Towards Full Cognitive Autonomy in Technical Systems

Ricardo Sanz

Autonomous Systems Laboratory

Universidad Politécnica de Madrid

1st CHIST-ERA Conference, Rome 27-28 May 2010

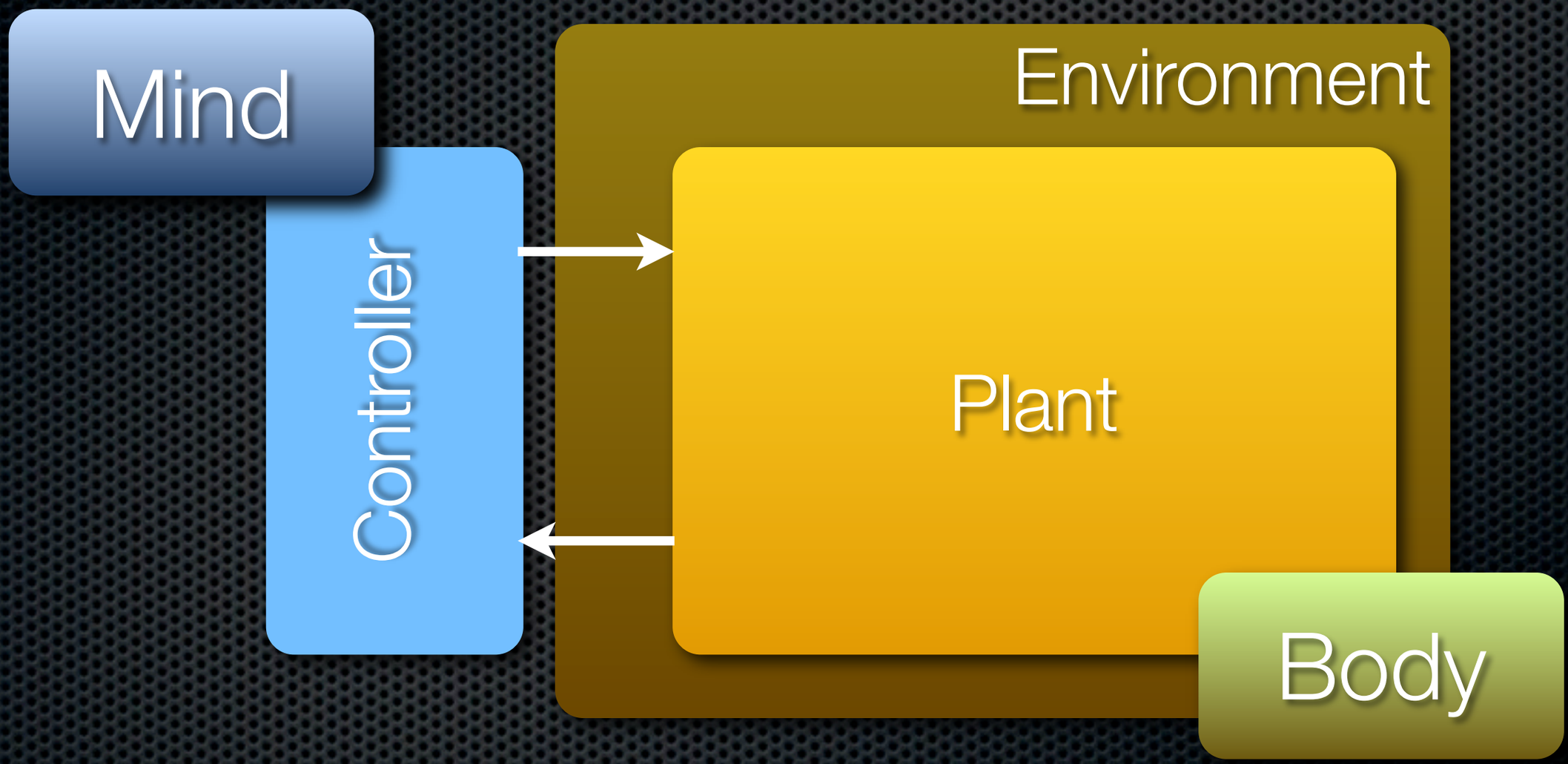
Two up-front issues

- Say thanks
- Apologise
- Topic: **Self-awareness and self-consciousness**
- *“A fuzzy topic which sprouts into cognitive psychology, neuroscience and philosophy.”*
- I feel myself more **incompetent** than **expert**.

A bit of context

ICT in control systems

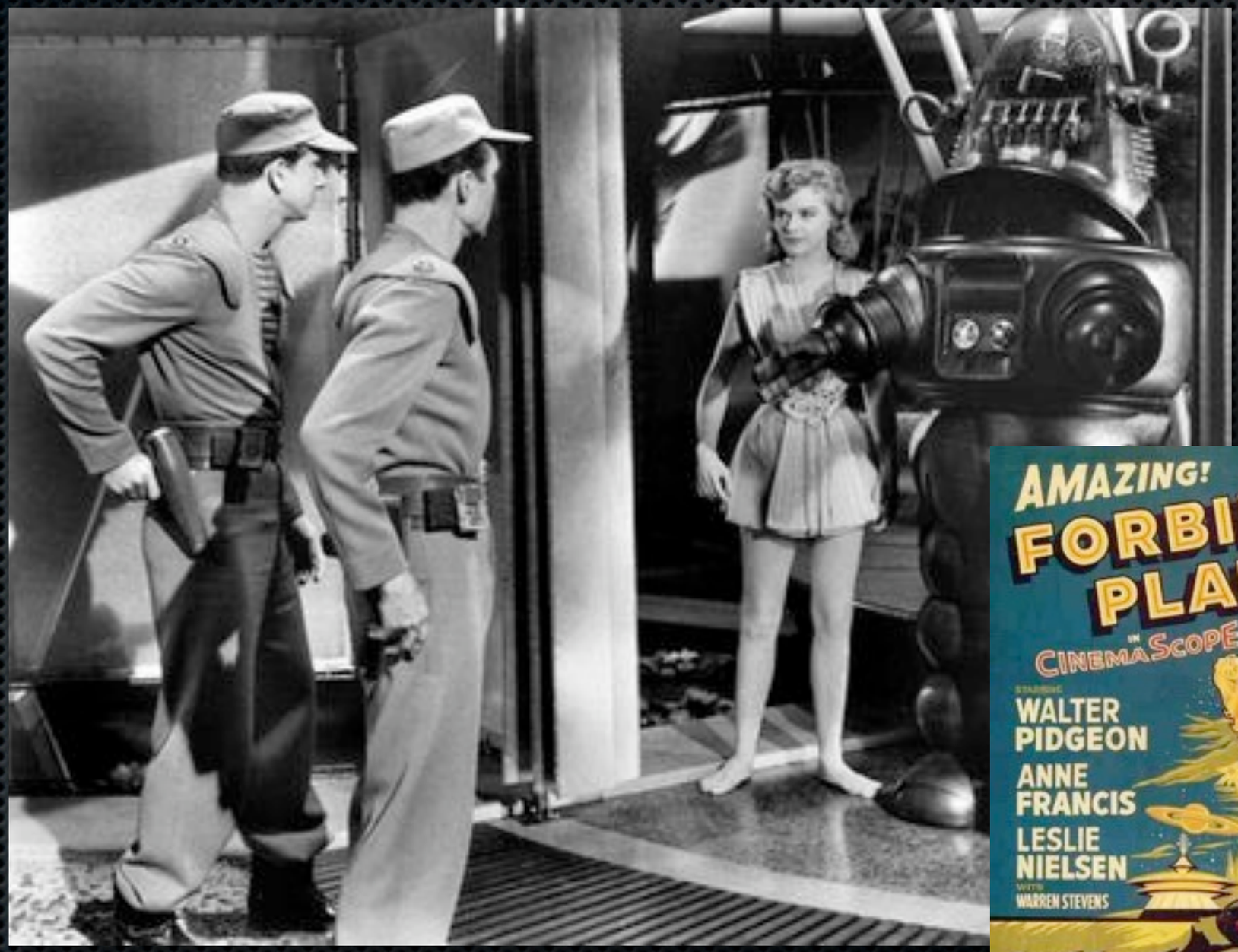
Control Engineering



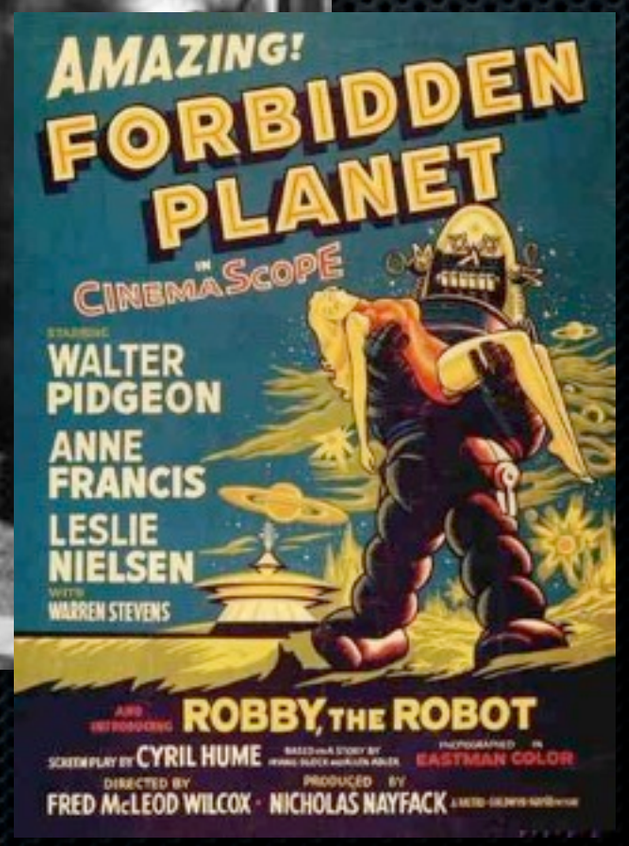
From dreams ...

Forbidden Planet

Robbie

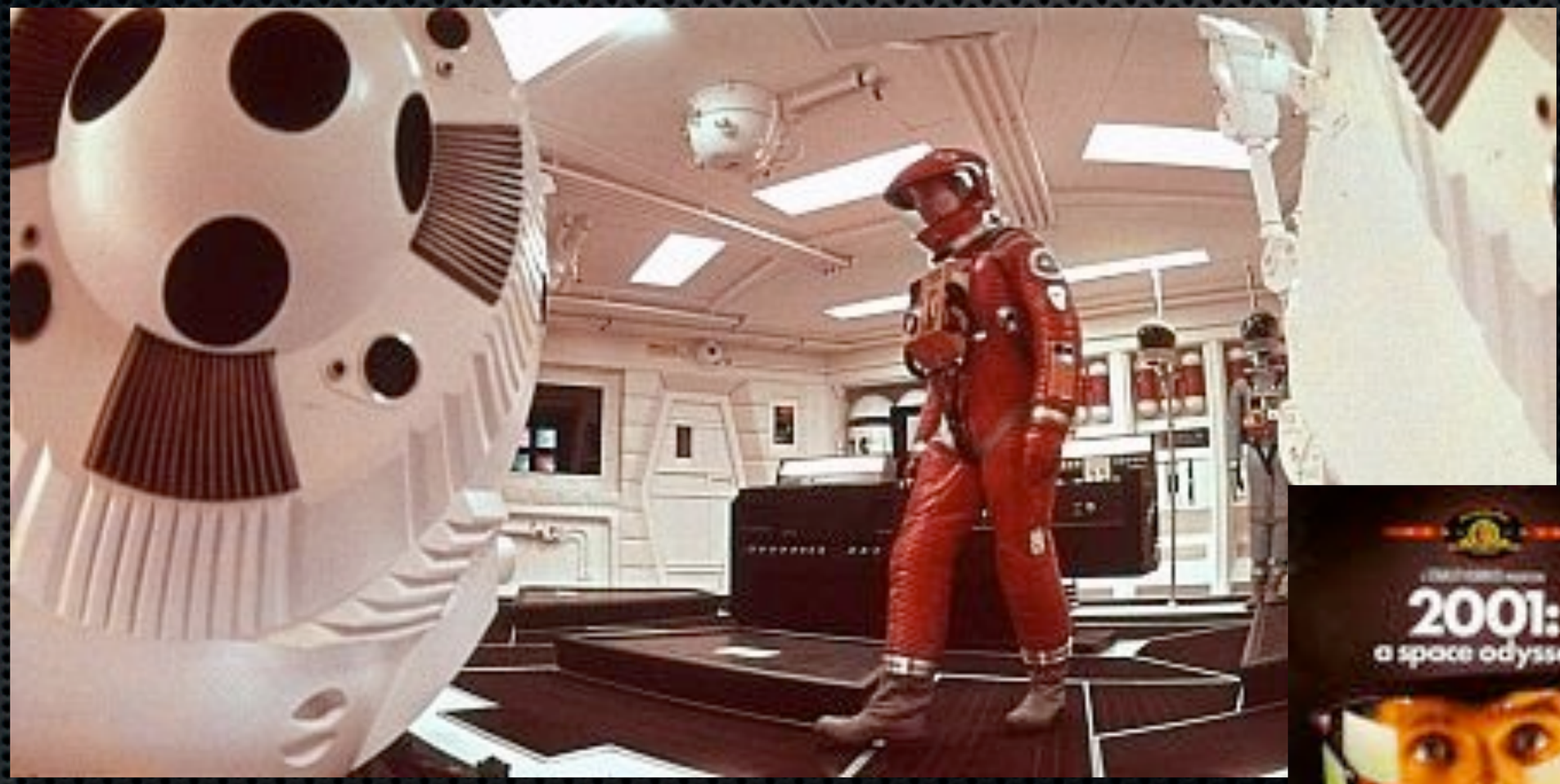


(1956)



