Robots, neuroscience and artificial cognition

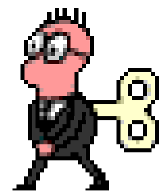
Giorgio Metta

Robotica Antropomorfa
Lezione 1-3

[Part 1]

Goals

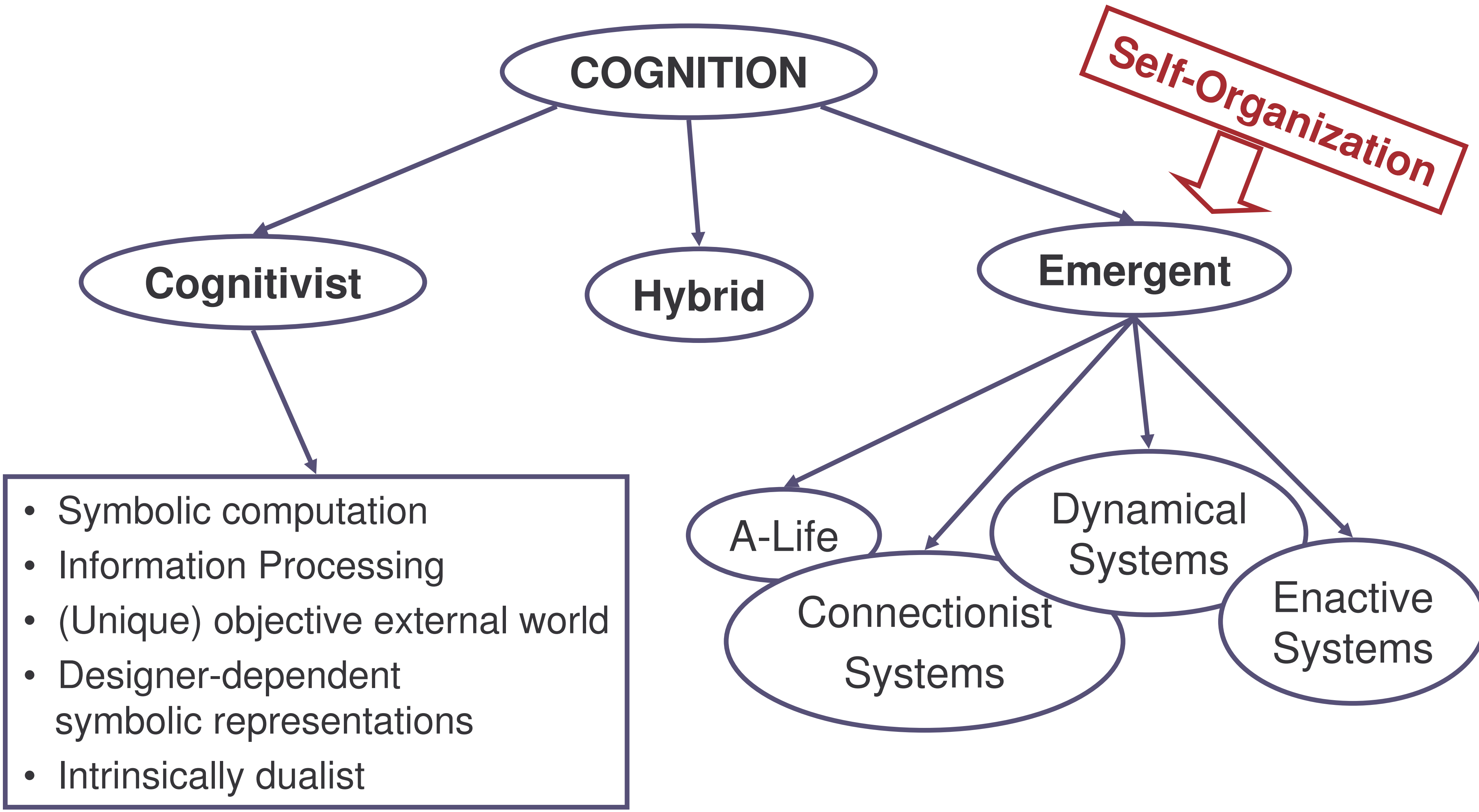
General-purpose artificial interactive systems with the robustness and resilience of a human



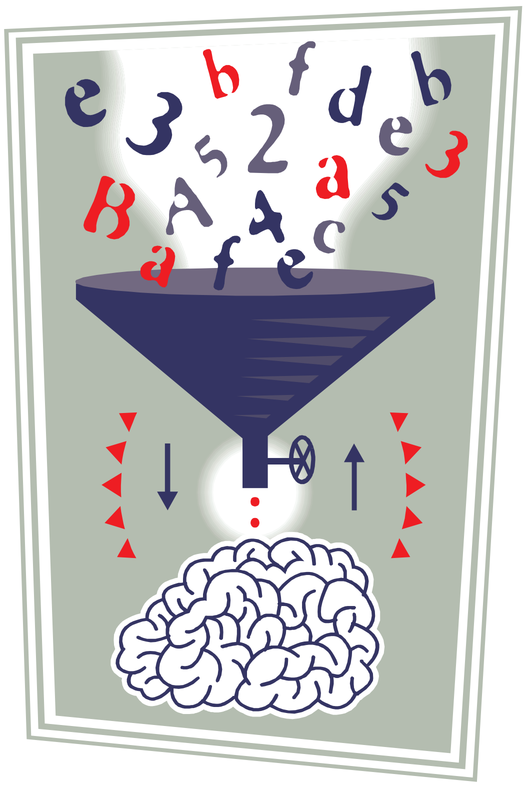
Cognitive Systems:

- Adaptive
- Anticipatory
- Autonomous
- Interactive
- Robust
- Intelligent

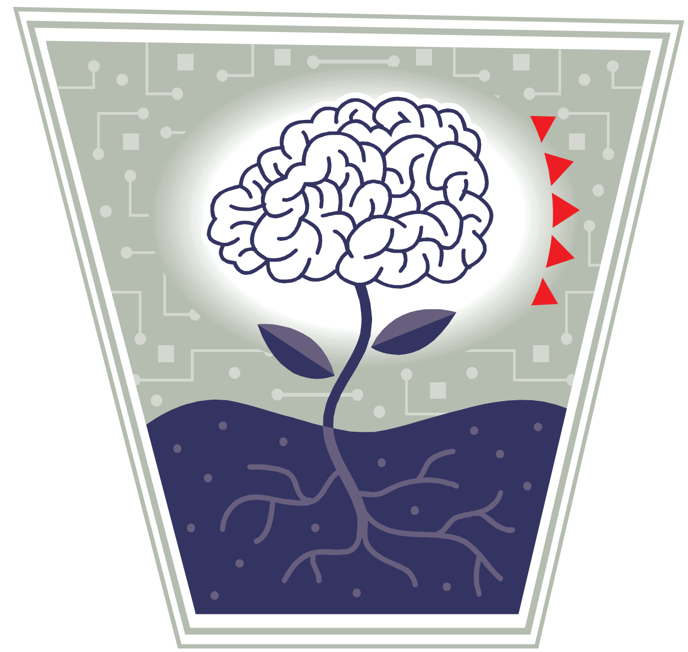
Paradigms of Cognitive Systems



The kernel of the problem



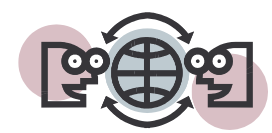
OR

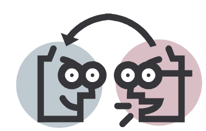


Cognitive Systems:

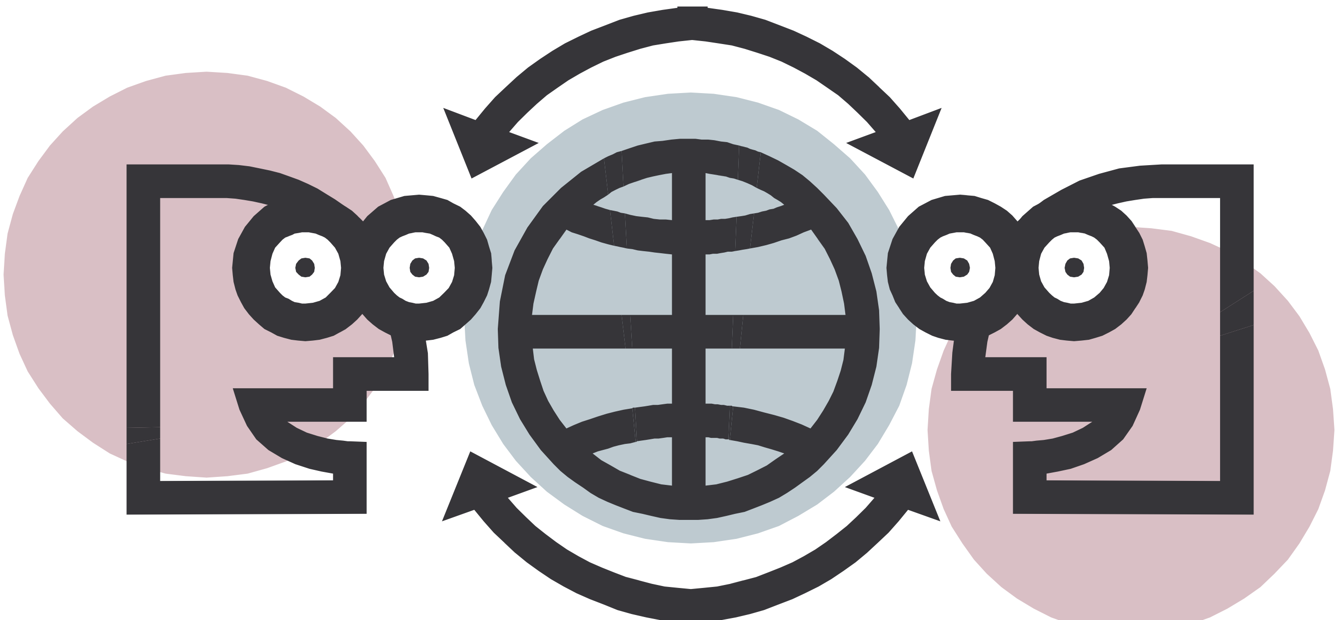
- *Adaptive*
- *Anticipatory*
- *Autonomous*
- Interactive
- Robust
- Intelligent

- In short, act in their environment
- We can pursue two strategies to tackle the problem

1. Study action, cognition is a byproduct of sophisticate acting, prediction, and long-term prospection 

2. Assume there's an independent cognitive module that eventually determines action 

Our approach



Our approach

Guiding Philosophy

- Cognition cannot be hand-coded
- Is necessarily the product of a process of embodied development
- Initially dealing with immediate events
- Increasingly acquiring a predictive capability

Cognition and perception are functionally-dependent on the richness of the action interface

Our Approach (contd.)

Emergent embodied cognitive systems:

- Given a rich set of innate action and perception capabilities
- Develop over time an increasing range of cognitive abilities
- Recruiting ever more complex actions
- Achieving an <u>increasing</u> degree of <u>prospection</u> (and, hence, adaptability and robustness)

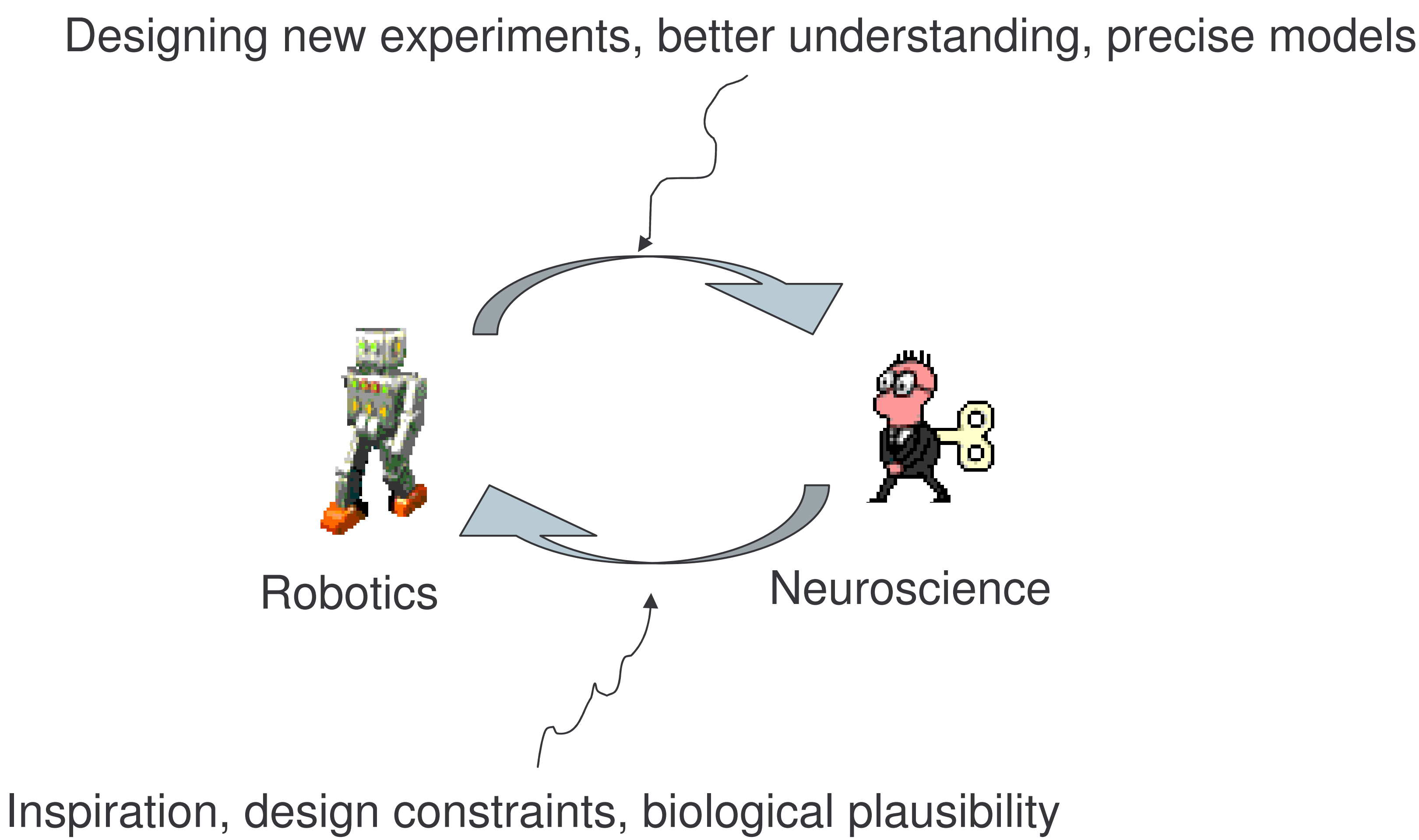
Our Approach (contd.)

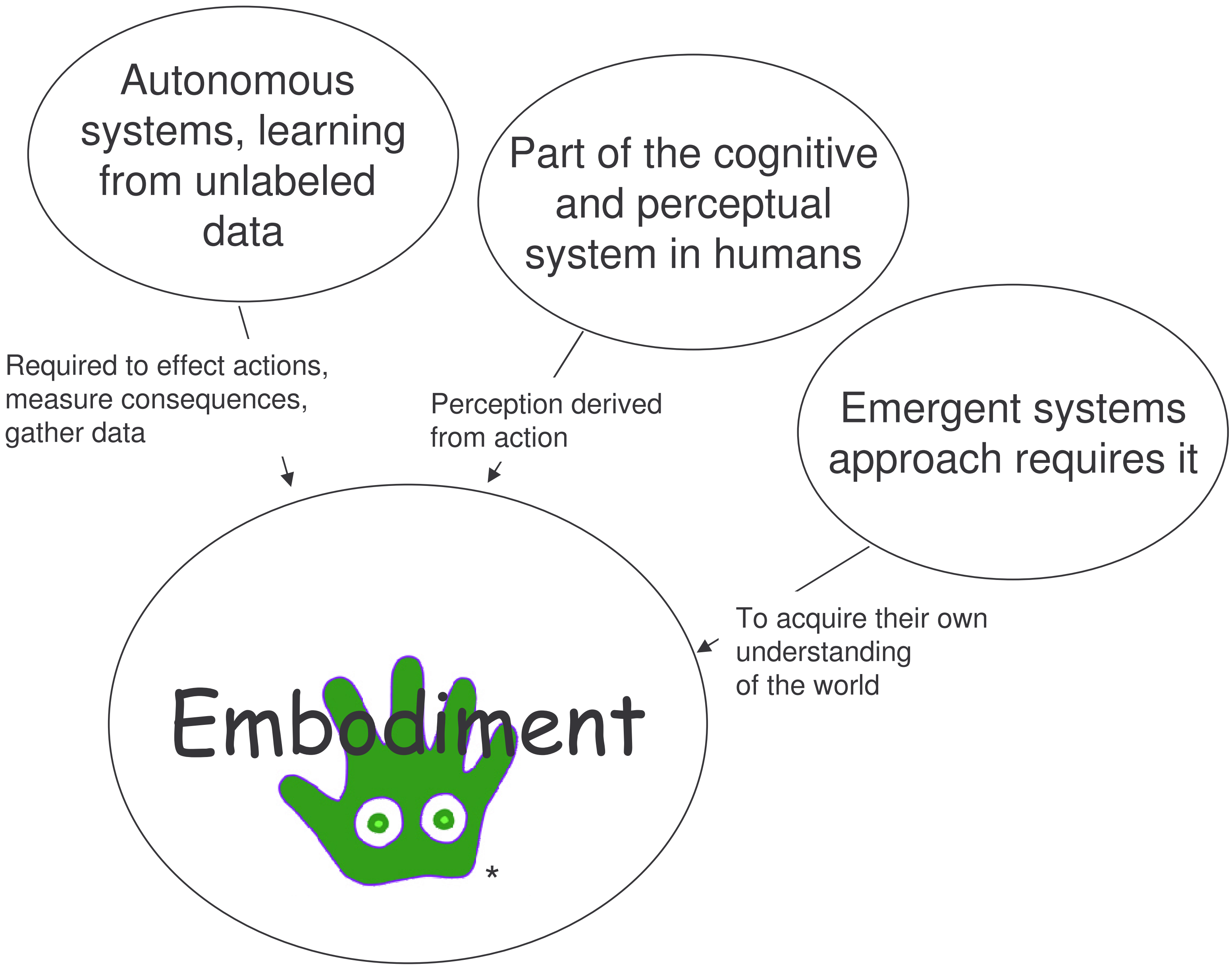
Cognitive development involves several phases

1. Coordination of eye-gaze, head attitude, hand placement when reaching
2. Dexterous manipulation of the environment: learn the affordances in the context of one's own developing capabilities
3. Ideally, communicate through gestures, simple expressions of its understanding, achieved through
   - rich manipulation-based exploration & social contact
   - imitation
   - multi-agent social interaction

More complex and revealing exploratory use of action

…but even more ambitiously

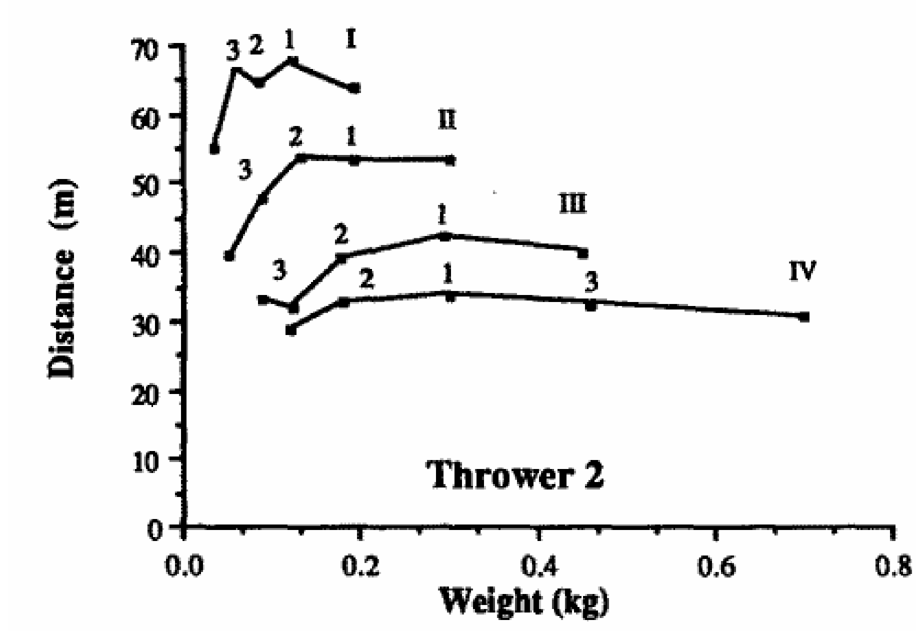




* http://www.chessstetson.com/eagleeye.html

Throwing stones…

Random source from the Internet



4 sizes, 5 weights, 3 choices

Because the kinetic energy developed during a throw determines flight distance and because the kinetic energy coupling term relates to one of the means by which kinetic energy is developed and transferred eventually to a projectile, the interjoint phase relations [elbow-wrist] were interpreted as providing information for the perceived property, "optimal object for throwing to a maximum distance".

