# AI and the Impending Revolution in Brain Sciences

AAAI Presidential Address

Tom M. Mitchell
Carnegie Mellon University
August 2002

Collaborators: Marcel Just, Radu Stefan Niculescu, Francisco Pereira, Xuerui Wang

# Thesis of This Talk

The synergy between AI and Brain Sciences will yield profound advances in our understanding of intelligence over the coming decade, fundamentally changing the nature of our field.

The synergy between AI and Brain Sciences will yield profound advances in our understanding of intelligence over the coming decade.

1. Common goal: understand intelligence

2. Significant correspondences between AI methods and brain organization

3. New instrumentation is causing a revolution

# Outline

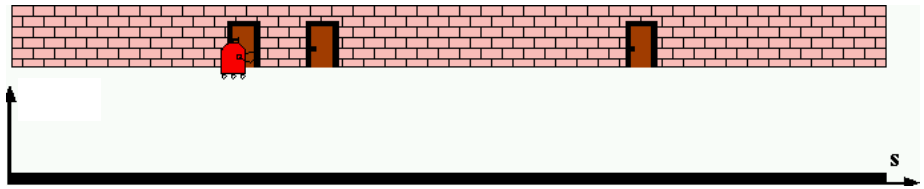
1. The Thesis
2. AI processes and brain processes
3. New instrumentation, discoveries, methods
4. Reflections and Projections

# Spatial Localization

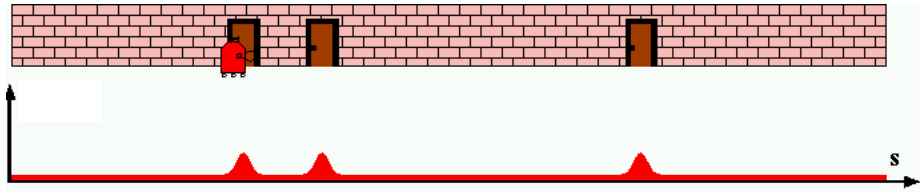
How can a mobile agent track its location as it moves about?

# Probabilistic Localization in Robots

$p(x_0 \mid m)$

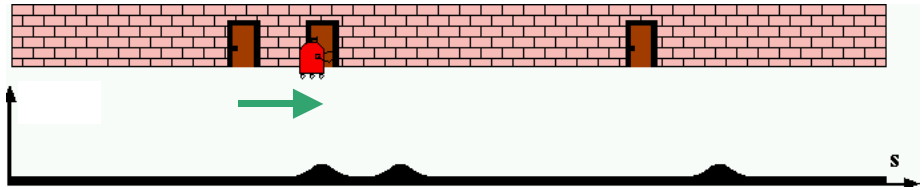



$p(z_0 \mid x, m)$

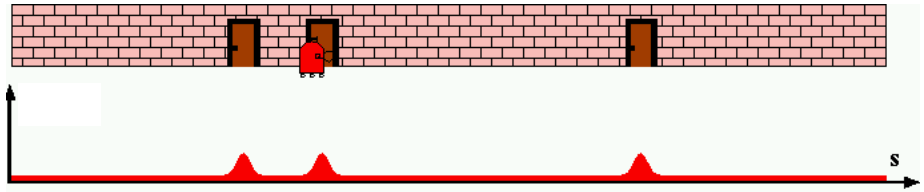



$p(x_0 \mid z_0, m)$


$p(x_1 \mid u_1, z_0, m)$




$p(z_1 \mid x, m)$




$p(x_1 \mid , z_1 , u_1, z_0, m)$


$m$ = *map*

$x$ = *state*

$z$ = *observation*

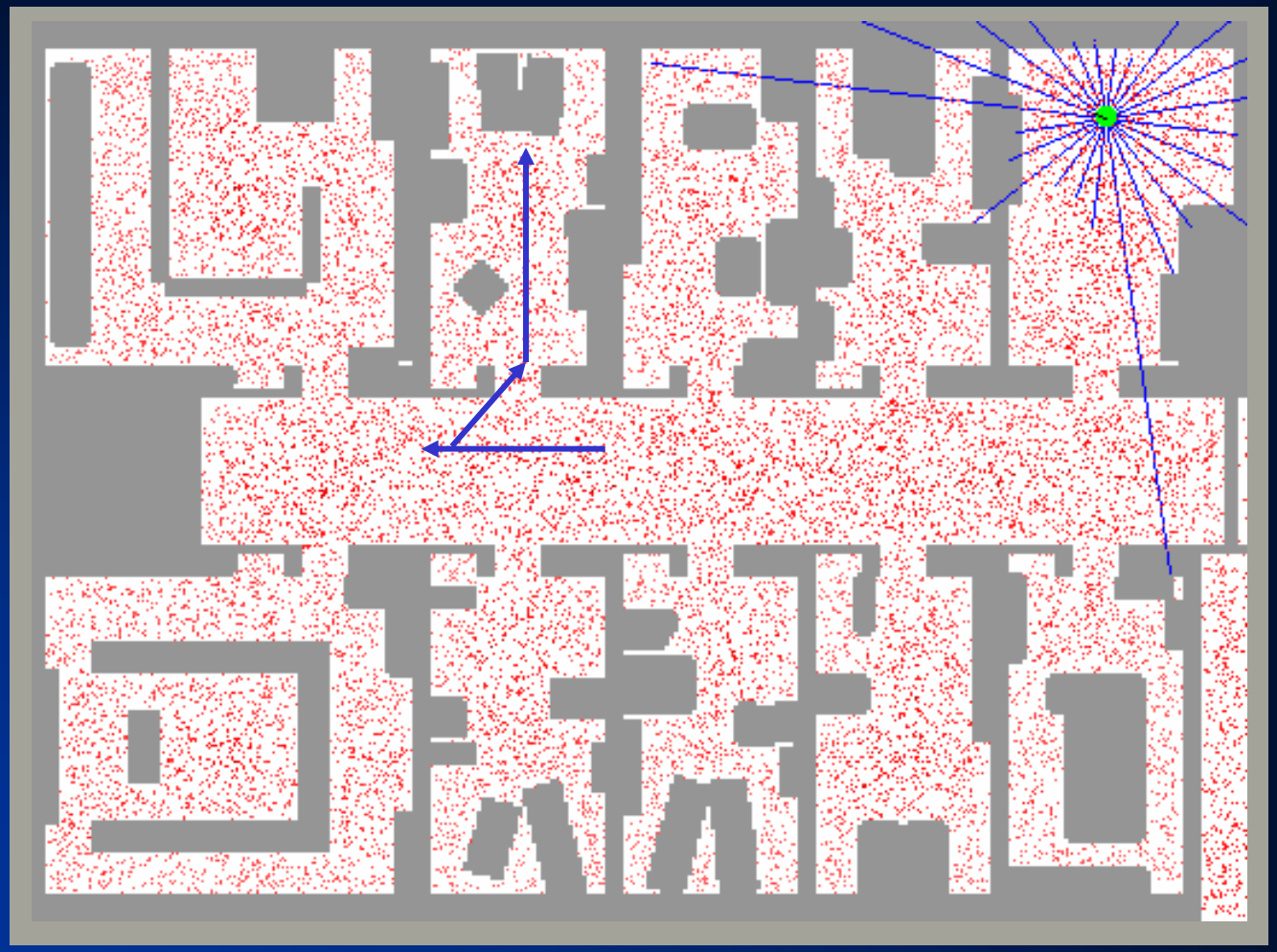
$u$ = *control*

[Courtesy S. Thrun]

[Simmons/Koenig 95]
[Kaelbling et al 96]
[Burgard et al 96]
[Thrun et al 96]

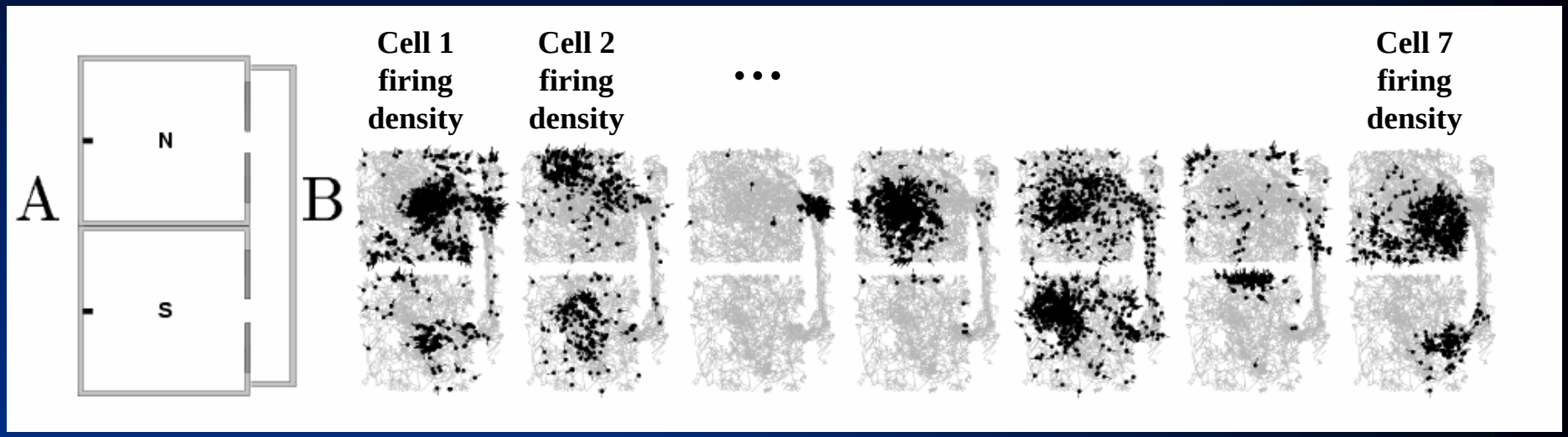
# Monte Carlo Localization (MCL)

[Thrun, Burgard, Fox, Dellaert]



