Cognitive semantics and cognitive theories of representation:

Session 9: Meanings in artificial systems

Martin Takáč
Centre for cognitive science
DAI FMFI Comenius University in Bratislava

Príprava štúdia matematiky a informatiky na FMFI UK v anglickom jazyku
ITMS: 26140230008

Approaches to AI

- Philosophical
  - Can an artificial system understand?
  - What is the nature of meaning in it – can they be made intrinsic?
- Modeling
  - How to model mind / cognitive functions?
  - “Is a model of” is a weaker claim than “is”.
  - Usually partial models
- Applicational
  - How to make effective cognitive technologies?
  - Little interest in the “nature” of meaning and understanding

Intelligence

- the ability to behave appropriately under unpredictable conditions
- the ability to achieve complex goals in complex environments
  - Behaving “appropriately” is a matter of achieving organismic goals, such as getting food, water, sex, survival, status, etc. Doing so under unpredictable conditions is one thing that makes the achievement of these goals complex.
  - Source: Goertzel & Pennachin (2007)

Artificial intelligence

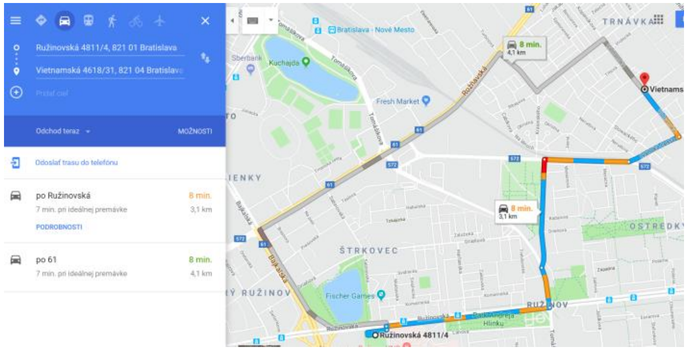
- “Artificial intelligence (AI) refers to systems that display **intelligent behaviour** by **analysing their environment** and **taking actions** – with some degree of **autonomy** – to achieve specific **goals**.
- AI-based systems can be purely software-based, acting in the virtual world (e.g. **voice assistants, image analysis software, search engines, speech** and **face recognition systems**) or AI can be embedded in **hardware devices** (e.g. advanced **robots, autonomous cars, drones** or **Internet of Things** applications).” (European Commission’s Communication on AI, 2018)

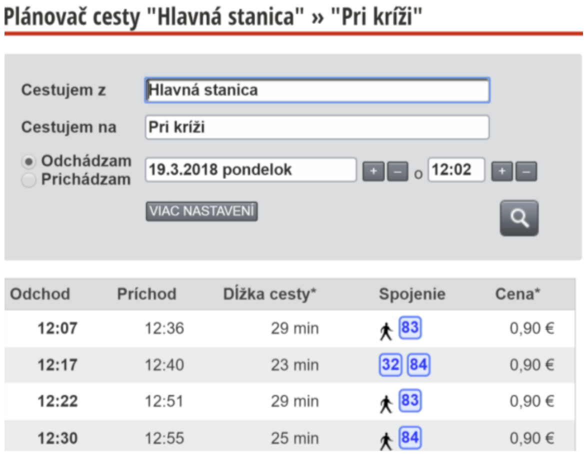
Artificial intelligence

- ‘AI system’ means a system that is either software-based or embedded in hardware devices, and that displays **behaviour simulating intelligence** by **collecting and processing data**, analysing and interpreting its environment, and by **taking action**, with **some degree of autonomy**, to achieve specific goals (EU parliament)

Intelligent technologies are ubiquitous

- Predictive texting in sms
- Automatic translation
- Intelligent web search
- Route/connection planners
- GPS navigation
- Intelligent hoover
- Computer viruses and antiviruses
- …





Odchod	Príchod	Dĺžka cesty*	Spojenie	Cena*
12:07	12:36	29 min	83	0,90 €
12:17	12:40	23 min	32 84	0,90 €
12:22	12:51	29 min	83	0,90 €
12:30	12:55	25 min	84	0,90 €

Intelligent technologies are ubiquitous

- … and much more

Types of AI

- **“narrow AI”**: programs **carrying out specific tasks** like playing chess, diagnosing diseases, driving cars and so forth (most contemporary AI work falls into this category.)
- **AGI:** systems that possess a reasonable degree of **self-understanding** and **autonomous self-control**, and have the **ability to solve a variety of complex problems** in a variety of contexts, and **to learn to solve new problems** that they didn’t know about at the time of their creation.

Artificial general intelligence

- AGI has ability to:
  - Autonomously and interactively **acquire new knowledge** and skills in real time, **make decisions** with limited information and in uncertainty.
  - **Use language, judge contextually**, logically and abstractly, explain its conclusions.
  - Remember recent events and interactions, understand the meaning of actions (even observed – **theory of mind**).
  - Dynamically **manage multiple potential conflicting objectives**, be able to select relevant stimuli and focus on relevant tasks.
  - Recognize and **respond adequately to human emotions** (EQ) and understand one's own cognitive states (**introspection**).

Approaches to achieve AGI

- symbolic
- probability- or uncertainty-focused
- neural net-based
- evolutionary
- artificial life
- program search based
- integrative

Symbolic approach

- **Good old-fashioned AI (GOFAI)** (1950-80
  - Rule based systems, predicate logic

INROOM(ROBOT,R1)

CONNECTS(D1,R1,R2)

CONNECTS(D2,R2,R3)

BOX(BOX1)

INROOM(BOX1,R2)

$(\forall x \forall y \forall z)[\text{CONNECTS}(x, y, z) \Rightarrow \text{CONNECTS}(x, z, y)]$

Operator:

GOTHRU(d, r1, r2)

Preconditions:

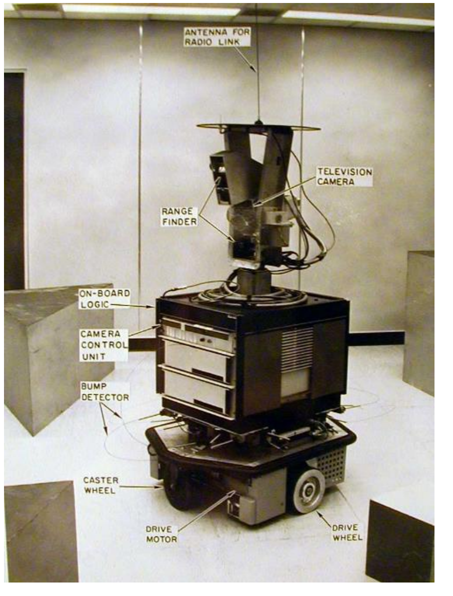
INROOM(ROBOT, r1) $\wedge$ CONNECTS(d, r1, r2)

Delete List:

INROOM(ROBOT, $)

Add List:

INROOM(ROBOT, r2)



Robot SHAKEY (SRI, 1966-72)

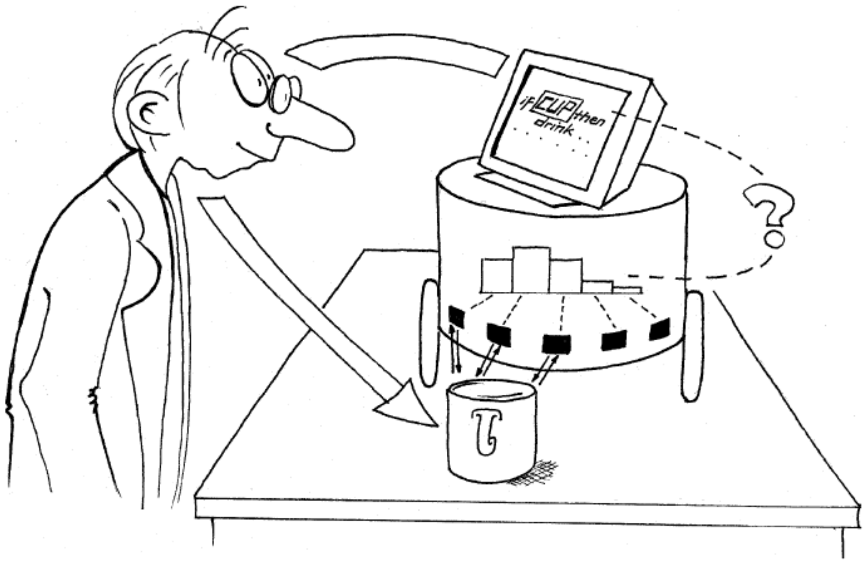
Symbolic approach

- General problem solver (Simon, Shaw, Newell, 1959)
- CYC (Lenat, since 1984)
- SOAR cognitive architecture (Laird et al, since 1983)