→

The answer entailed by the approach taken in the [experiments] is that the scaling relations are based on the same or on closely related dynamics.

Bingham, Schmidt, Rosenblum. *Hefting for a maximum distance throw: a smart perceptual mechanism.* Journal of experimental psychology: human perception and performance. 15(3) 1989

Speech listening…

European Journal of Neuroscience, Vol. 15, pp. 399–402, 2002

SHORT COMMUNICATION

Speech listening specifically modulates the excitability of tongue muscles: a TMS study

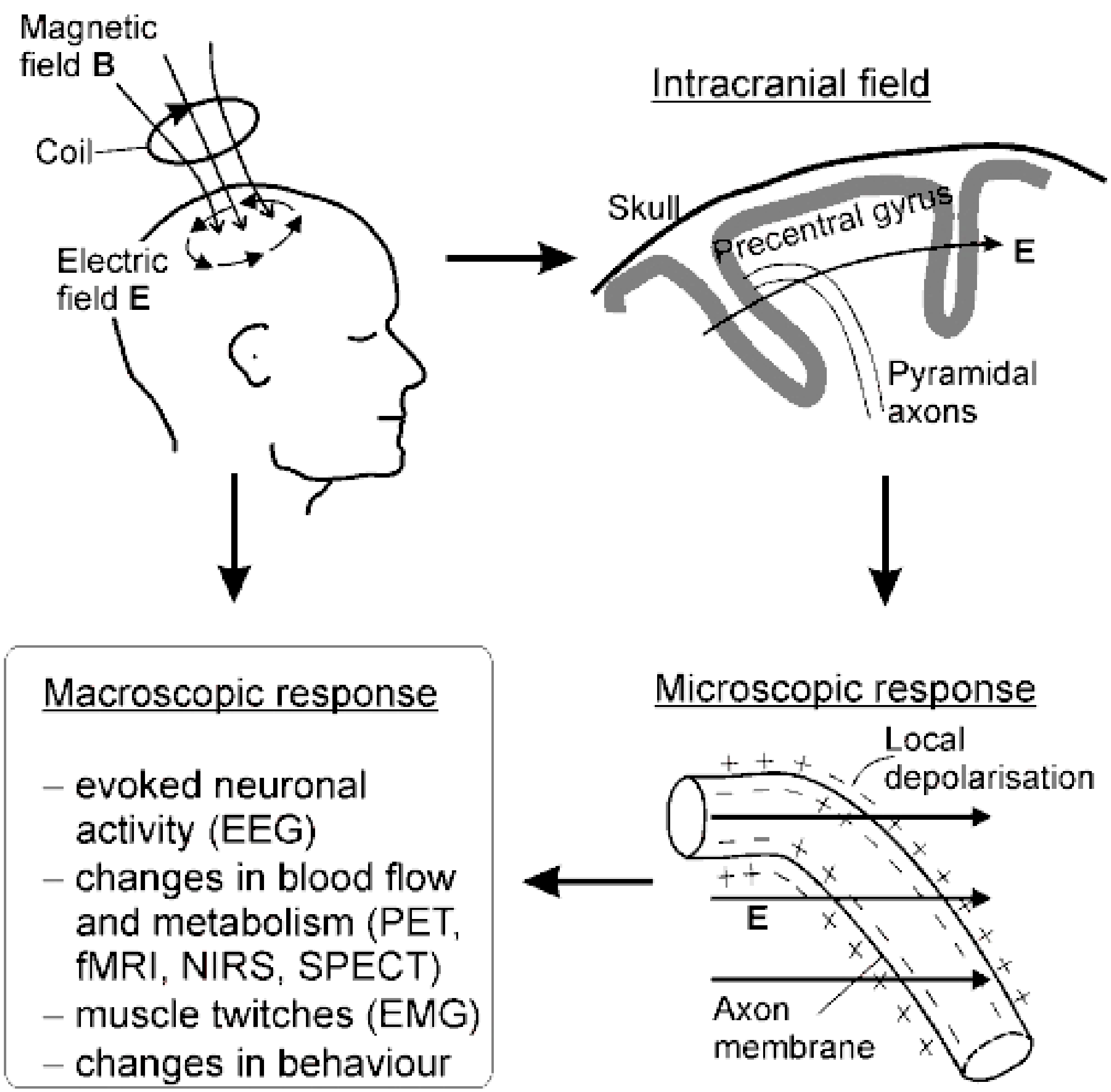
Luciano Fadiga,[1] Laila Craighero,[1,2] Giovanni Buccino[2] and Giacomo Rizzolatti[2]
[1]Dipartimento di Scienze Biomediche e Terapie Avanzate, Sezione di Fisiologia Umana, Università di Ferrara, via Fossato di Mortara 17/19, 44100 Ferrara, Italy
[2]Istituto di Fisiologia Umana, Università di Parma, via Volturno 39, 43100 Parma, Italy

Keywords: mirror neurons, motor-evoked potentials, motor system, motor theory of speech perception

What we get from the paper

- Listening: three categories of stimuli (words, pseudo-words, bi-tonal sounds).
- Two phonemes 'rr' requires strong tongue tip movement, 'ff' requires slight tongue tip movement.
- TMS of the under-threshold motor cortex.
- Recording of the MEP (motor-evoked potential) from the tongue muscles.

What is TMS



Intracellular more negative:
-70mV

The electric field depolarizes the cells.

It's more likely that action potentials are generated.

Numbers

- 3KVolts
- 5.5KA
- Raising time 2μs
- Impulse duration 100μs
- Activation region 2-3cm diameter
- Magnetic field 2.5T (40000 times the field of the earth)
- Neurons involved in the order of 100K

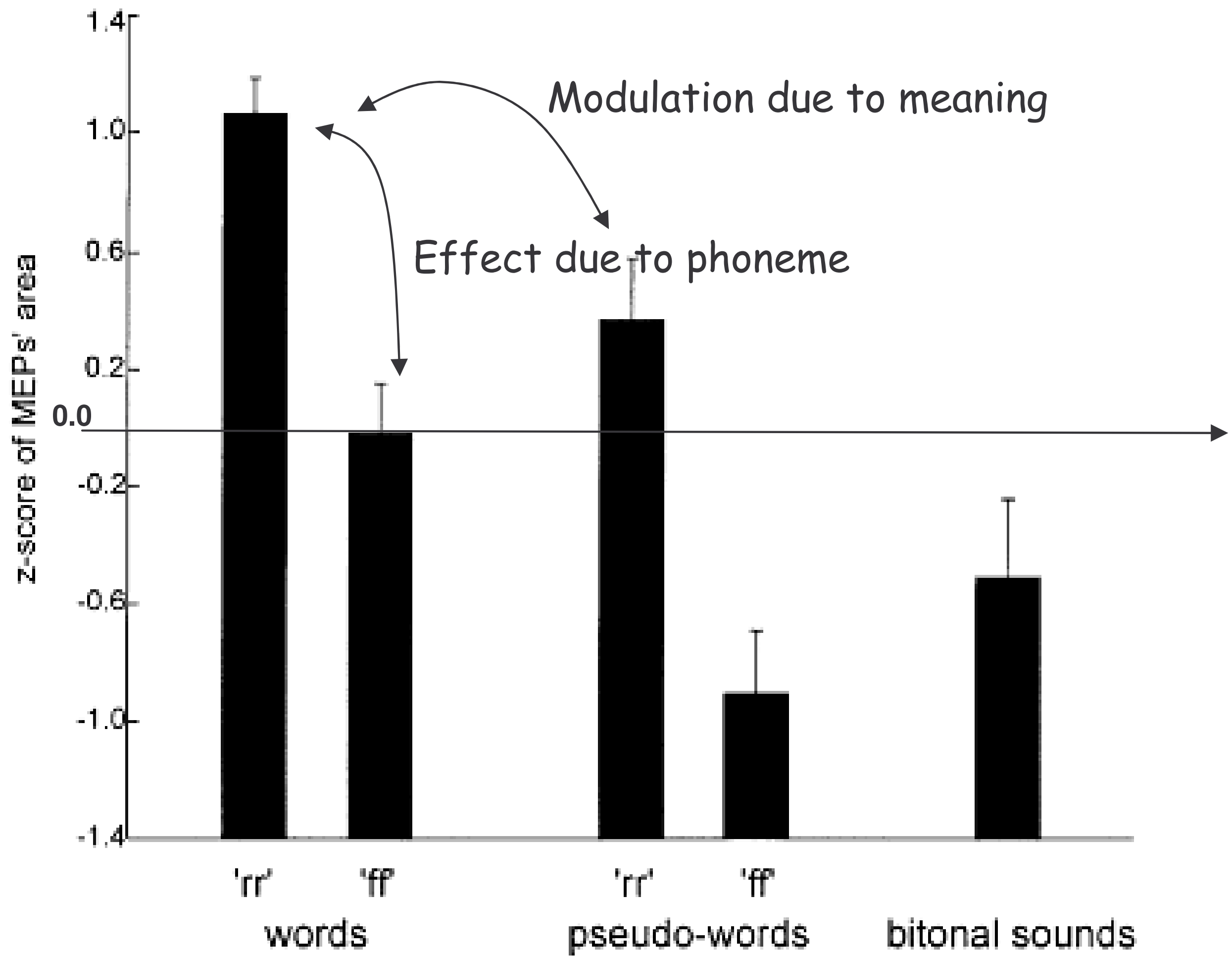
What it is nice about TMS an example

- In blind people a tactile discrimination task activated also "visual" and association areas.
- PET or fMRI cannot prove though that this activation in functional to the task.
- TMS was used to disrupt the functioning of the visual areas.
- Blind people had a greater chance to make mistakes compared to non-blind controls.

Examples of word/pseudo-words

Labiodental fricative consonant, 'rr'		Lingua-palatal fricative consonant, 'ff'	
Words	Pseudo-words	Words	Pseudo-words
birra (bier)	berro	baffo (moustache)	biffo
carro (cart)	firra	beffa (hoax)	ciffo
cirro (cirrus)	forro	buffo (funny)	leffa
farro (spelt)	furra	ceffo (snout)	meffa
ferro (iron)	marro	coffa (crow's nest)	paffo
mirra (myrrh)	merro	goffo (clumsy)	peffa
morra (morra)	parro	muffa (mold)	poffa
porro (leek)	perro	puffo (smurf)	seffa
serra (greenhouse)	vorro	tuffo (dive)	viffa
terra (ground)	vurro	zaffo (plug)	voffo

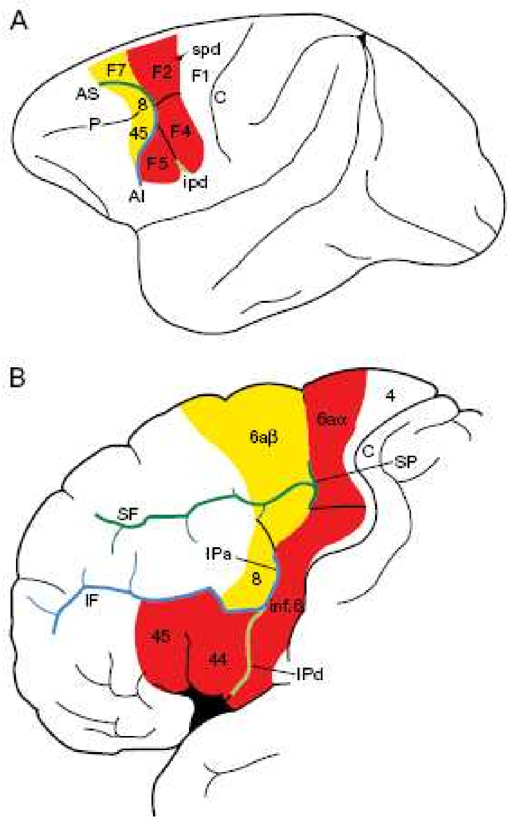
Results (in short)



Language within our grasp

Giacomo Rizzolatti and Michael A. Arbib

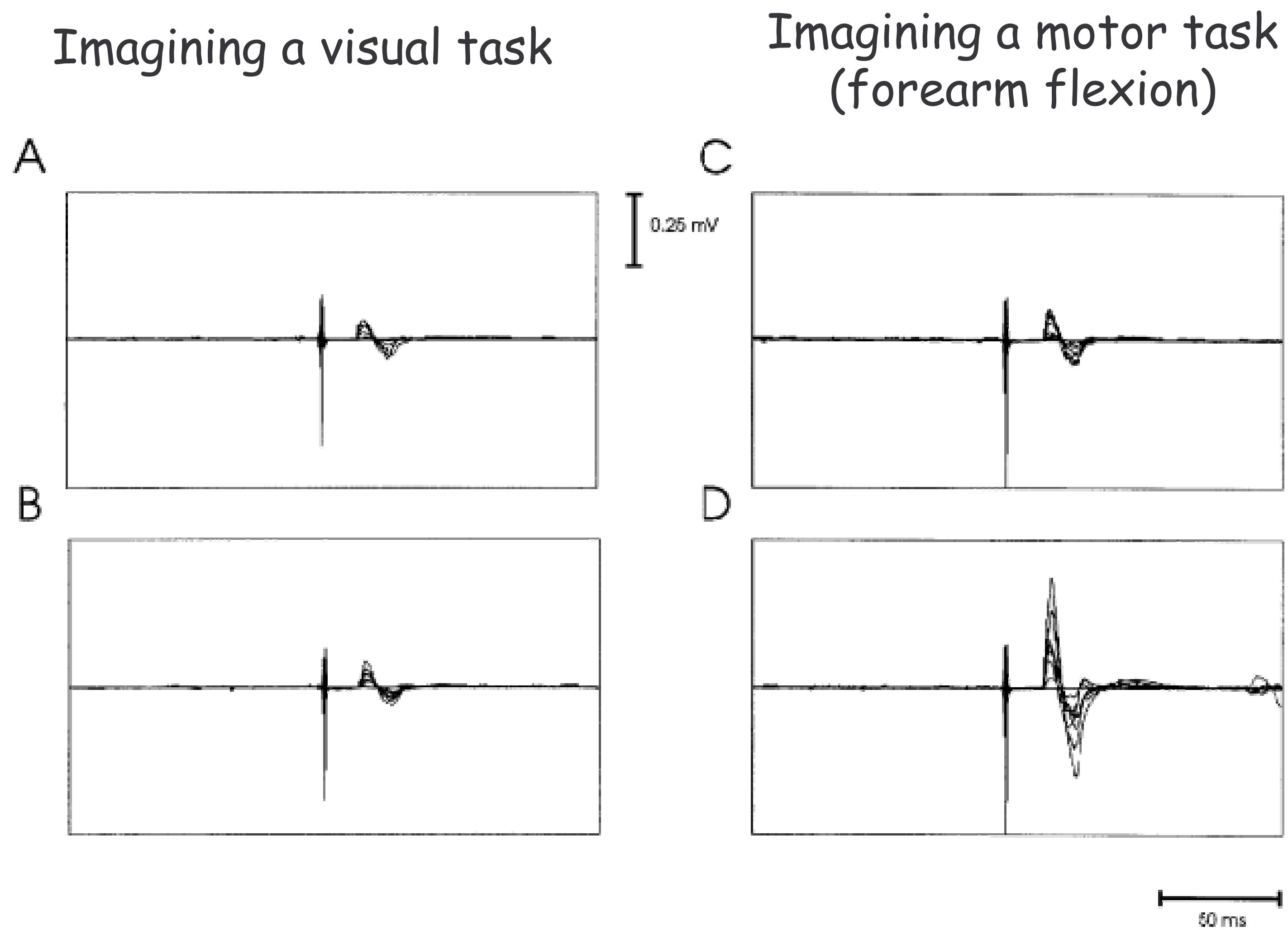
In monkeys, the rostral part of ventral premotor cortex (area F5) contains neur
both when the monkey grasps or manipulates objects and when it observes
making similar actions. These neurons (mirror neurons) appear to represent a sy



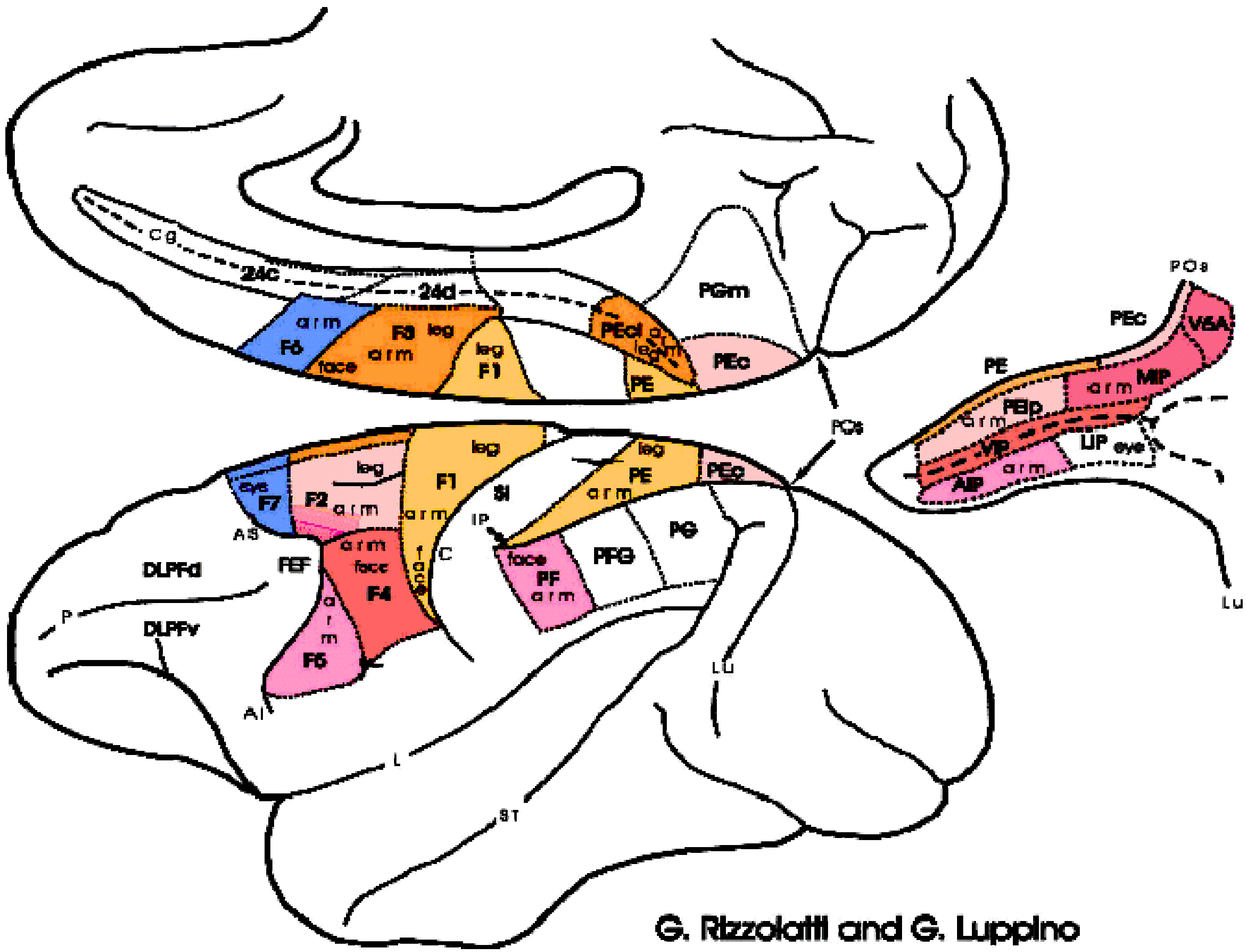
"In all communication, sender and receiver must be bound by a common understanding about what counts; what counts for the sender must count for the receiver, else communication does not occur. Moreover the processes of production and perception must somehow be linked; their representation must, at some point, be the same."