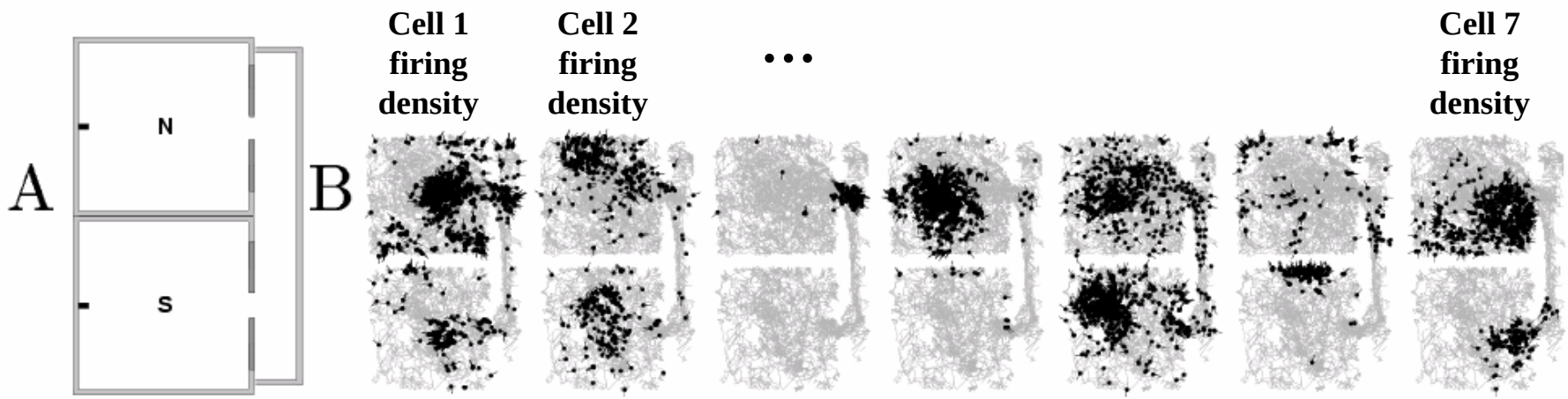
# Rat Place Cell Firing Patterns

**[Skaggs & McNaughton, *J. Neurosci.*, Oct 1998, 18(20)]**



# Rat Place Cell Firing Patterns

**[Skaggs & McNaughton, *J. Neurosci.*, 1998]**

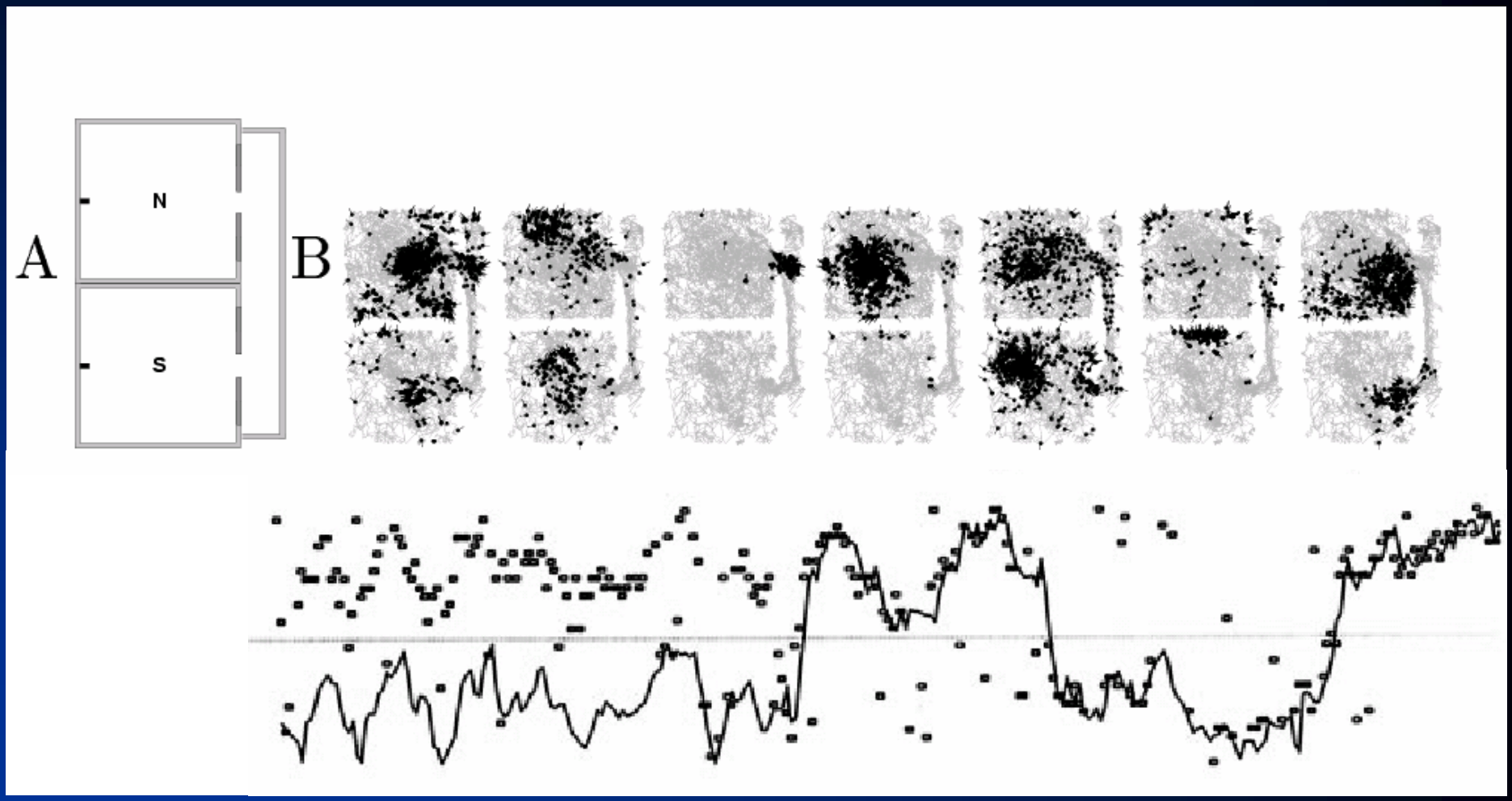



$$\text{define } P(x(t)) = \frac{1}{N} \prod_{i \in \text{cells}} P_i(x(t))$$

$$\text{where } P_i(x(t)) = \text{Poisson}(\text{density}_i(x), \text{spike\_rate}_i(t))$$

# Rat Place Cells Encode Location Distribution

**[Skaggs & McNaughton, *J. Neurosci.*, 1998]**



# Navigation and Localization

Robots

- Probabilistic representation of location and orientation
- Bayesian update of estimated position
- Practical using MC methods
- Related Bayesian methods for simultaneous mapping/localization
- Calculate x,y,θ jointly

Rats

- Place cells: encode location
- Place cell firing rates encode probability distribution over locations
- Method for updating location estimate unknown
- Unknown how simultaneous mapping/localization works
- Other cells reflect head orientation θ

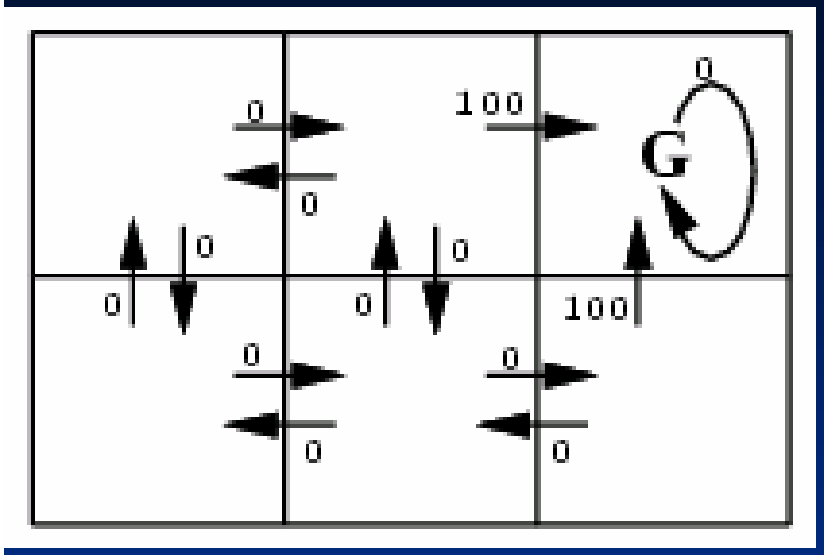
# Spatial Reasoning

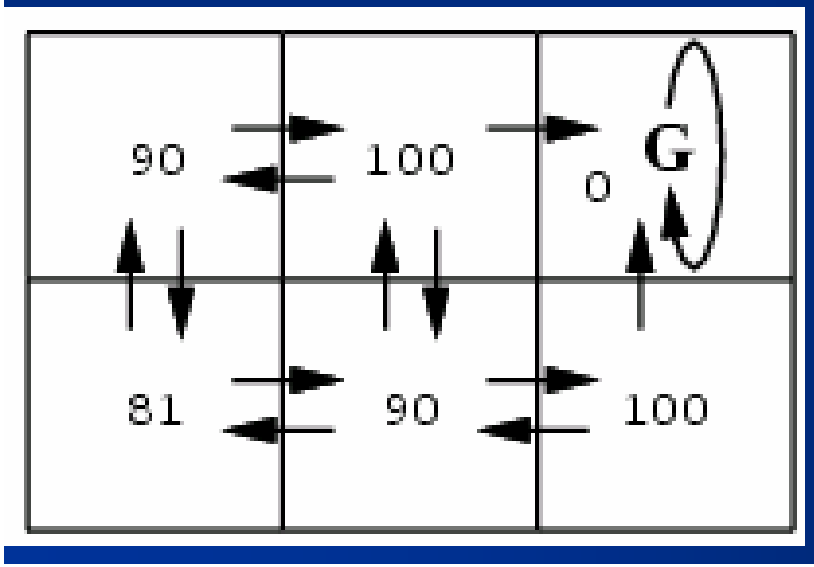
- Do mammals use probabilistic representations for location, orientation?
- And for other things?
- Why use their encoding of belief state?
- Do they make Bayesian updates?
- Do they explicitly model $P(x'|x,u)$? $P(obs|x)$?
- How do they learn, store, invoke the right map?

# Reinforcement Learning

# Temporal Difference Learning

[Sutton and Barto 1981; Samuel 1957]

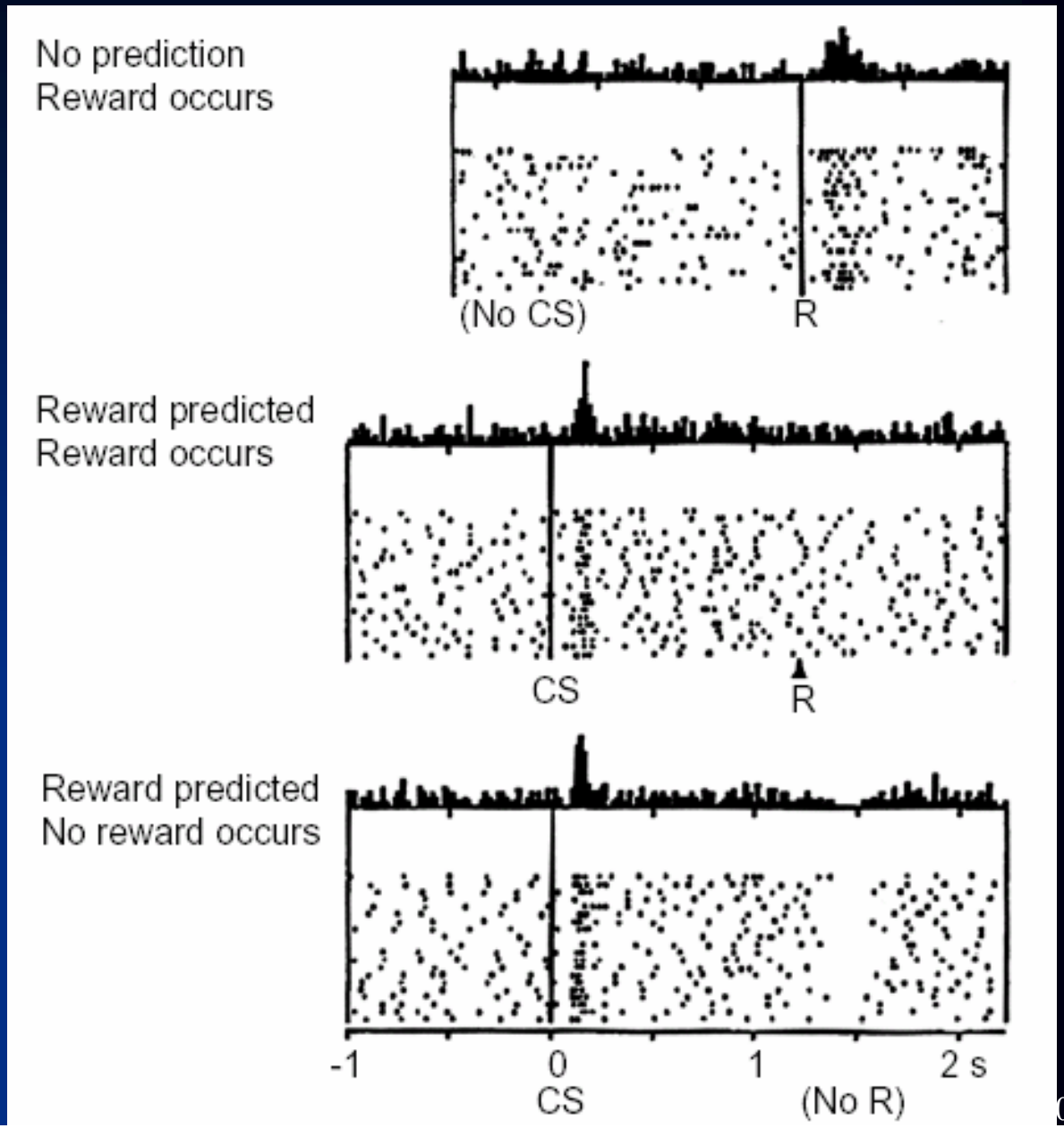



$$V^*(s) = E[r_t + \gamma\, r_{t+1} + \gamma^2 r_{t+2} + \ldots]$$

# Dopamine As Reward Signal

[Schultz et al., *Science*, 1997]