Critique of the symbolic approach

- Harnad: Symbol grounding problem
- Searle: Chinese Room argument

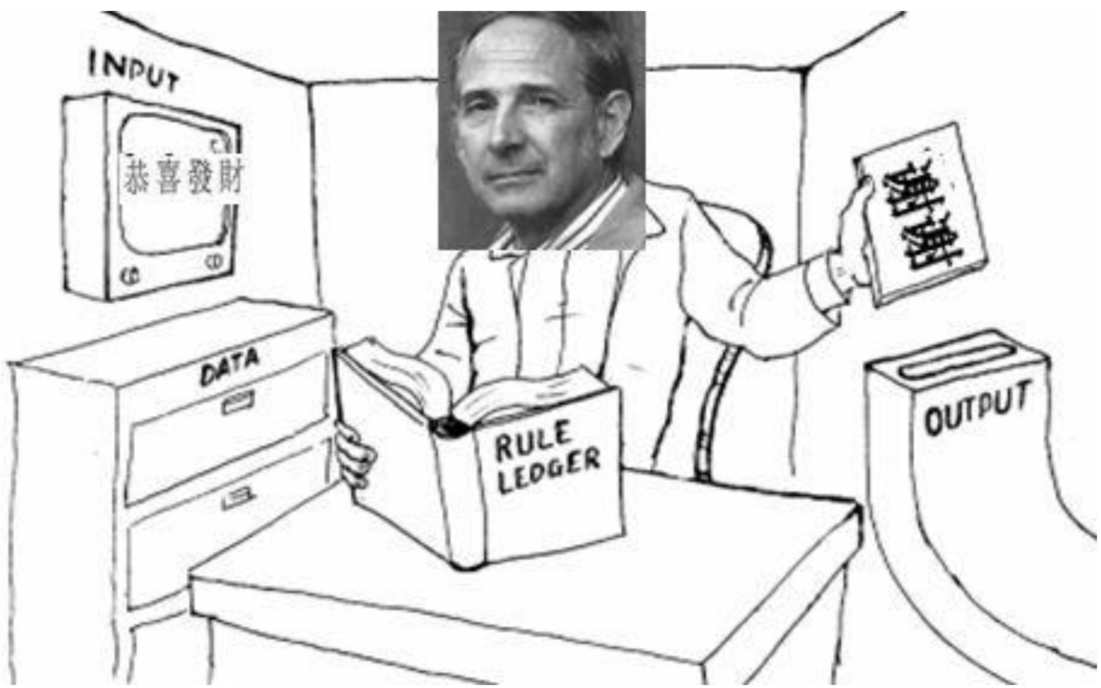
Symbol grounding problem



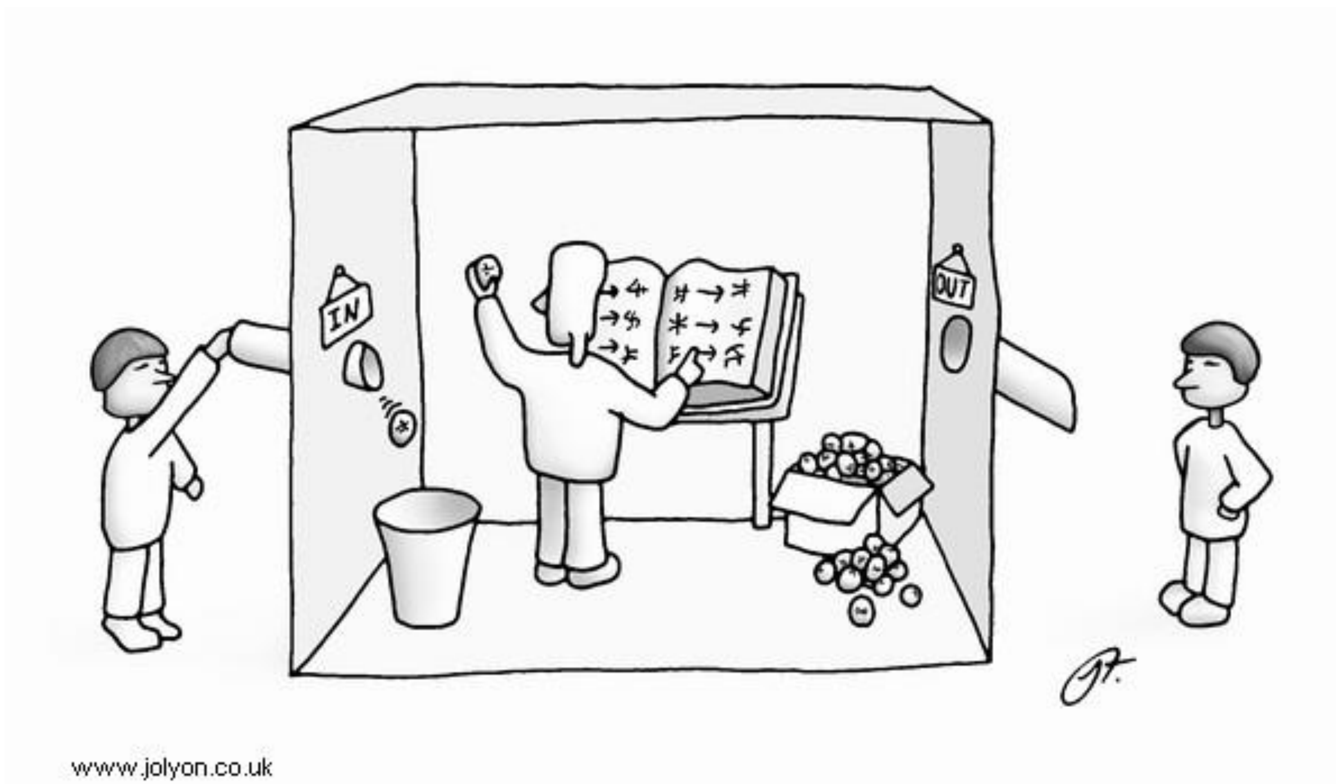
How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads? How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?

Chinese room argument (CRA)

- Searle (1980)
- Argument against strong AI



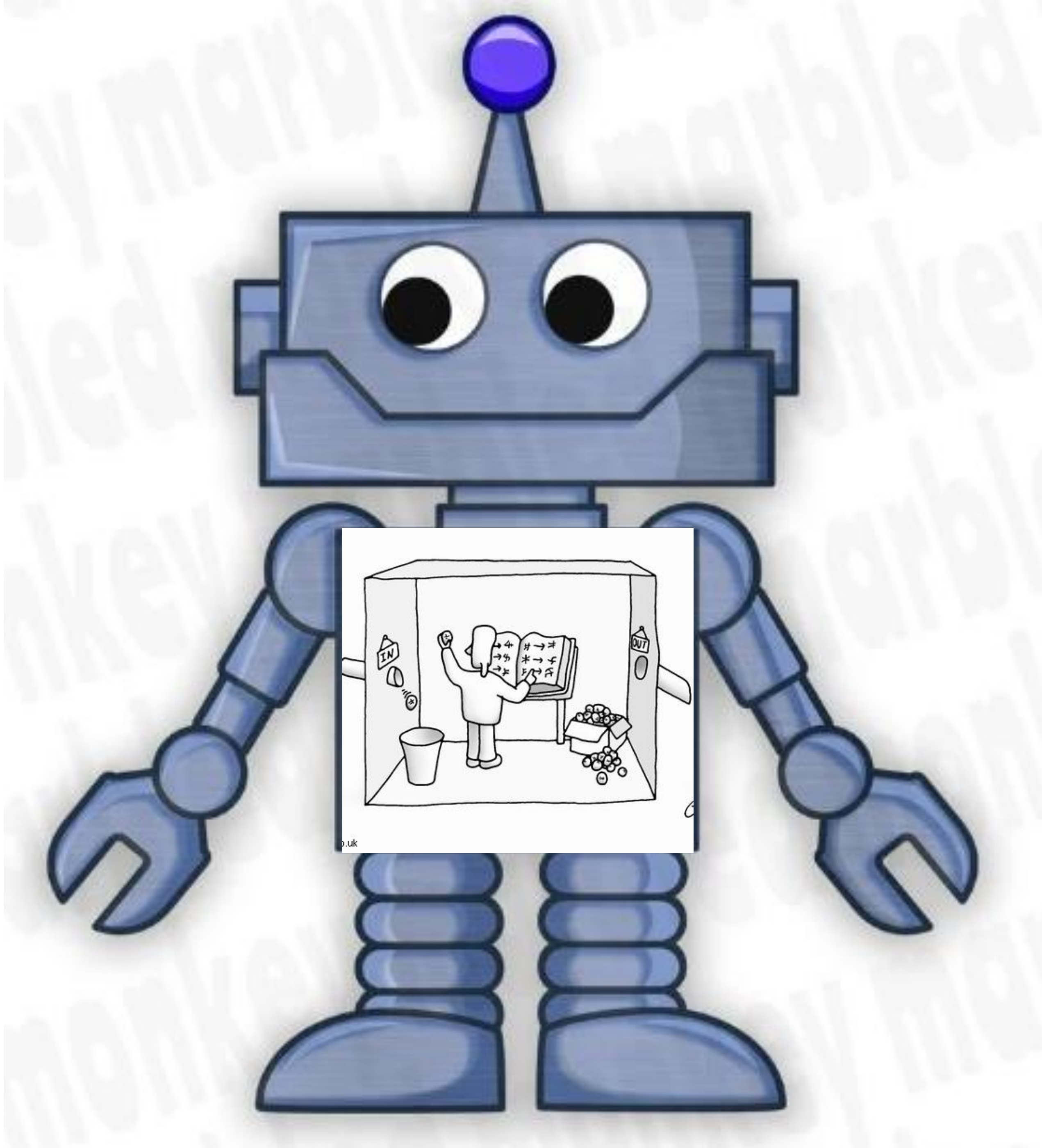
Chinese room argument (CRA)



CRA - The robot reply (Yale)

- "Suppose we wrote a different kind of program from Schank's program. Suppose we put a computer inside a robot, and this computer would not just take in formal symbols as input and give out formal symbols as output, but rather would actually operate the robot in such a way that the robot does something very much like perceiving, walking, moving about, hammering nails, eating drinking -- anything you like. The robot would, for example have a television camera attached to it that enabled it to 'see,' it would have arms and legs that enabled it to 'act,' and all of this would be controlled by its computer 'brain.' Such a robot would, unlike Schank's computer, have genuine understanding and other mental states."