2001: A Space Odyssey

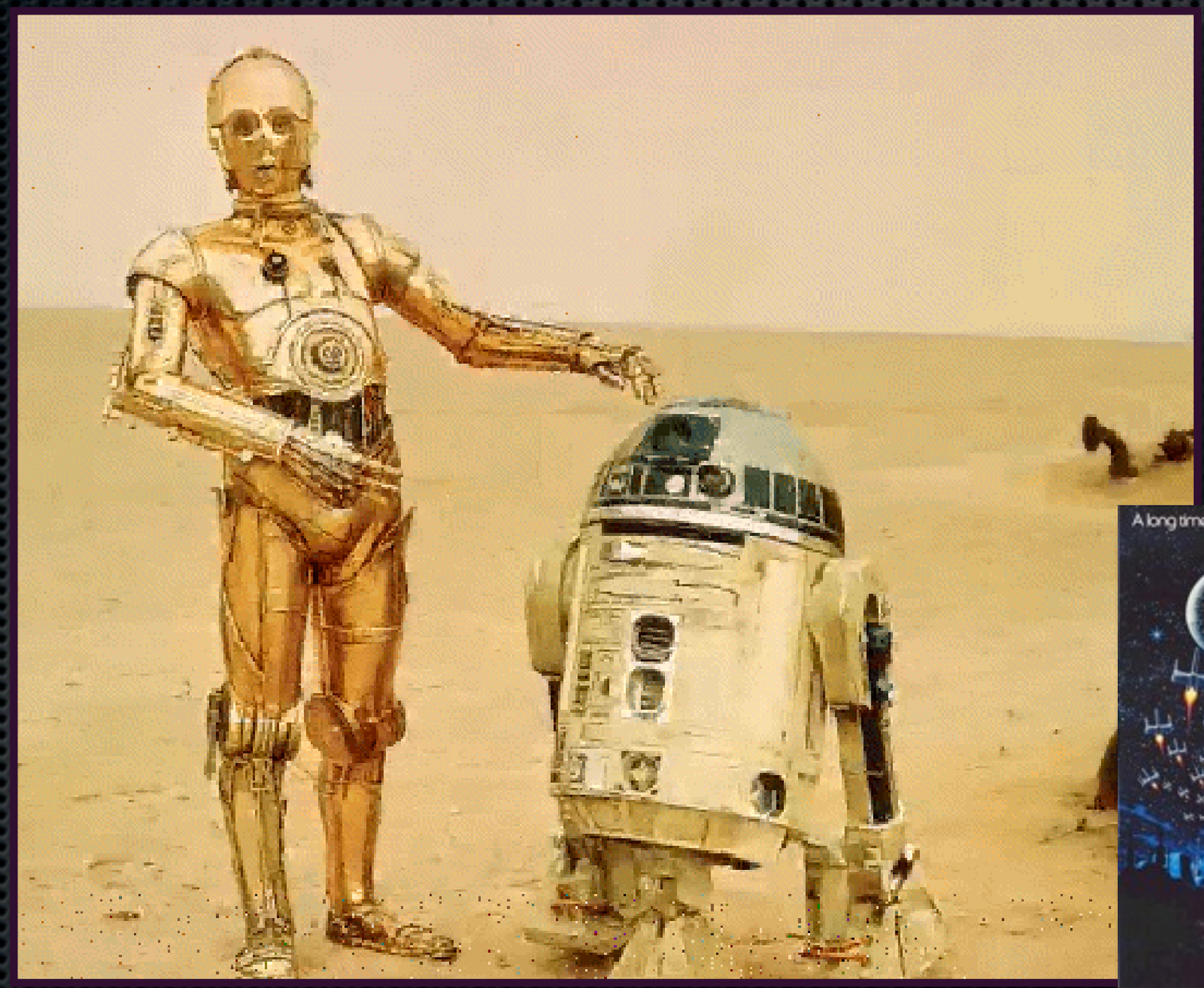
HAL



(1968)

Star Wars

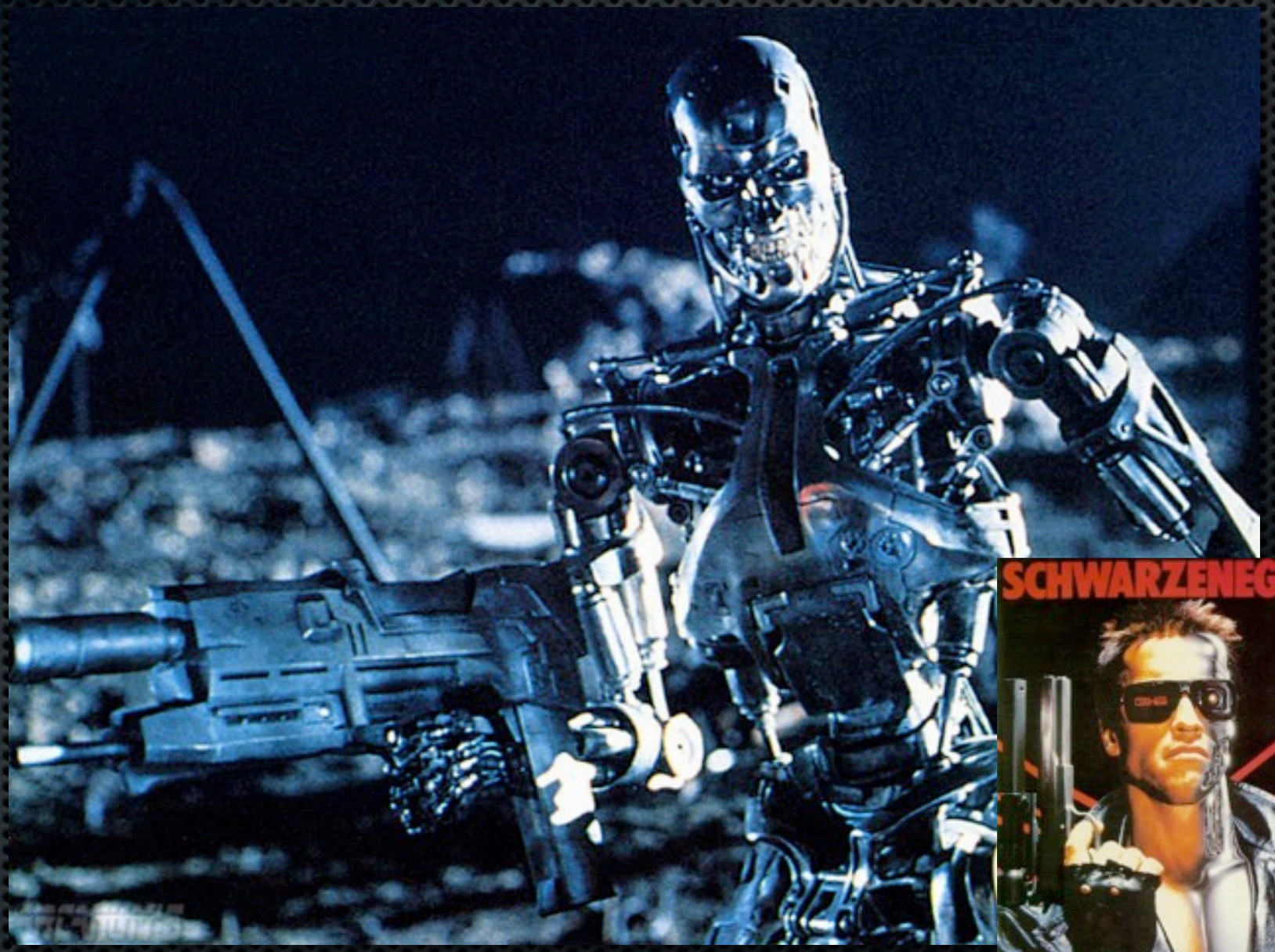
C3PO
R2D2



(1977)

Terminator

T800



(1984)

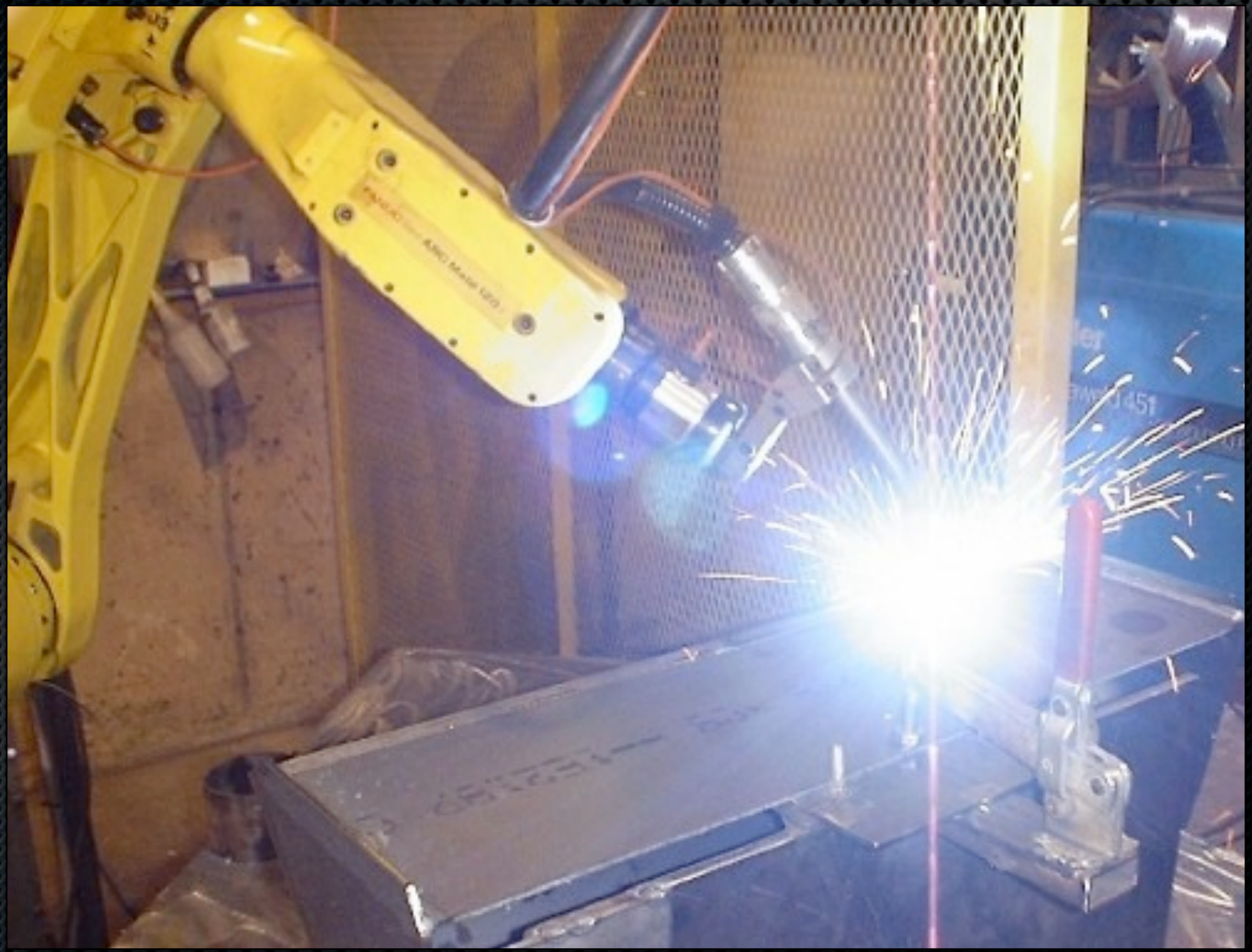
I Robot

Sonny

(2004)