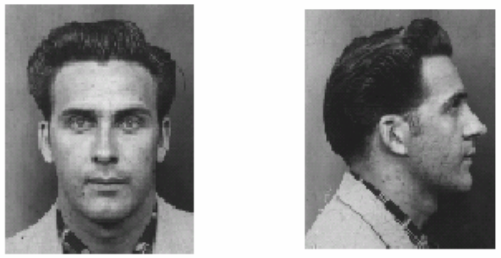


# Reinforcement Learning

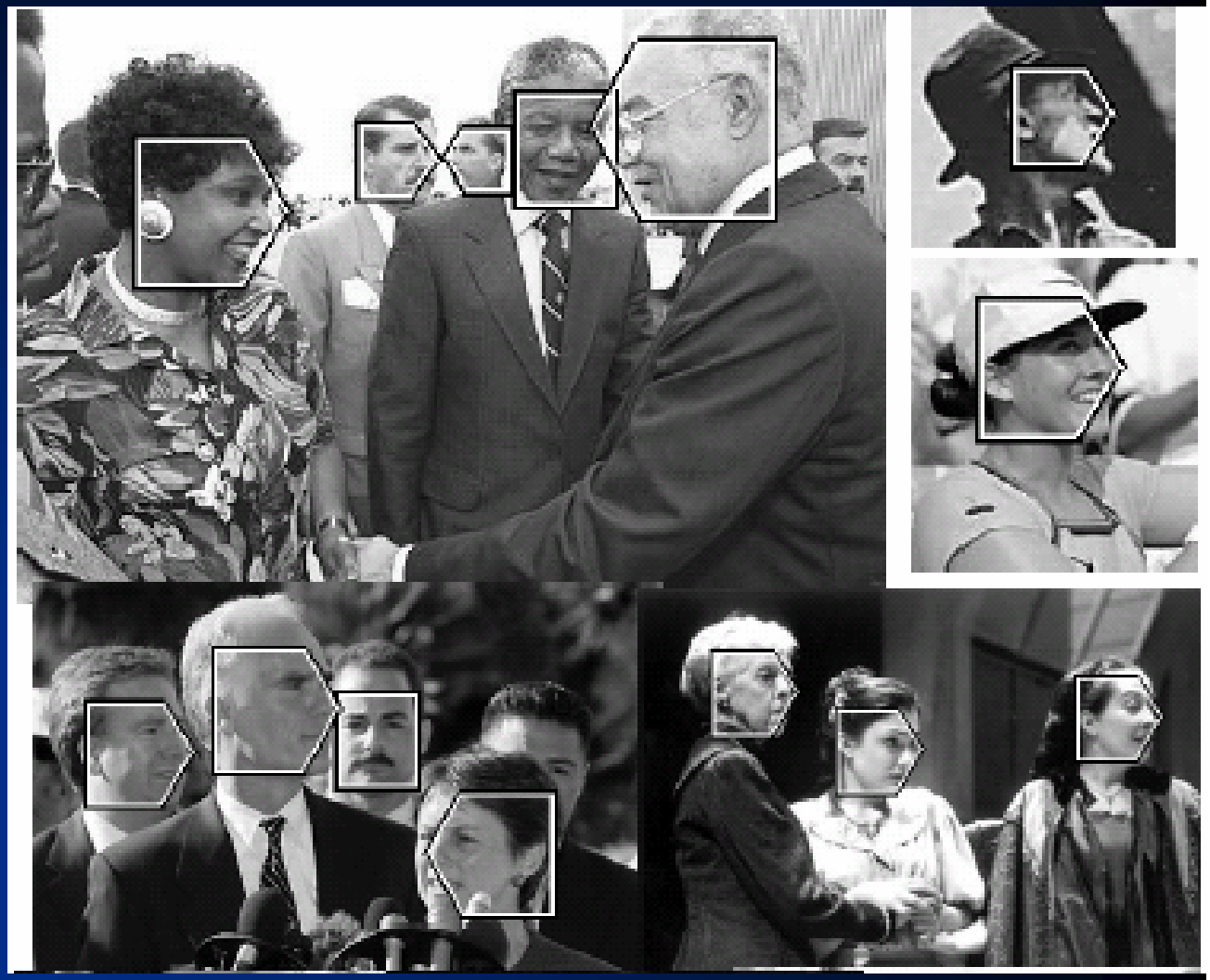
- How close is biological learning to the temporal difference learning algorithm?
- What learning strategy do primates use?
- Are “positive” (appetitive) and “negative” (aversive) rewards handled differently?
- One learning mechanism, or many?

# Object Detection

[Schneiderman, 2000]



training images for each orientation



# New instrumentation enables scientific revolutions - Kuhn

- Individual neuron recordings (100's)
- New dyes to observe brain metabolism
- 'Knock out' experiments (genetic engineering)
- Brain imaging (fMRI, PET, ERP, …)