- Robot picture source



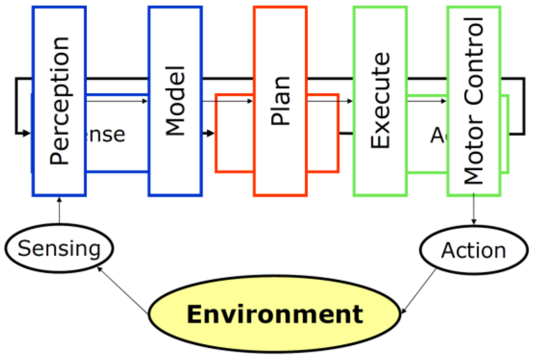
[Robot picture source](#)

Nouvelle-AI

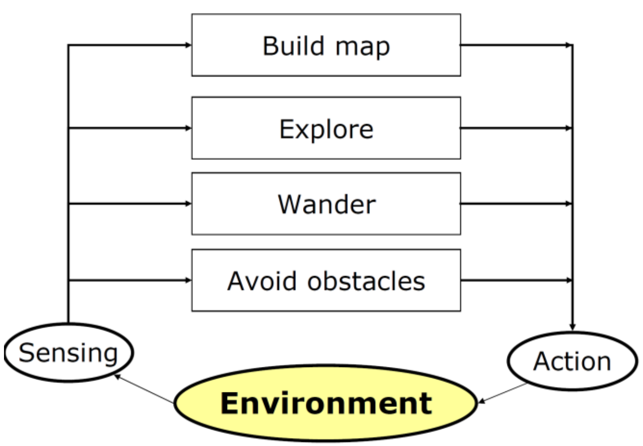
- Shift from GOFAI to nouvelle-AI
  - Complete agents: Autonomy, embodiment and situatedness
- Intelligence without representation (Brooks, 1991). "*The world is its own best model.*" R. Brooks
  - example: Mechanical lady bug
  - subsumption architecture (Brooks, 1992) – designer approach

Nouvelle-AI

- HorizontalDecomposition – centralised control

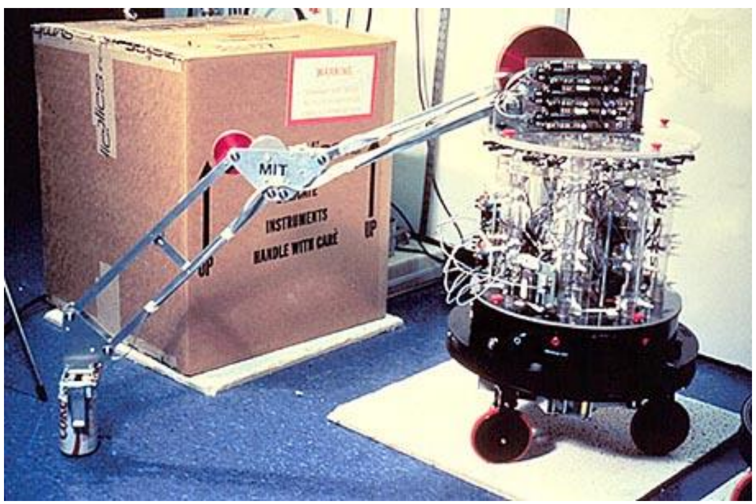


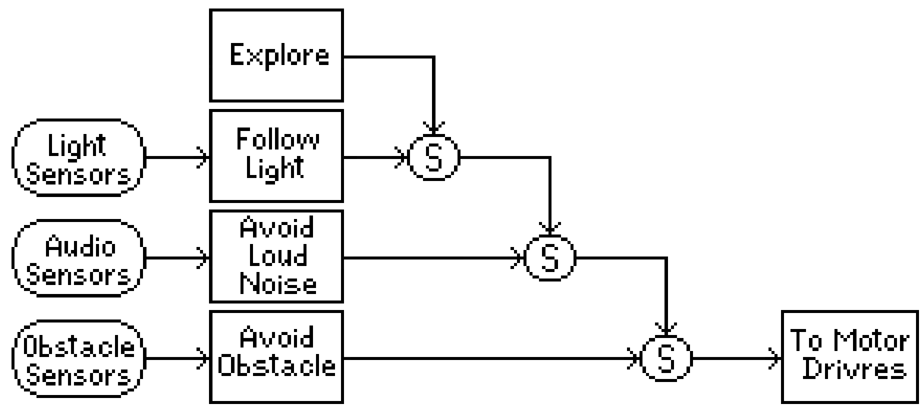
- Vertical decomposition – layered design



Nouvelle-AI

- Subsumption architecture (Brooks, 1992)



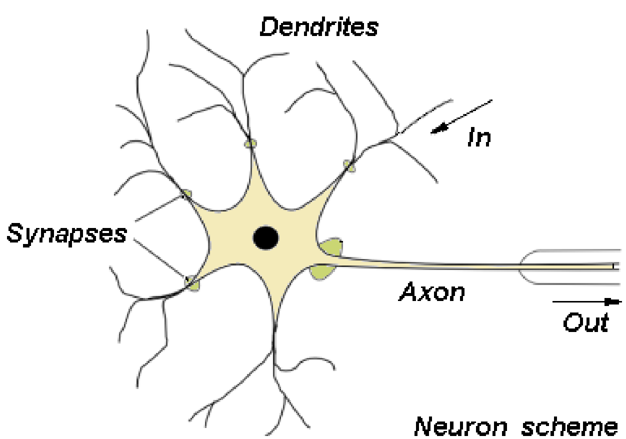


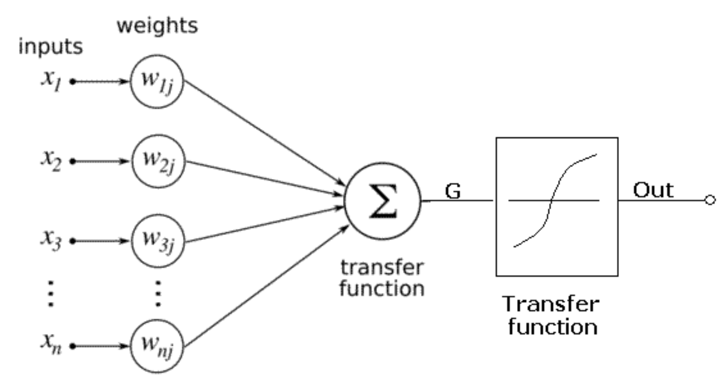
CRA - Developmental reply

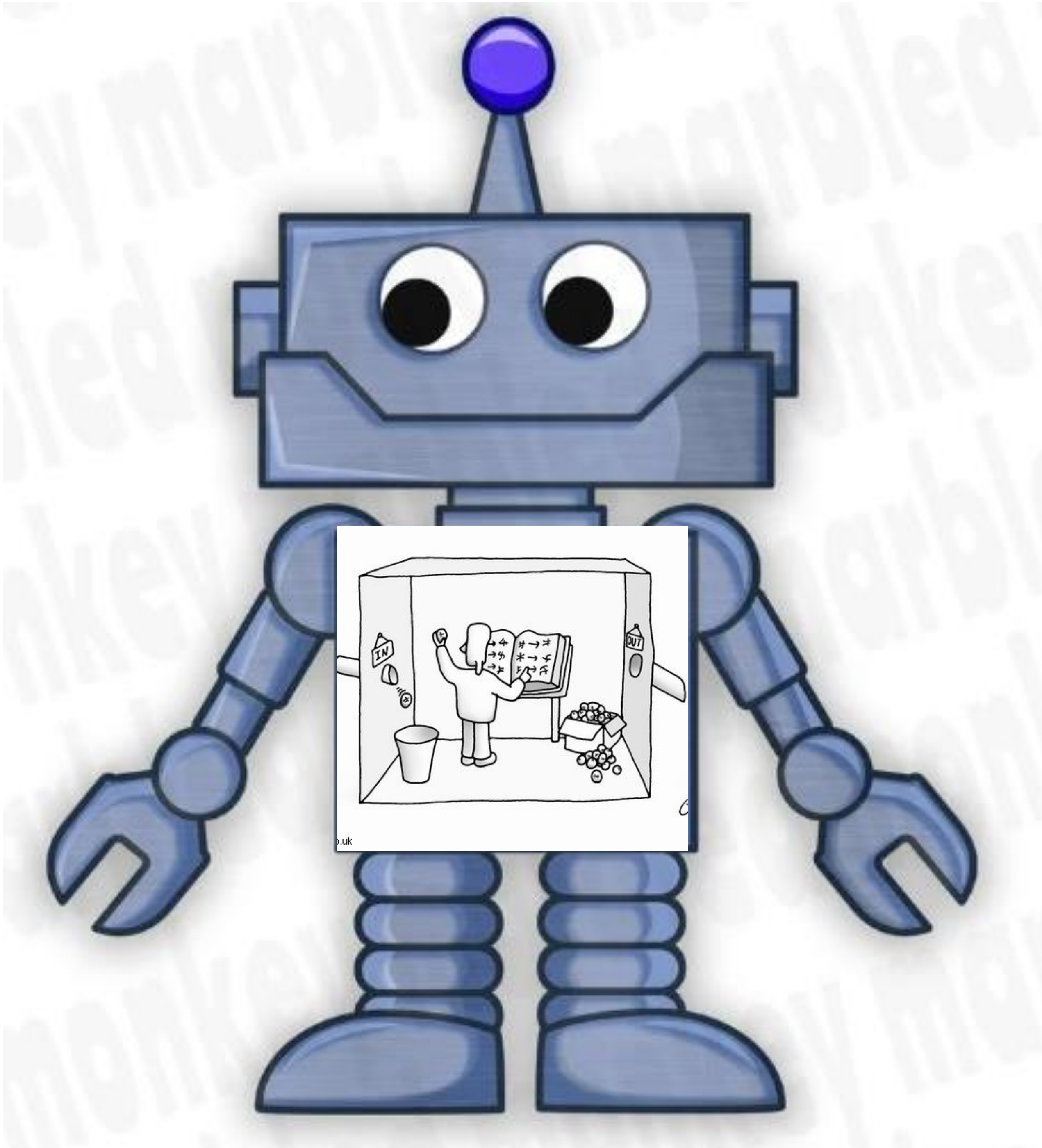
- What if Searle – baby is put in the room (or in a robot) and gradually acquires the rules of interactions?

