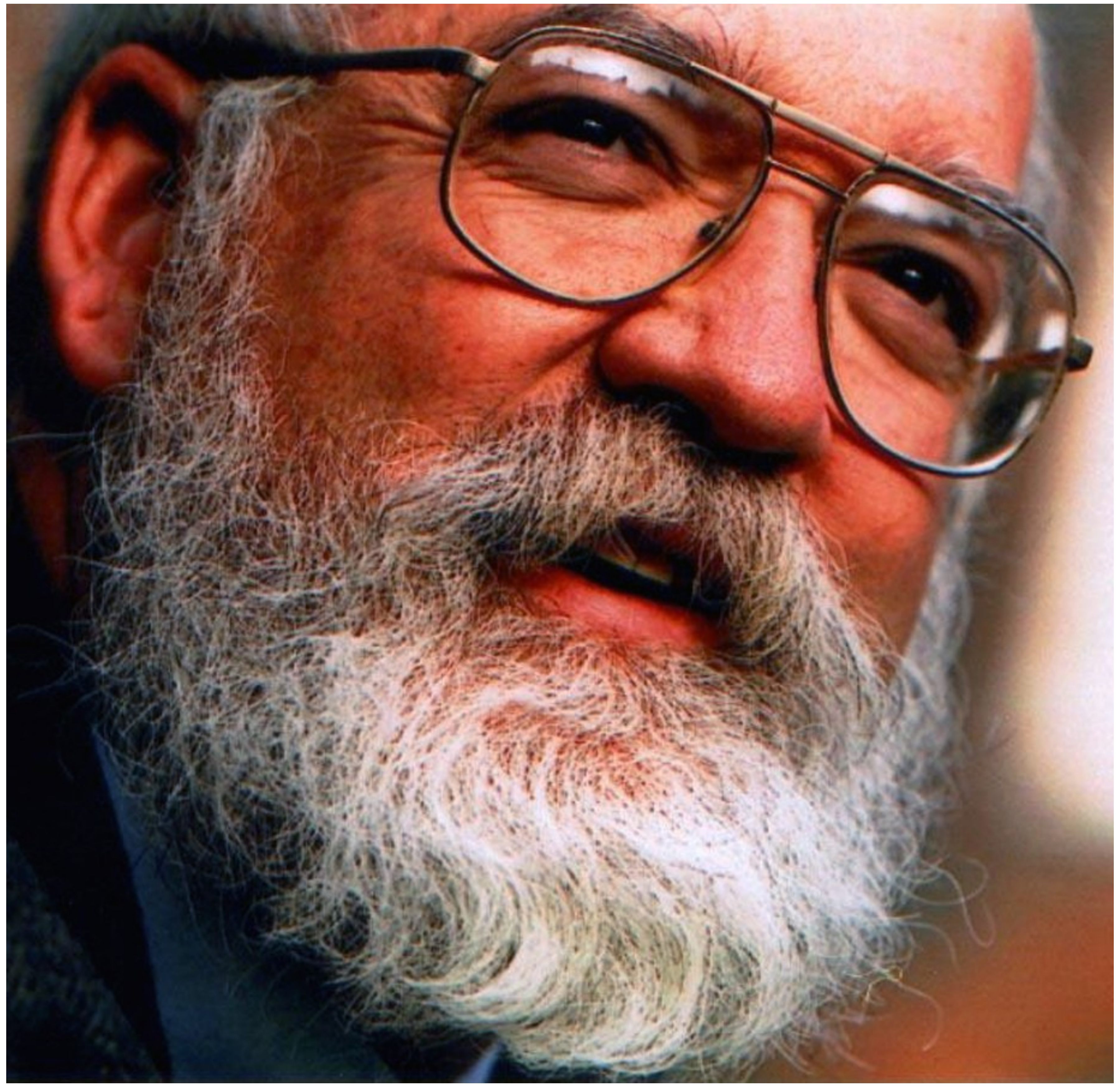
- The philosopher Daniel Dennett coined the term intentional system to describe entities:

> *"... whose behaviour can be predicted by the method of attributing belief, desires and rational acumen..."*