# functional Magnetic Resonance Imaging

~1 mm resolution

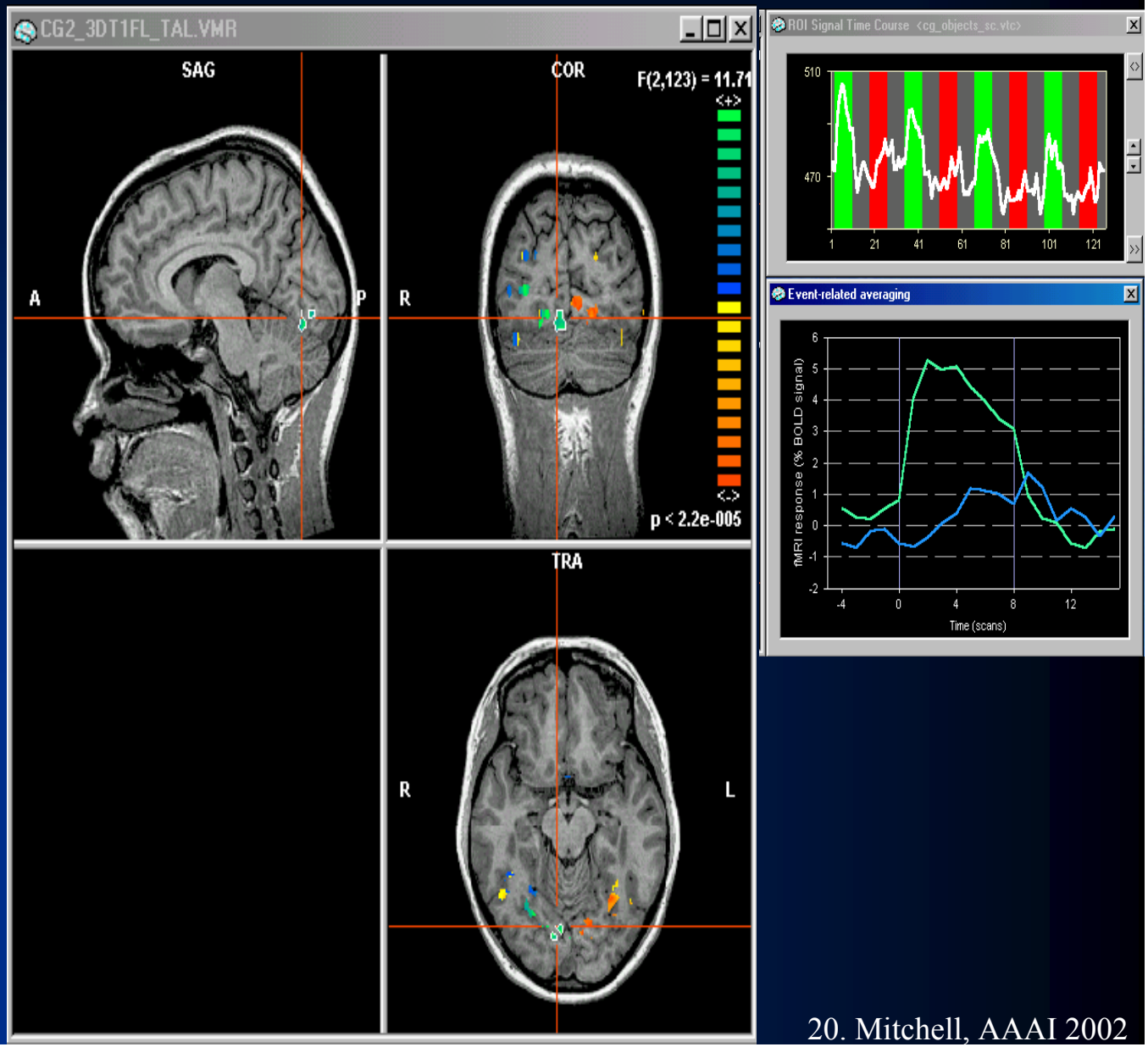
~1 sec temporally

30,000 voxels/image

two images per sec

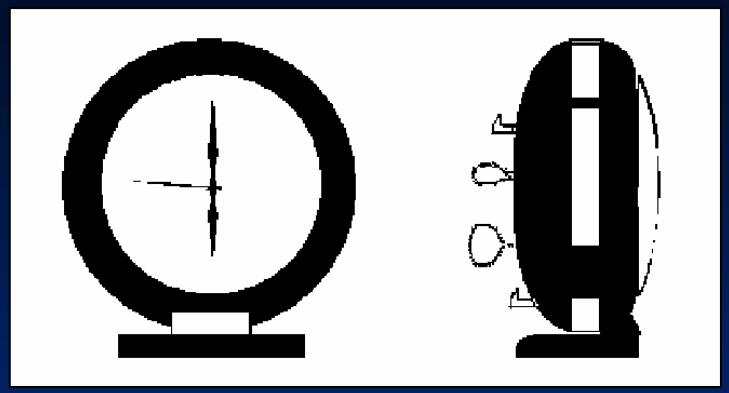
non-invasive, safe

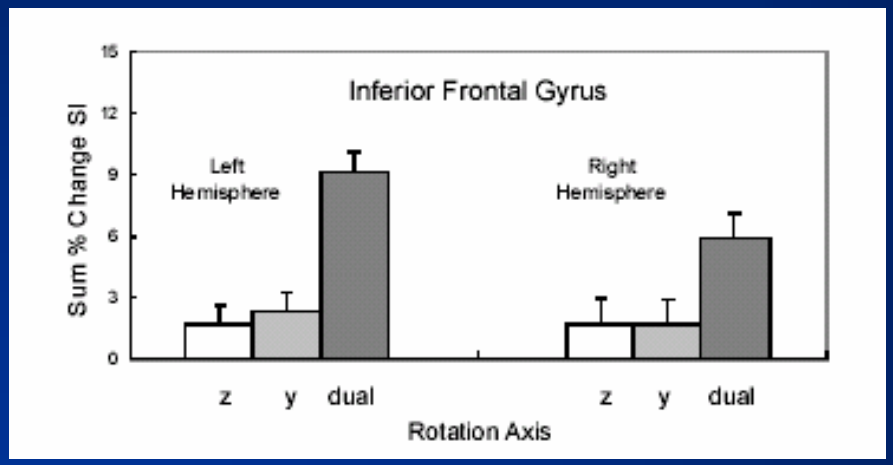
measures blood oxygen fluctuations

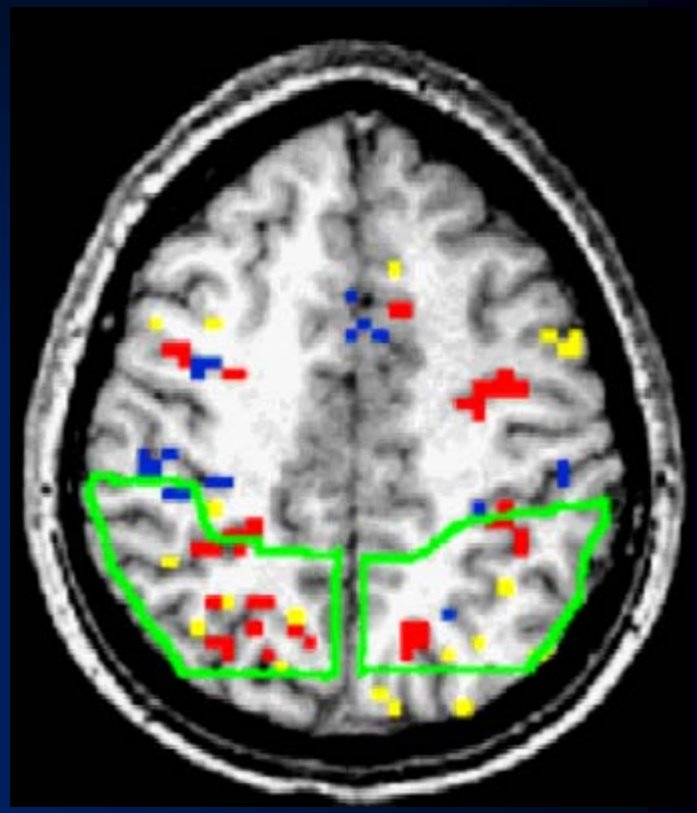



20. Mitchell, AAAI 2002

# Mental Rotation of Imagined Objects





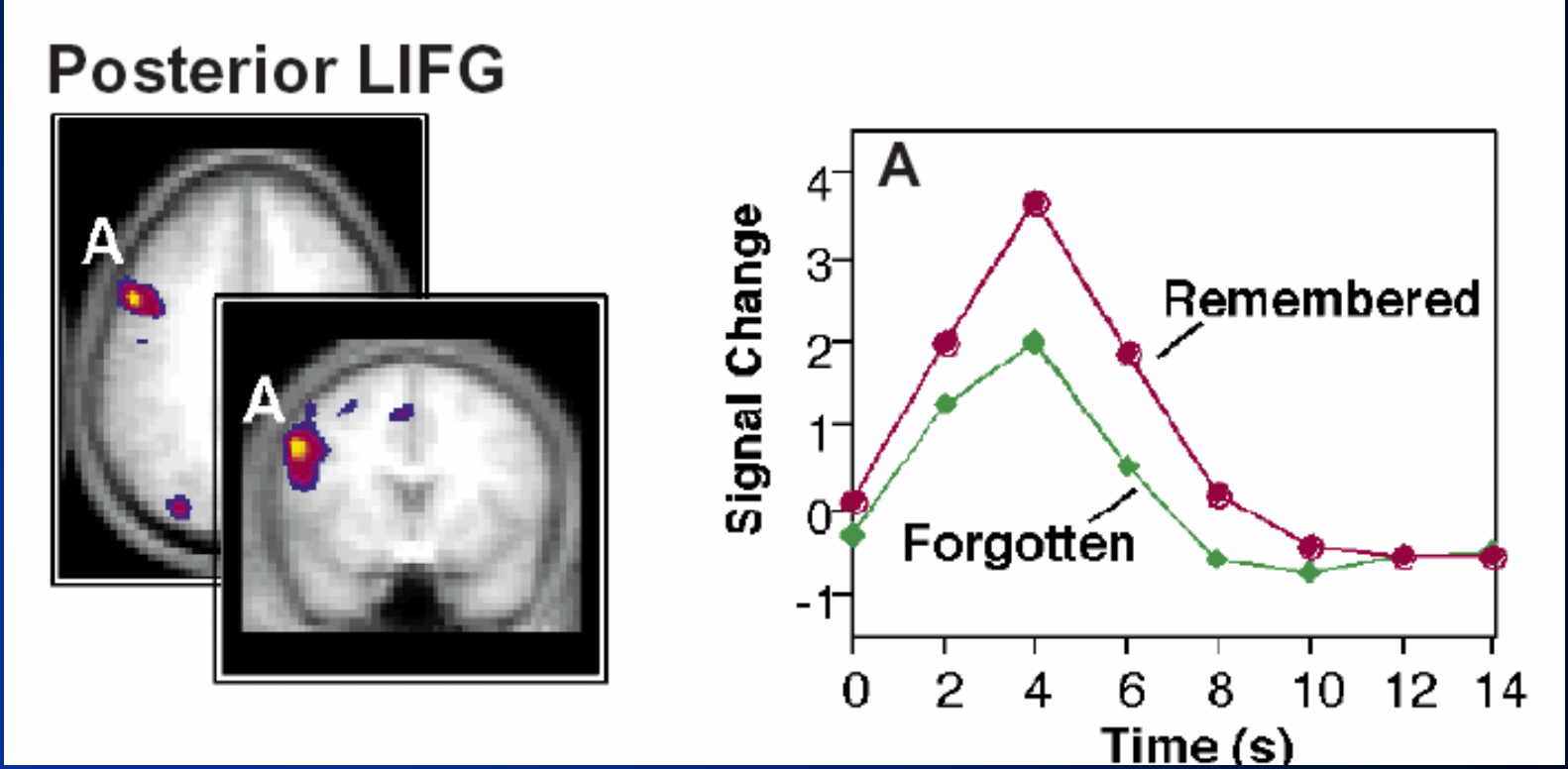



- Clock rotation
- Shephard-Metz rotation
- both

[Just, et al., 2001]

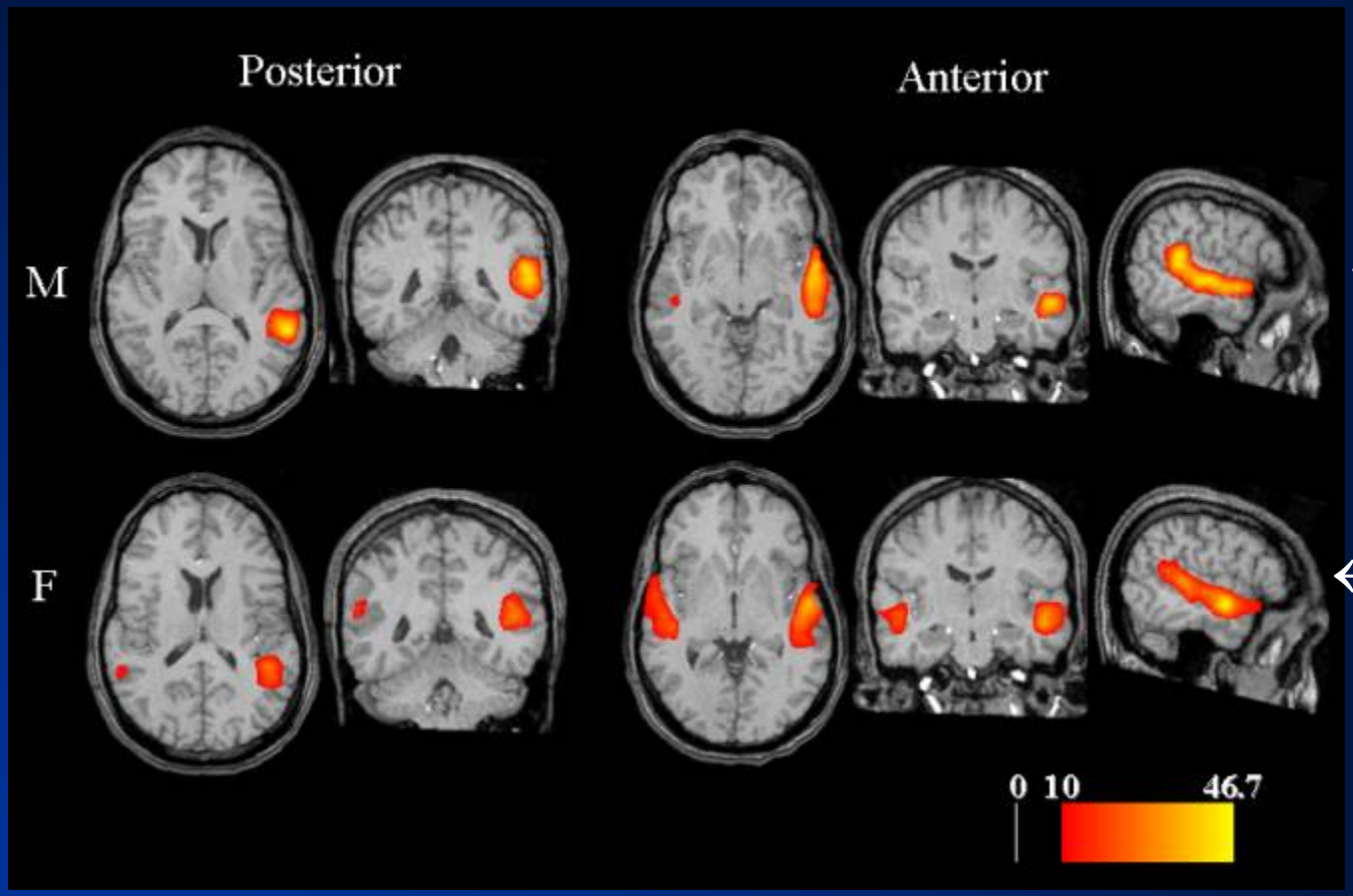
# Verbal Remembering and Forgetting Predicted by fMRI Activity

[Wagner et al., *Science*, 1998]



# Study of Men and Women Listening

"Men listen with only one side of their brains, while women use both"

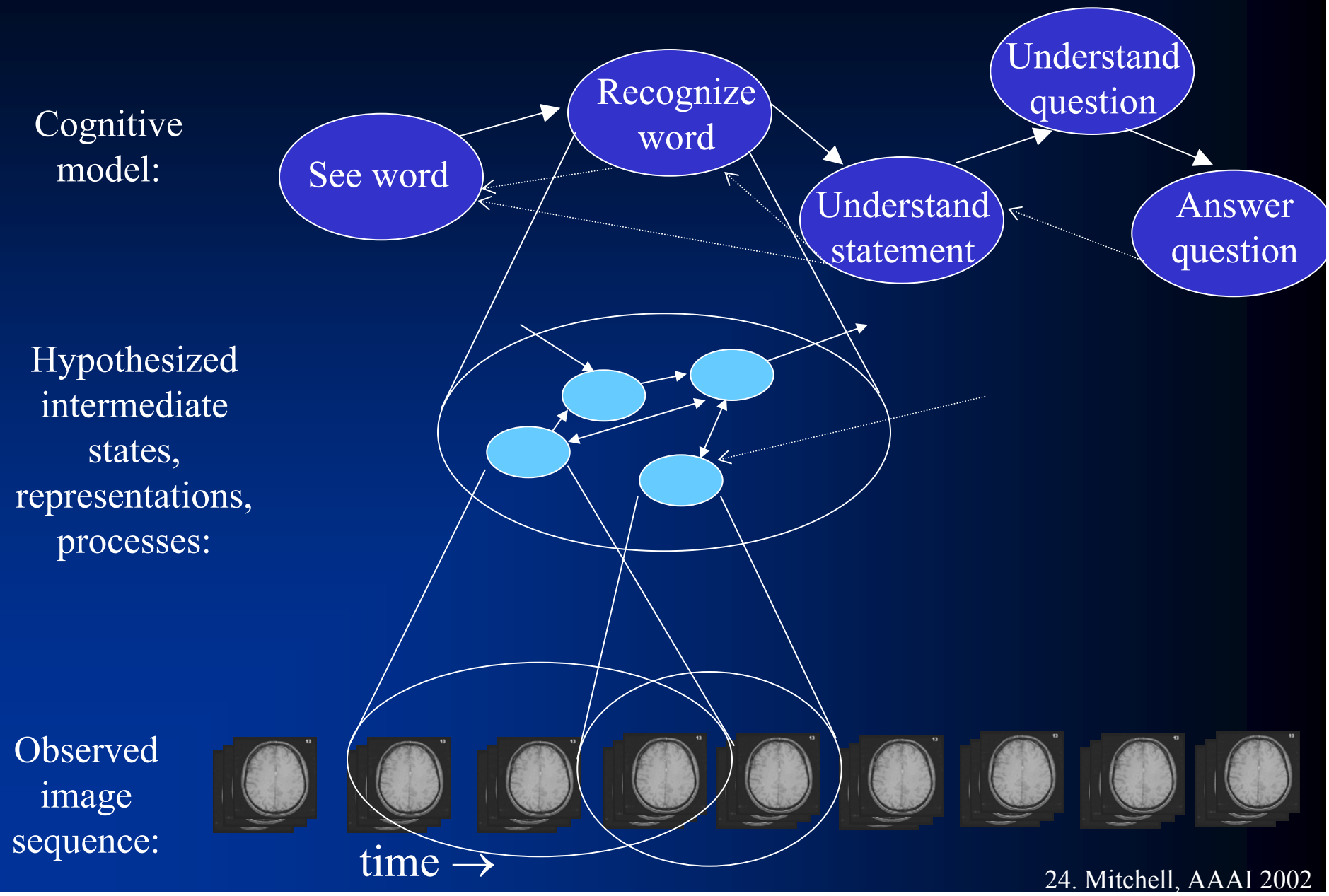



← Men listening

←Women listening

(IU School of Medicine Department of Radiology)