[Alvin Liberman, 1993]

Motor imagery

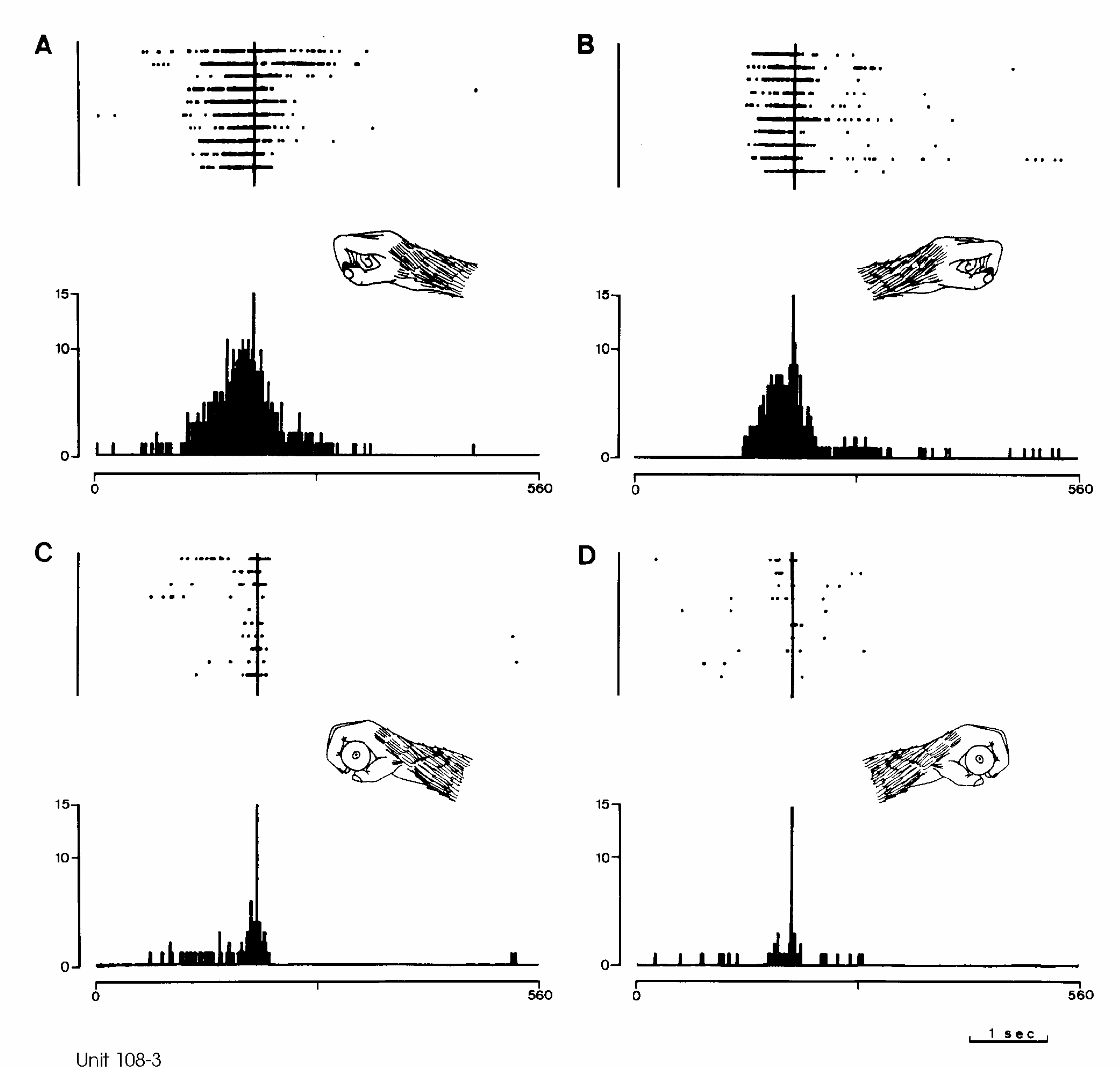


[Part 2]

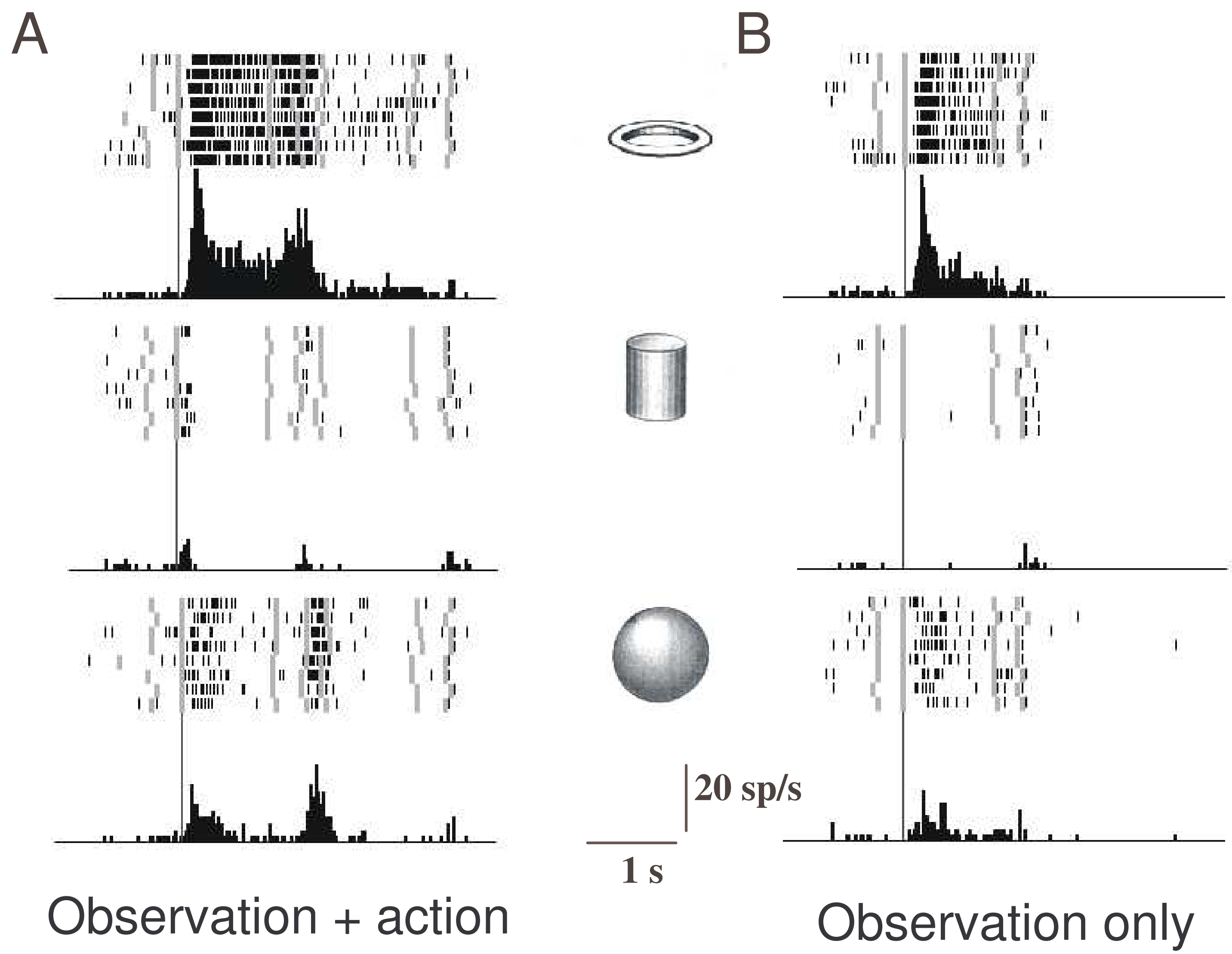


G. Rizzolatti and G. Luppino

Grasping neurons

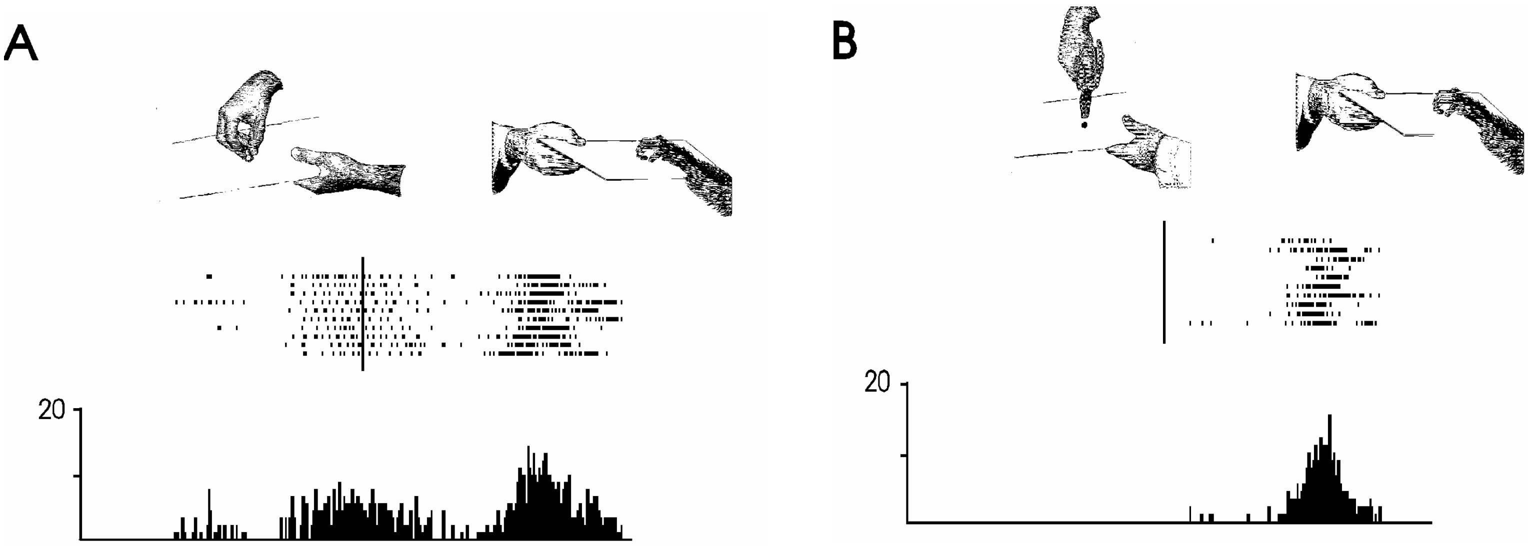


F5 canonical neurons



Mirror Neurons

The neuron is activated by “seeing” someone else’s hand performing a manipulative action **and** while the monkey is performing the same action

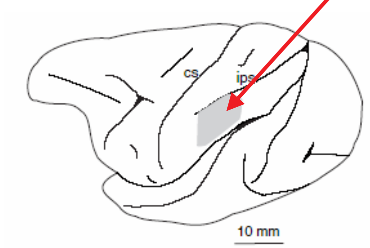


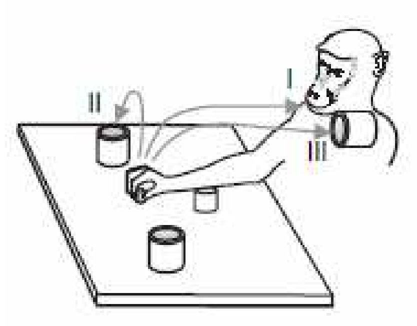
The type of action seen is relevant

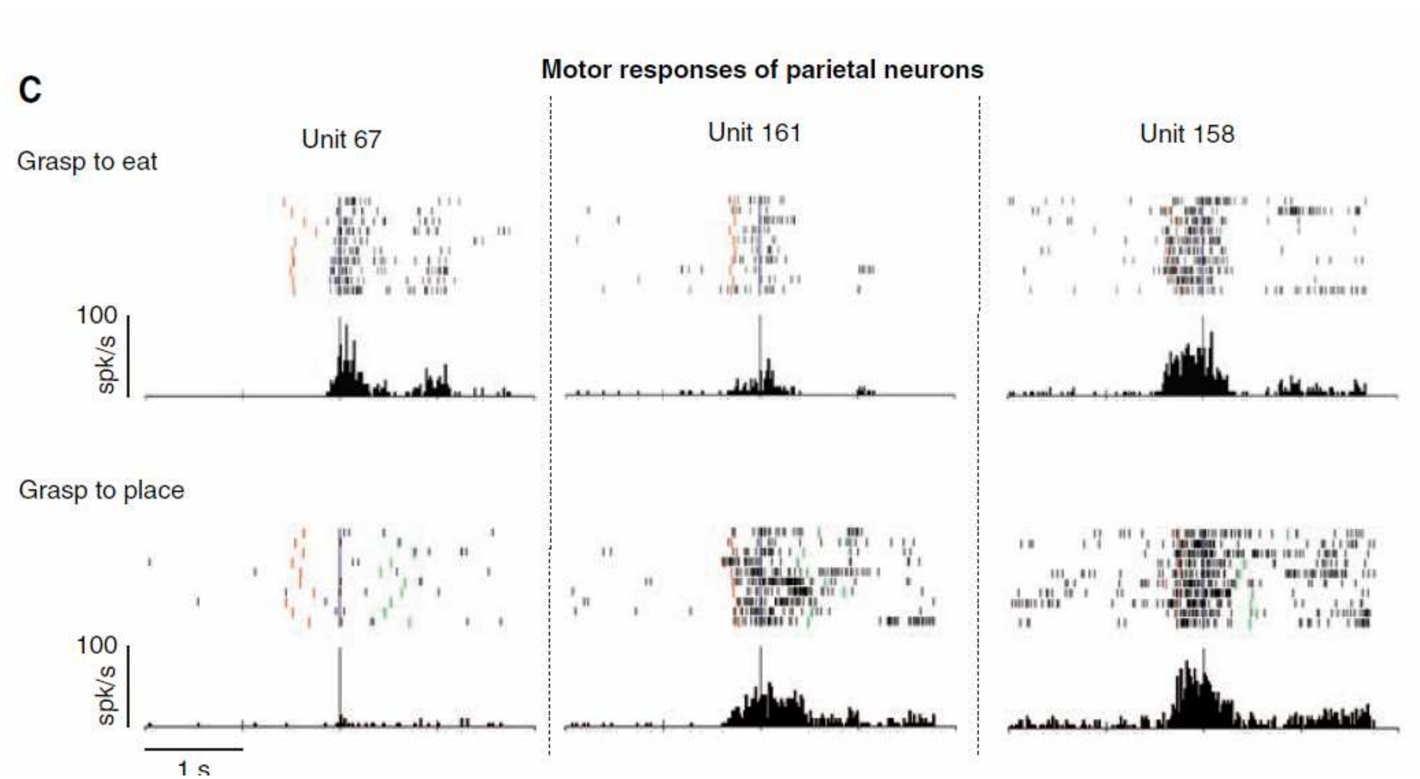
From: Fadiga, L., L. Fogassi, V. Gallese, and G. Rizzolatti, *Visuomotor Neurons: ambiguity of the discharge or "motor“ Perception?* Internation Journal of Psychophysiology, 2000. **35**: p. 165-177.

More mirror neurons

Parietal Mirror Neurons

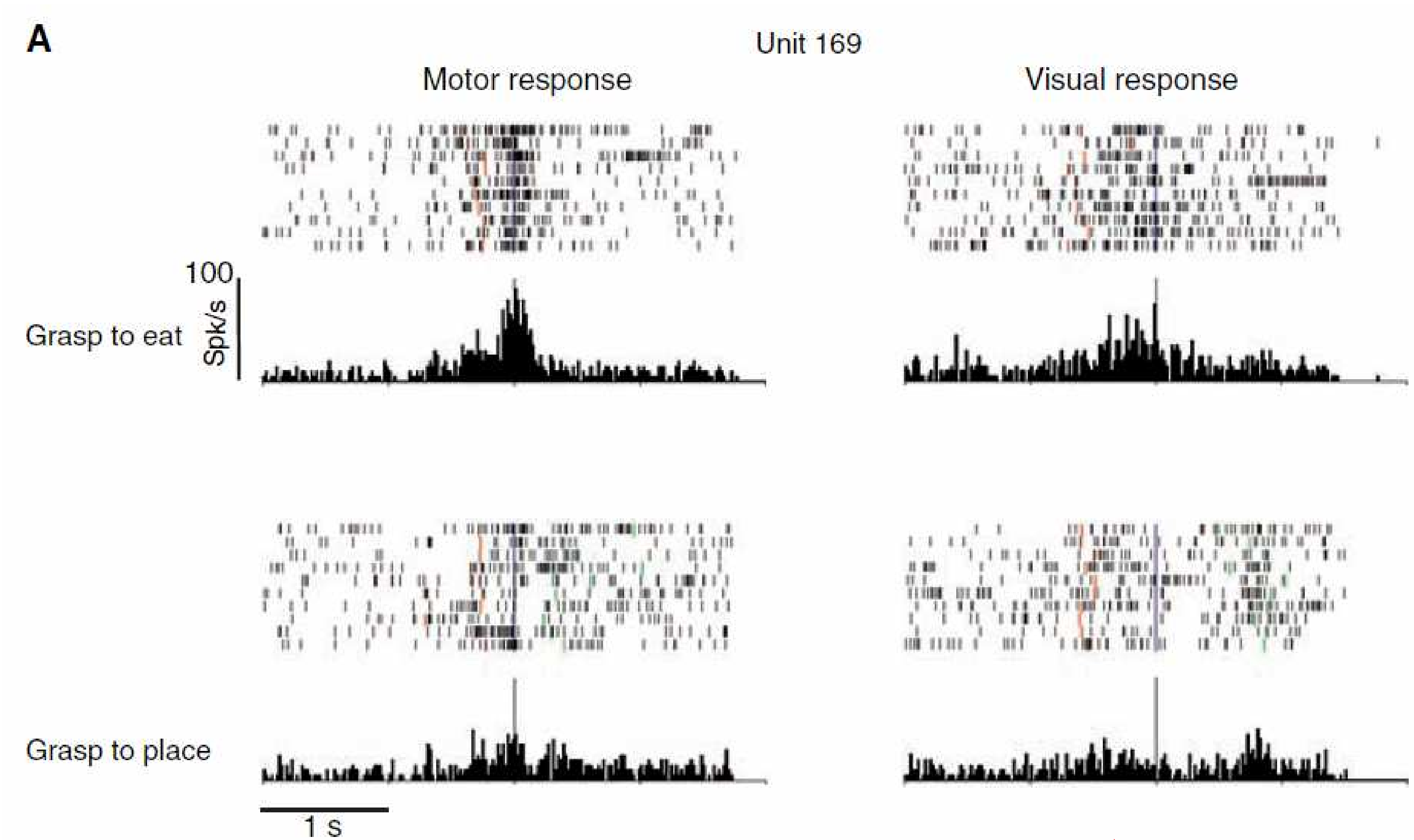






From: L. Fogassi, P.F. Ferrari, B. Gesierich, S. Rozzi, F. Chersi, G. Rizzolatti: Parietal Lobe: From Action Organization to Intention Understanding – Science - Vol. 308 pp. 662- 667 - 29 April 2005

Neurons code intentions of others



Action made

Action seen

Same Neuron

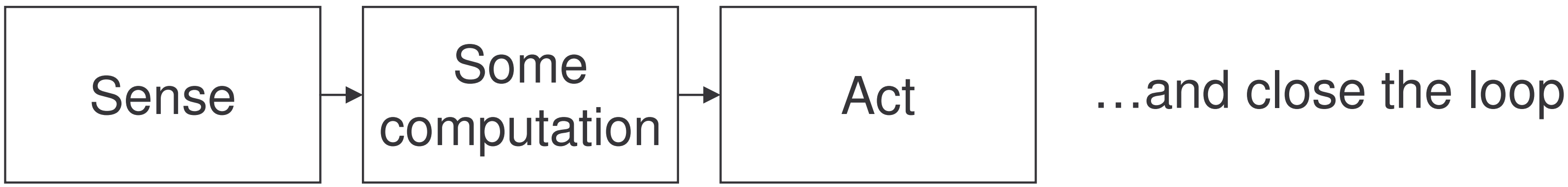
Mirror neurons?