... to realities

Industrial Robots



Automatic Planes

Chemical Plants

Distribution Grids

Computerised cars

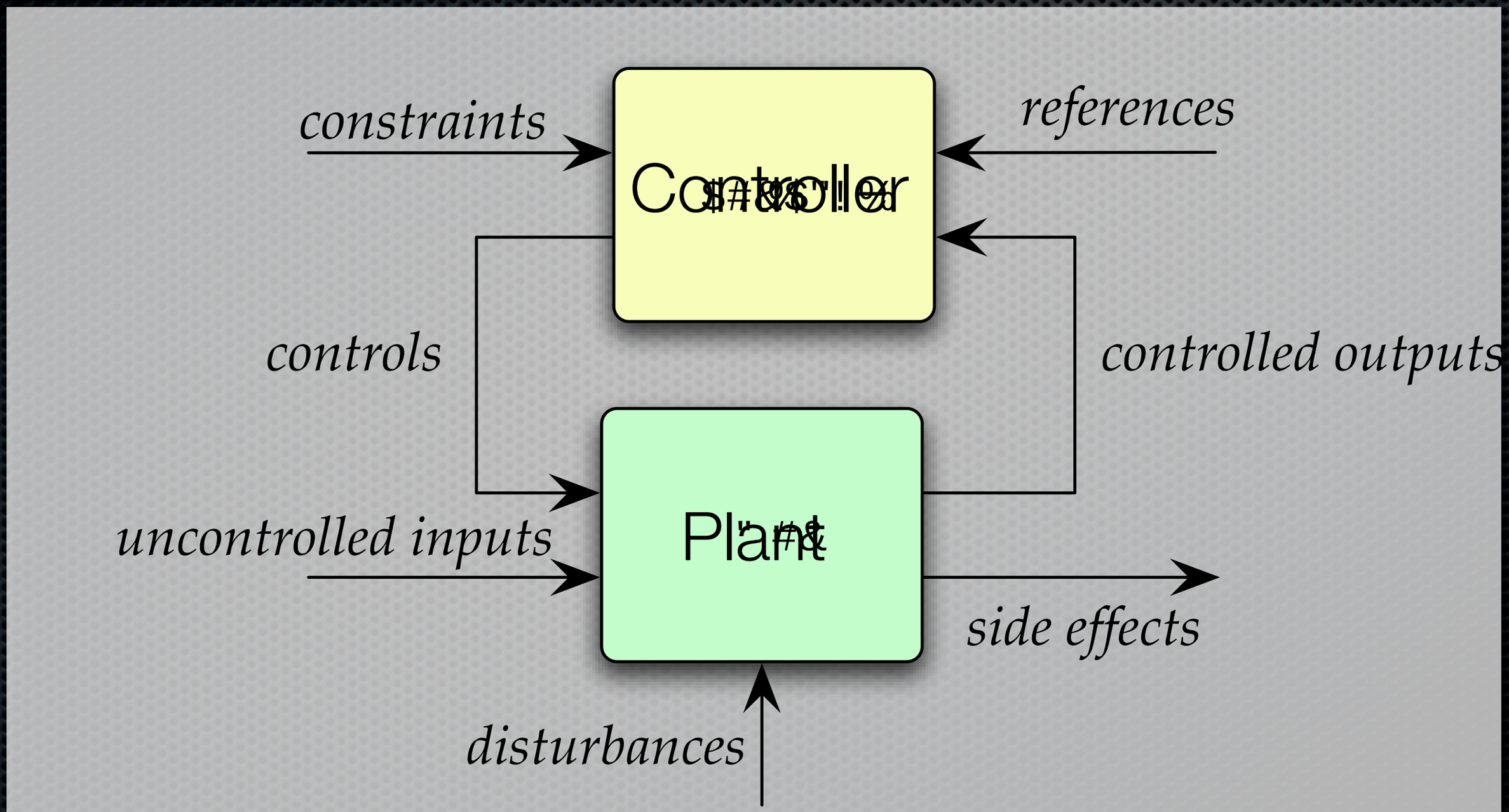
Mars Rovers

Control ... for what ?

Pieces in a controlled system

- The **Plant**
  - The system that performs the task we're interested in
    - A plane, a chemical reactor, a router, etc.
- The **Environment**
- The **Controller**
  - The system that forces the plant to follow desirable trajectories
    - The FCS, the PID, etc.

Control ... for what ?



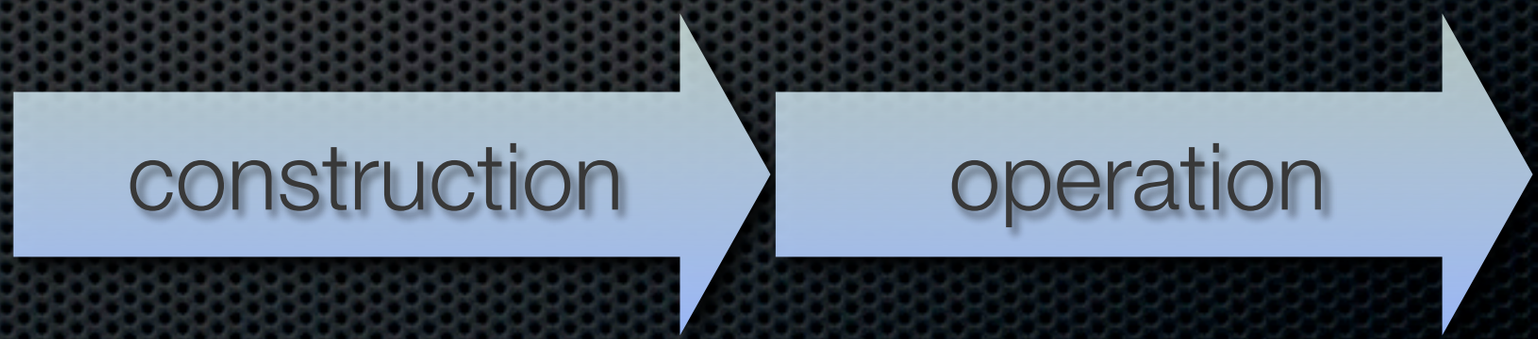
Control engineering process

1. **Specify** desired plant behaviour
2. **Model** body and environment
3. **Design** controller
4. **Build** (normally using ICT)
5. **Deploy**
6. **Tune** (continuously !)
7. **Exploit**
8. **Decommission** when obsolete

Two classes of problems

- **Run-time** problems
  - What shall the controller do to handle disturbances?

- **Construction-time** problems
  - How to design and build such controller?



Two knowledge batteries

engineering knowledge

problems
problems
problems

construction

operation knowledge

problems
problems
problems

operation

The basic desideratum is achieving

Robust autonomy

Make a system able to **solve its own problems**

Self-X

Self-Control

Controllers that Control Themselves

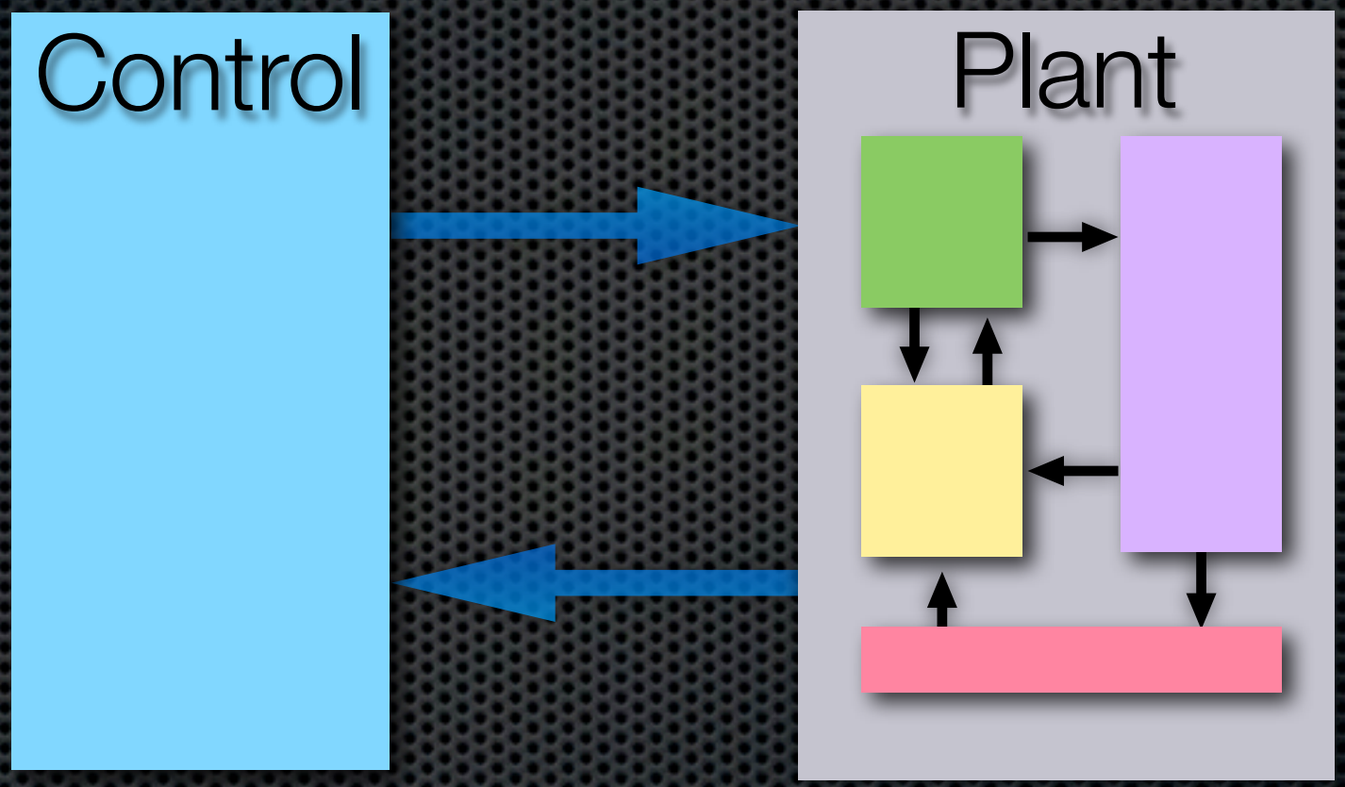
Increasing resilience

- Make **better** things (problem avoidance)
- Make **robust** things (problem tolerance)
  - Passive (Masking)
    - Transparent, fast, cost, deteriorating
  - Active (Compensating)
    - Cheaper, flexible, overhead, adaptable
- Controllers do **also suffer faults**, errors and failures