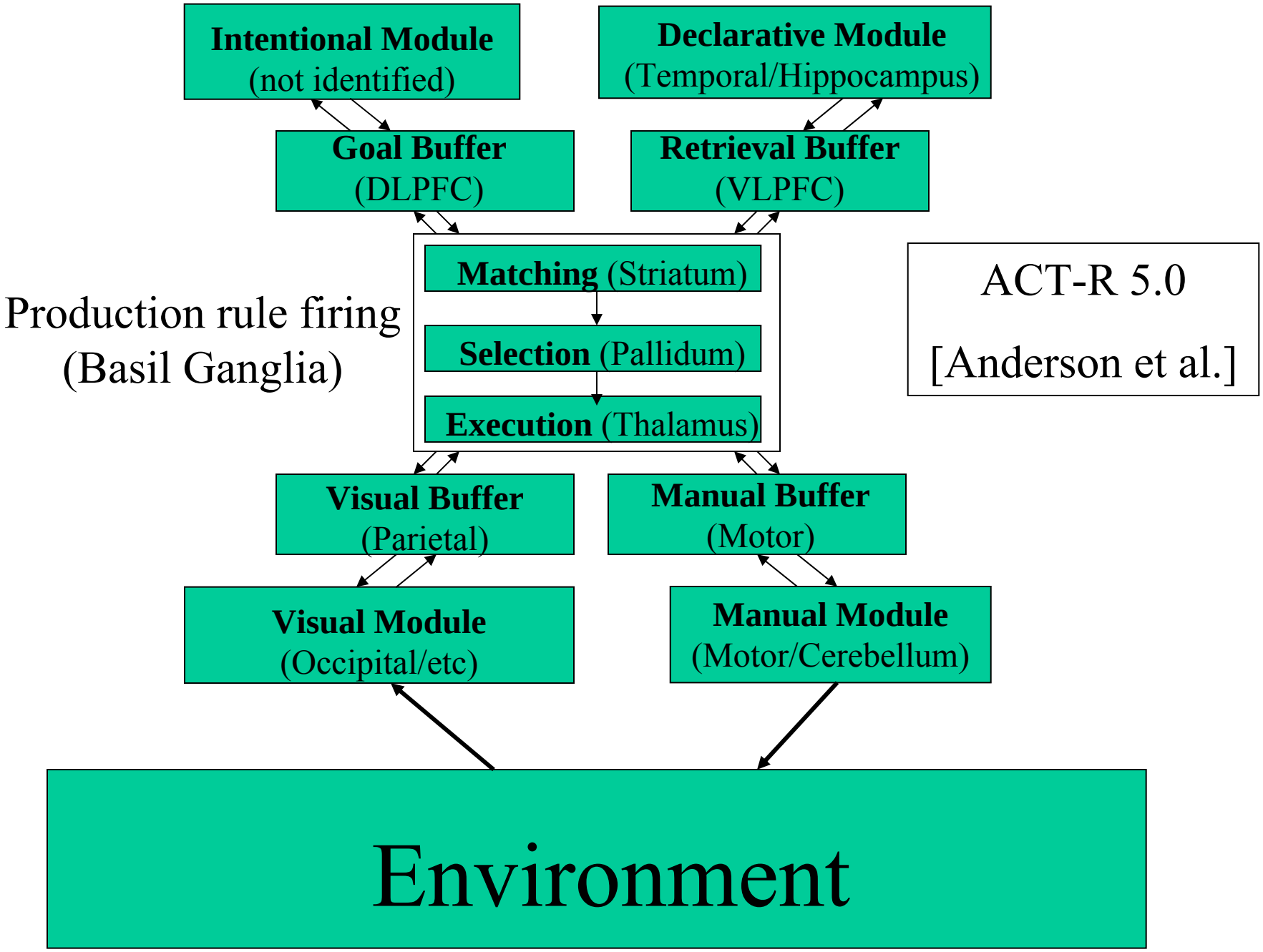
# What We'd Like



# ACT-R Cognitive Architecture

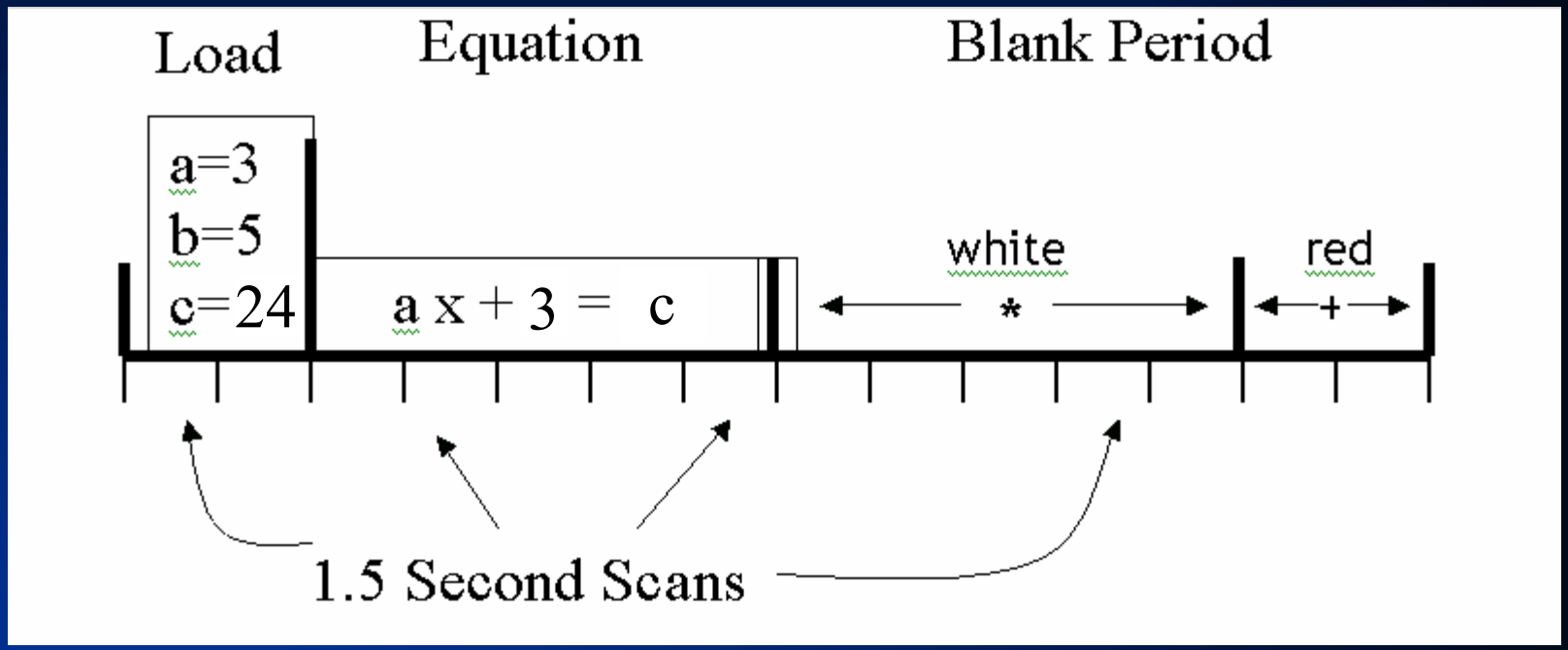
[Anderson et al.]

- Production rule, inductive and analytical learning methods
- Successfully models some human cognitive functions, e.g.
  - Predicts response times, error rates, learning rates
  - E.g., Learning of arithmetic and multiplication tables [Lebiere]


Intentional Module
(not identified)
Declarative Module
(Temporal/Hippocampus)
Goal Buffer
(DLPFC)
Retrieval Buffer
(VLPFC)
Matching (Striatum)
Selection (Pallidum)
Execution (Thalamus)
Production rule firing
(Basil Ganglia)
ACT-R 5.0
[Anderson et al.]
Visual Buffer
(Parietal)
Manual Buffer
(Motor)
Visual Module
(Occipital/etc)
Manual Module
(Motor/Cerebellum)
Environment

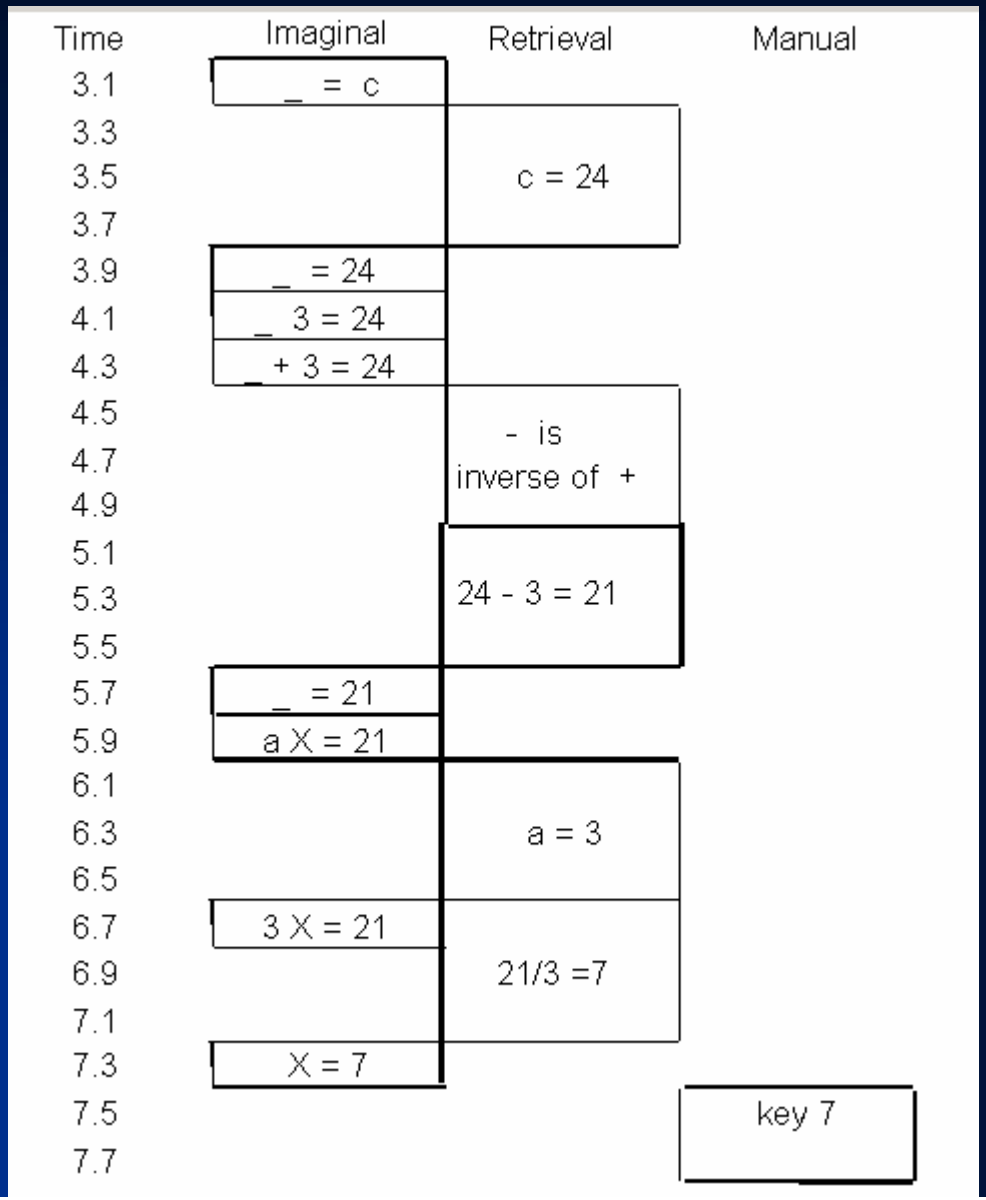
# Mental Algebra Task

[Anderson, Qin, & Sohn, 2002]