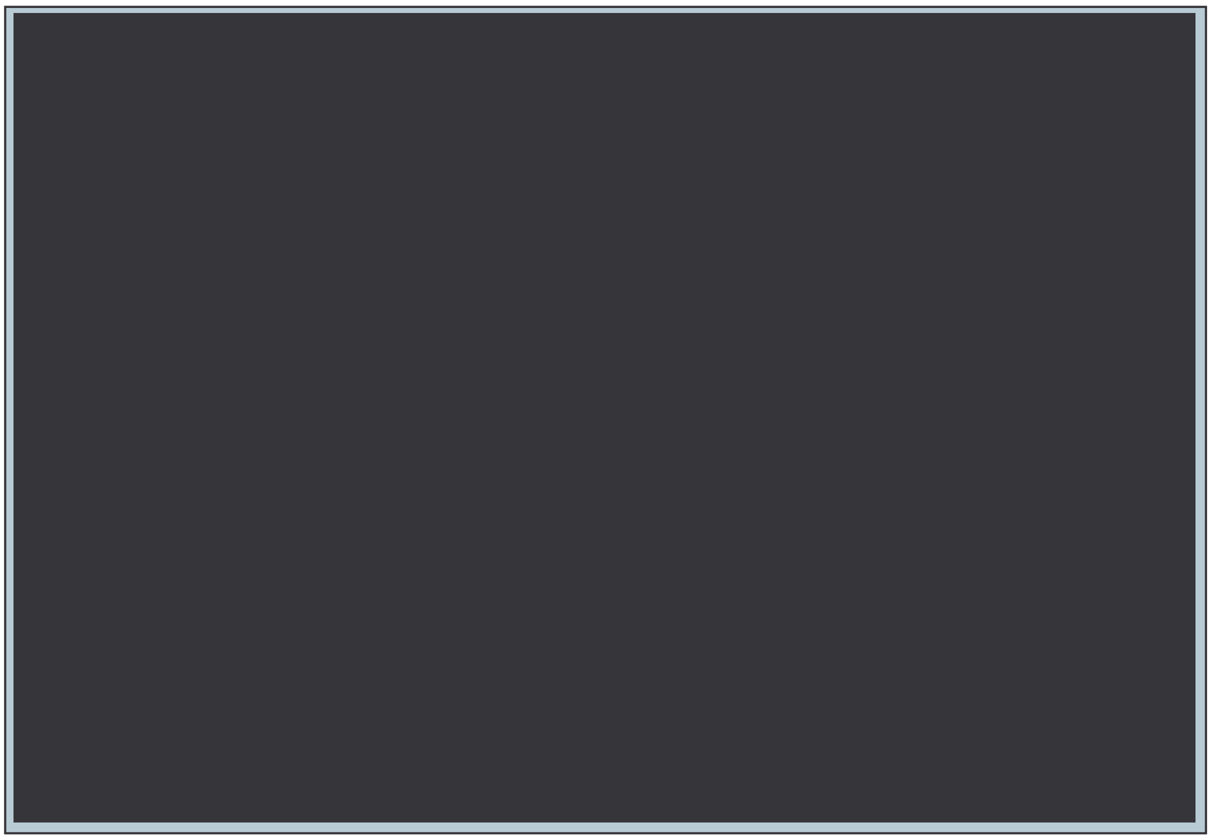
Vision	Acoustic
Manipulation	Speech
Motor	Motor
Watching others	Listening to others

Summary

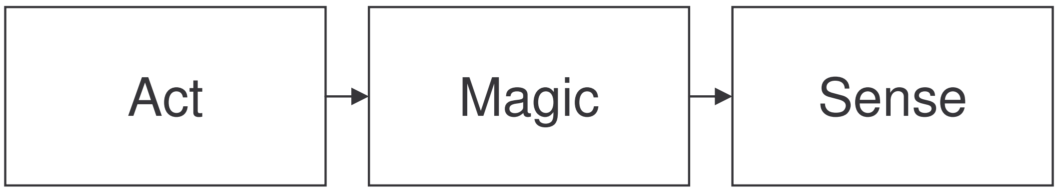
- Complex structure might be required
  - If we want to investigate complex interactions
- Complex skills emerge from the very structure of the motor interface
  - E.g. will mirror neurons be the same without manipulation?
- Interestingly, perceptual skills seem to be intertwined with the motor system
  - E.g. canonical neurons

One view of the problem





Another view



…and close the loop

- That is: invert the loop, action first!
- A motto: act to sense (better)

- I'll spend the next 50 slides or so in trying to show that there's something to be gained by taking this view

Sensory information for:

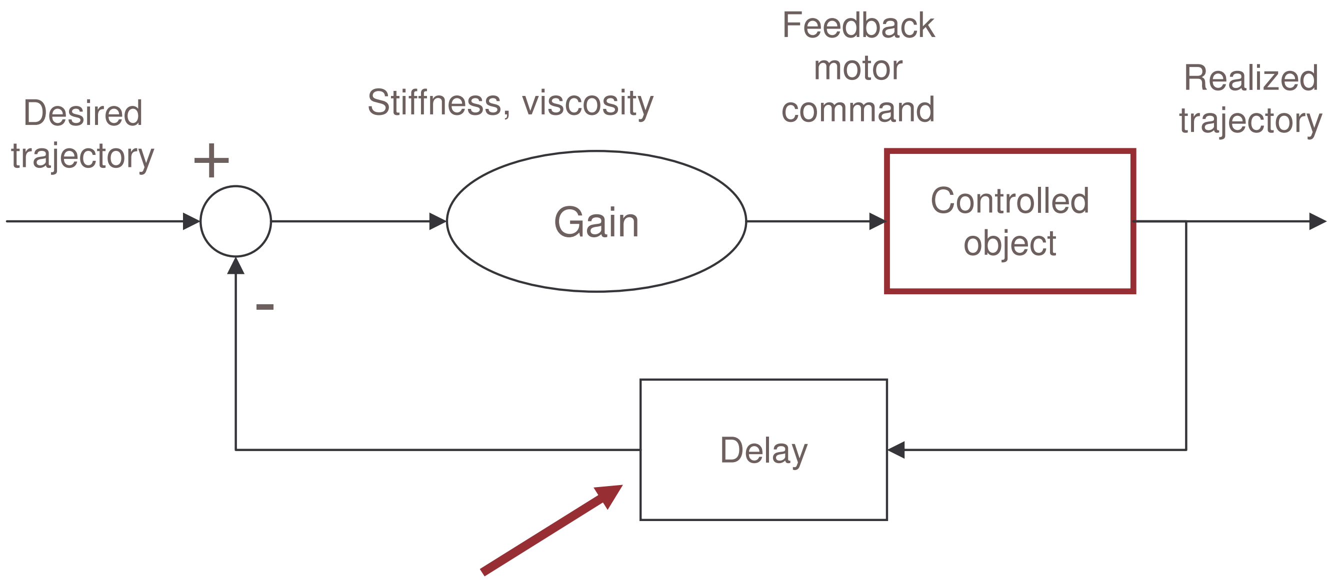
- Planning
  - Forming the desired trajectory/behavior
- Learning
  - How to adapt to changes in the physical conditions
- Prediction
  - Forecast the outcome of actions
- State estimation
  - Recovering the actual body state (position, speed)

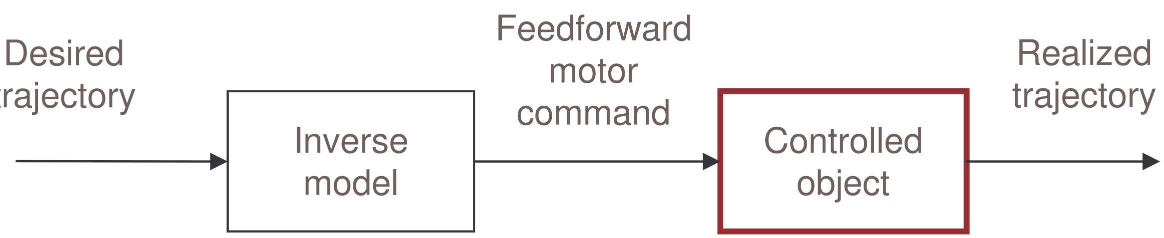
- …suggest that internal models might be required

Internal models come in two flavors

- Direct
  - From motor commands to executed behaviors
    - They can be used to predict the expected outcome of actions
    - …and to estimate the current state in spite of feedback delays
- Inverse
  - From desired behavior to motor commands
    - Given the desired state/behavior, they produce the motor commands required to attain that state/behavior

- A note: Internal models could be either functional mappings (input-output) or dynamical systems (differential equations)





From: M. Kawato. *Internal models for motor control and trajectory planning*. Current Opinion in Neurobiology 1999, 9:718–727

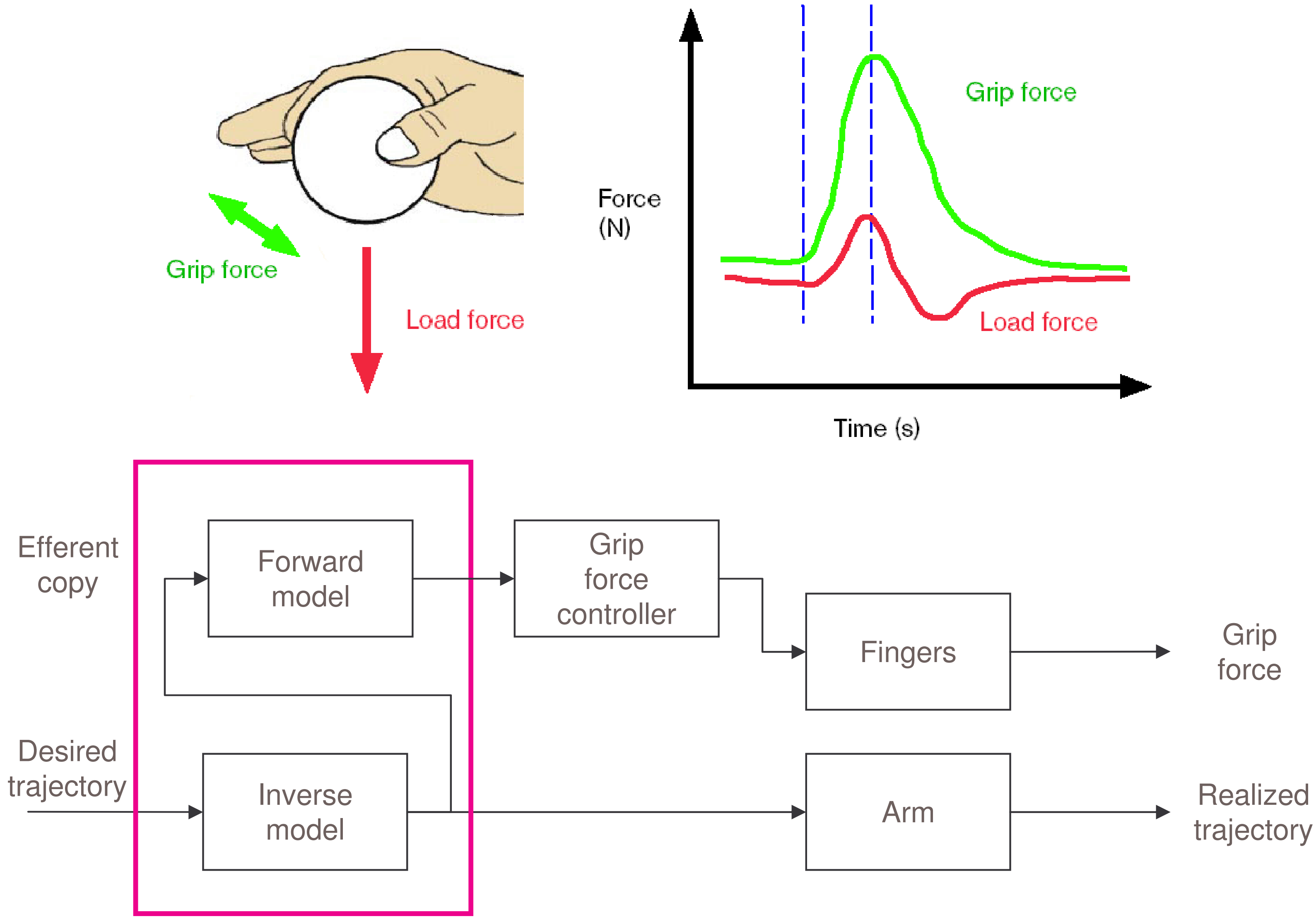
Effect of delays

Make feedback control either poor or unstable altogether

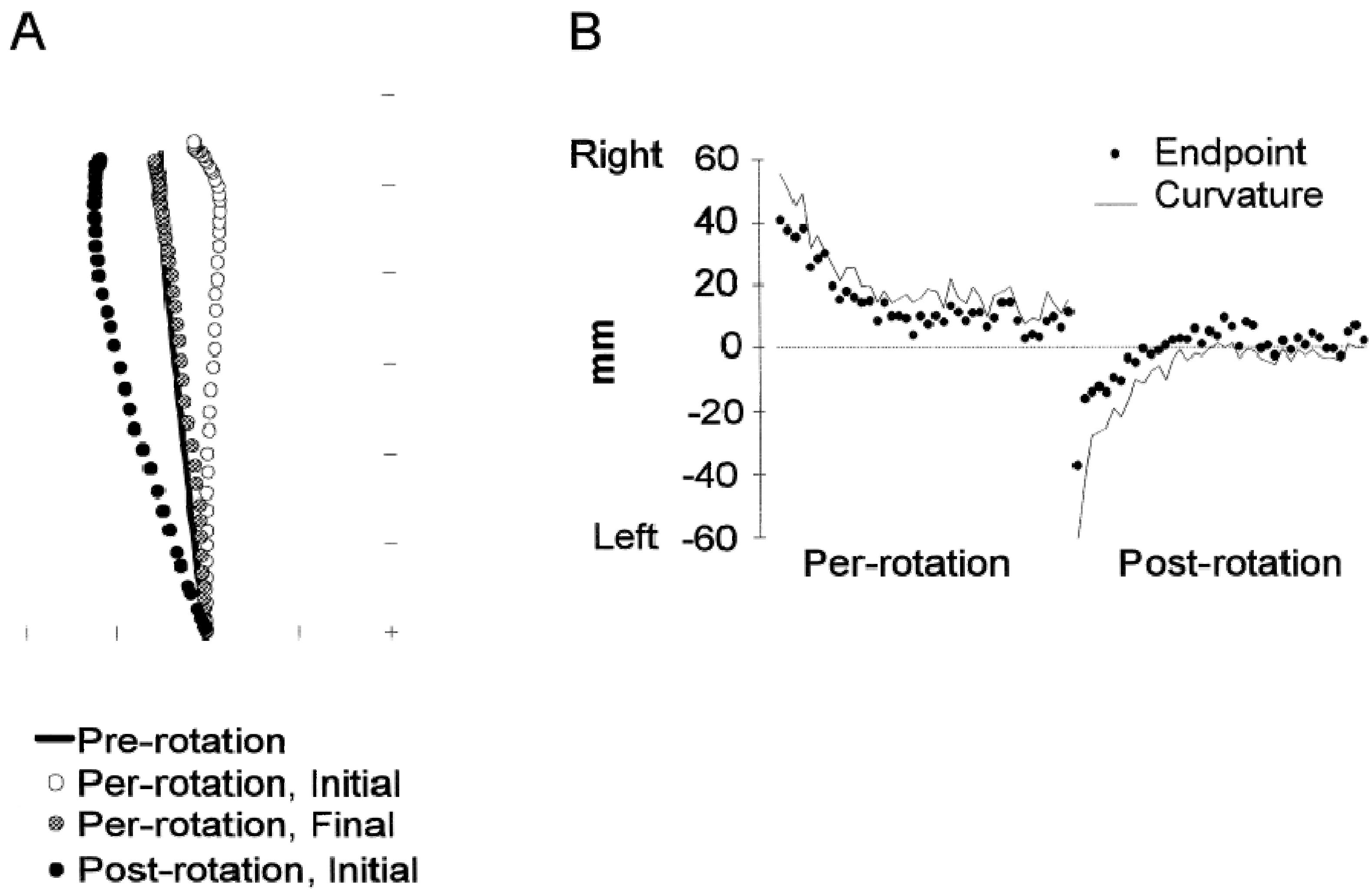
- Engineering control systems
  - Delays: 500μs
  - Movement duration: seconds
  - Gain of the controller: can be made high

- BTW: maintaining fast control loops is not an easy feat

- Humans
  - Delays:
    - 20-50ms (spinal)
    - 150-250ms (vision)
  - Movement duration: 150-500ms
  - Gain of the controller: stiffness and muscle viscoelastic properties (comparatively low)



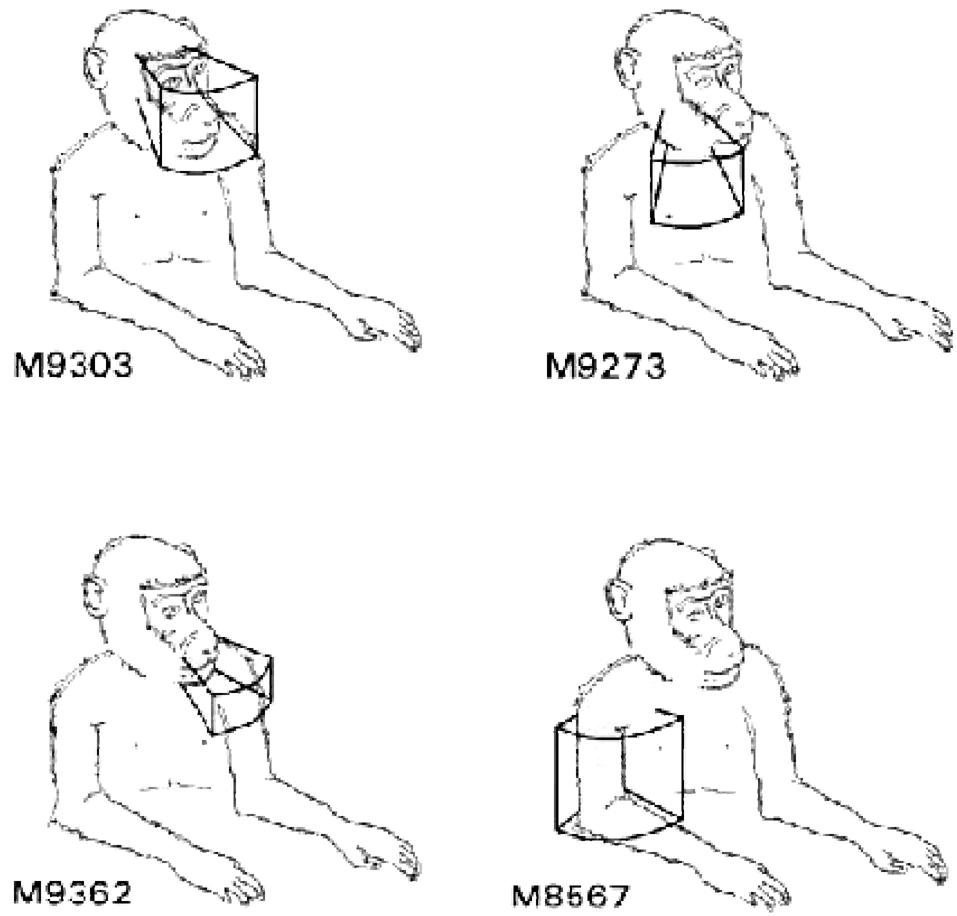
<u>From:</u> **Flanagan JR, Wing AM.** ***The role of internal models in motion planning and control: evidence from grip force adjustments during movements of hand-held loads.*** **Journal of Neuroscience 1997, 17:1519-1528.**



From: Lackner JR, Dizio P. *Gravitoinertial force background level affects adaptation to Coriolis force perturbations of reaching movements.* Journal of Neurophysiology 1998, 80:546-553.

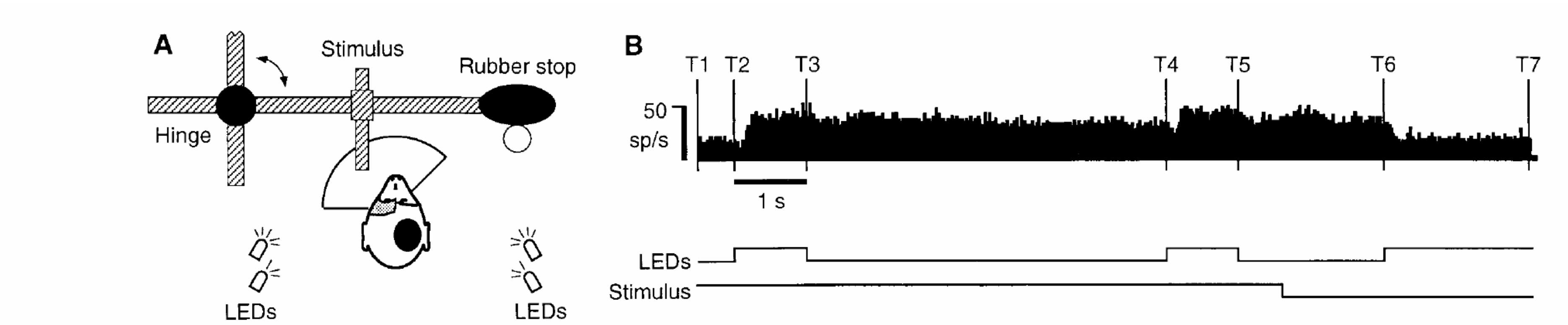
F4 neurons

- Visual and tactile responsive neurons.
- Representation of peripersonal space carved out of visual information because of motor activity.
- Receptive field linked to body part.