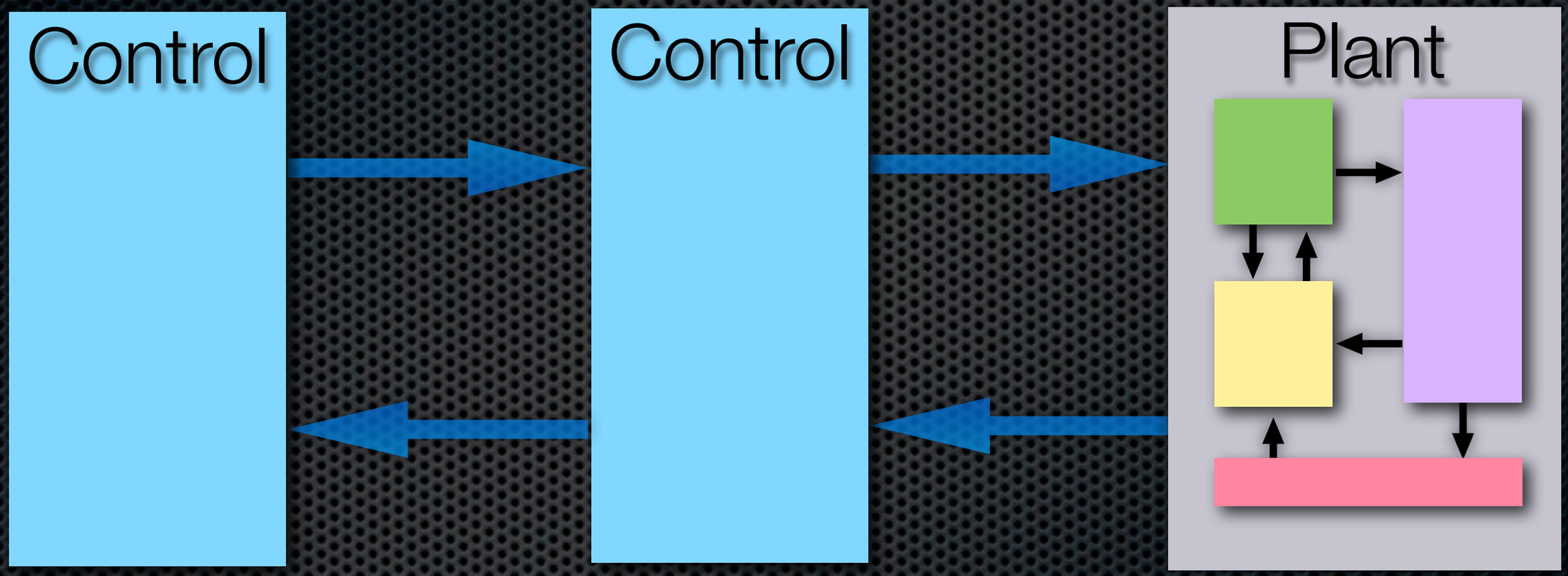
Controller resilience

- One way: Make passive-robust controllers
- Other way: Control the controller
  - Monitor, diagnose, modify the controller
- Second level controllers
  - Nested controllers
  - Adaptive controllers
  - Redundant fault-tolerant controllers
  - ... **conscious** controllers?

Control a plant

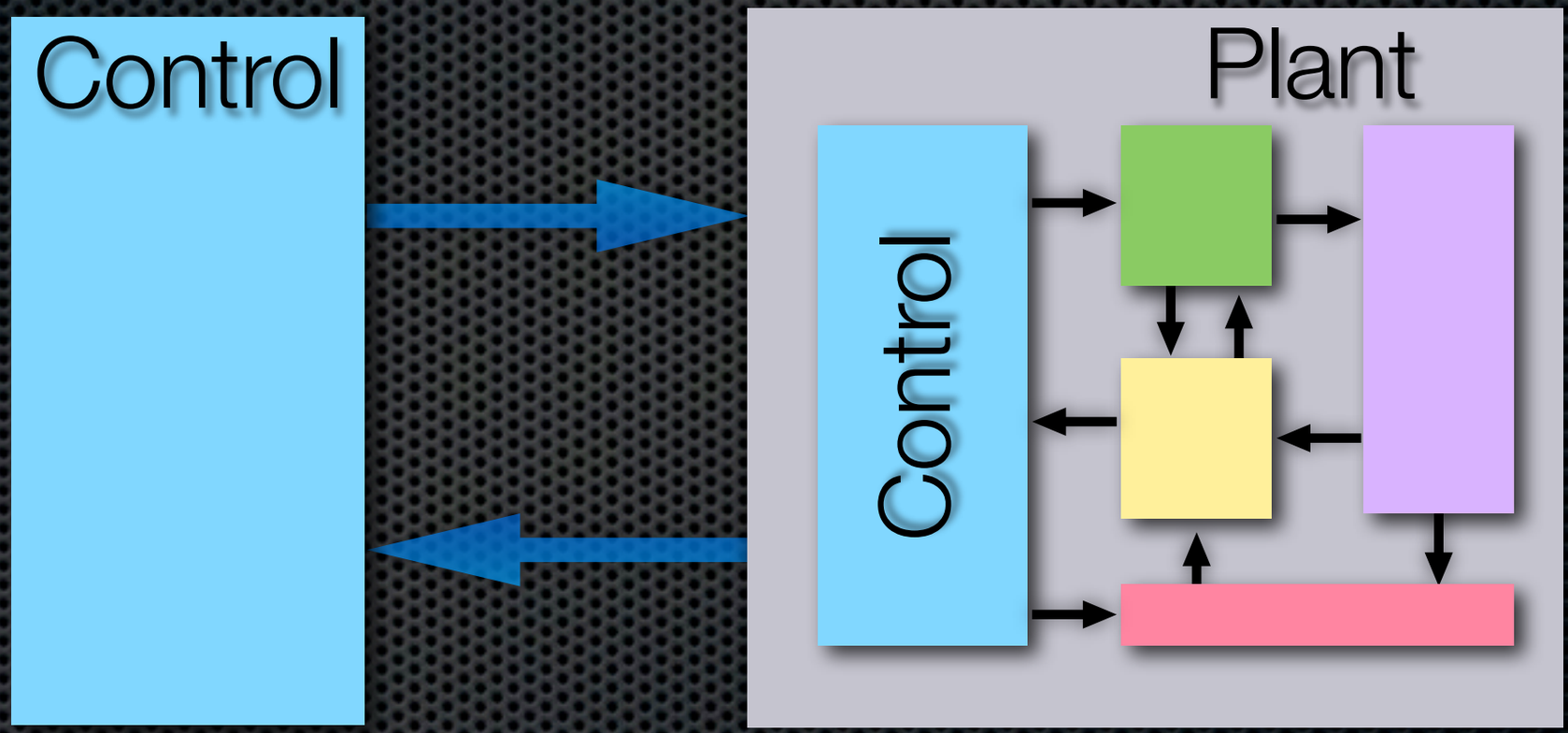


Control a controller ?



Metacognition?
Metacontrol?