- Dennett identifies different 'grades' of intentional system:

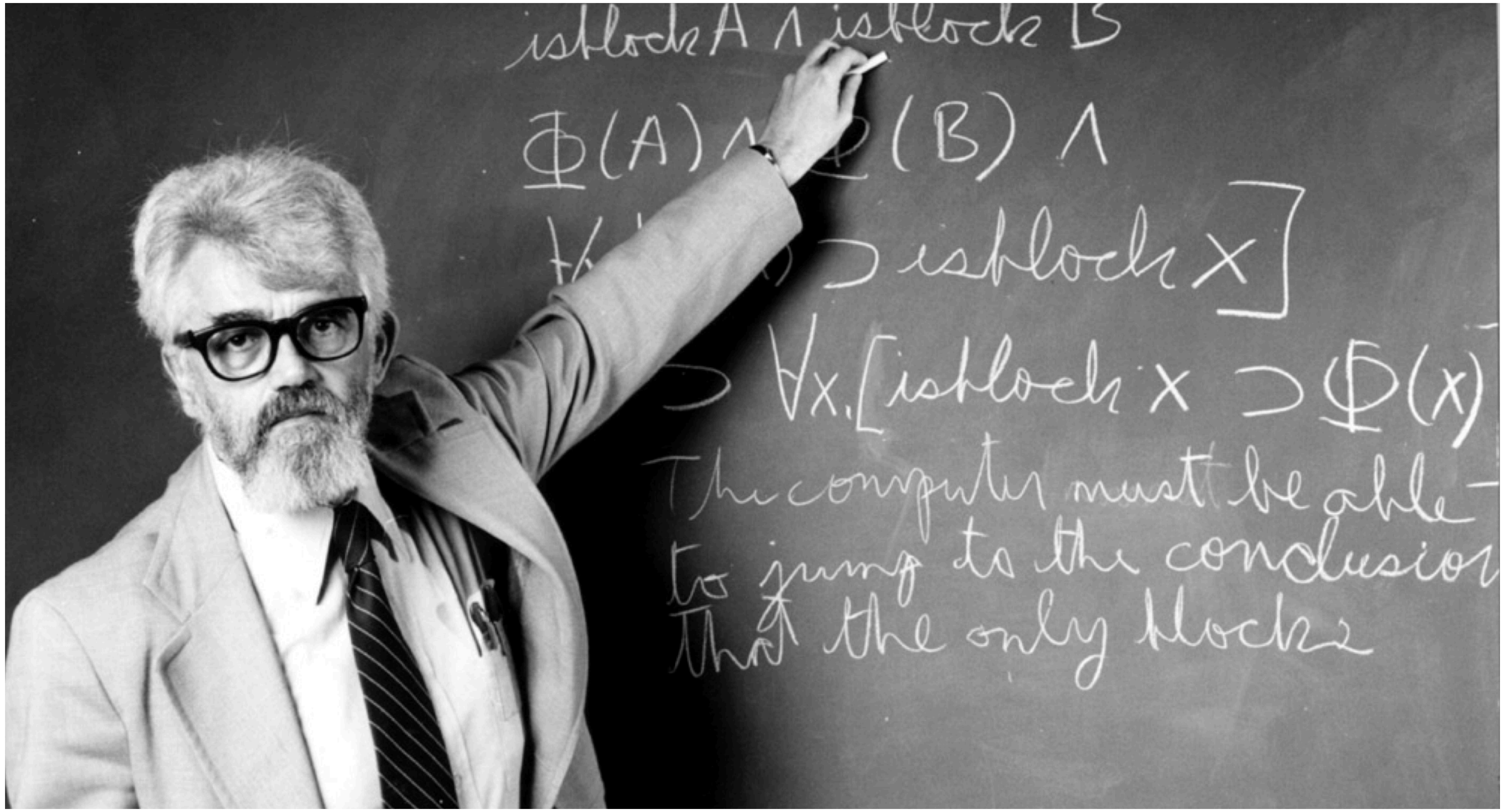
> *"...A* ***first-order*** *intentional system has beliefs and desires (etc.) but* ***no beliefs and desires about*** *beliefs and desires...*
>
> *...A* ***second-order*** *intentional system is more sophisticated; it has beliefs and desires (and no doubt other intentional states) about beliefs and desires (and other intentional states) — both those of others and its own..."*

- Is it legitimate or useful to attribute beliefs, desires, and so on, to computer systems?



Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

McCarthy on Intentional Systems



- John McCarthy argued that there are occasions when the ***intentional stance*** is appropriate:

McCarthy on Intentional Systems

> *"...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 behaviour, or how to repair or improve it. It is perhaps never **logically required** even for humans, but expressing reasonably briefly what is actually known about the state of the 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 ..."*

- John McCarthy argued that there are occasions when the ***intentional stance*** is appropriate:

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

What can be described with 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?

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

Intentional Systems

- It provides us with a familiar, non-technical way of understanding and explaining agents.

What can be described with the intentional stance?

- 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 behaviour ..." (Yoav Shoham)*

- Put crudely, the more we know about a system, the less we need to rely on animistic, intentional explanations of its behaviour.

- But with very complex systems, a mechanistic, explanation of its behaviour 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.

Agents as Intentional Systems

- So agent theorists start from the (strong) view of agents as intentional systems: one whose simplest consistent description requires the intentional stance.
- This ***intentional stance*** is an ***abstraction tool***...
  - ... a convenient way of talking about complex systems, which allows us to predict and explain their behaviour without having to understand how the mechanism actually works.
- Most important developments in computing are based on new abstractions:
  - procedural abstraction, abstract data types, objects, etc
- Agents, and agents as intentional systems, represent a further, and increasingly powerful abstraction.

So why not use the intentional stance as an abstraction tool in computing — to explain, understand, and, crucially, ***program*** *computer systems, through the notion of "agents"?*

Abstractions

- Remember: most important developments in computing are based on new abstractions.
- Just as moving from machine code to higher level languages brings an efficiency gain, so does moving from objects to agents.
  - The following 2006 paper claims that developing complex applications using agent-based methods leads to an average saving of 350% in development time (and up to 500% over the use of Java).
    - S. Benfield, *Making a Strong Business Case for Multiagent Technology*, Invited Talk at AAMAS 2006.

Programming has progressed through:

- *machine code;*
- *assembly language;*
- *machine-independent programming languages;*
- *sub-routines;*
- *procedures & functions;*
- *abstract data types;*
- *objects;*

to

- *Agents, as intentional systems, that represent a further, and increasingly powerful abstraction.*

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

Agents as Intentional Systems

- There are other arguments in favour of this idea...
  1. Characterising Agents
     - It provides us with a familiar, non-technical way of understanding and explaining agents.
  2. Nested Representations
     - It 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.
     - "If you think that Agent B knows x, then move to location L".

North by Northwest

*Eve Kendell **knows** that Roger Thornhill is working for the FBI. Eve **believes** that Philip Vandamm **suspects** that she is helping Roger. This, in turn, leads Eve to **believe** that Philip **thinks** she is working for the FBI (which is true). By pretending to shoot Roger, Eve **hopes** to convince Philip that she is not working for the FBI*

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

Agents as Intentional Systems

- There are other arguments in favour of this idea...