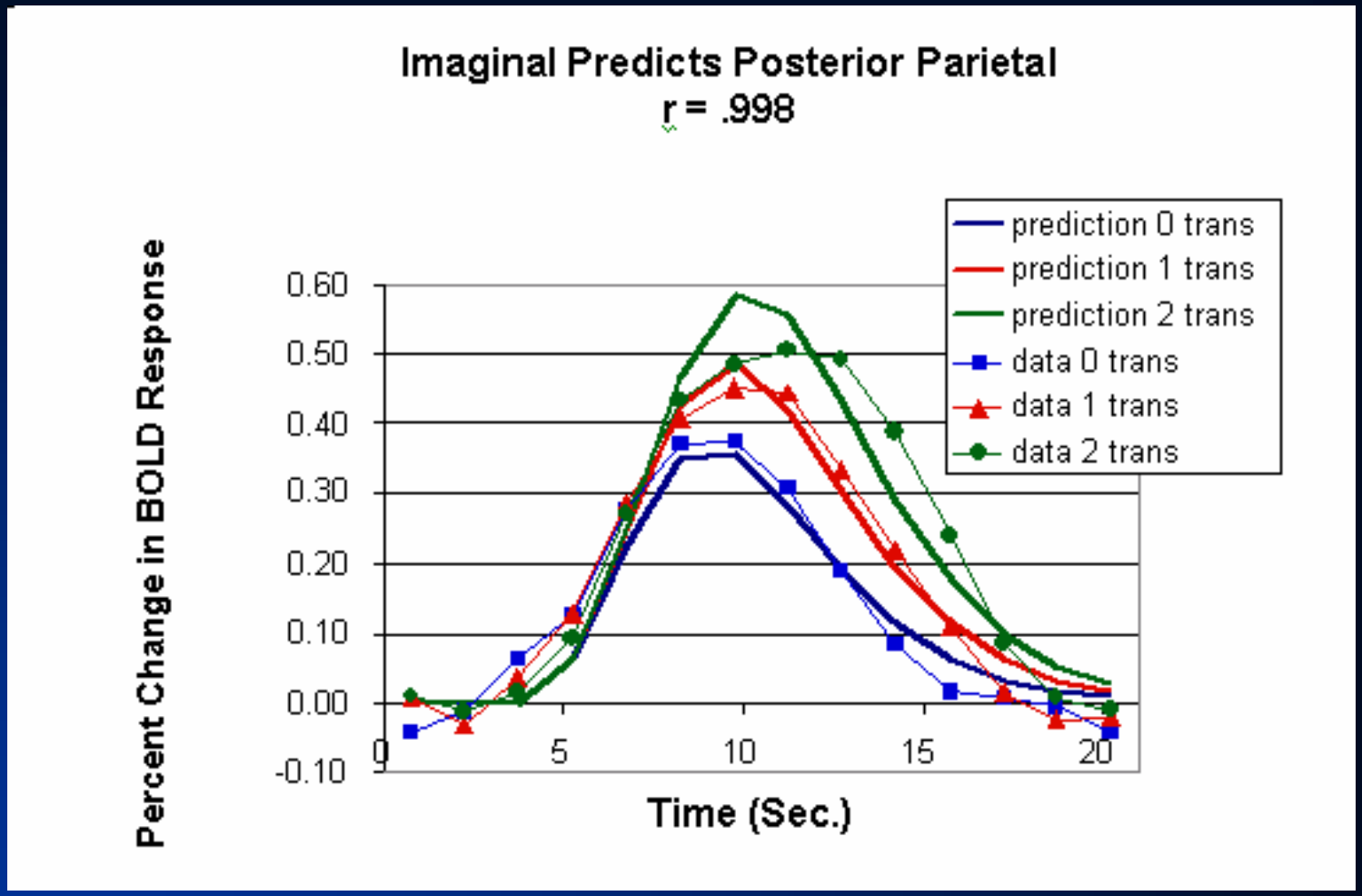


# Activity Predicted by ACT-R Model

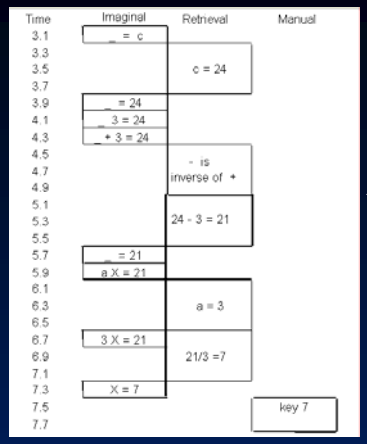
[Anderson, Qin, & Sohn, 2002]

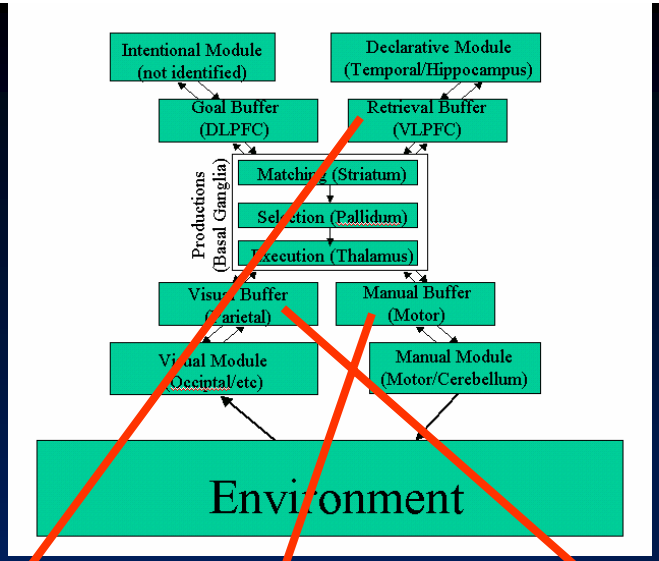


# From [Anderson, Qin, & Sohn, 2002]

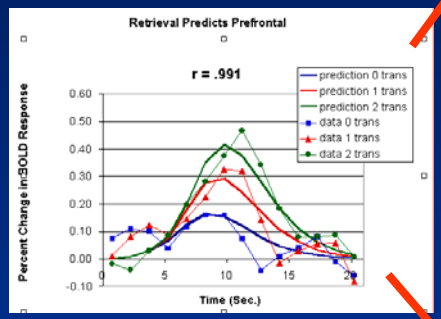


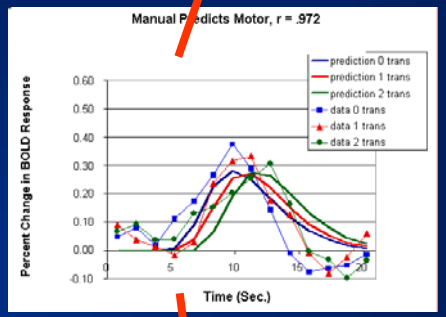

Imaginal buffer predicts posterior parietal activity: effect of number of transformations

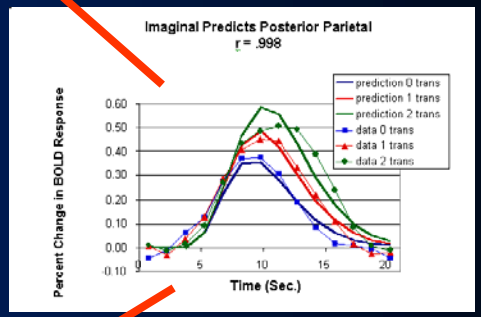


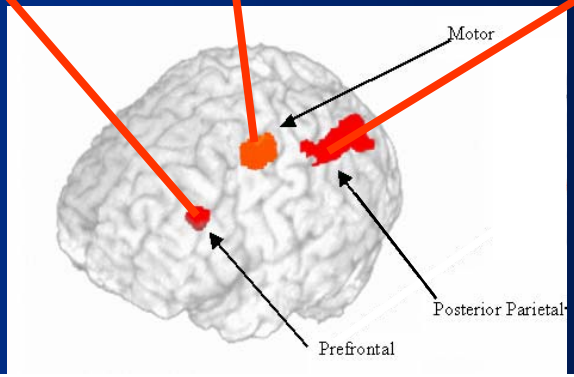




[Anderson, Qin, & Sohn, 2002]



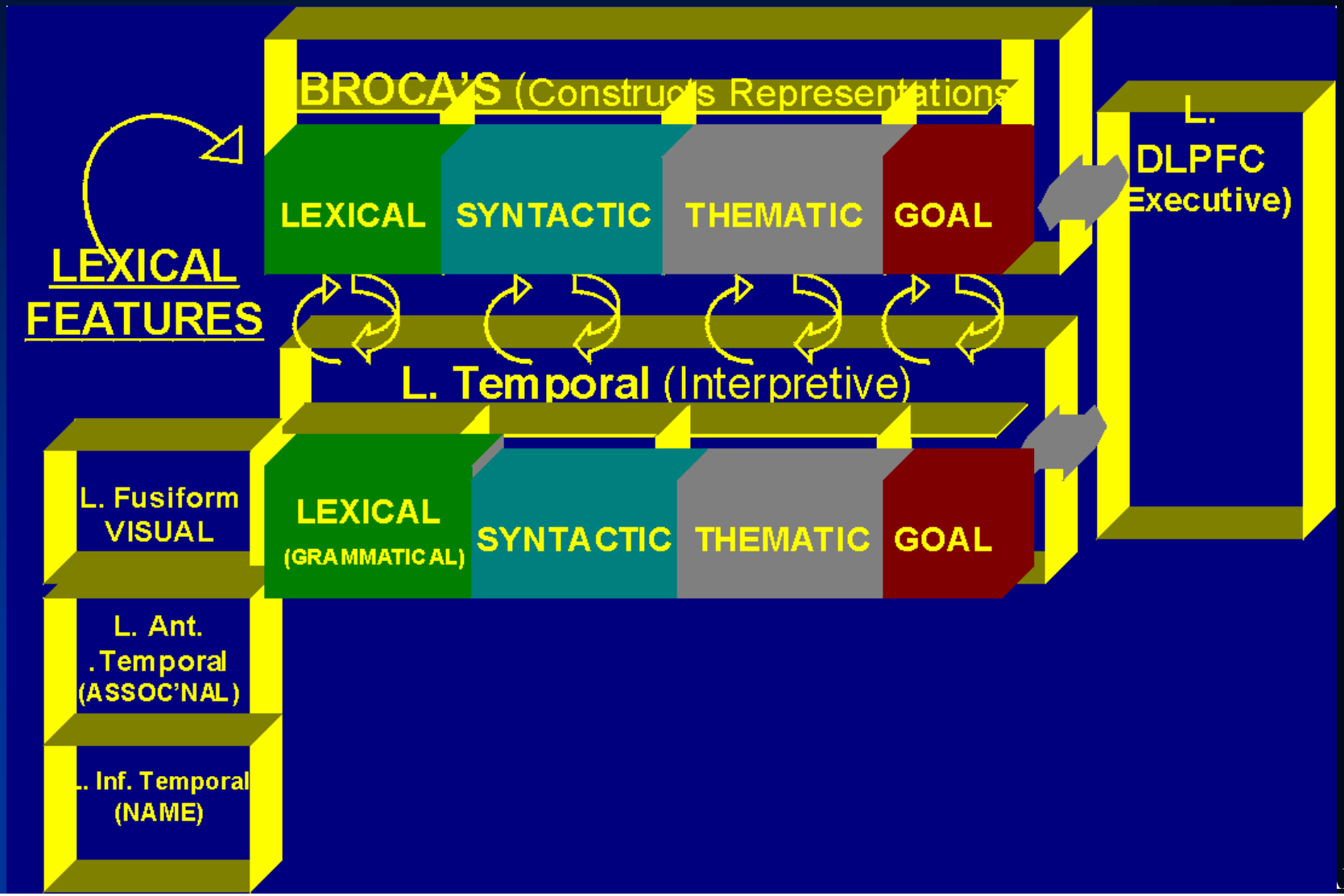



# 4CAPS Model of Language Processing

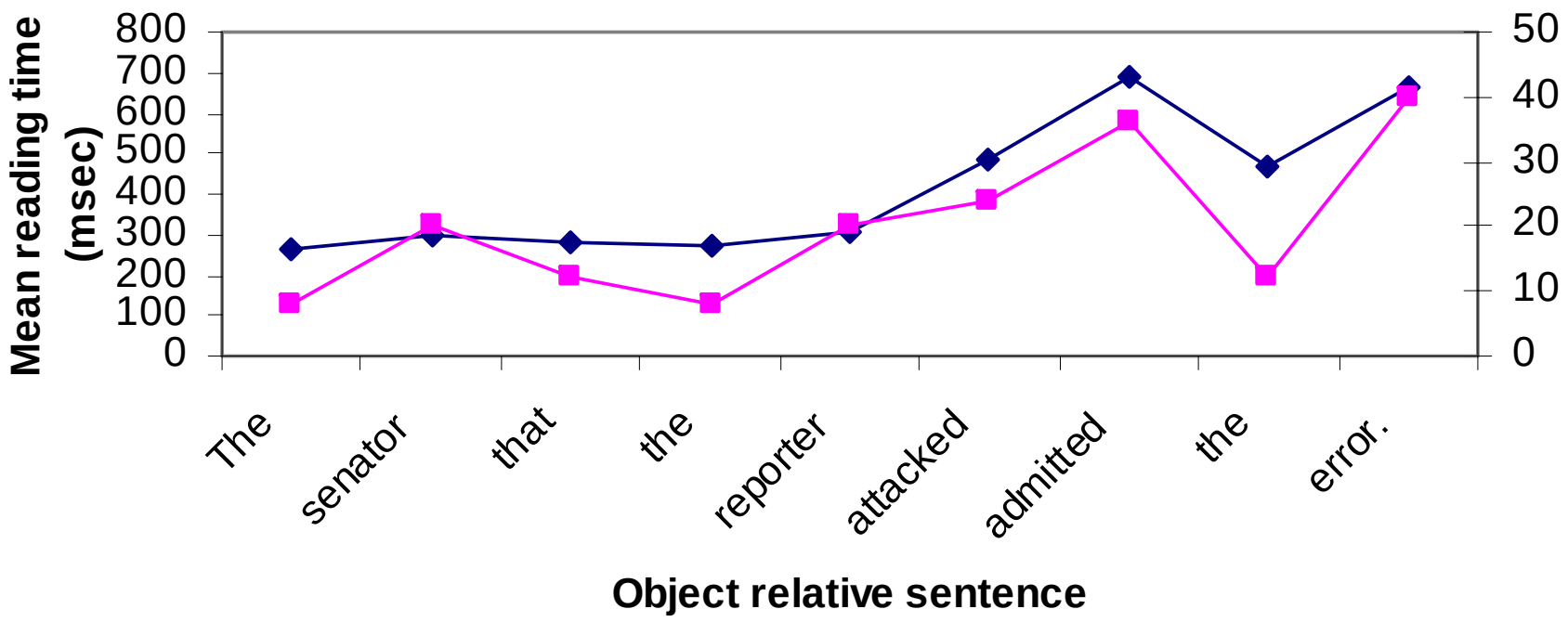
[Just, et al., 2002]