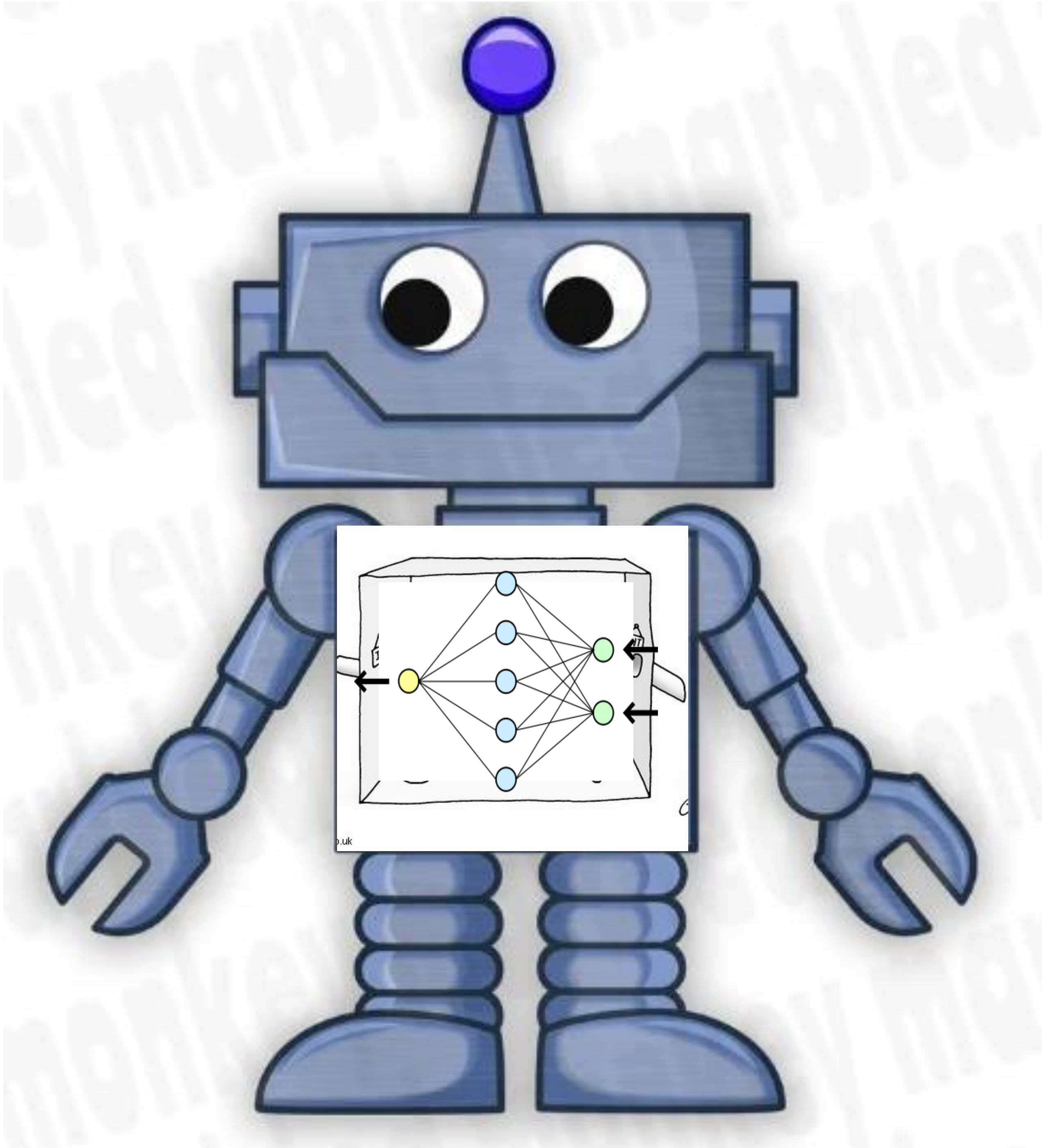
Neural networks

- **Connectionism** – since 1990s
  - Learning from experience/examples
  - Ability to generalize
  - Robustness against noise and damage






IN
OUT



Developmental approach

- Key idea: We are not born (super)smart – we **become** smart
- Endow a system with powerful **learning** mechanisms
- Let it **interact** with the environment and other systems

Six lessons from babies

(Artificial Life 11: 13–29, 2005)

The Development of Embodied Cognition: Six Lessons from Babies

Linda Smith
Psychology Department
Indiana University
Bloomington, IN 47405
smith4@Indiana.edu

Michael Gasser
Computer Science Department
Indiana University
Bloomington, IN 47405
gasser@Indiana.edu

Abstract The embodiment hypothesis is the idea that intelligence emerges in the interaction of an agent with an environment and as a result of sensorimotor activity. We offer six lessons for *developing* embodied intelligent agents suggested by research in developmental psychology. We argue that starting as a baby grounded in a physical, social, and linguistic world is crucial to the development of the flexible and inventive intelligence that characterizes humankind.

Keywords
Development, cognition, language, embodiment, motor control

1 Introduction

Traditional theories of intelligence concentrated on symbolic reasoning, paying little attention to the body and to the ways intelligence is affected by and affects the physical world. More recently, there has been a shift toward ideas of embodiment. The central idea behind the embodiment hypothesis is

Six lessons from babies
(Artificial Life 11: 13–29, 2005)

1. Be multi-modal
2. Be incremental
3. Be physical
4. Explore
5. Be social
6. Learn a language

https://sites.google.com/view/developing-minds-series/home

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 05 December 2017
doi: 10.3389/fpsyg.2017.02124

Check for updates

A Developmental Approach to Machine Learning?

Linda B. Smith and Lauren K. Slone*

Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IN, United States

Visual learning depends on both the algorithms and the training material. This essay considers the natural statistics of infant- and toddler-egocentric vision. These natural training sets for human visual object recognition are very different from the training data fed into machine vision systems. Rather than equal experiences with all kinds of things, toddlers experience extremely skewed distributions with many repeated occurrences of a very few things. And though highly variable when considered as a

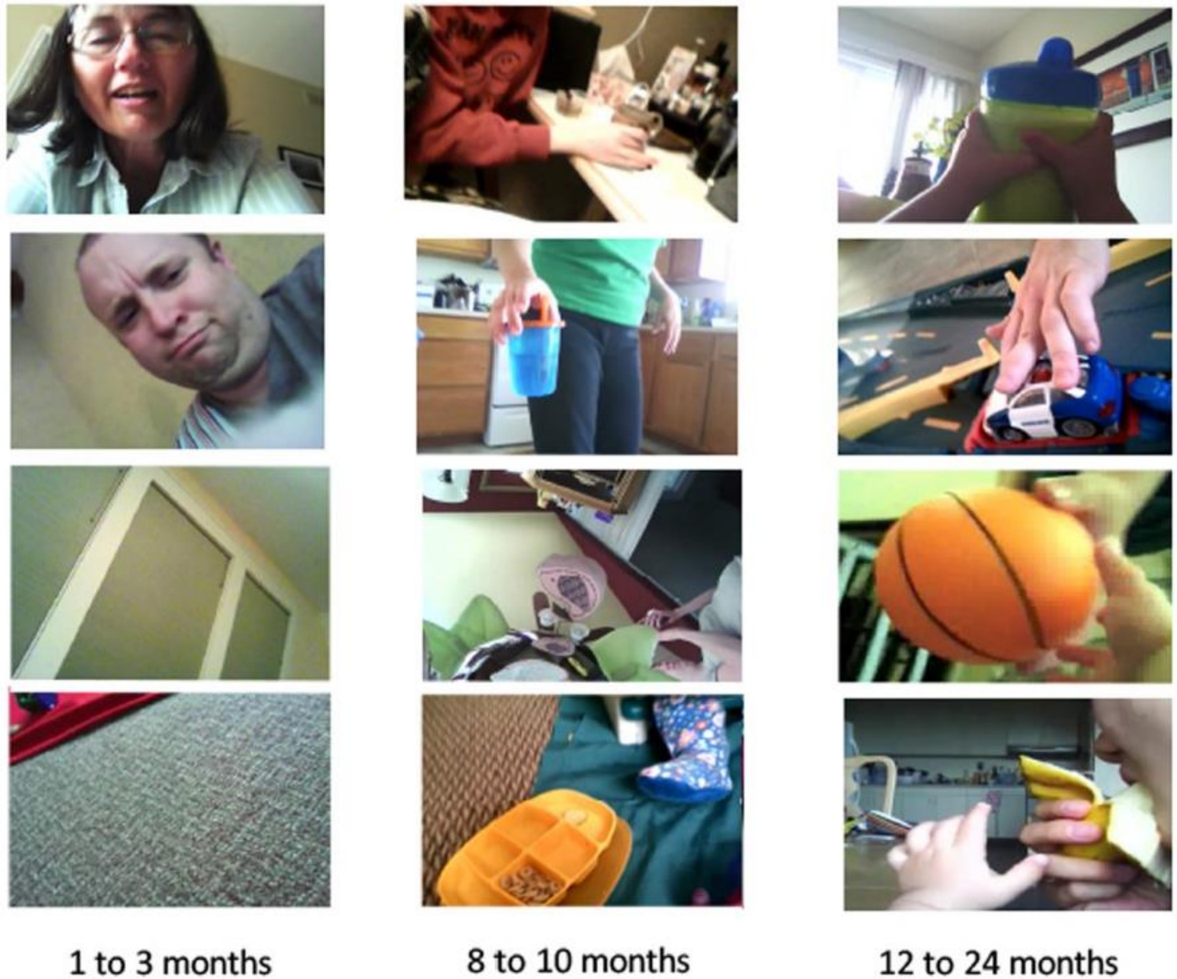


FIGURE 2 | Sample head-camera captured images for three different ages of infants.