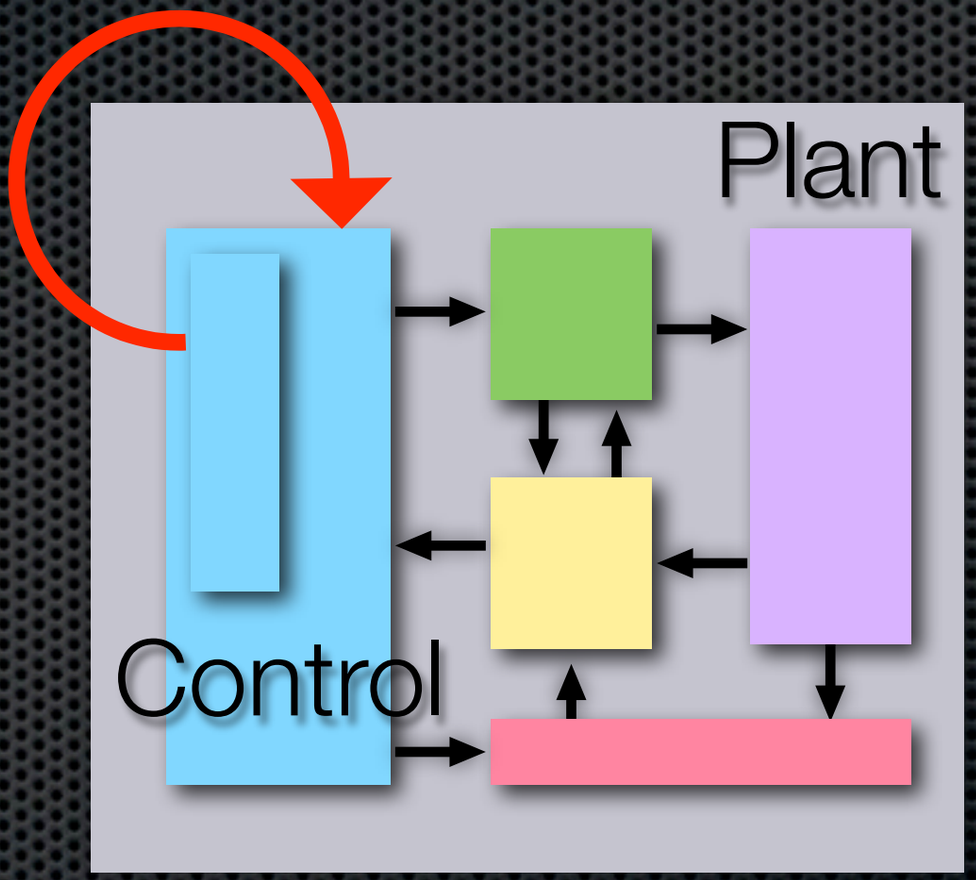
Control a controlled plant



A controlled plant is just **another kind** of plant

The key step

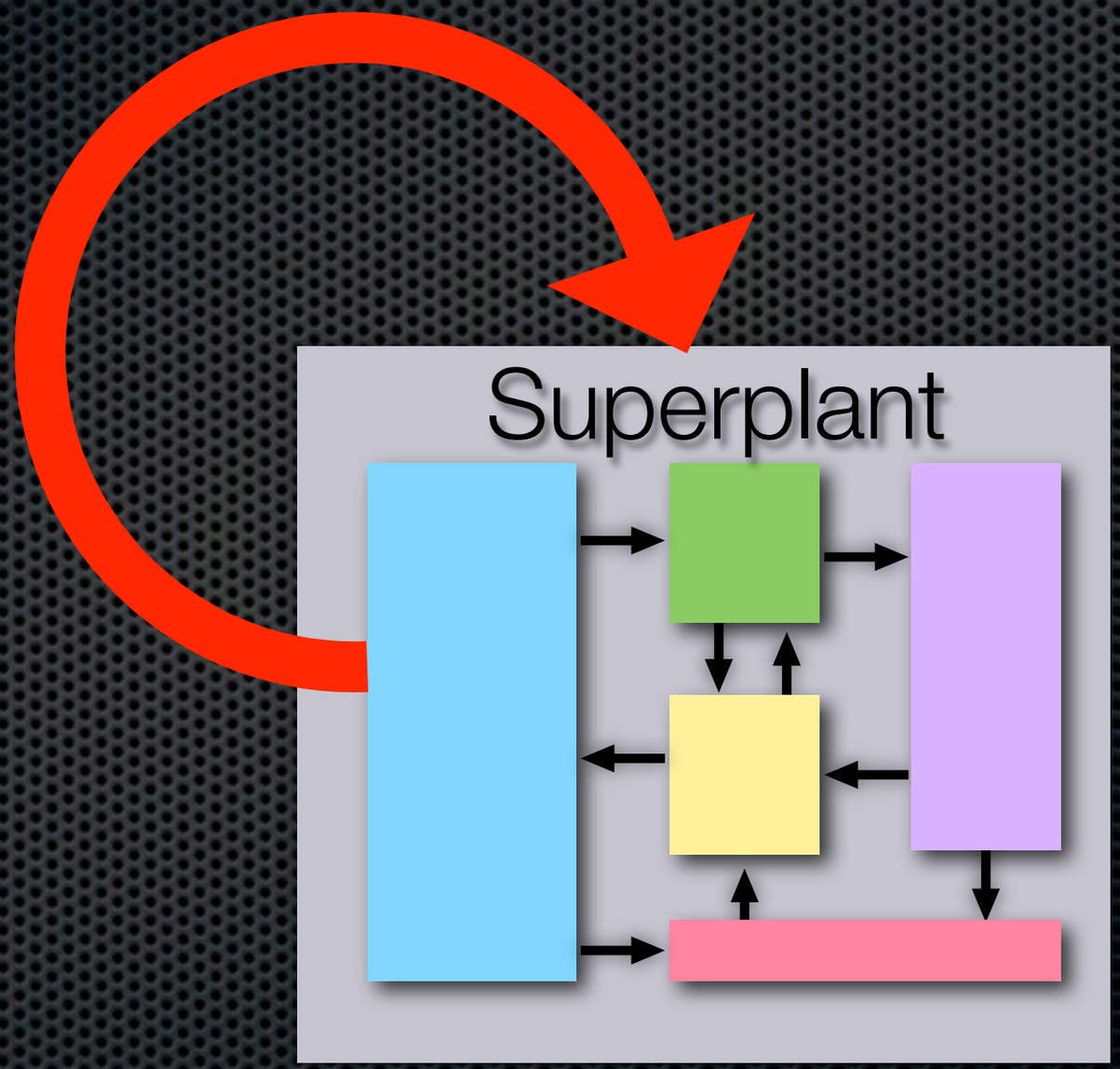
Self-loop



A way of bounding recursion in metacontrol

The key step

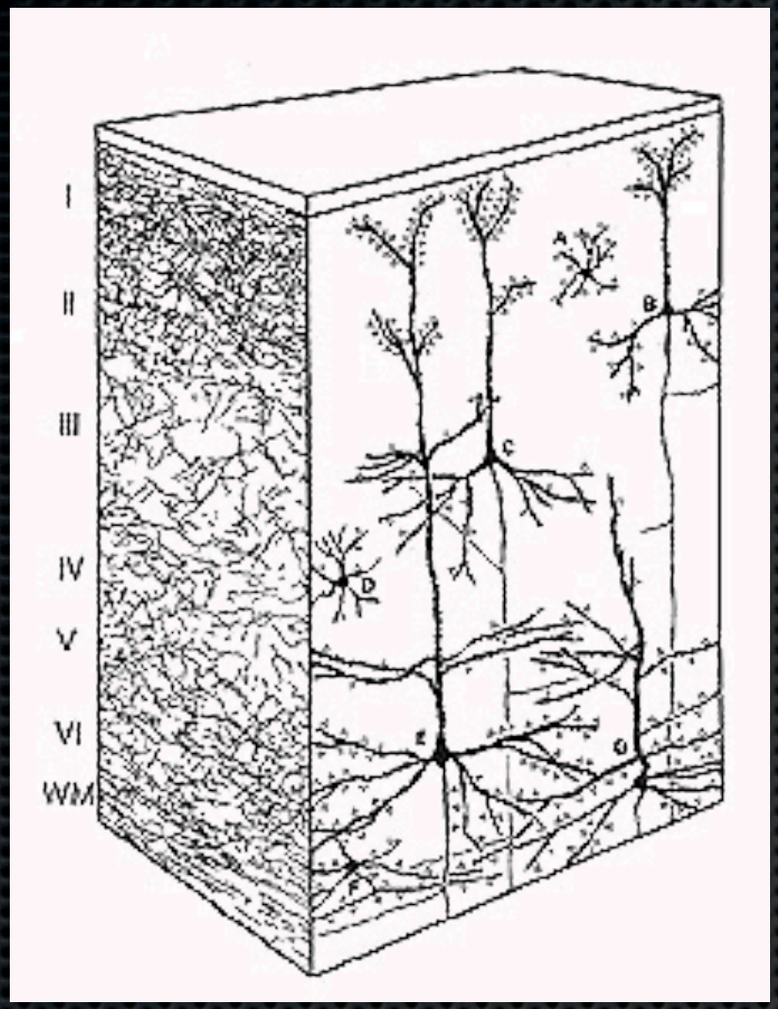
Self-loop

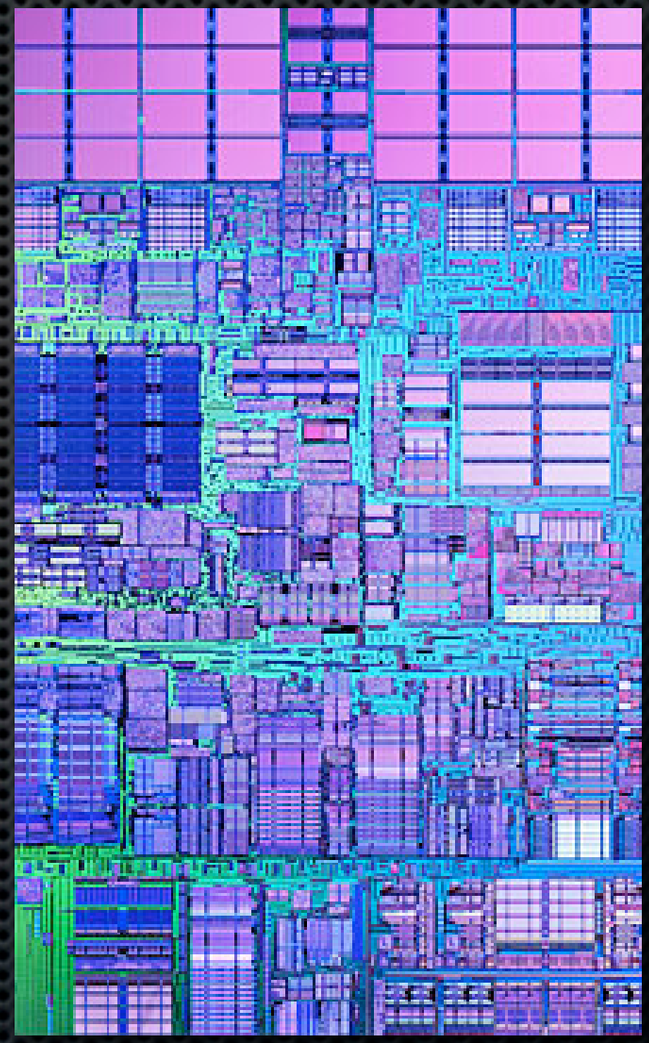


Where to look for solutions?

Cognitive science, neuroscience, etc.

Brain-inspired systems



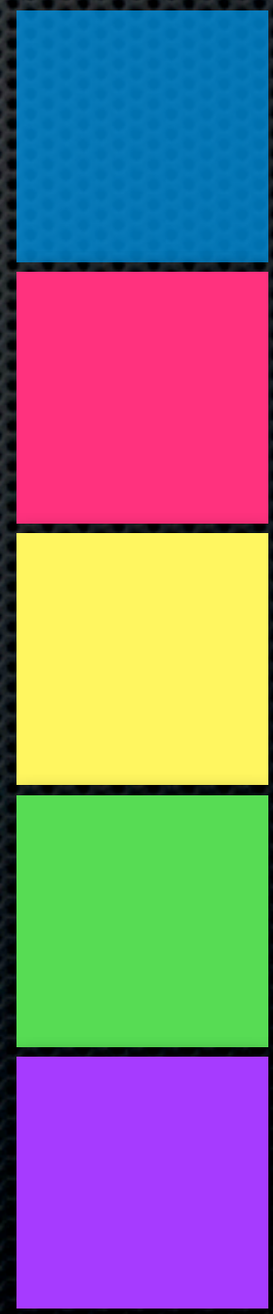


Accidental

Intentional

Consciousness ?

- Measuring the **world**
  - Access consciousness
- Experiencing the **world**
  - Phenomenology (feeling qualia)
- Experiencing the **self**
- Knowing **others'** minds



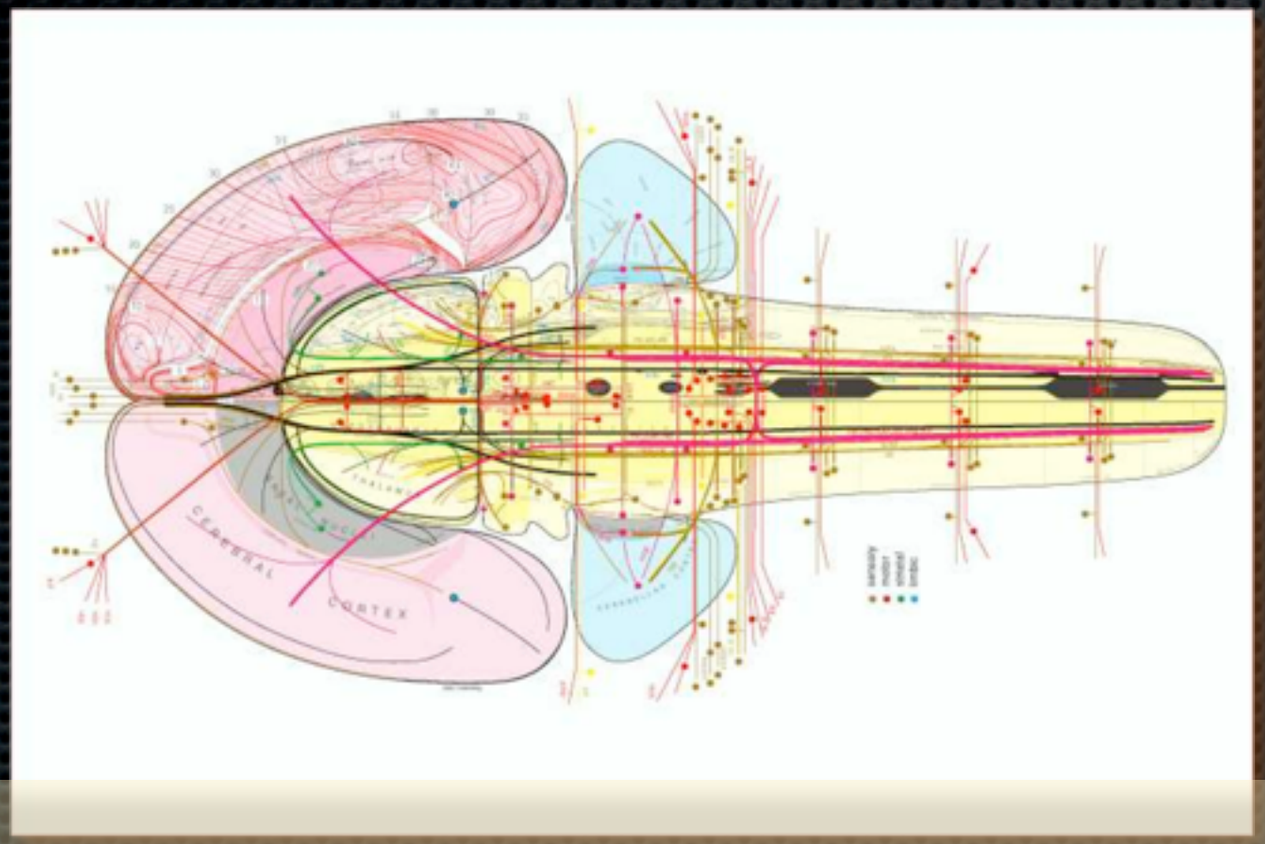
Some BIG problems

- Too many scattered foci on **biological self-awareness**
  - Psychology
  - Neuroscience
  - Philosophy
  - Anthropology
  - …

Cognitive Science

- Cognitive science **does not have a clear grasp** of the issue of consciousness;
- Less of its brain architecture