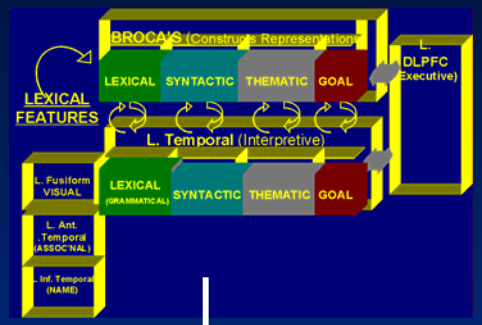
# 4CAPS Model of Human Sentence Processing

## [Just et al.]

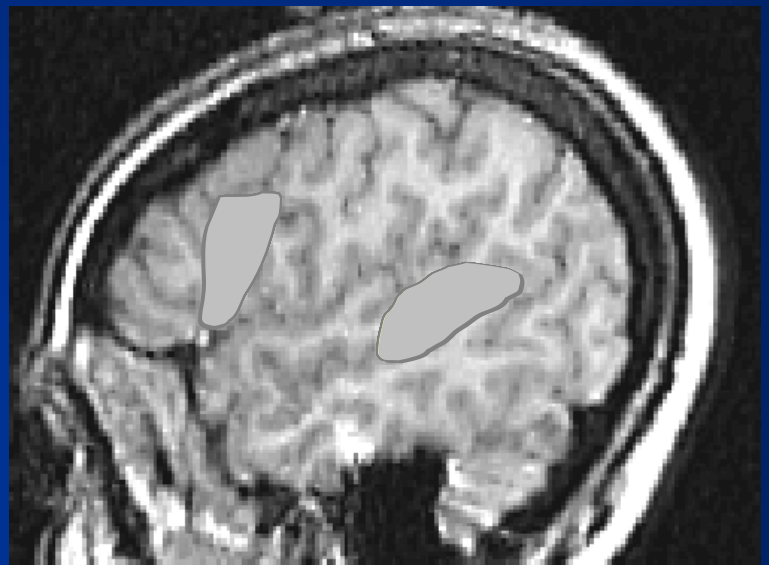
**Figure 5. Human reading time (diamonds) and 4CAPS predictions (purple squares)**



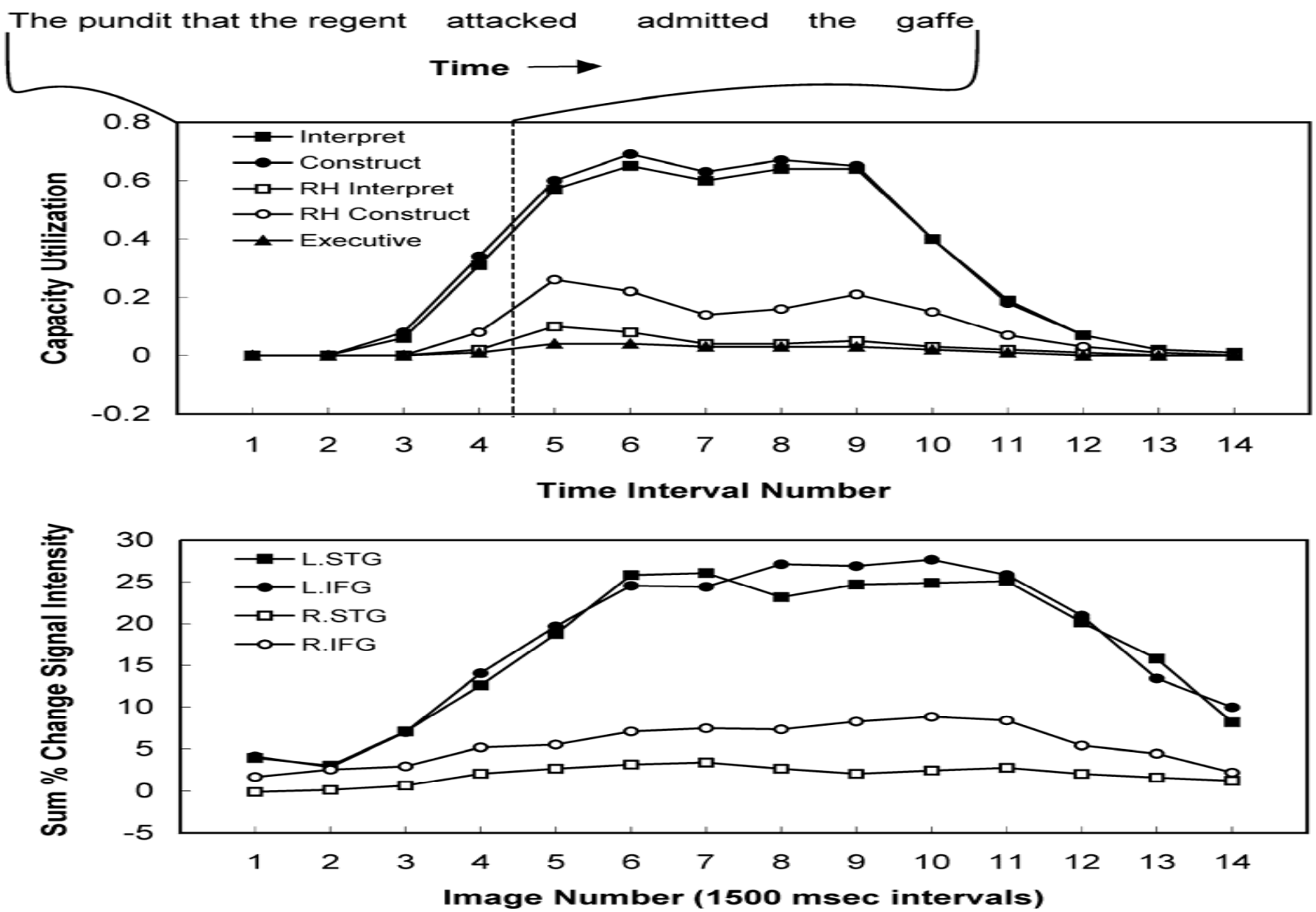
**The player was followed by the parent.**




[Just, et al., 2002]



# 4CAPS Prediction of fMRI Activity




CU in 4CAPS comprehension model components

**Model prediction**

Model CU transform

**fMRI data**

Figure 10

# Semantic Word Category Experiment

- Animal (4 legged, fish)
- Nature (trees, flowers)
- Food (fruits, vegetables)
- People (family members, occupations)
- Artifact (tools, kitchen items)
- Building (human dwellings, parts of buildings)

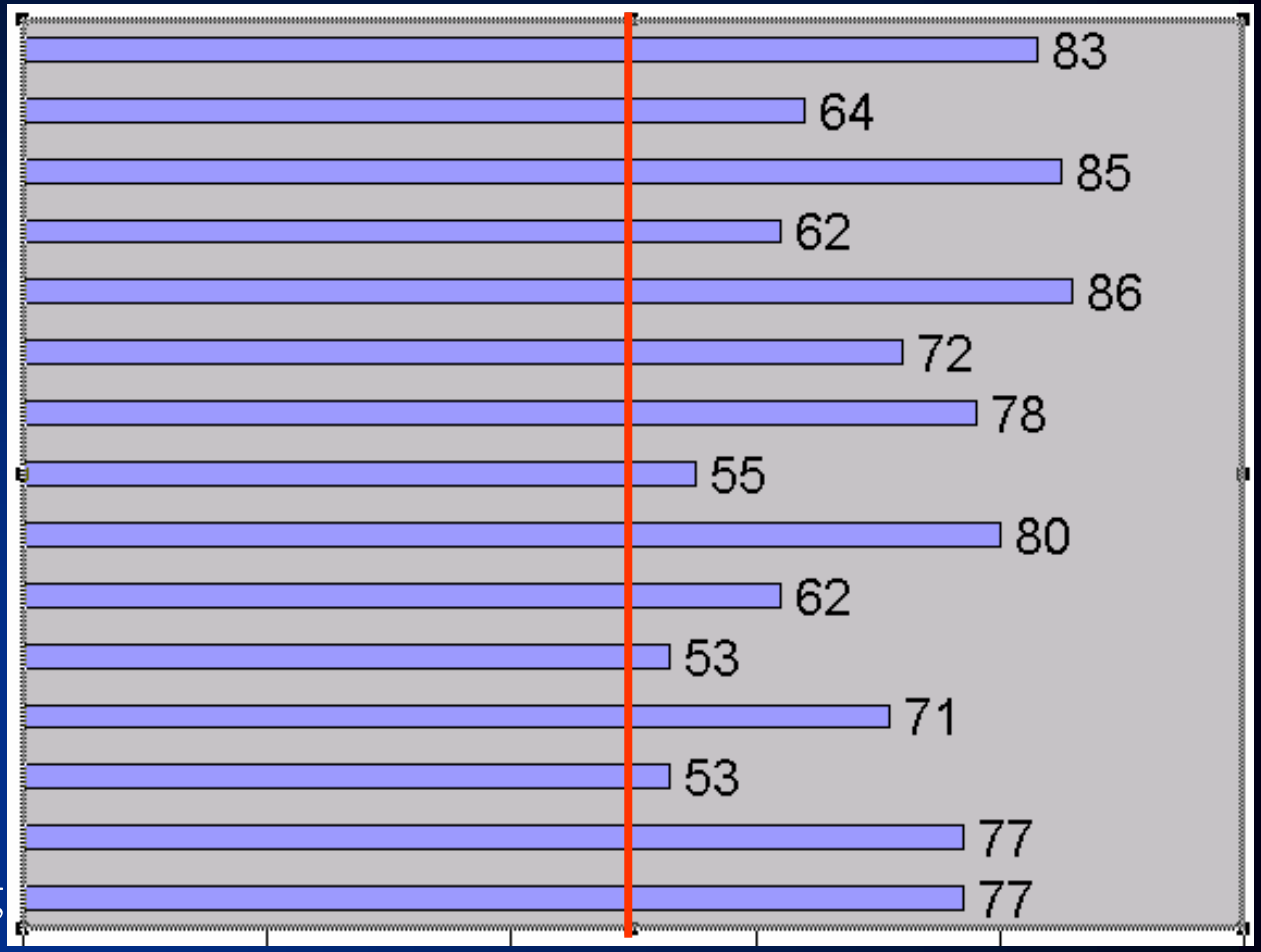
# Is Word from Category i or j?