From: Fogassi L., Gallese V., Fadiga L., Luppino G., Matelli M., Rizzolatti G. *Coding of peripersonal space in inferior premotor cortex (area F4)*. Journal of Neurophysiology **76** (1) 1996.

Object constancy



T1: in the dark
T2: LEDs on, w/ object
T3: LEDs off

T4: LEDs on
T5: LEDs off, object removed
T6: LEDs on

From: Graziano M.S.A., Hu, X., Gross C.G. (1997) *Coding the location of objects in the dark*. Science **277** (July): 239-241.

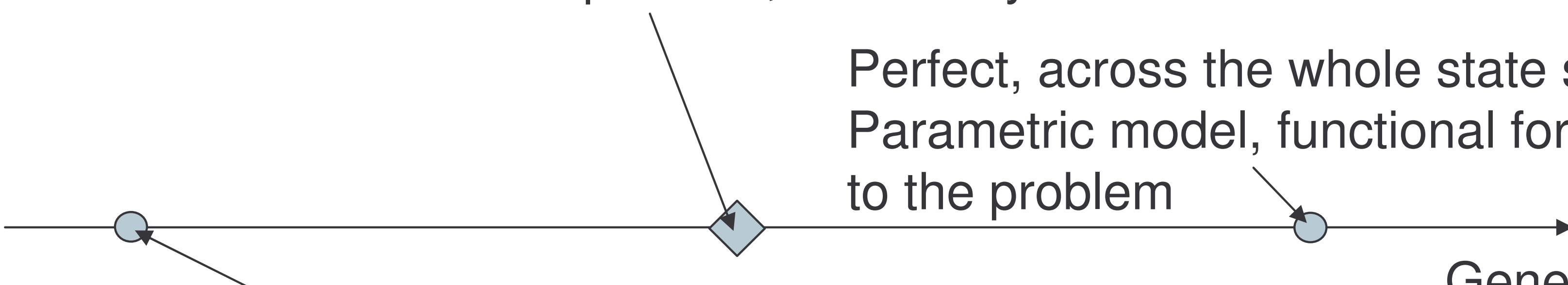
Building the internal models

- Rote-learning would be unpractical:
  - Too many possible actions (dof) for the available number of neurons (although they're quite a lot!)
- Generalizing past experience:
  - Past experience is bound (unfortunately) to "represent" only a portion of the whole state space
- Developing and extending the control structure to new behaviors
  - Sequencing and combining primitive behaviors appropriately
- Predicting the future course of action
  - It might turn out to be useful!

On generalization

Something in between
Generic components, modularity

Perfect, across the whole state space
Parametric model, functional form tuned to the problem



Generalization

None, apart from the explored portion of the state space
Lookup table approach

$$e_{generalization} \leq S_{z,H} + A_H$$

VC dimension (complexity of the Hypothesis space of functions *H*)

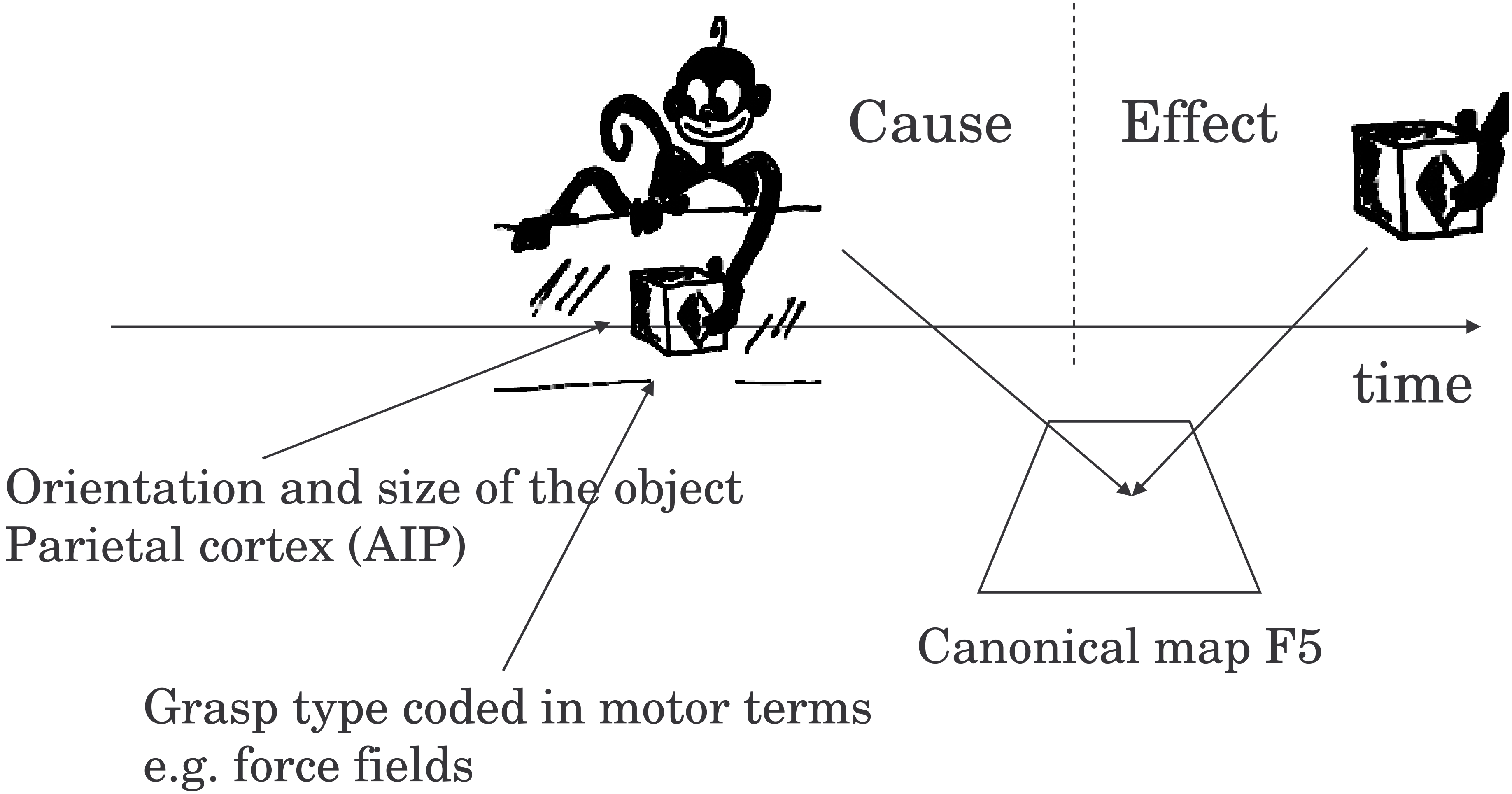
Remaining questions

- Learning when *the number of samples* is small
  - Like humans do!
- Hierarchical organization, partitioning big problems, reusing components
  - e.g. modular controllers
- Learning across developmental time scales
  - May be worth looking at human development
  - Exploiting large quantity of data when available

[Part 3]

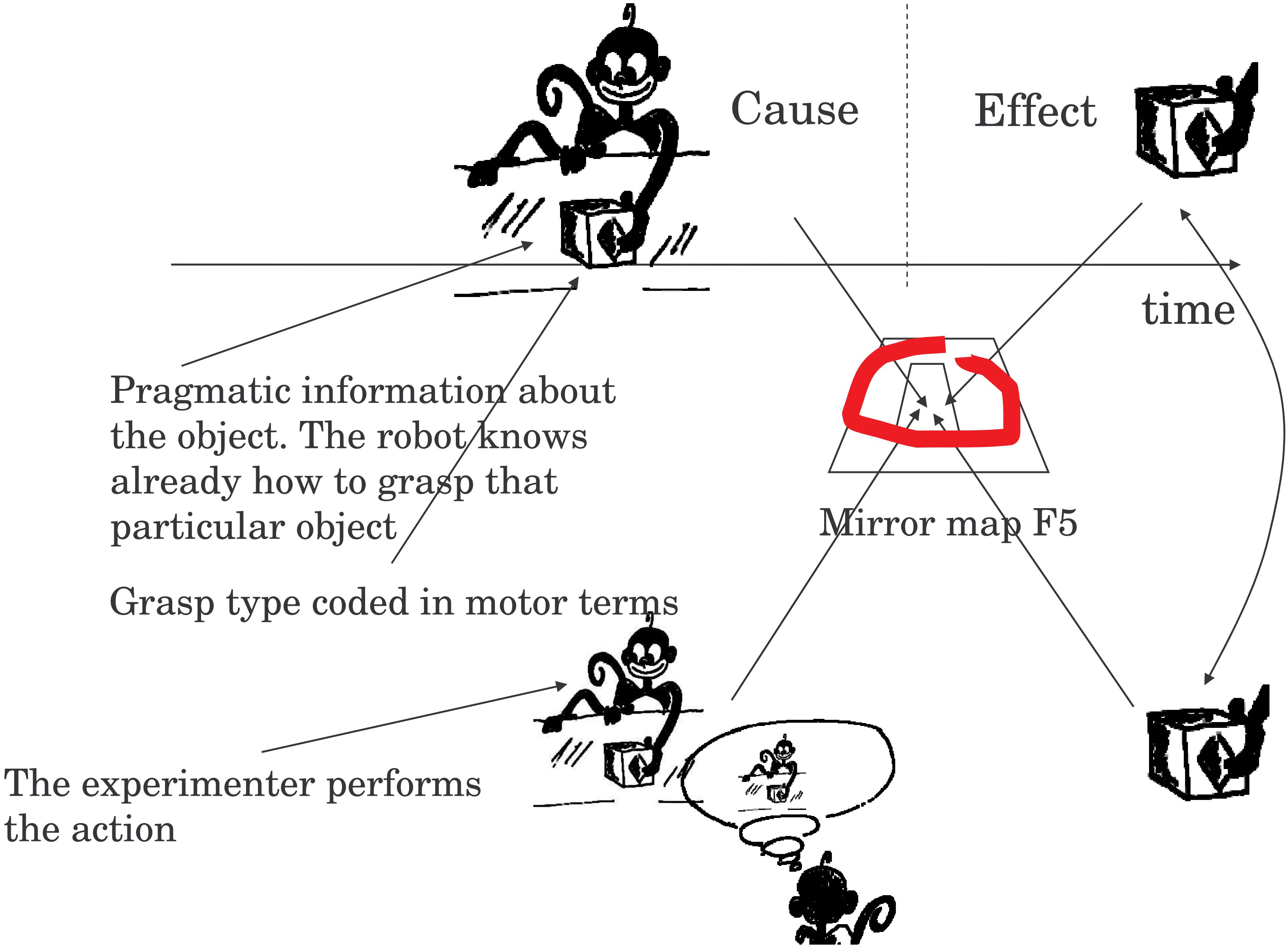
Model of the mirror neurons system

- **Developmental aspect:** context information specifies the goal – a prerequisite for linking action execution to action observation
  - canonical neurons and the detection of object affordances are a prerequisite for mirror neurons
- **Mirror neuron response:** *A* feedback system co-opted in recognizing actions
  - action recognition in motor terms



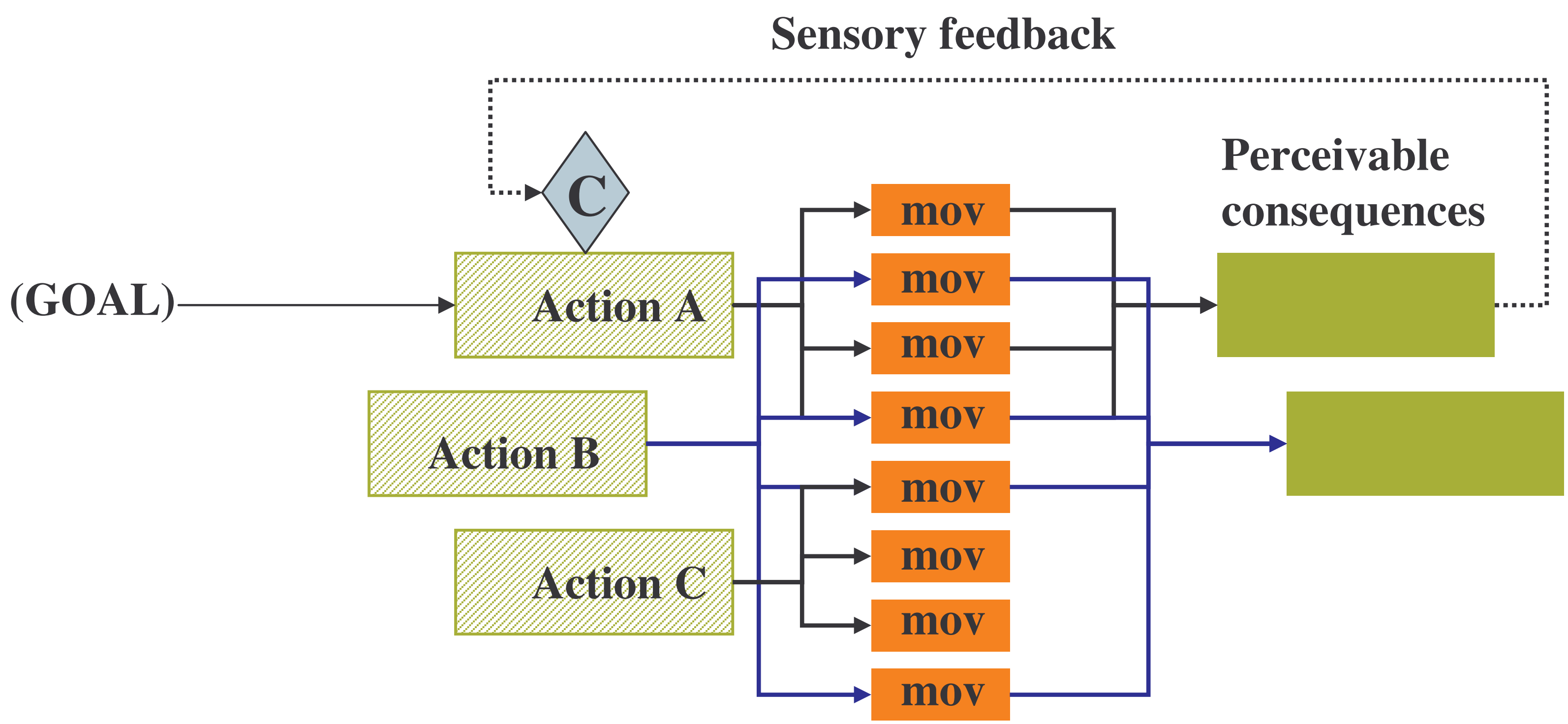
The neural map can be imagined as a table filled with object-grasp type pairs.

The teaching signal is the successful grasping of the object which can be reinforced by both tactile and visual stimuli.



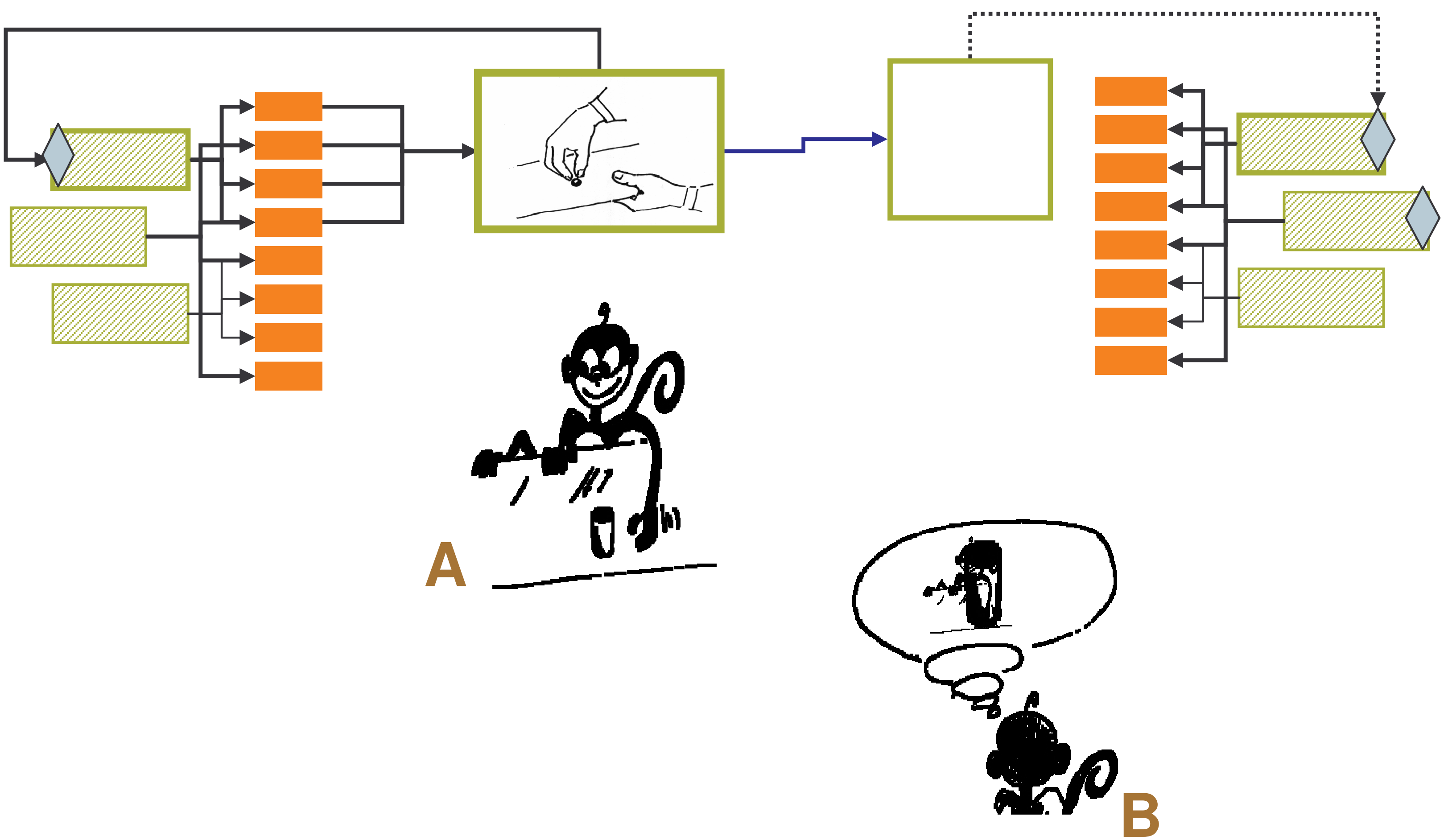
Cause
Effect
time
Pragmatic information about the object. The robot knows already how to grasp that particular object
Mirror map F5
Grasp type coded in motor terms
The experimenter performs the action