Looking at brains

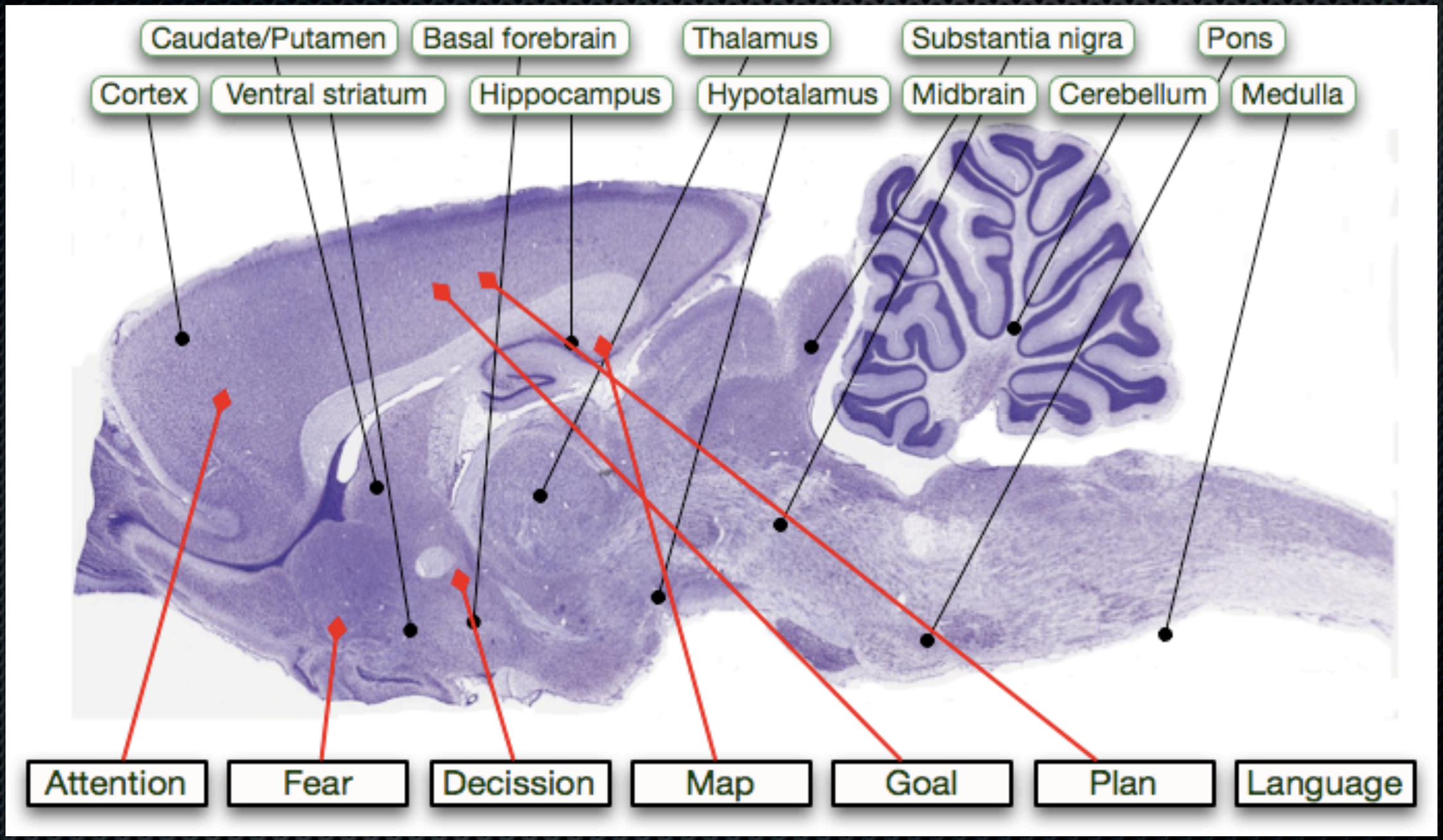


Rat brain

1 $3x10^2$ 10^5 10^{10}

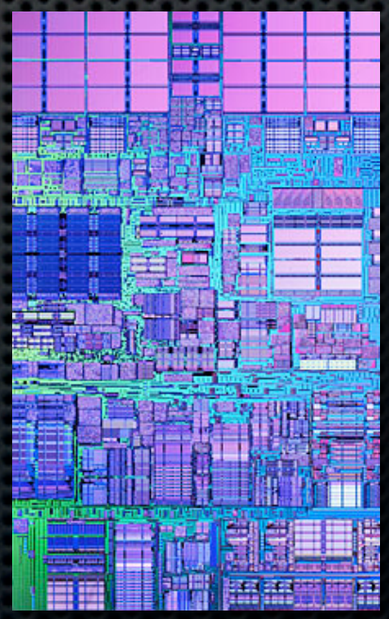
E. Coli C. Elegans D. Melanogaster H. Sapiens

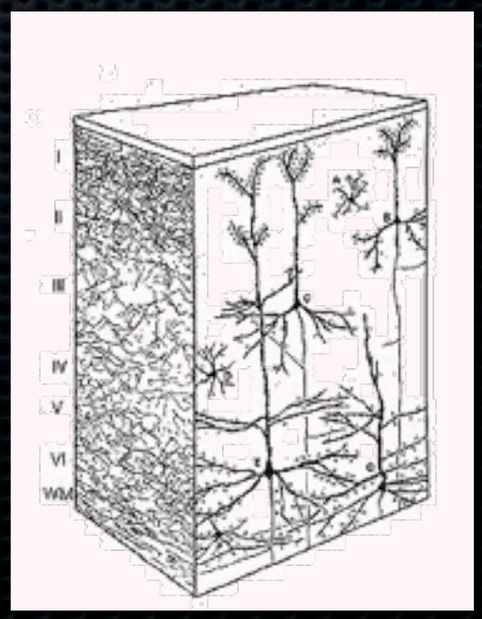
Function vs. structure



The other way round

- It seems that bio-inspiration is yet far from providing solid insights into the **architecture of self**
- At least to the level of resilience requiered by technical systems (esp. safe-critical ones)
- My impression is that what will happen is that self-aware **ICT will illuminate** cognitive science concerning these issues





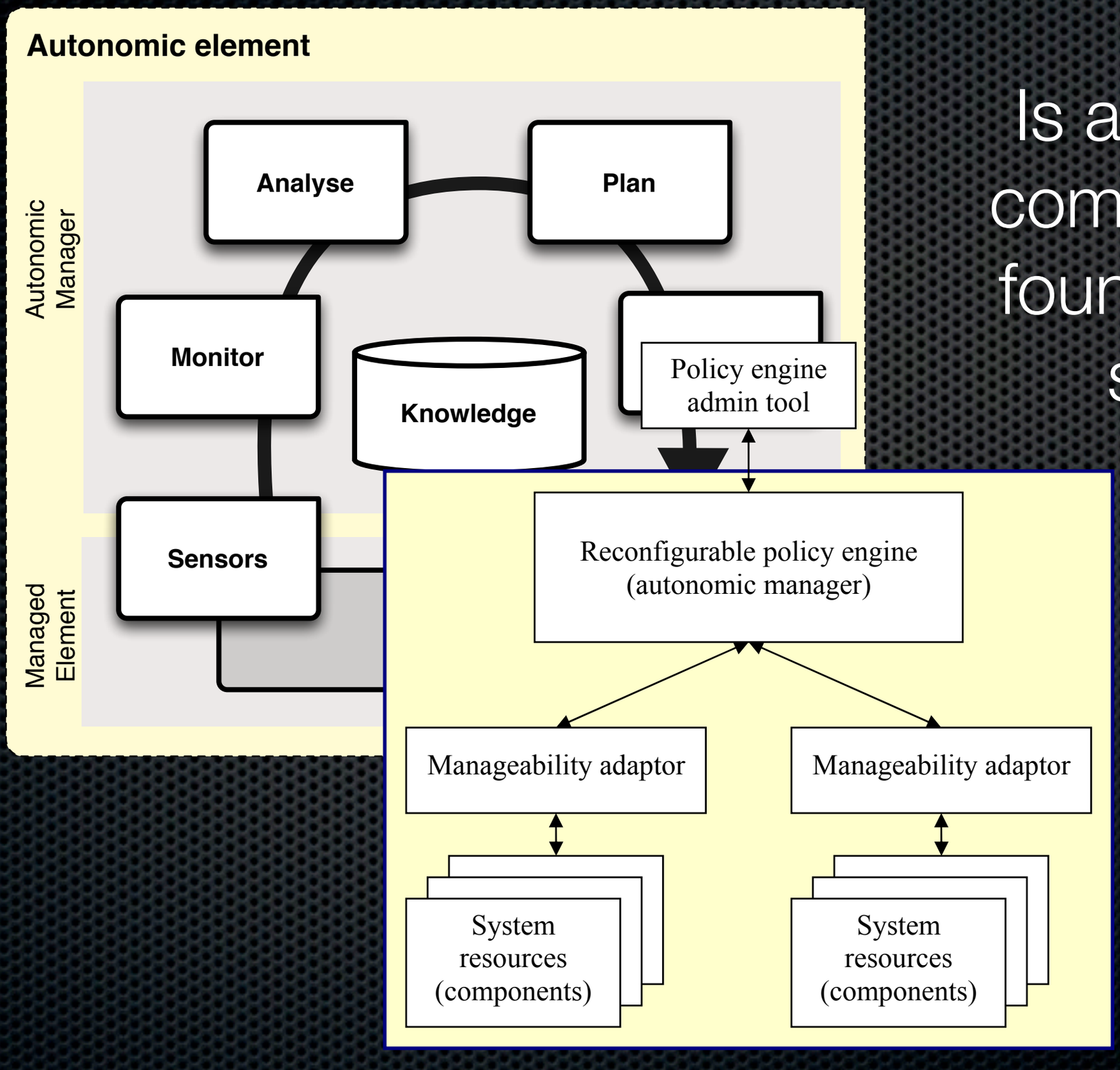
Fundamentals for Self

Basic cognitive patterns for self-awareness

An Approach to “Self”

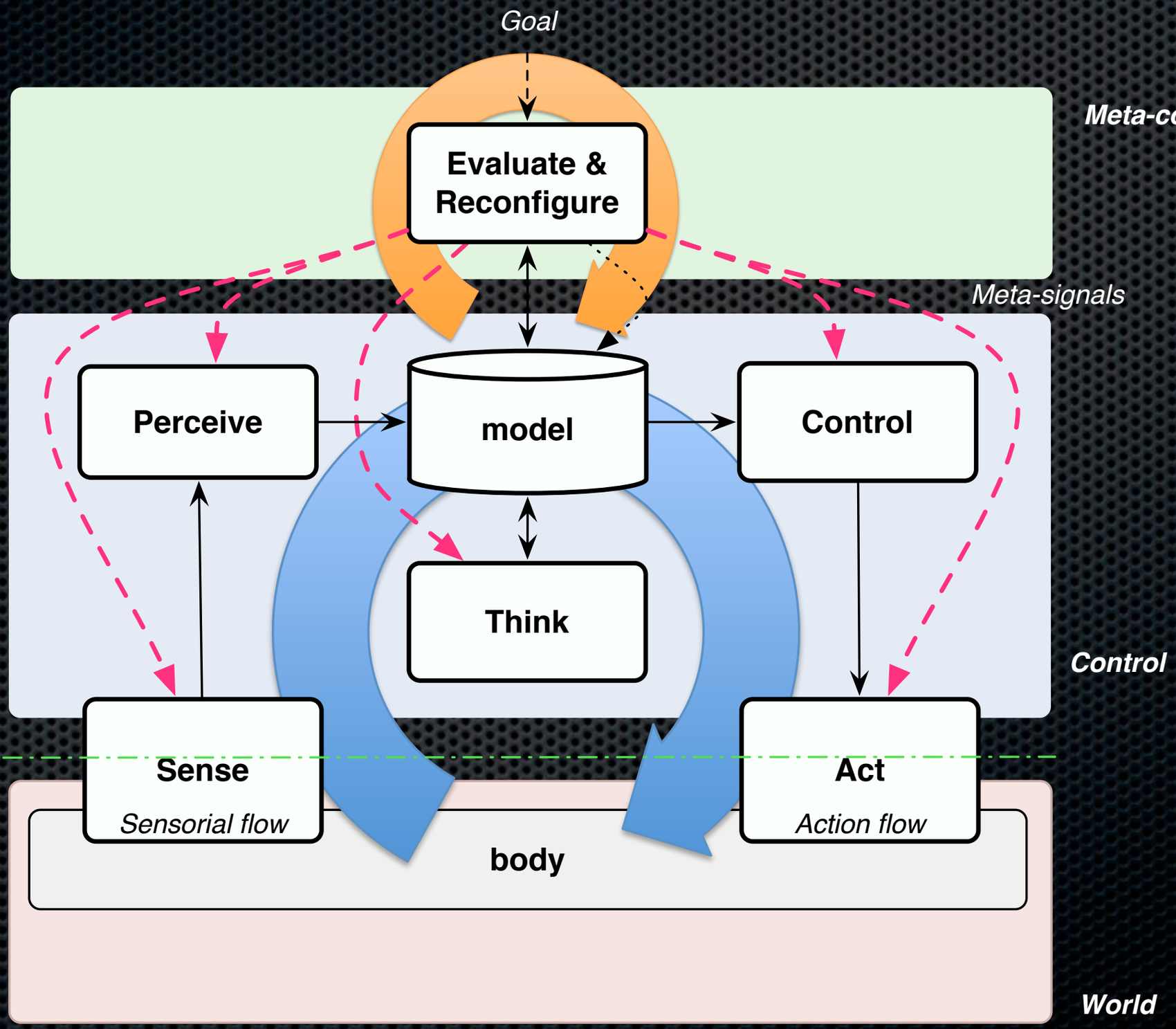
- Enhance the control capabilities -robustness- of autonomous agents by means of exploitation of self representations (both in perception and action)
- Sophisticated cognitive competences -self-consciousness- shall emerge from a generalisation and integration of metacontrol mechanisms

Autonomic computing



Is autonomic computing the foundation for self-X?

A general epistemic pattern ?