(Source: Smith & Slone, 2017)

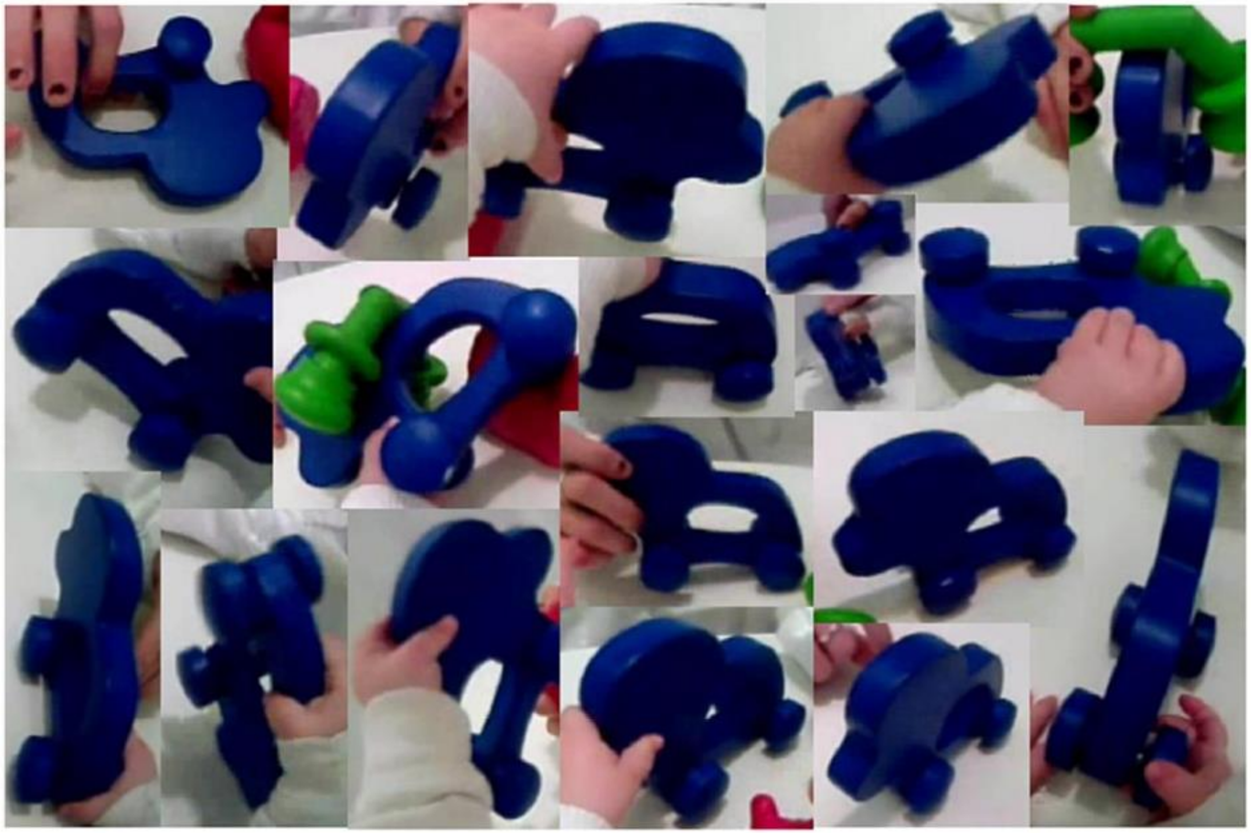
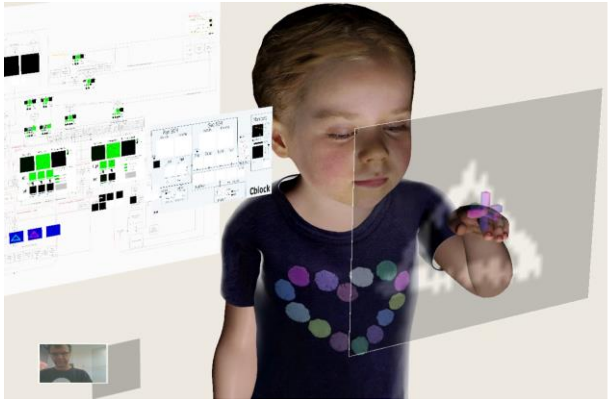


FIGURE 4 | Sample images of a single object captured by a 15-month-old infant's head-camera during play.

(Source: Smith & Slone, 2017)

Baby-X

- Complex hyper-realistic 3D simulation of a human baby
- Driven by artificial nervous system
- Artificial emotions
- A network of interconnected dynamical systems – potential for emergent behaviour
- Learning in interaction
- Developmental approach to artificial cognition



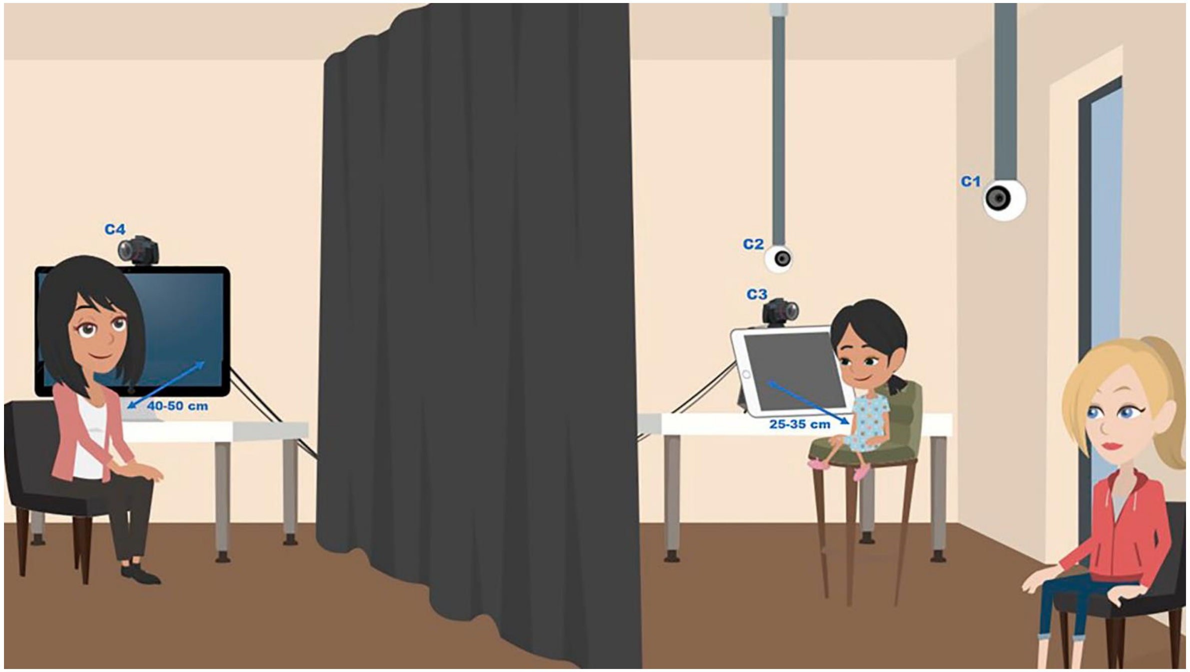
BabyX

BabyX includes:

- A graphical simulation of a baby’s **physical body**, including face, torso, limbs, and autonomic functions (breathing, heartbeat).
- **Sensory systems** (vision, hearing, touch, proprio/interoception).
- Models of visual **attention**, visual **object classification, category learning**, **reaching-to-grasp, visual object tracking.**
- A model of the **emotional system**, focussing on subcortical pathways, neurochemicals and emotional behaviours.
- Models of **working memory** (WM), **episodic long-term memory** (LTM), **emotional memory**, and building on these, **language.**

- Its cognitive architecture is a network of interconnected neural network modules, dynamical systems and state machines – potential for **emergent behaviour**
- Specially tailored for **interaction**

Experimental setup – real babies



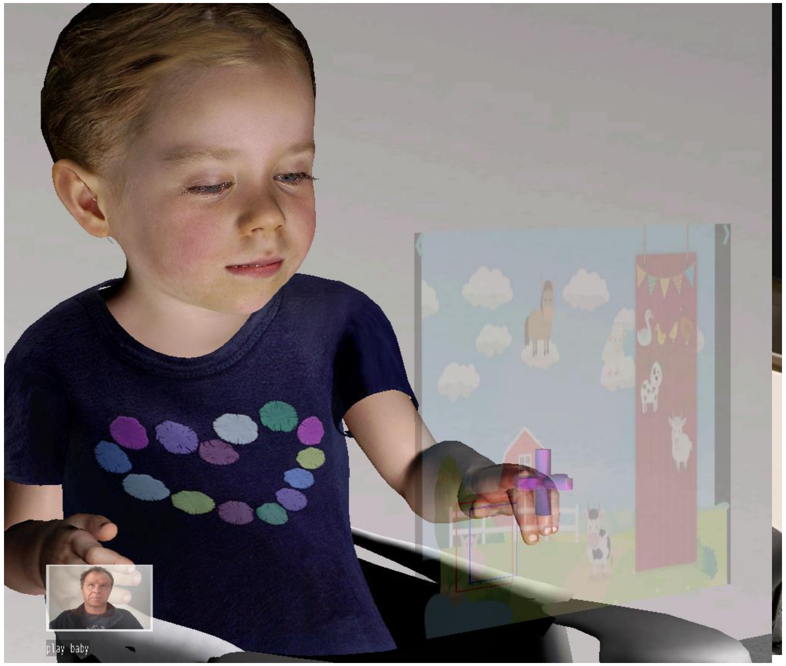
Picture from Sagar *et al.*, JRSNZ, 2023

Real baby scenario

Pictures from Sagar et al., TDCS, 2022

[video 1]