A cartoon-like explanation…



Exchange of information

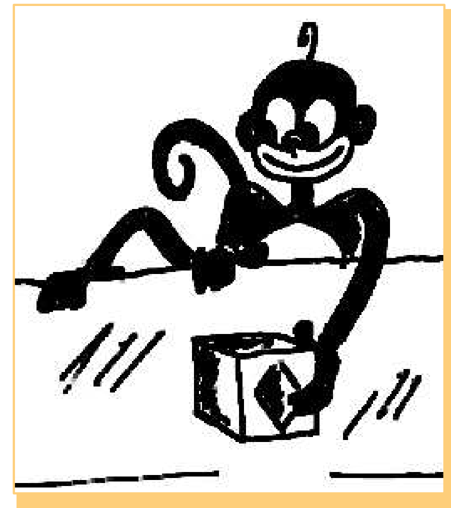
Individual A | Individual B



F5 neurons

Canonical neurons

Active when manipulable objects are presented visually



Mirror neurons

Active when another individual is seen performing manipulative gestures

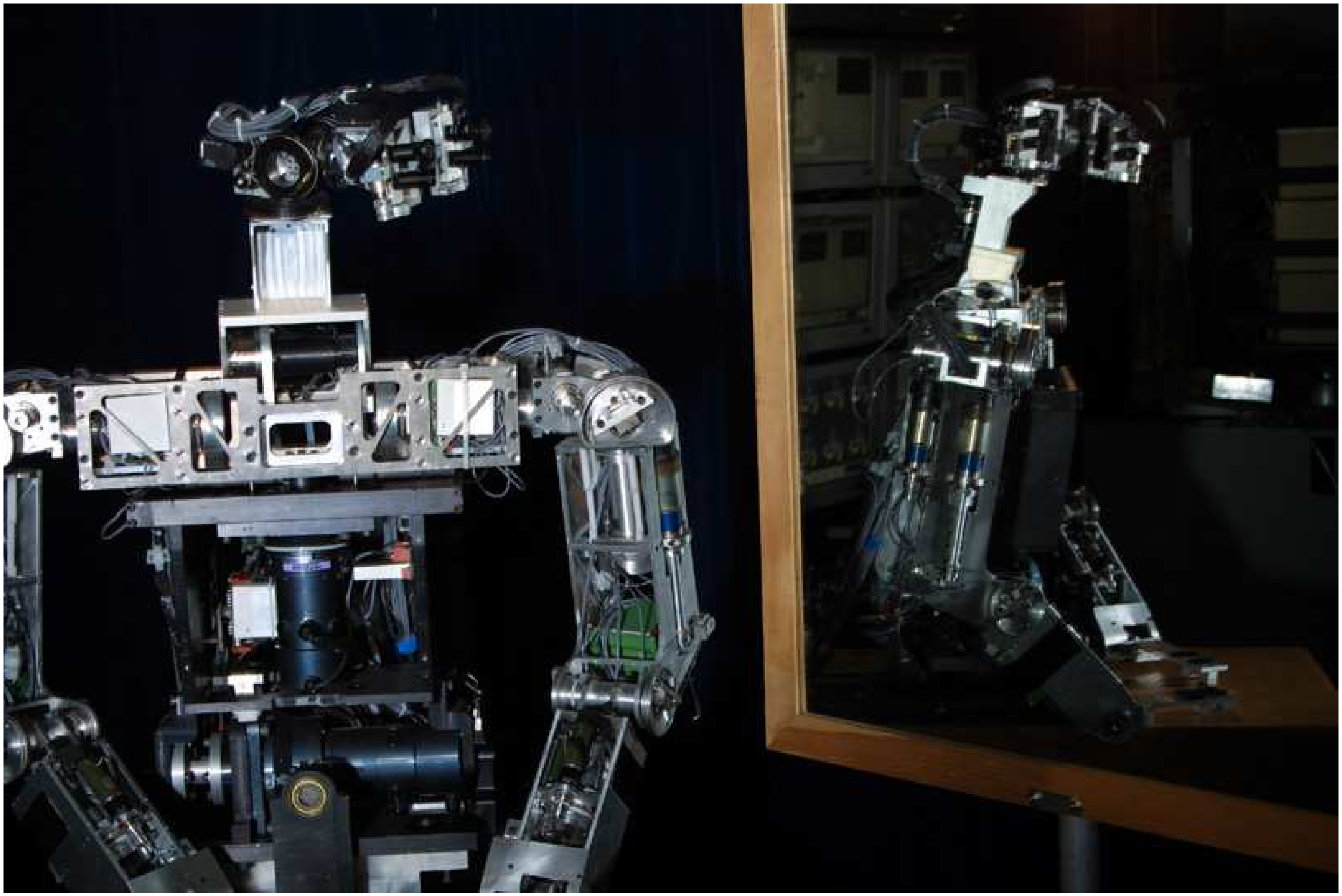
Development in a two-stage model

- **First:** learn to interact with objects.
  - Manipulate objects.
  - Acquire the motor primitives.
  - Generate the *canonical neurons* representation.
  - Learn by interacting with the environment (the error is measured directly).
- **Second:** learn the *mirror* representation.
  - Correlate the experimenter's action with the performed action (goal).
  - This requires *canonical neurons* to be constructed in advance.

Simplest Form of Manipulation

- What is the simplest possible manipulative gesture?
  - Contact with object is necessary; can't do much without it
  - Contact with object is sufficient for certain classes of affordances to come into play (e.g. rolling)
  - So can use various styles of poking/prodding/tapping/swiping as basic manipulative gestures
  - (if willing to omit the *manus* from manipulation…)

Experimental setup…



A bit more lively…

Objects come to existence because they are manipulated

Fixate target — *Track visual motion…* — *(…including cast shadows)* — *Detect moment of impact* — *Separate arm, object motion* — *Segment object*

Which edge should be considered?

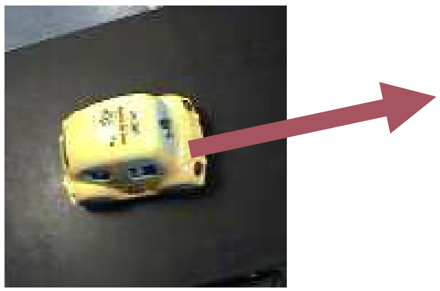
Maybe some cruel grad-student glued the cube to the table

Color of cube and table are poorly separated

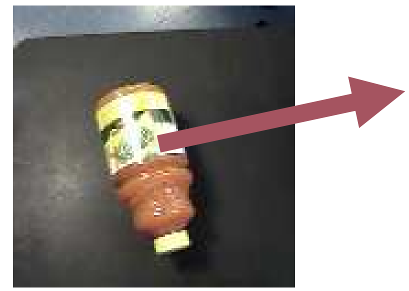
Cube has misleading surface pattern

by Paul Fitzpatrick

Exploring an affordance: rolling

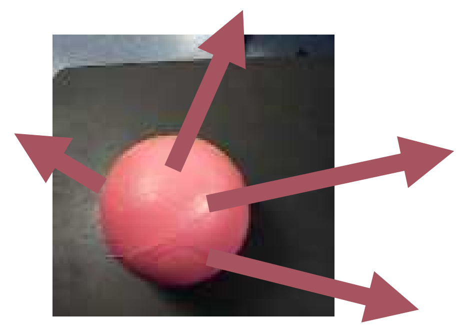


A toy car: it rolls in the
direction of its principal axis



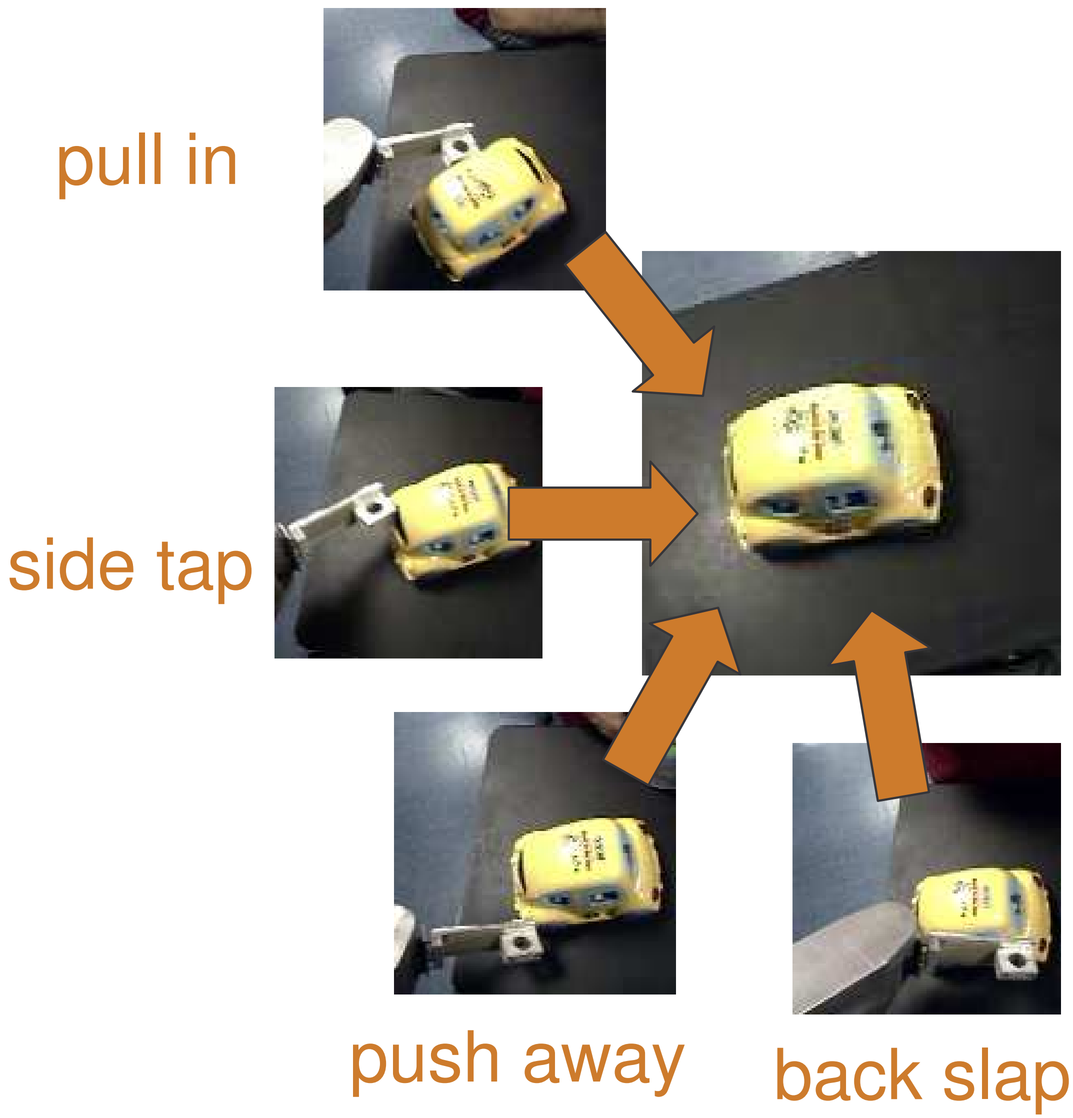
A bottle: it rolls orthogonal to the
direction of its principal axis

A toy cube: it doesn’t roll,
it doesn’t have a principal axis



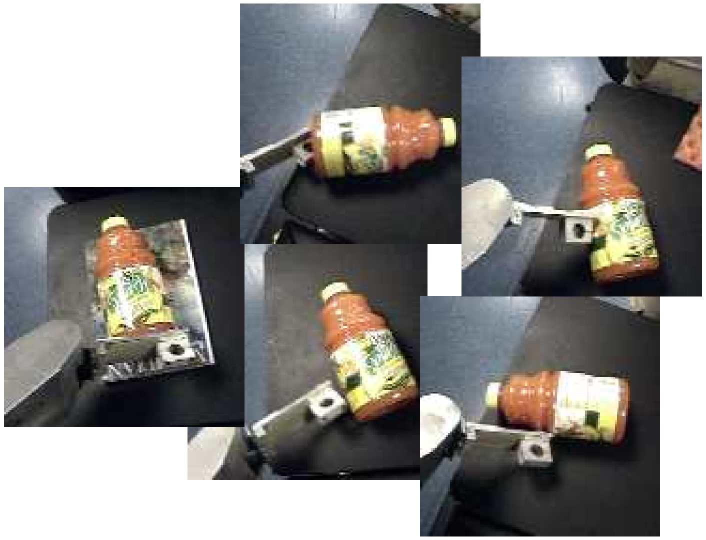
A ball: it rolls,
it doesn’t have a principal axis

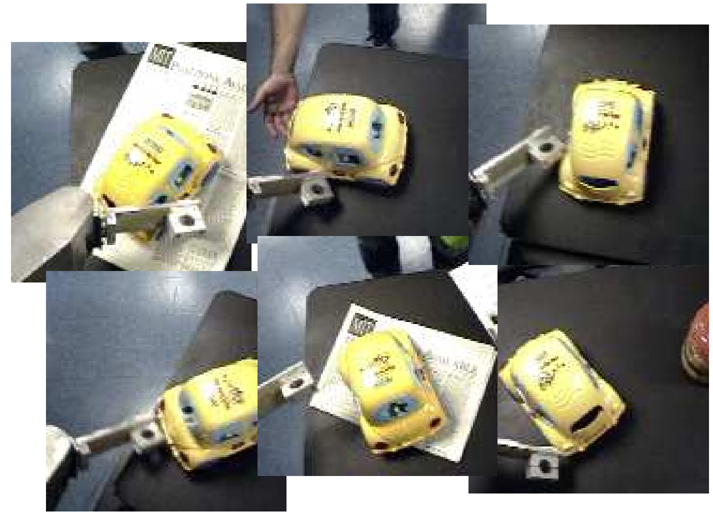
Gesture “vocabulary”