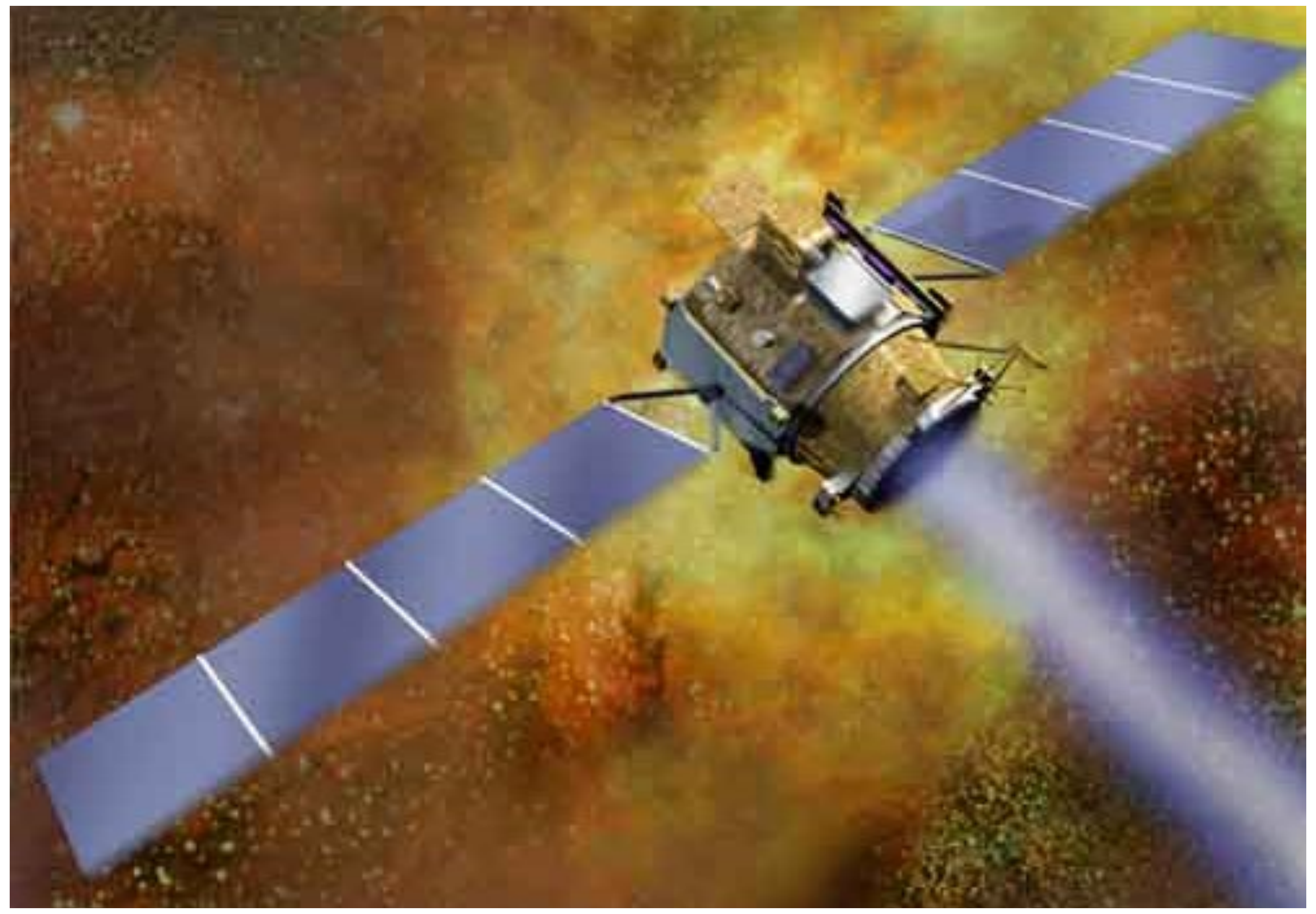
  3. Post-Declarative Systems

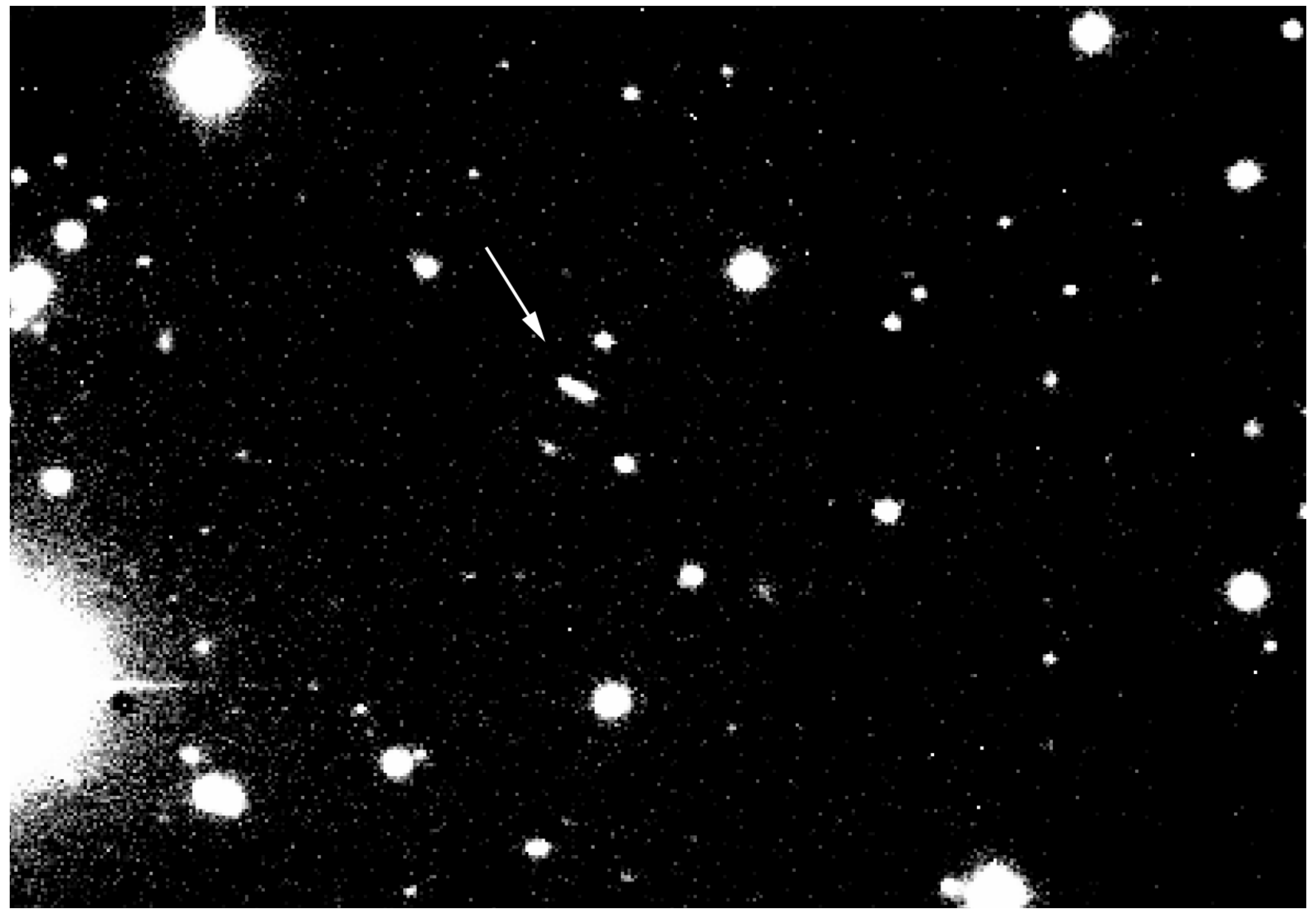
  - In ***procedural programming***, we say exactly ***what*** a system should do;
  - In ***declarative programming***, we state something ***that we want to achieve***, give the system general info about the relationships between objects, and let a built-in control mechanism (e.g., goal-directed theorem proving) figure out what to do;
  - With agents, we give a ***high-level description of the delegated goal***, and let the control mechanism figure out what to do, knowing that it will act in accordance with some ***built-in theory of rational agency***.

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017

Post-Declarative Systems

- What is this built-in theory?
- Method of combining:
  - What you ***believe*** about the world.
  - What you ***desire*** to bring about
- Establish a set of ***intentions***
  - Then figure out how to make these happen.





DS1 seen 2.3 million miles from Earth

Summary

- This lecture reflected on the idea of anagent.
  - It discussed briefly the properties of the environments in which the agents operate
  - It also introduces the intentional stance.
    - And describes why this idea is
- Next time we will look at practical reasoning and the ***Belief, Desire, Intention*** model

Original Source: M. Wooldridge, S.Parsons, D.Grossi - updated by Terry Payne, Autumn 2016 & 2017