BabyX scenario

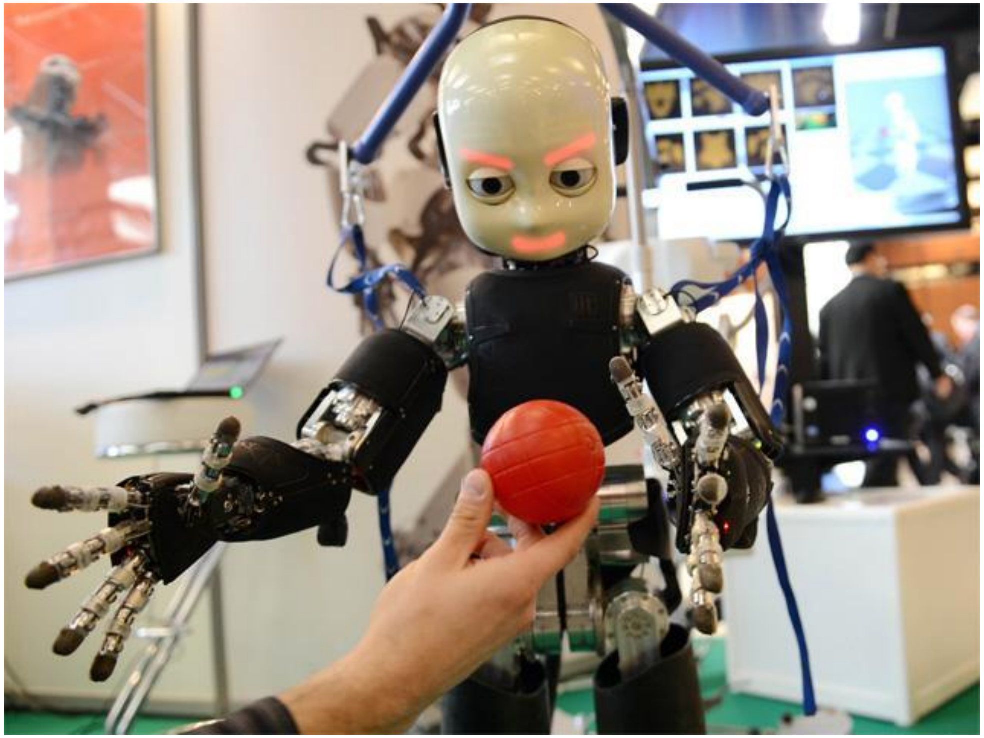


- Camera, microphone
- Simulated touch sensors
- Shared virtual environment
- Baby and the user both manipulate the objects in VE by touching/moving

Pictures from Sagar et al., TDCS, 2022 video 2 video 3

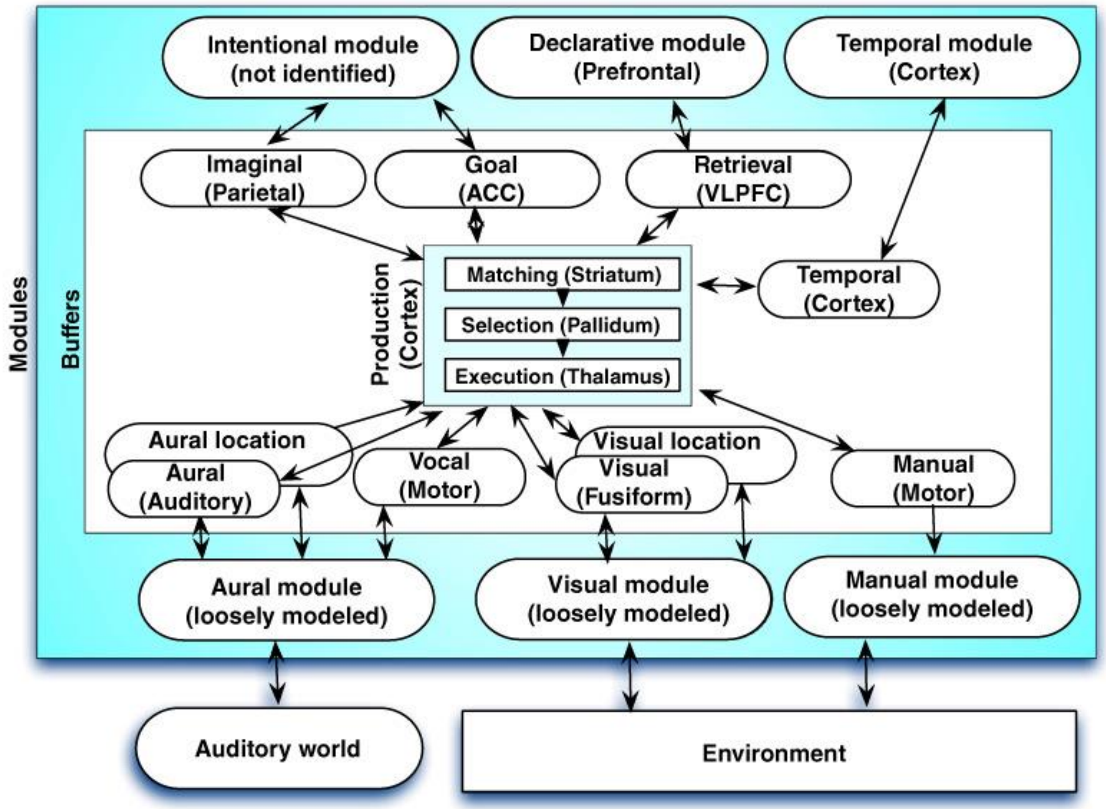
Developmental robotics

[iCub]



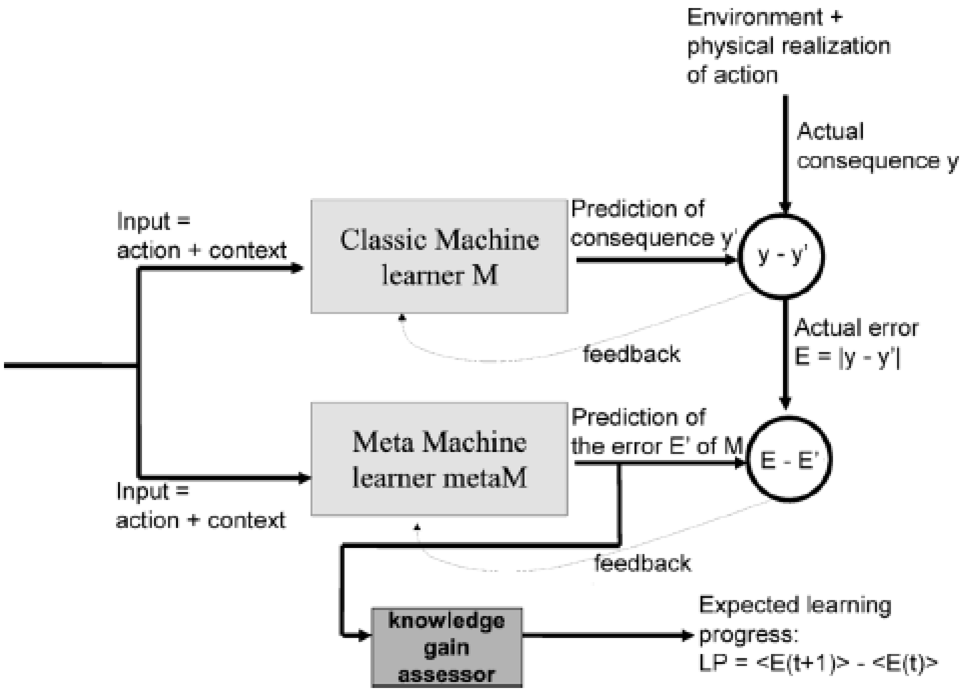
Hybrid cognitive architectures

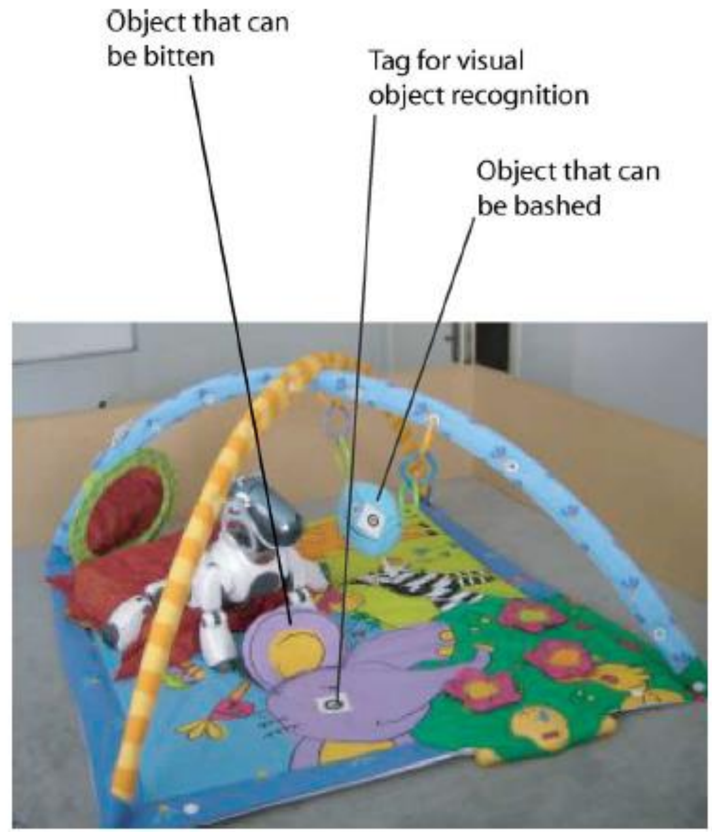
- [ACT-R](ACT-R) (Anderson, since 1973)



Autonomous goals

- Intrinsic motivation (Oudeyer, Kaplan, Hafner, 2007)



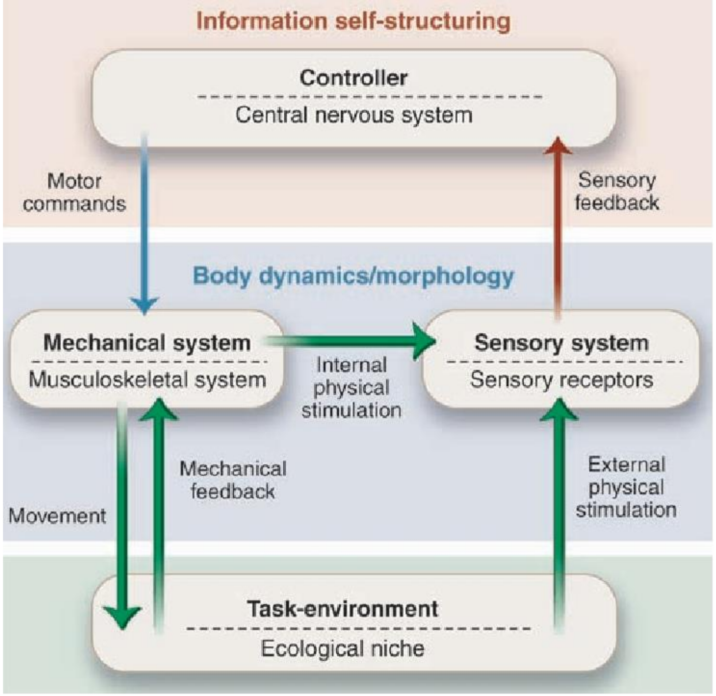


Searle’s conclusion (1980)

- “I see no reason in principle why we couldn't give a **machine** the capacity to understand English or Chinese, since in an important sense **our bodies with our brains are precisely such machines**. But I do see very strong arguments for saying that we could not give such a thing to a machine where the operation of the machine is defined solely in terms of **computational processes over formally defined elements**; that is, where the operation of the machine is defined as an instantiation of a **computer program**.”