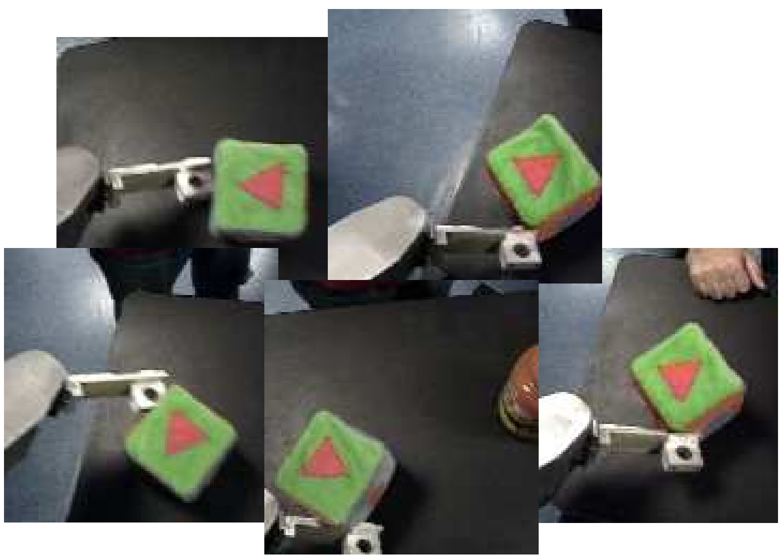


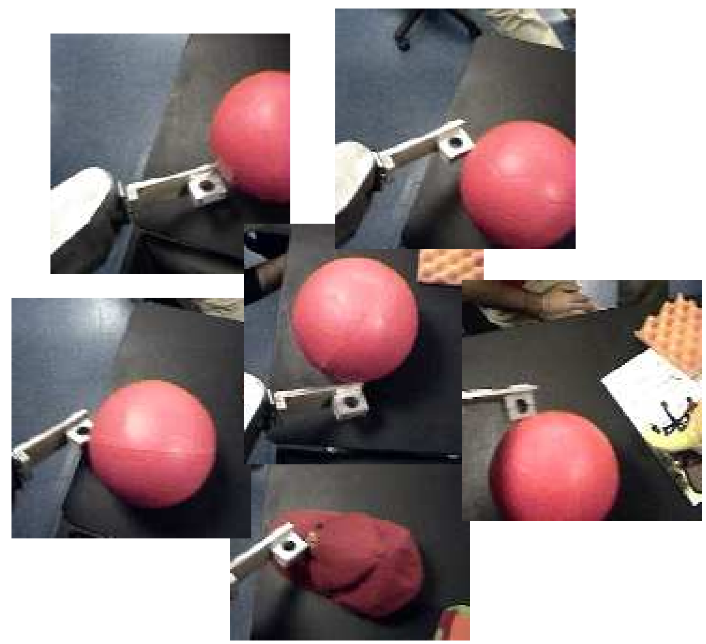
Exploring an affordance: rolling

Forming object clusters

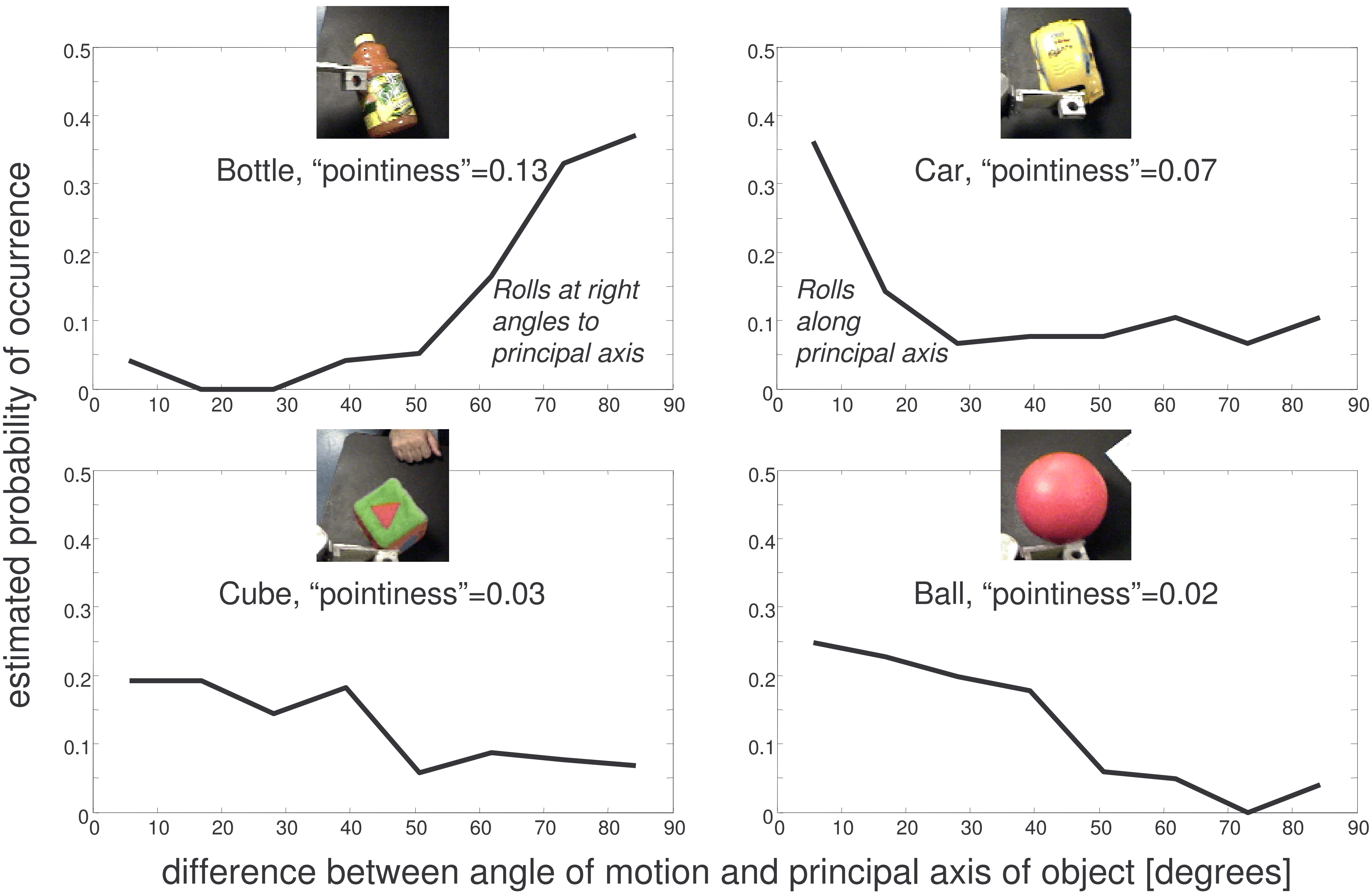




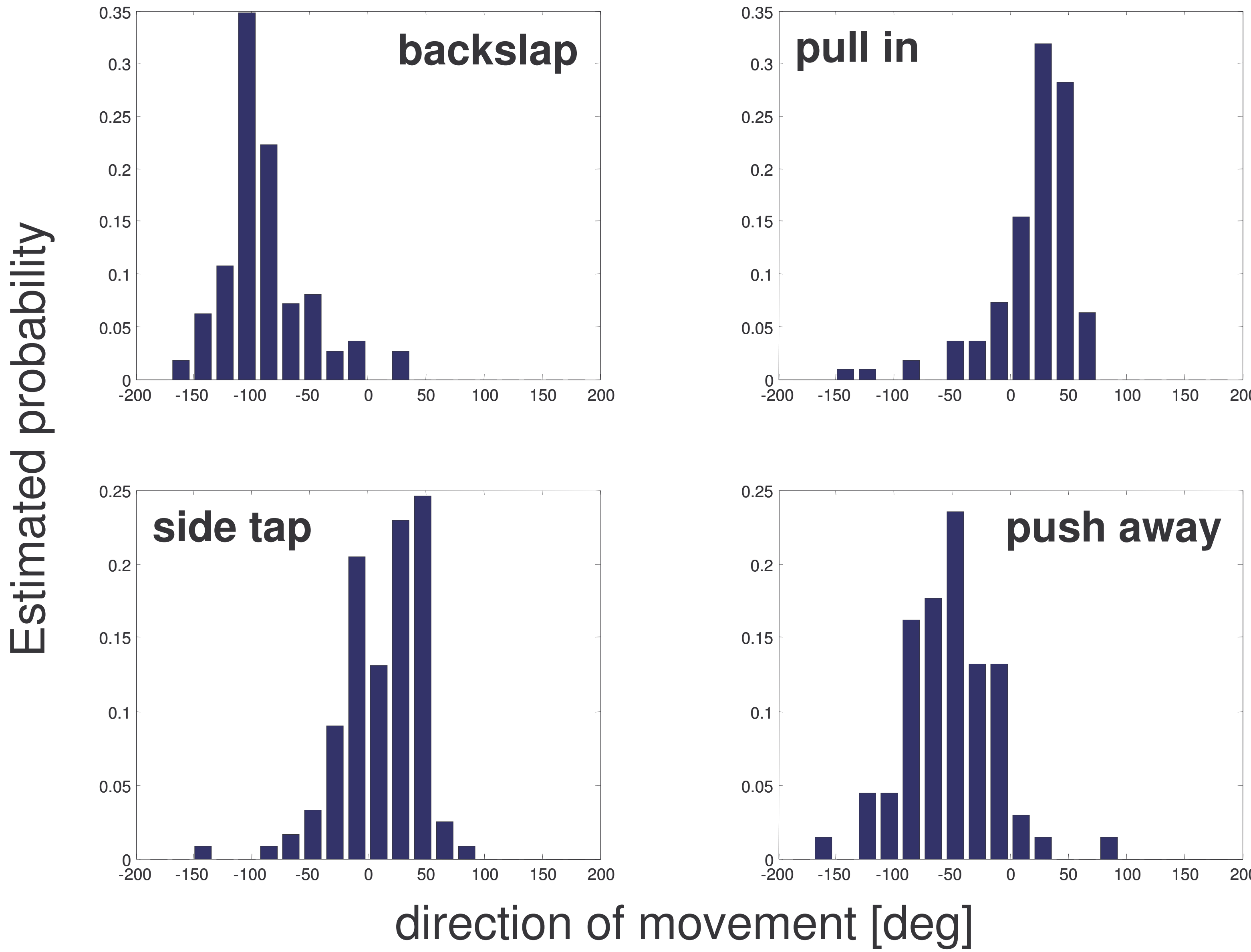




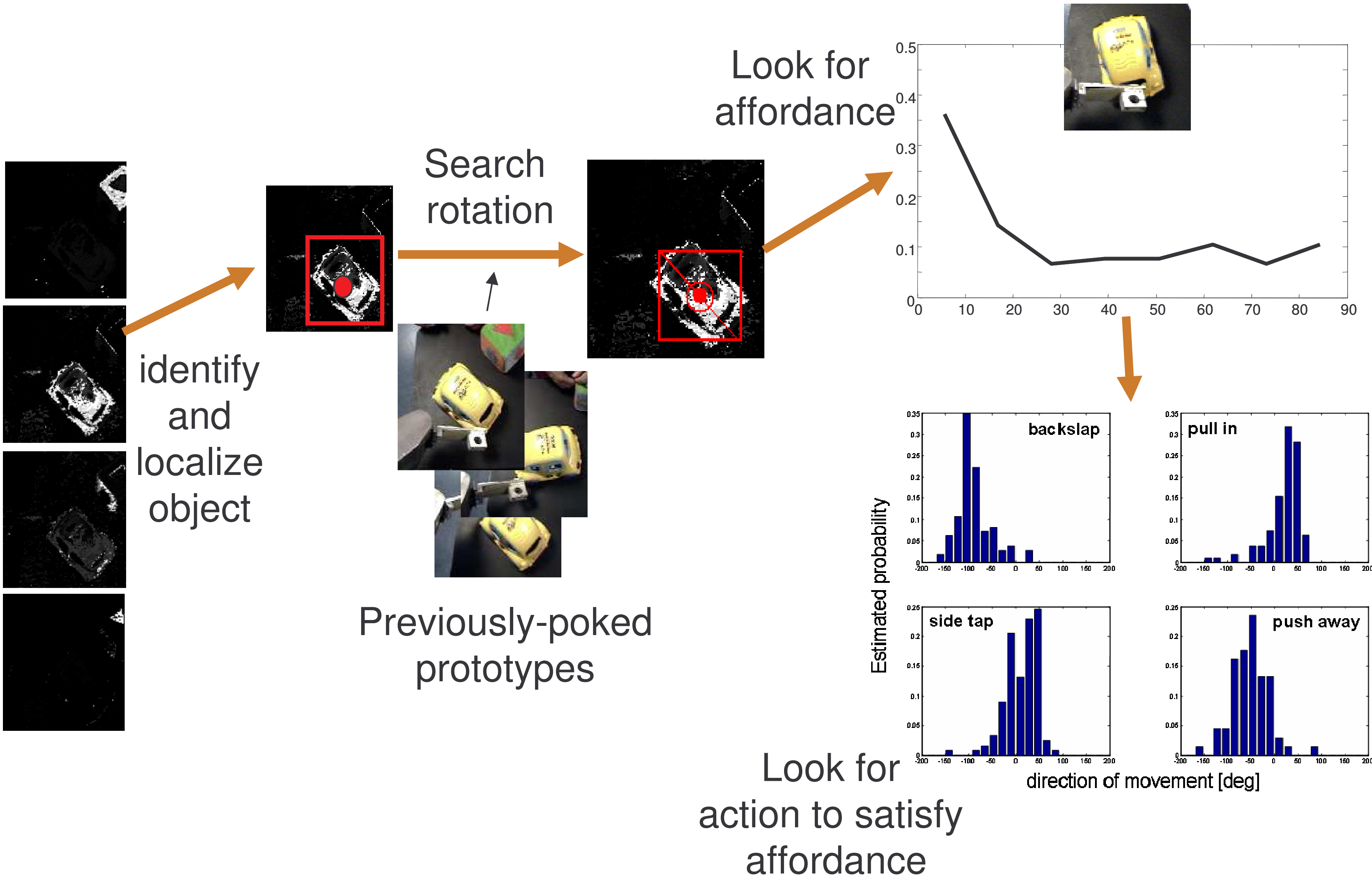
Into object affordances…



The geometry of poking

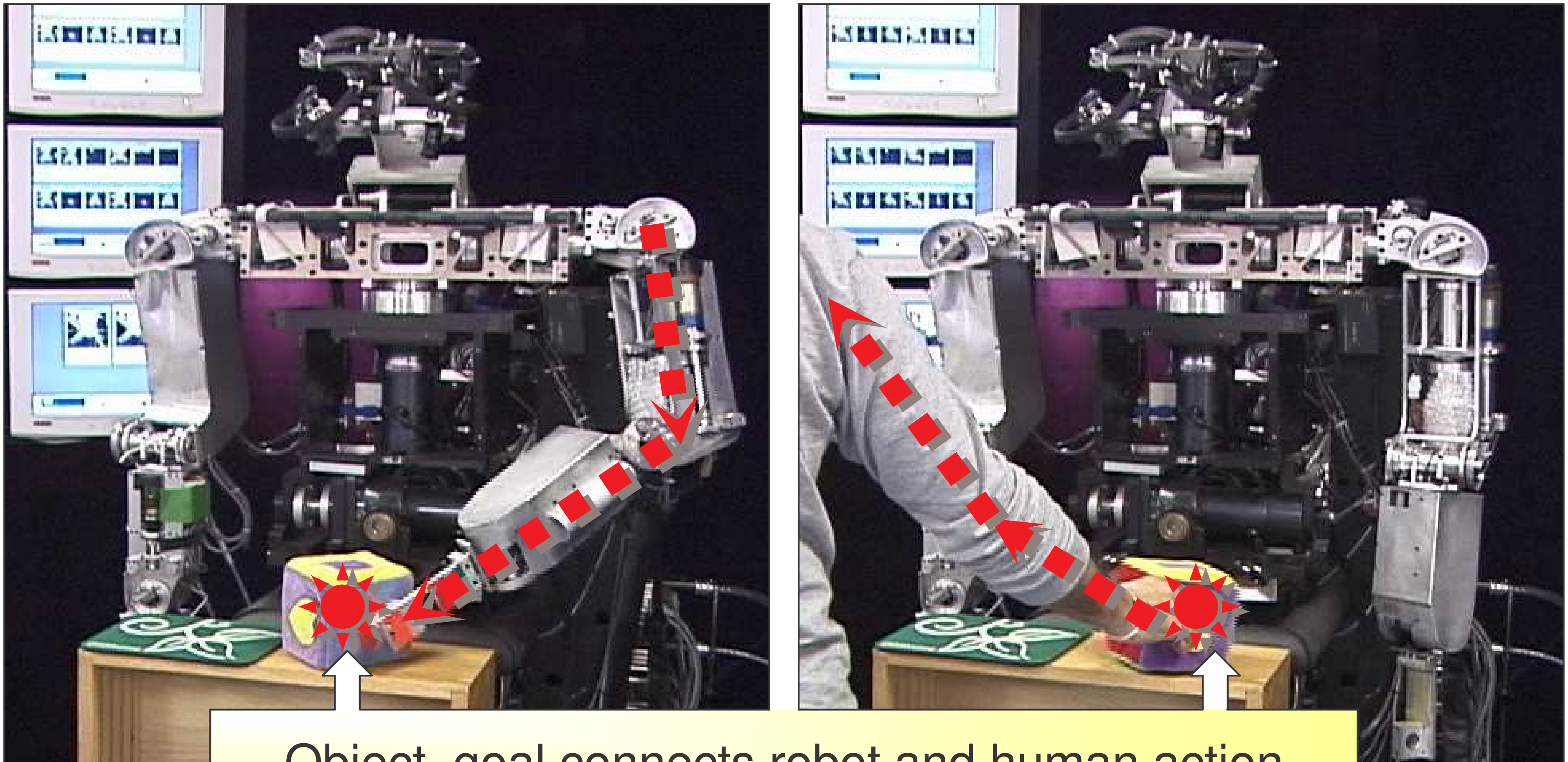


Behavior: poking according to affordance



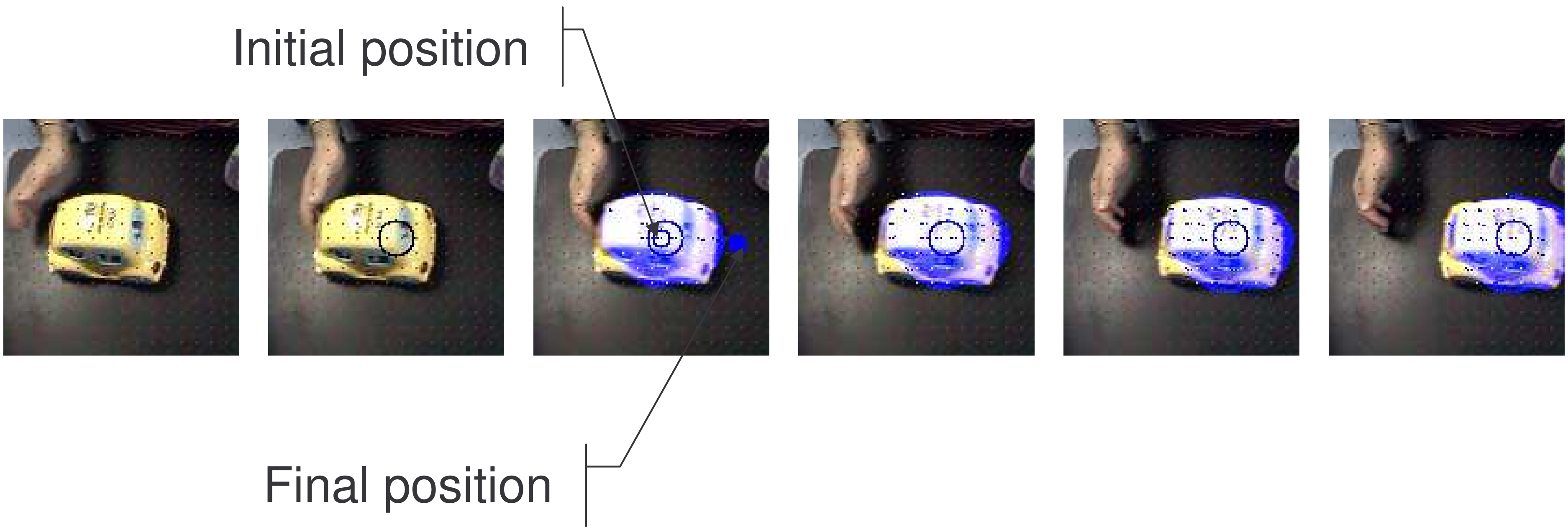
Behavior: poking according to affordance

Understanding a foreign manipulator



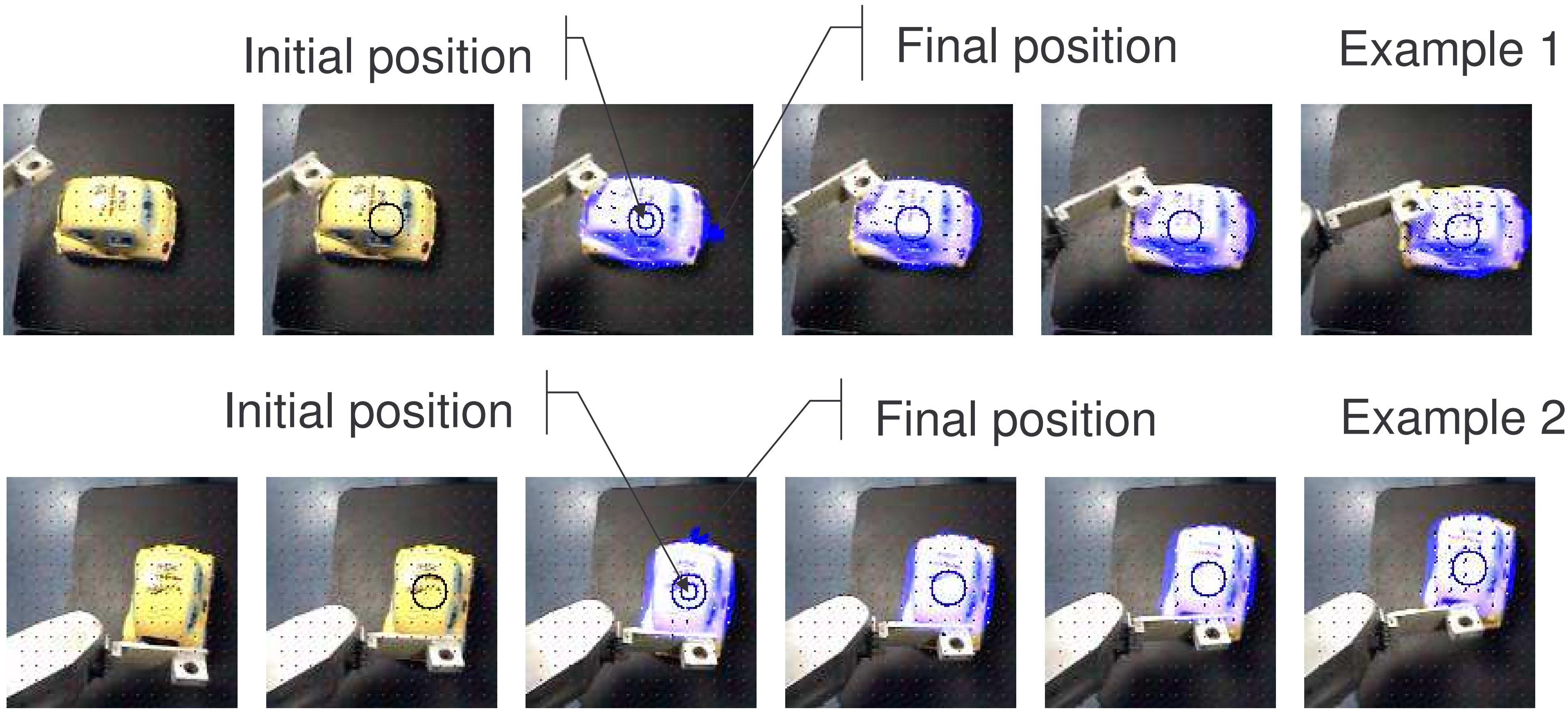
Object, goal connects robot and human action

Interpreting observations



A foreign manipulator (human) pokes an object
The direction of movement is compared with the object affordance

Behavior: mimicry



The robot mimics the observed action trying to fulfill the goal rather than an actual movement

Interpreting observations

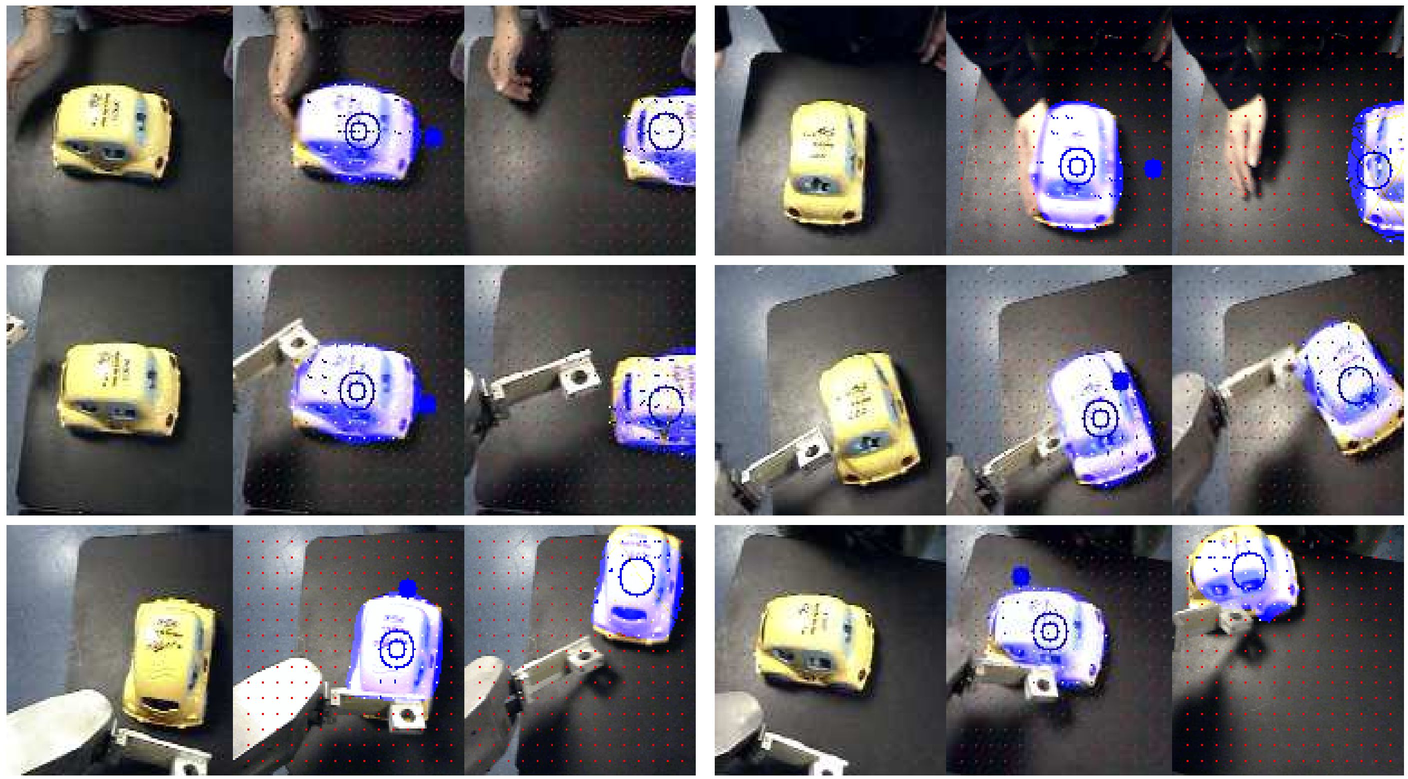
Invoking the object's natural rolling affordance