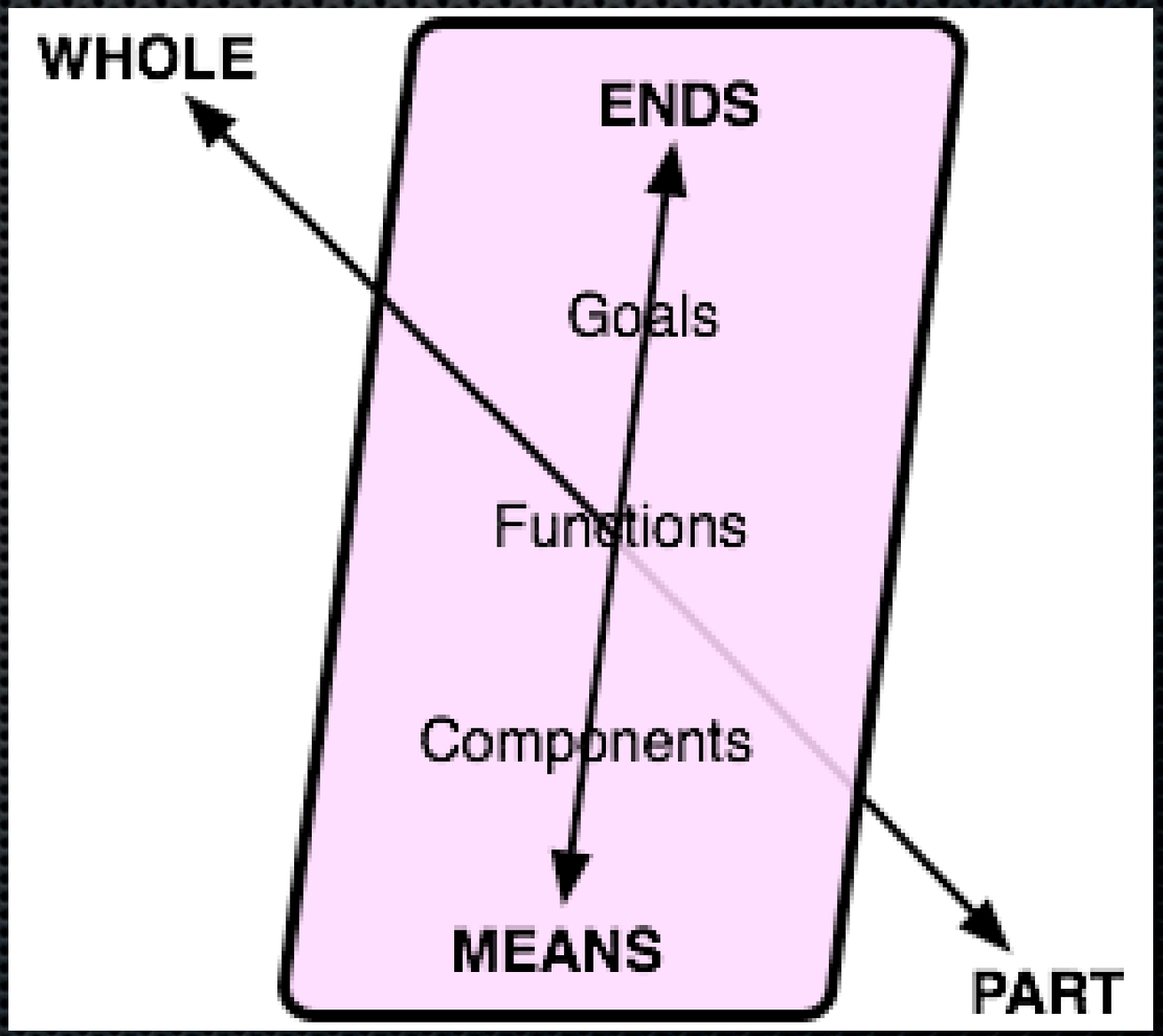
Structure and Function

- Focus on **function** instead of **structure**

- *“an object with a set of goals, in a well defined environment, exerts an activity (function) and at the same time experiences how its internal structure evolves through time keeping its identity”.*
  [Le Moigne]

The Concept of “Function”

“Some define function to be equivalent to behavior whereas others … define functions as purely intentional concepts” [Lind 1994]



Dual **Causal** / **Intentional** viewpoints on function

The core objective

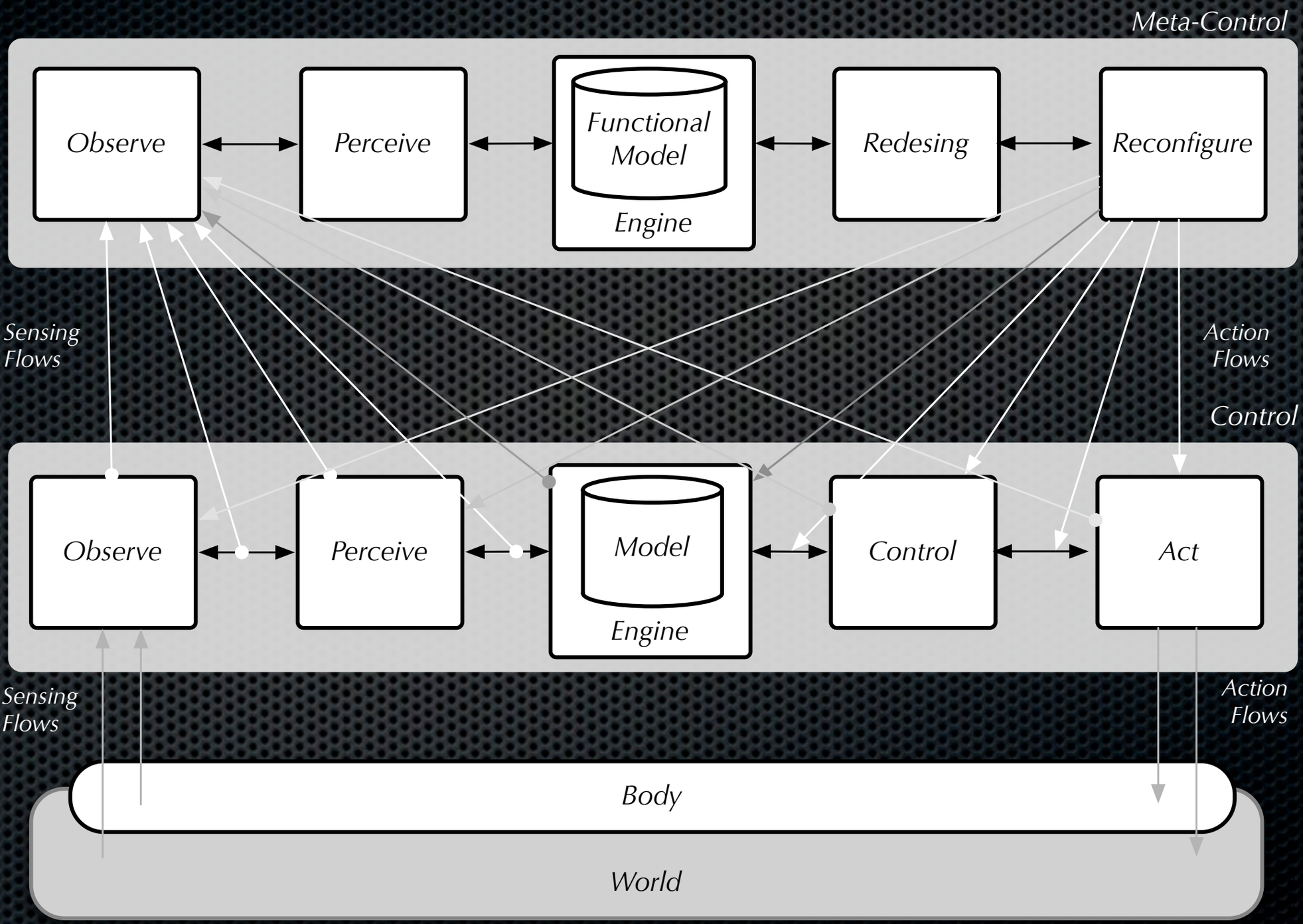
- The core objective of self-aware systems must be the provision of:

Functional robustness
in the intentional sense

The vision

- The vision is that:
  - Model-based perception of functional state + Model-based action on functional state
- Can render a system that can be
  - **Self-aware** (using a model of self in self- perception)
  - **Robust** (based on world/self model integrated use)
  - **Autopoietic** (using the self-model in a model-driven run-time synthesis)

Cognitive functional metacontrol



Consciousness as function

- Control mechanisms organise into hierachies/ heterarchies of functional assemblies
- Layering happens: e.g. Reflexes-Drives-Instincts-Cognitions
- Metacontrollers (functionally) perceive and (functionally) organise assemblies to maximise value expectancy (the core of emotion understanding)
- (Emotional) metacontrol mechanisms scale up to consciousness

Principles

- **Model-based cognition**: A cognitive system exploits models of other systems in their interaction with them.
- **Model isomorphism**. An embodied, situated, cognitive system is as good performer as its models are.
- **Anticipatory behavior**. Maximal timely performance is achieved using predictive models.

Principles

- **Unified cognitive action generation**. Generate action based on an integrated, scalable, unified model of task, environment and self in search for global performance maximisation.
- **Model-driven perception**. Perception is realised as the continuous update of the integrated models used by the agent in a model-based cognitive control architecture by means of real-time sensorial information.

Principles

- **System awareness**. An aware system is continuously perceiving and computing meaning - future value- from the continuously updated models.
- **System self-awareness**. A conscious system is continuously generating meanings from continuously updated self-models in a model-based cognitive control architecture.

Principles

- **Emotional metacontrol**: Emotional systems help reconfigure the system to maximise present/future value using control system patterns.
- **System attention**: Attentional mechanisms allocate both physical and cognitive resources for system perceptive and modelling processes so as to maximise performance.

The Concept of “Autonomy”

- This dual analysis of function [causal / intentional] gives a clue for the elusive understanding of “autonomy”

- **Autonomy** = Ability to keep intentional function while suffering causal function change (functional robustness)

- All control levels -autonomic, cognitive, emotional- may contribute to this.