Importance of the body

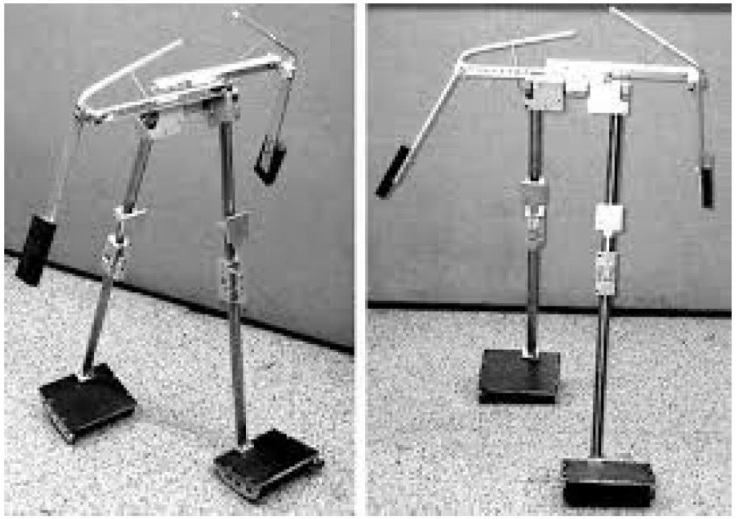
- Offloading intelligence to body



(Pfeifer et al., *Science*, 2007)

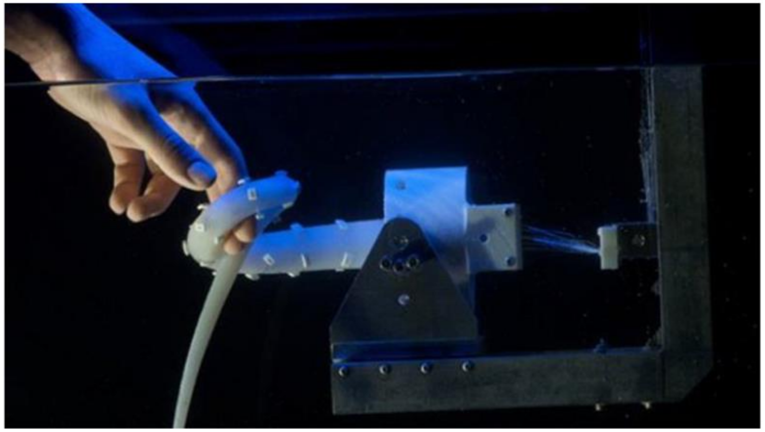
Intelligence in body mechanics

- Passive walkers (McGeer, 1990)
- Theo Jansen - Kinetic Sculptor (see also his TEDx Talk)



Intelligence in material

- Soft robotics
- Coffee-filled balloon gripper robot (Cornell Uni - iRobot)
- The Octopus Project



Body modification

– Self-modification – Evolvable hardware
– Artificial evolution: SIMS

Social AGI

- based on Multi-agent systems (MAS)
- **Creatures** (Steve Grand, 1992) are social agents with elaborate internal architecture based on a complex neural network which is divided into several lobes: for handling verbal input, managing the internal state (which was implemented as a simplified biochemistry, and kept track of feelings such as pain, hunger and others), adaptation, goal-oriented decision making and learning of new concepts.
- The original design had explicit AGI goals, with attention paid to allow for symbol grounding, generalization, and limited language processing.
- **Swarm intelligence**

Collective intelligence

- **Decentralized**
- **Self-organized**
- Simple agents interact locally with one another and with environment
  - Follow simple rules
  - No centralized control structure
  - Interactions lead to the **emergence** of intelligent behavior unknown to individual agents

- E.g. Ant colonies, bird flocking…

Emergence

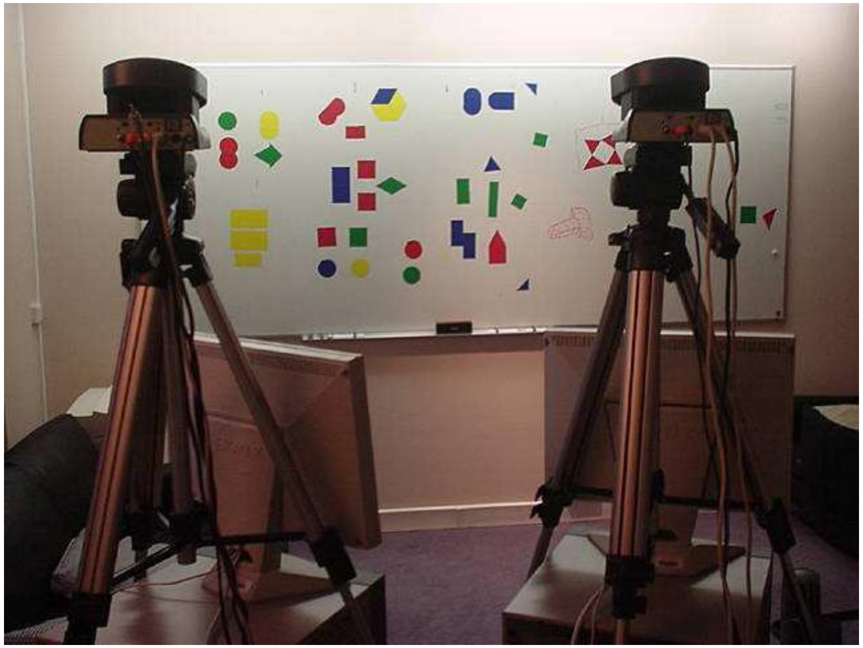
- Very simple rules can create very complicated behavior
- **Multi-agent systems**
- **Boids** (Reynolds, 1986) – flocking behavior of birds
  - **separation**: steer to avoid crowding local flockmates
  - **alignment**: steer towards the average heading of local flockmates
  - **cohesion**: steer to move toward the average position (center of mass) of local flockmates

- http://www.youtube.com/watch?v=GUkjC-69vaw
- http://www.red3d.com/cwr/boids/

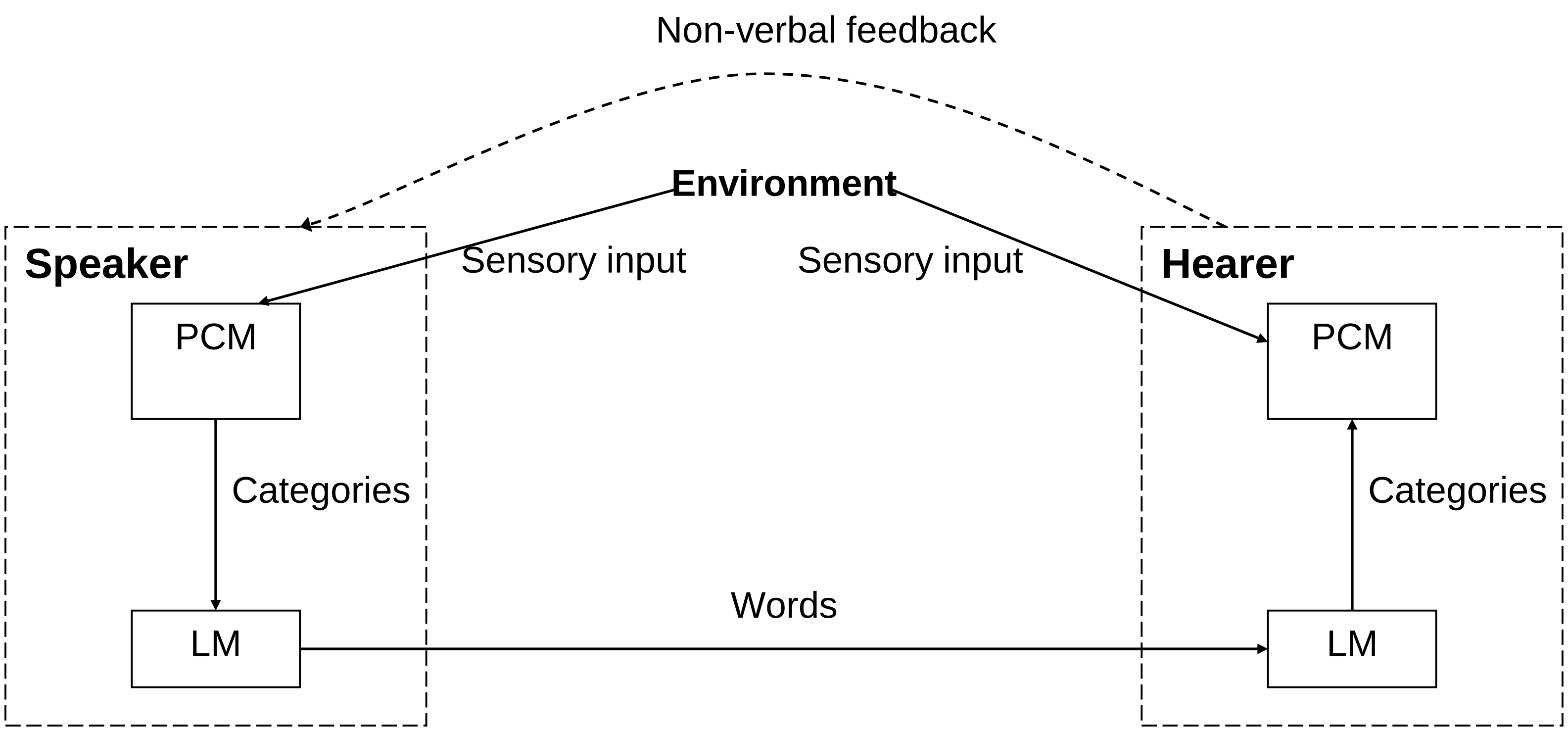
Intersubjective meanings

- Similar bodies
- Shared experience
- Social coordination
  - Multi-agent systems
  - Language games – Talking Heads (Steels, 1999)