Going against the object's natural rolling affordance

Demonstration by human

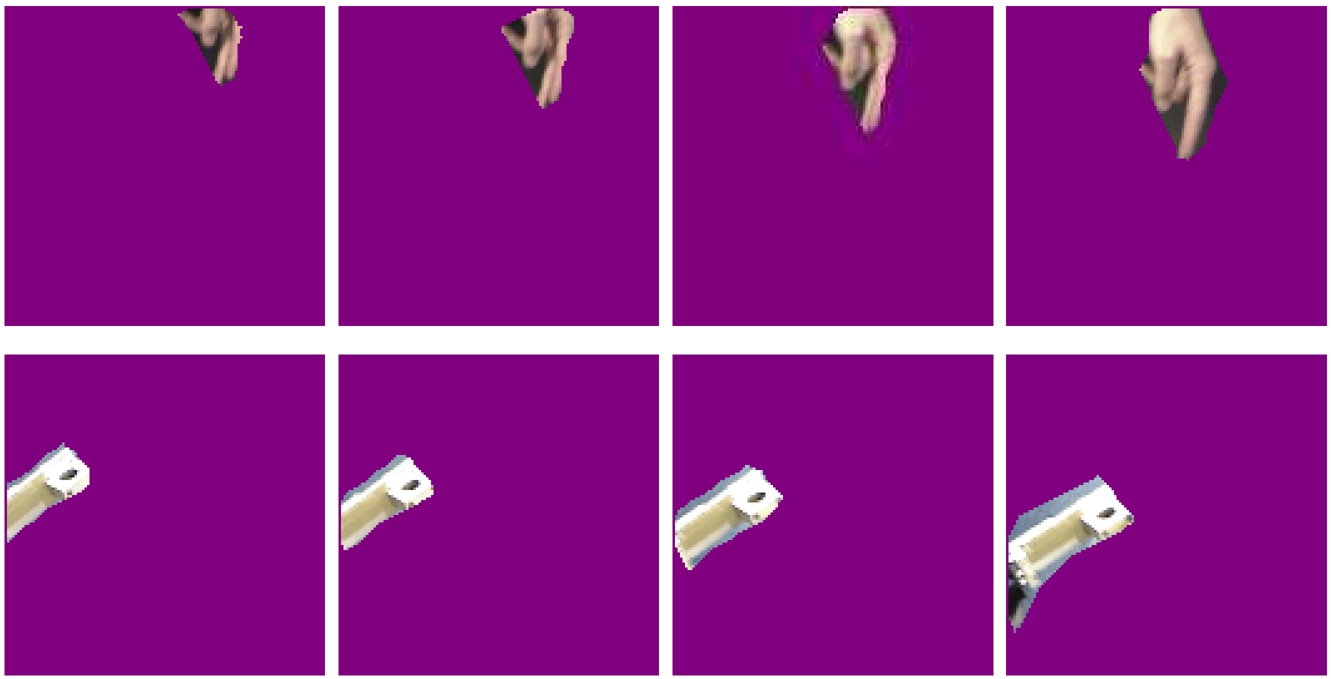
Mimicry in similar situation

Mimicry when object is rotated



Mimicry

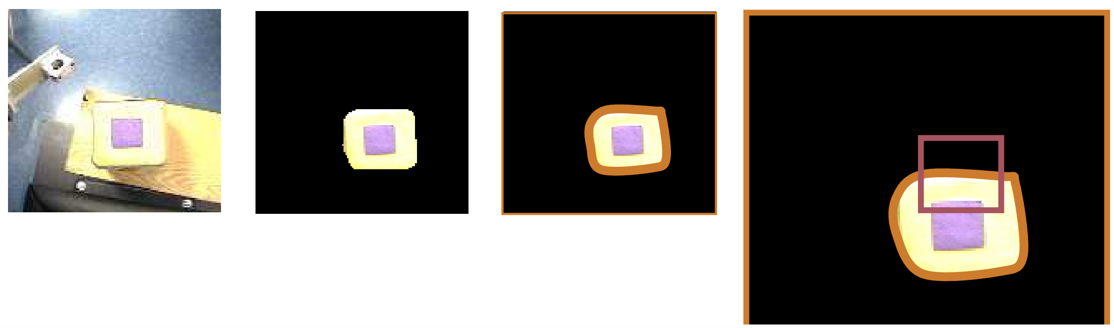
Manipulators...

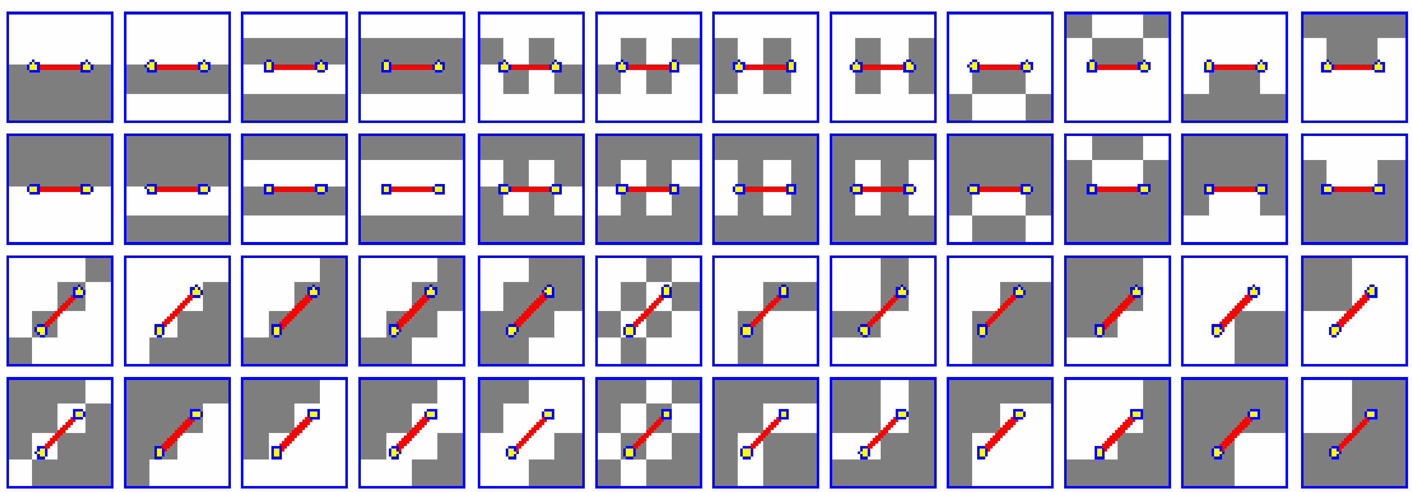


For details: Fitzpatrick 2003, From First Contact to Close Encounters: A Developmentally Deep Perceptual System for a Humanoid Robot

Bootstrap better vision

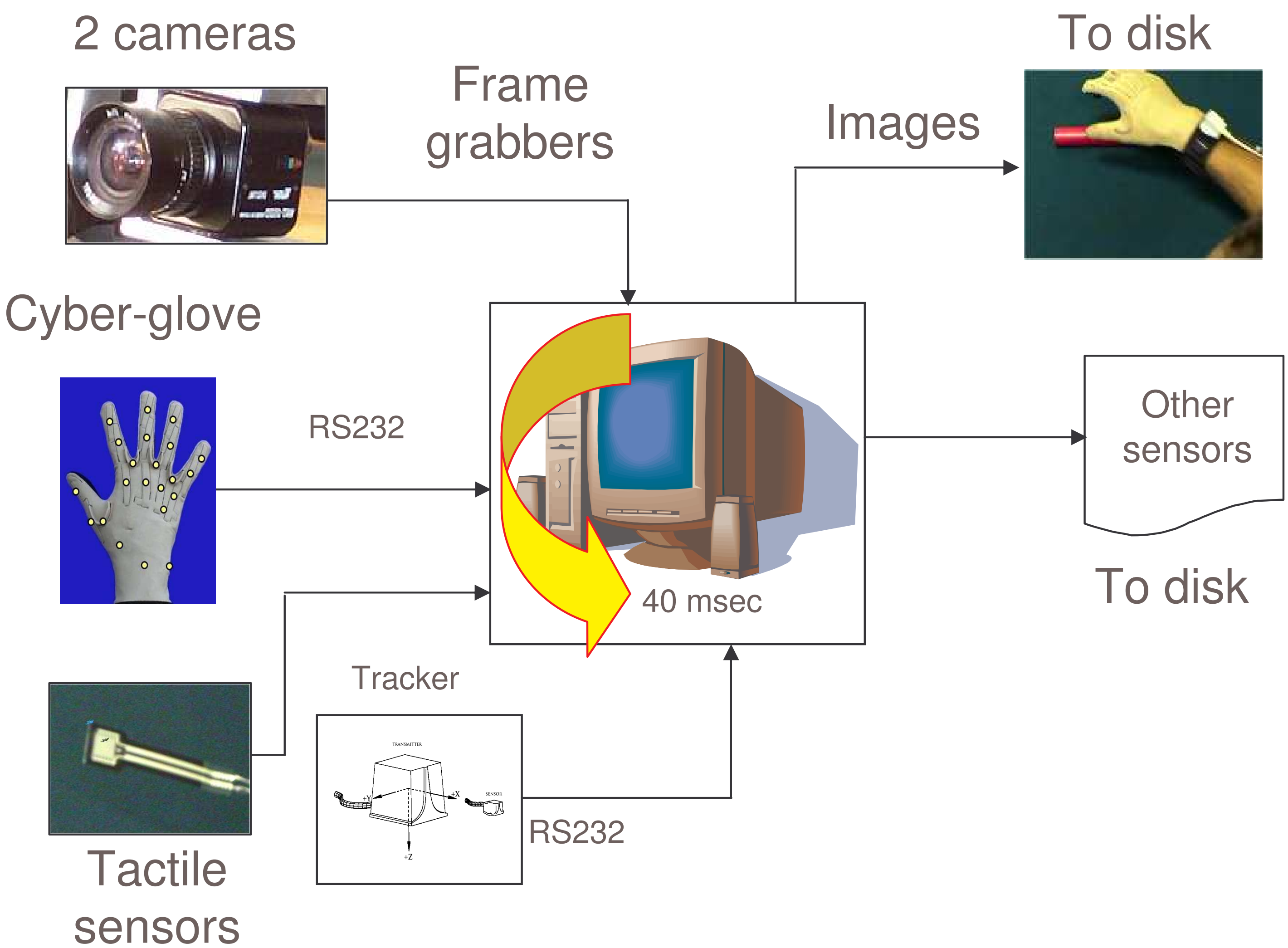
- Benefit of manipulation are non-trivial
  - e.g. training edge detectors





For details: Fitzpatrick 2003

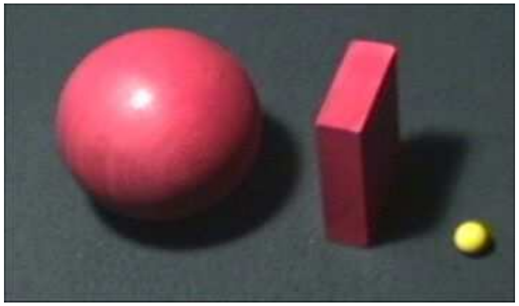
Data from human grasping

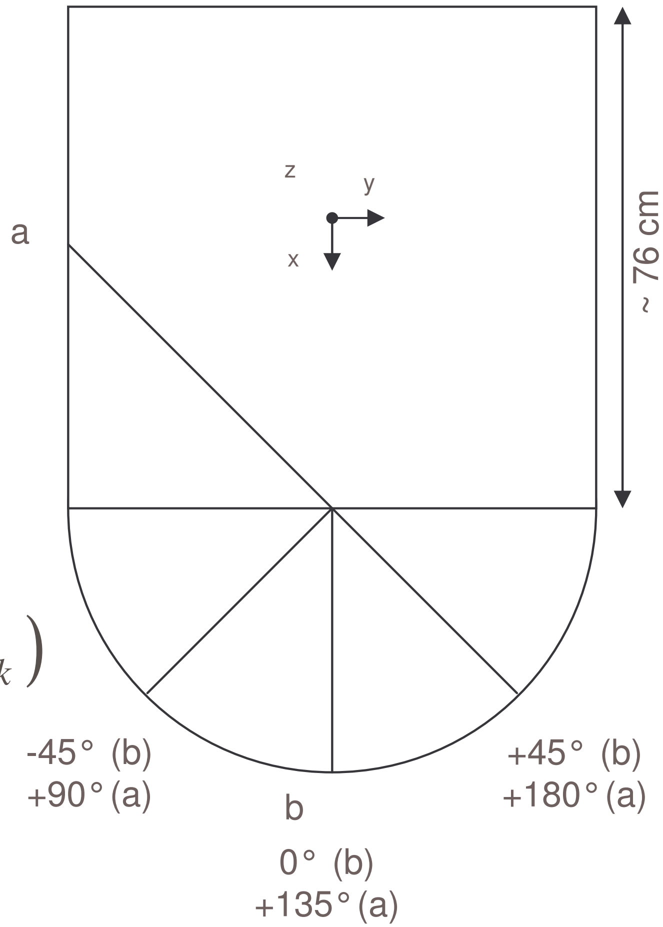


Bayesian classifier

168 sequences per subject
10 subjects
6 complete sets

{G_i}: set of gestures
F: observed features
{O_k}: set of objects



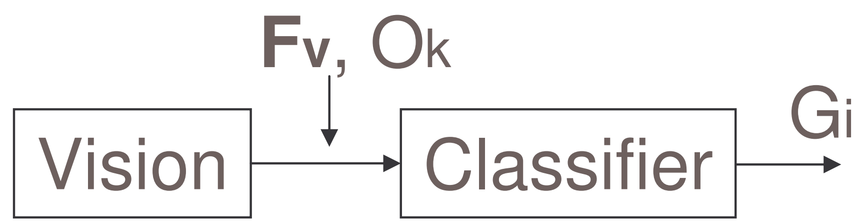


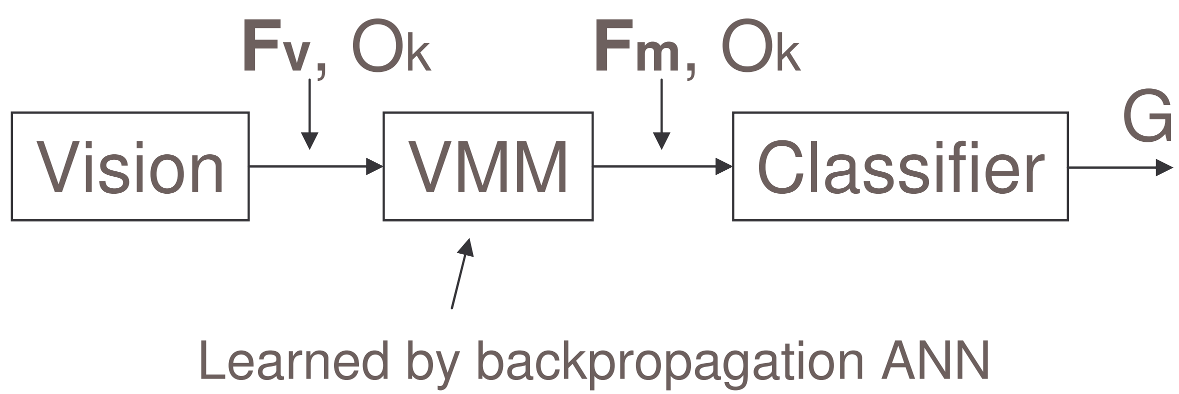
$p(G_i|O_k)$: priors (affordances)
$p(\mathbf{F}|G_i,O_k)$: likelihood to observe **F**

$$p\left(G_i \mid \mathbf{F}, O_k\right) = p\left(\mathbf{F} \mid G_i, O_k\right) p\left(G_i \mid O_k\right) / \, p\left(\mathbf{F} \mid O_k\right)$$

$$\hat{G}_{MAP} = \arg\max_{G_i} \left(G_i \mid \mathbf{F}, O_k\right)$$

Two types of experiments

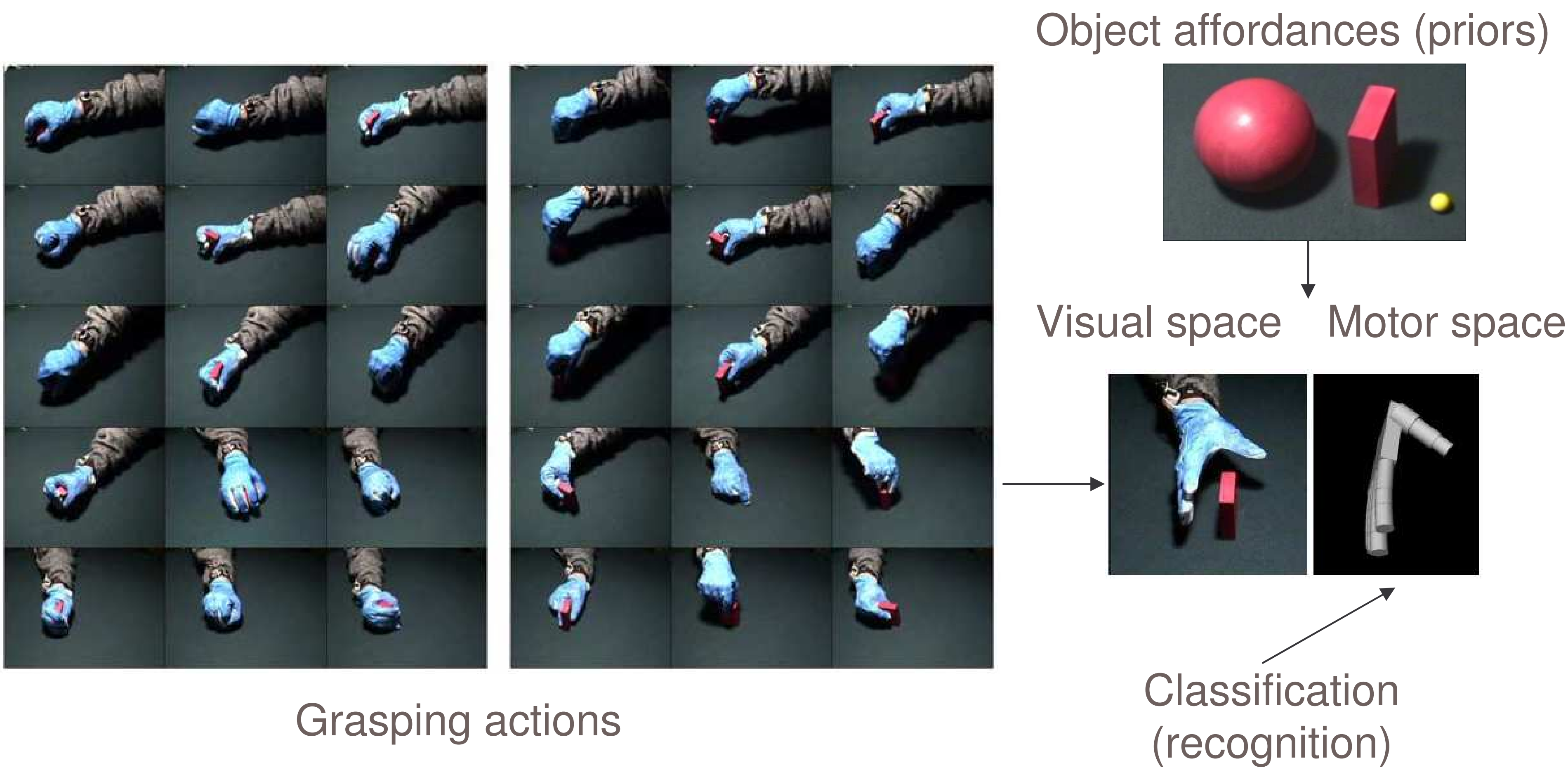




Estimation

- p(Gi|Ok): affordances, by counting, estimated on the whole database
- p(**F**|Gi,Ok): EM algorithm on the parameters of a mixture of Gaussians (from Matlab implementation)
- VMM: Neural network, sigmoidal activation units, linear output, trained on the whole database

Role of motor information in action understanding



Understanding mirror neurons: a bio-robotic approach. *G. Metta, G. Sandini, L. Natale, L. Craighero, L. Fadiga.* Interaction Studies. Volume 7 Issue 2. 2006

Some results…

	Exp. I (visual)	Exp. II (visual)	Exp. III (visual)	Exp. IV (motor)
	Training			
# Sequences	16	24	64	24
# of view points	1	1	4	1
Classification rate	100%	100%	97%	98%
# Features	5	5	5	15
# Modes	5-7	5-7	5-7	1-2
	Test			
# Sequences	8	96	32	96
# of view points	1	4	4	4
Classification rate	100%	30%	80%	97%

Pre-conclusion

[such tools] permit the [users] to do the tasks that need to be done while doing the kinds of things people are good at: recognizing patterns, modeling simple dynamics of the world, and manipulating objects in the environment. (Hutchins, 1995).