Two system epochs

problems
problems
problems

engineering knowledge

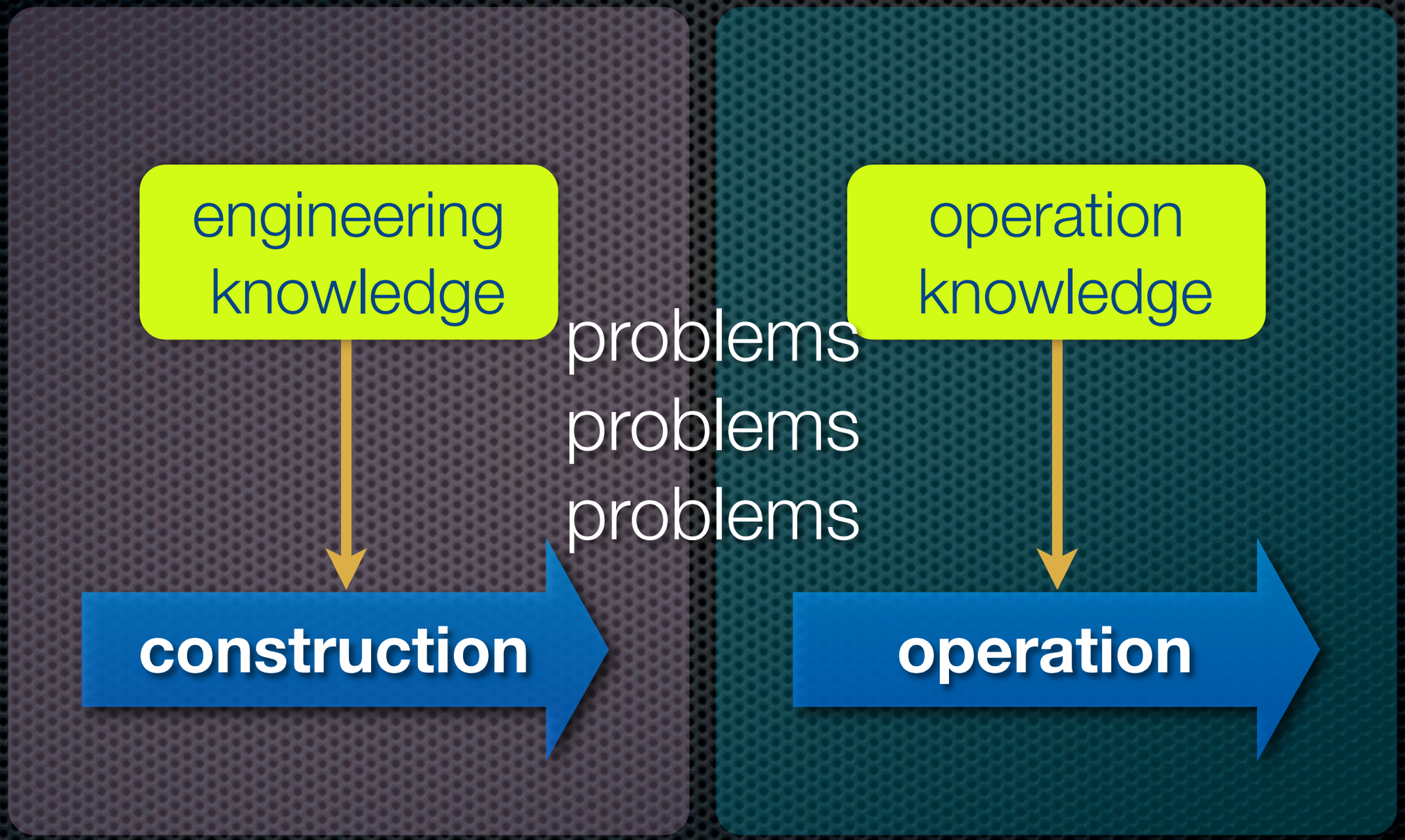
construction

operation knowledge

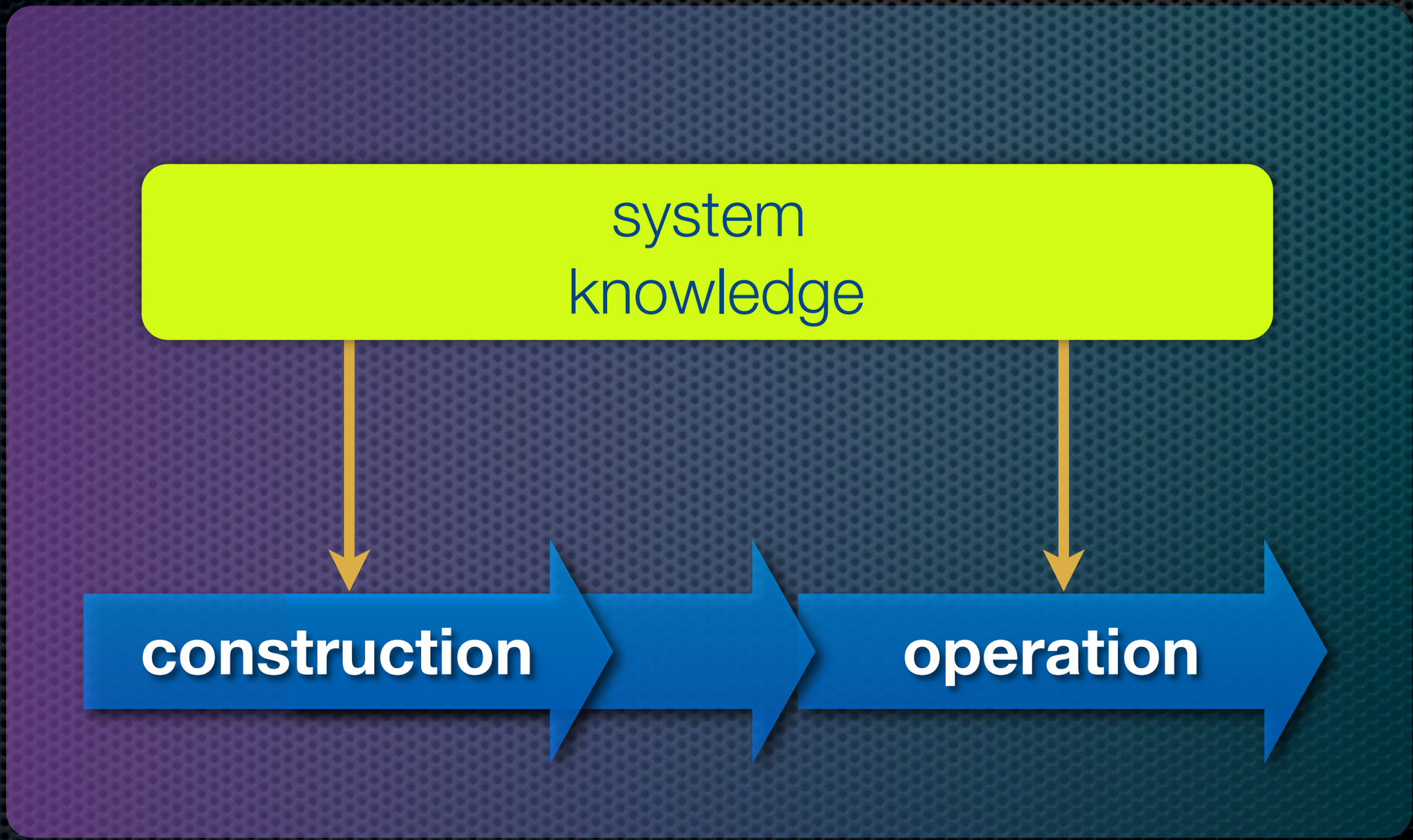
operation

problems
problems
problems

A single problem set



Removing the gap



Self-engineering systems (predictable autopioetic)

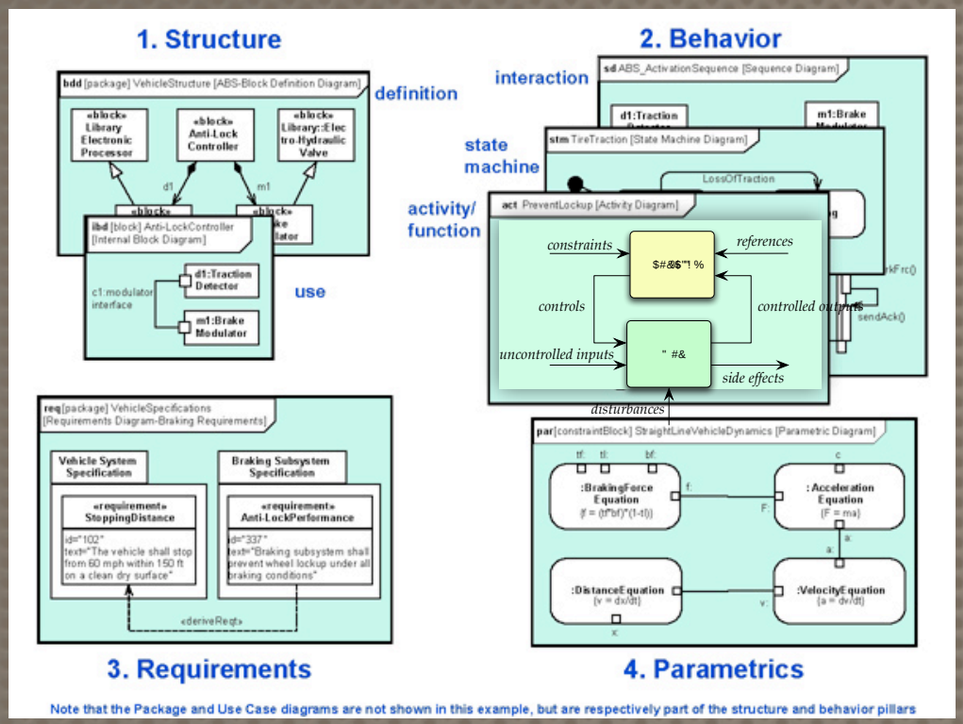
Formalise and reify architectural concepts

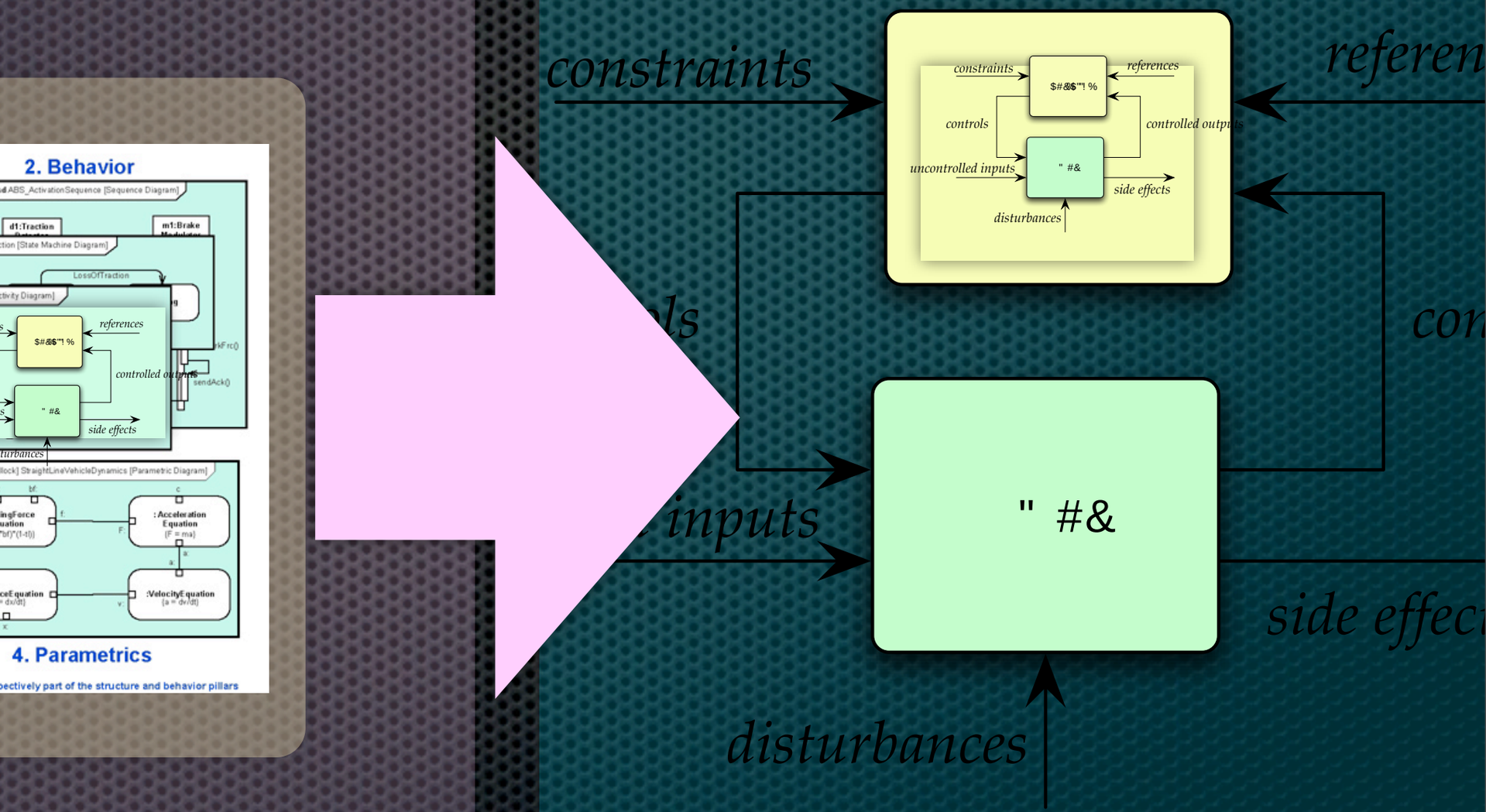
- This vision requieres the **formalistation of core architectural design concepts**
- E.g.: action, agent, algorithm, coupling, environment, component, function, grounding, object, goal, percept, organisation, property, relation, model, resource, structure, structure, system, time-invariant relation, effect, autonomy, pattern, state, role, behaviour, constraint, etc.

Themes to consider

- Real-time control (sense-perceive-act)
  - Ontology of control tasks/objects
- Deep/wide Integration
  - Sharing ontologies horizontally and vertically
- Self-management
  - Explicit self model-based control
- Model Centric Engineering
  - Software engineering for Self-awareness

from **system models** *to* **model-based reflective systems**





A final proposal

Robust Autonomy by Model-based Self-Awareness

Establish the grounds for predictable self-engineering of systems for functional robustness