Talking Heads (Steels, 1999)

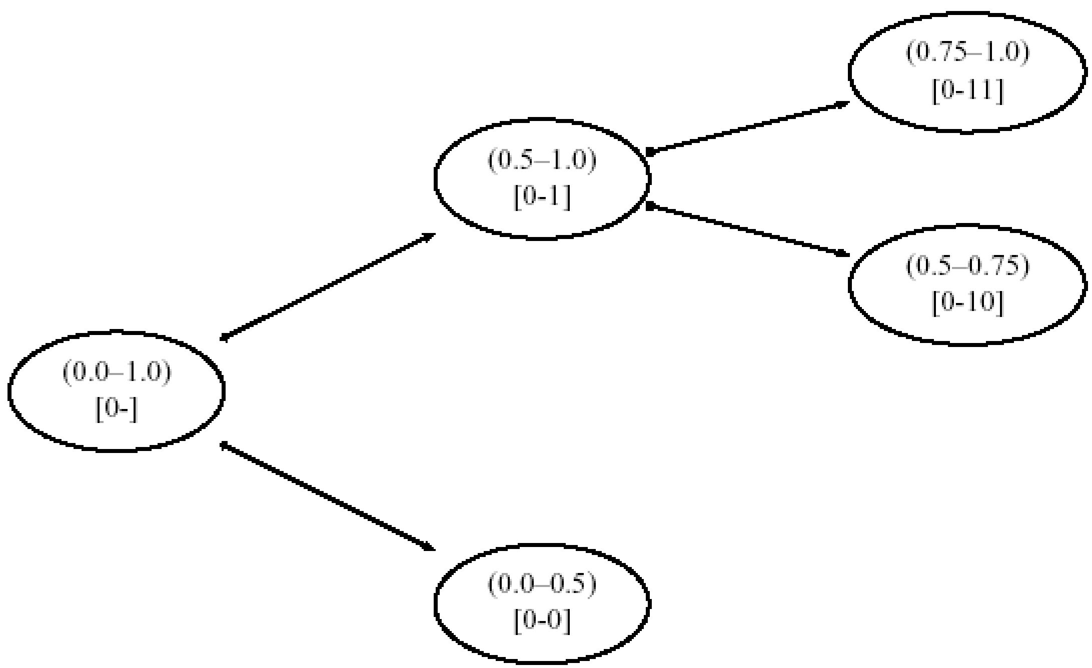


Talking Heads



Internal representations

- Discrimination trees



Lexicon

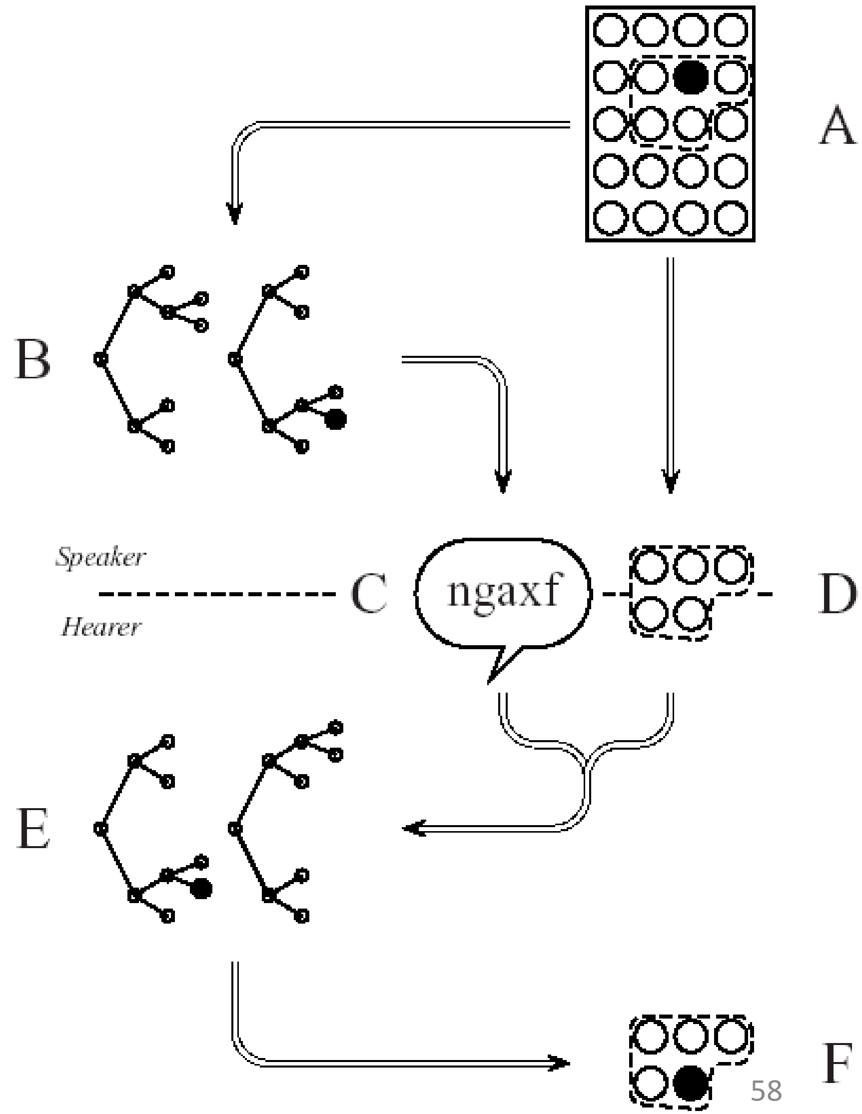
association table

{< signal, category, usage, confidence>}

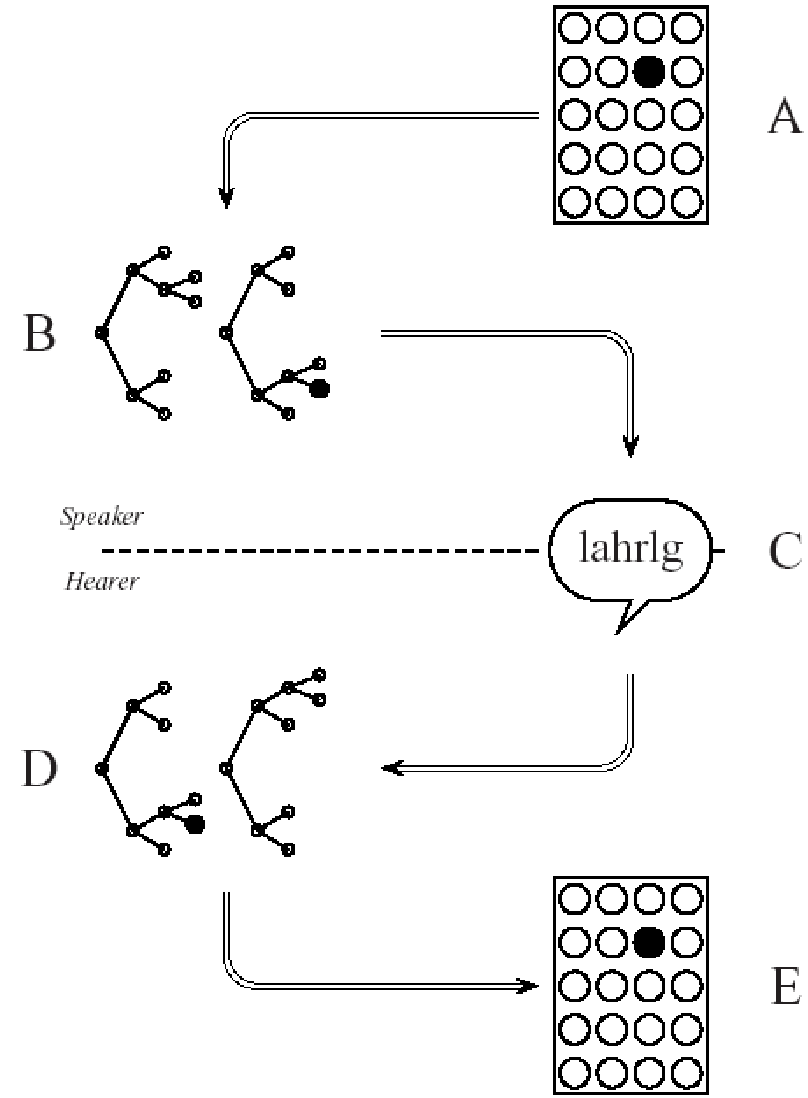
Signal	Meaning	Usage	Conf. Prob.
gttr	0-0	1	0.083
gttr	0-1	2	0.167
gttr	0-11	1	0.083
oij	1-0	9	0.600
gttr	2-0	4	0.333
oij	2-0	6	0.400
gttr	2-1	1	0.083
gttr	3-1	2	0.167
gttr	4-00	1	0.083

Communication

- **Speaker:**
  - context, topic, background
  - discriminating category
  - lexicalization (signal selection)
- **Hearer:**
  - signal interpretation



Communication successful if the referents match



Talking Heads – evaluation

- meanings created individually by each agent
- dependent on interaction history
- important demonstration of global coordination of meanings on the community level
- co-ordination decentralized – based on positive feedback loops and self-organization

Thank you