- Learn fMRI(t) → word-category(t)
  - Single subject
  - Classify based on single time instant
  - 2592 voxels used
  - Train on all category pairs (six categories total)

- Training method:
  - Gaussian Naïve Bayes classifier
  - P(fMRI | word-category)

# Accuracy Detecting Word Semantics from Single fMRI Snapshot

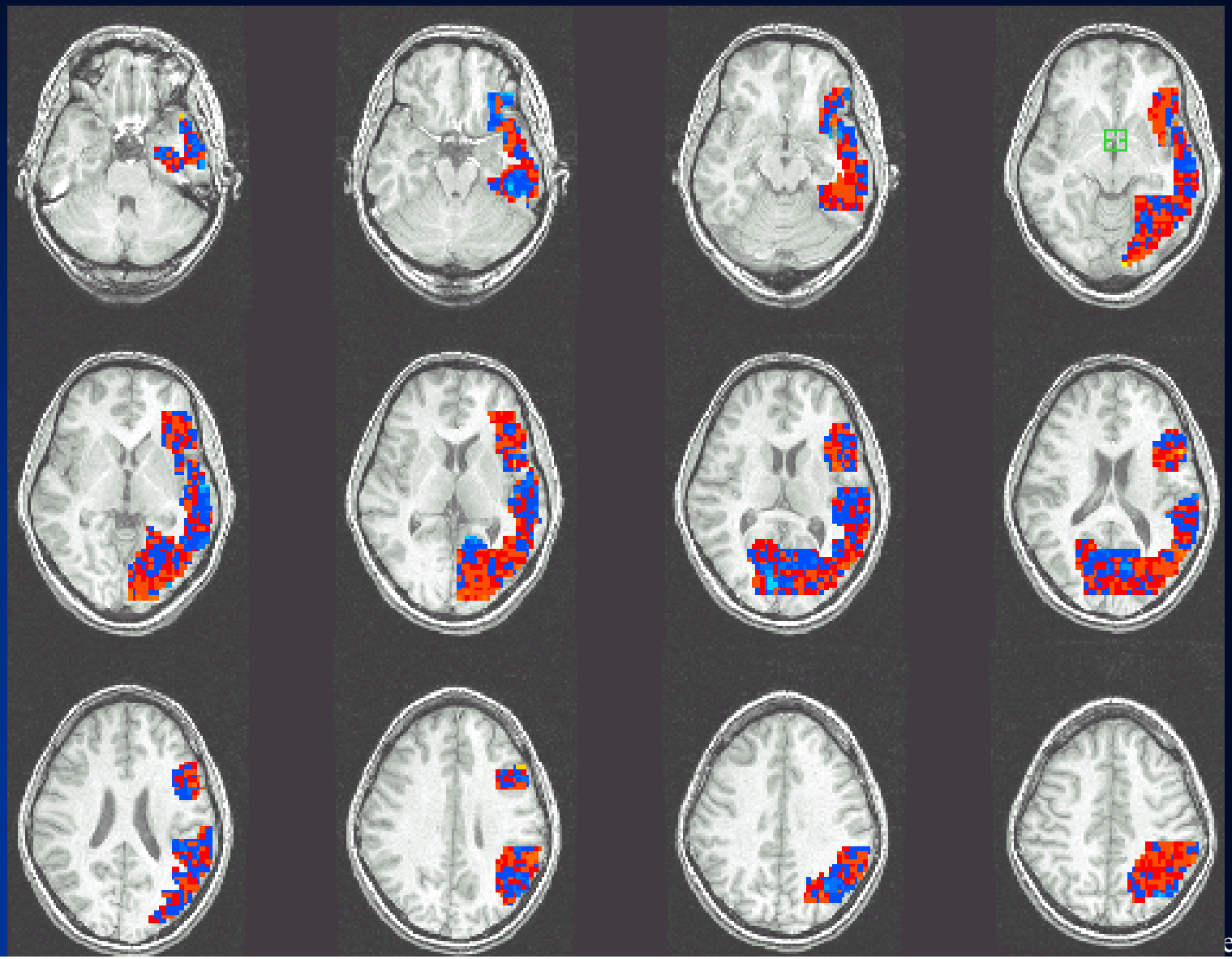
1. Animal-Nature
2. Animal-Food
3. Animal-People
4. Animal-Artifact
5. Animal-Building
6. Nature-Food
7. Nature-People
8. Nature-Artifact
9. Nature-Building
10. Food-People
11. Food-Artifact
12. Food-Building
13. People-Artifact
14. People-Building
15. Artifact-Building




**(subject5886, Gaussian Bayes classifier)**

37. Mitchell, AAAI 2002

# Learned Bayes Models - Means
# P(BrainActivity | WordCategory = People)

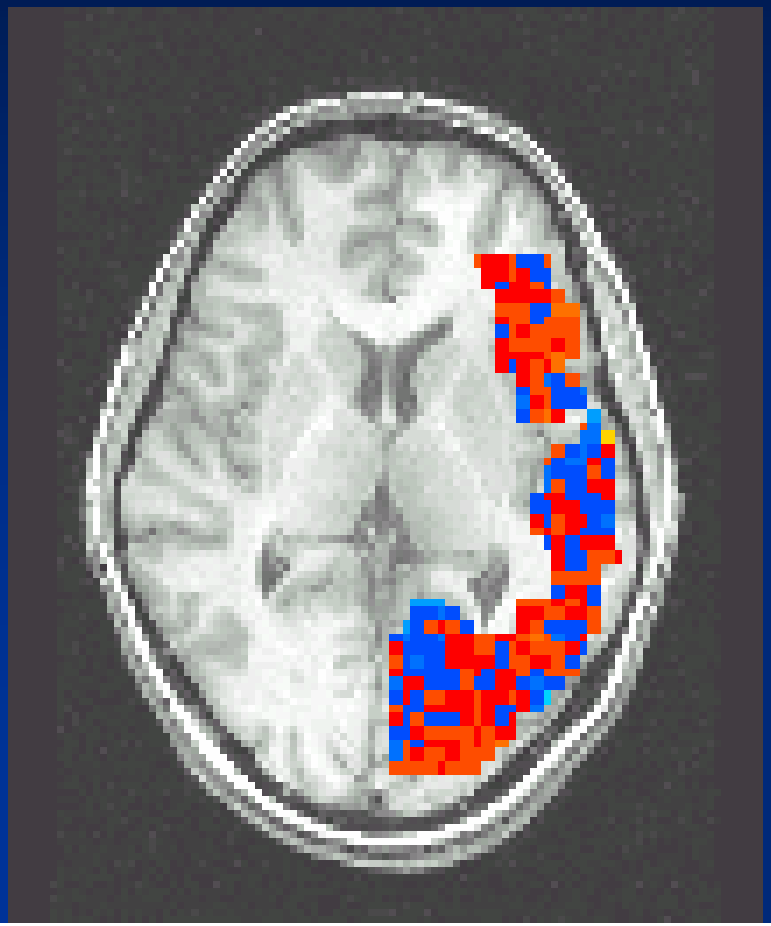


ell, AAAI 2002

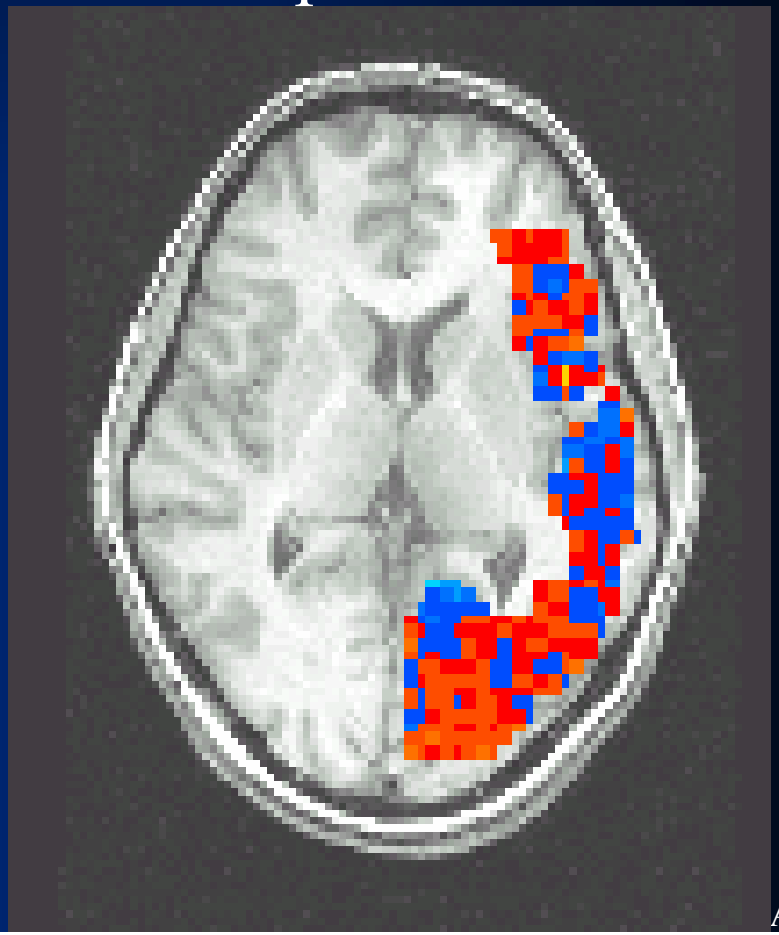
# Learned Bayes Models - Means
# P(BrainActivity | WordClass)

Accuracy: 85%

Animal words



People words

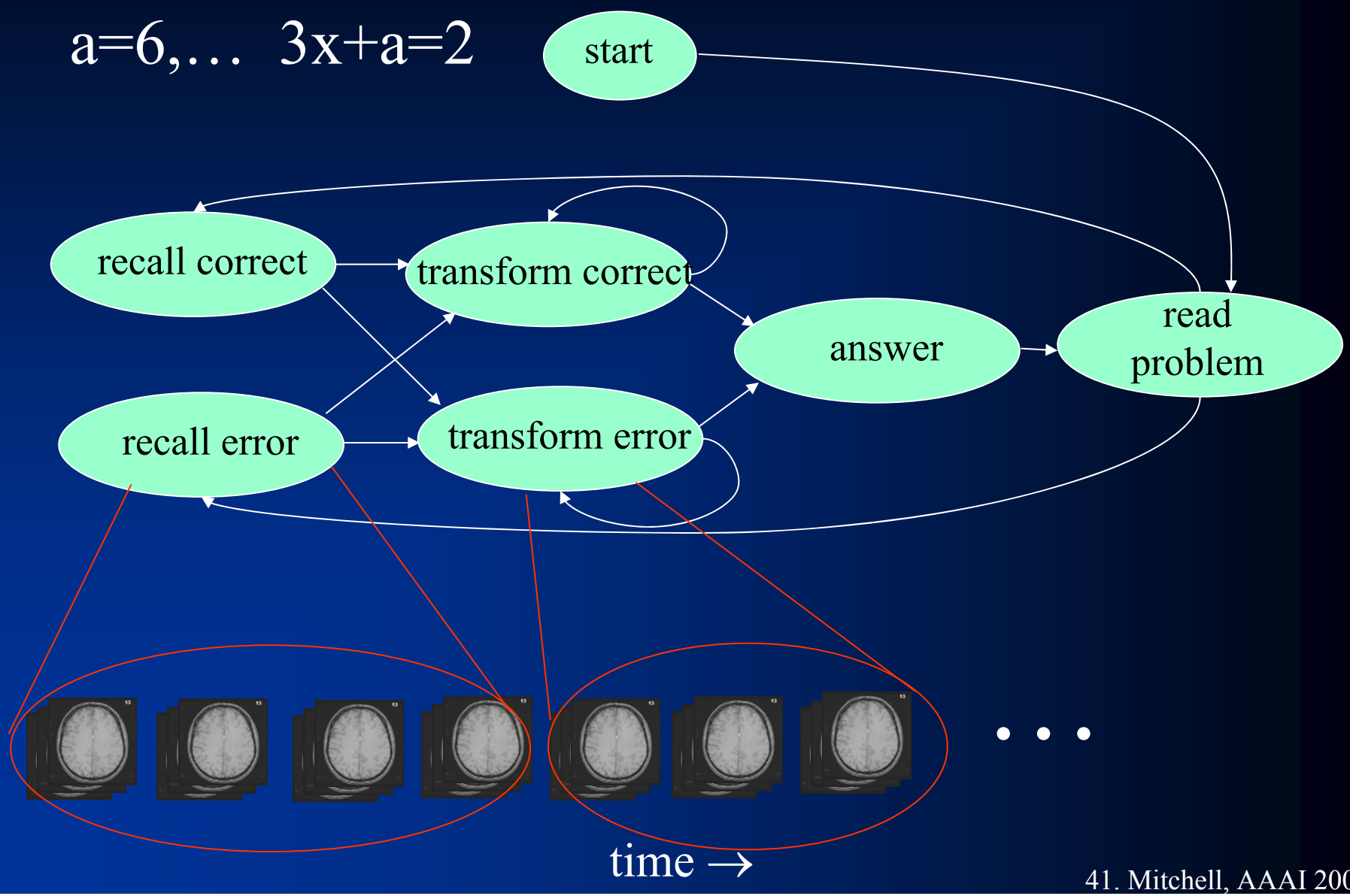


# *New instrumentation enables scientific revolutions* - Kuhn

- Virtual sensors of mental state
  - semantic word category
  - garden path versus standard sentence
  - examining picture versus sentence
  - rat's location belief state

- In future publish virtual sensors for broad use?
- See Dartmouth fMRI data repository

# Challenge: virtual sensors to track sequence of cognitive states



# fMRI Clues for Cognitive Architectures

- Most of the time, most of brain is idling
- Functions are not strictly local, but distributed
- Greater task difficulty results in recruiting more cortical regions (e.g., bilaterally)
- Greater task difficulty can result in greater synchronization among cortical regions
- Learning typically results in *decreased* activity, but often *increased* synchronization

# AI: Impact on Brain Sciences

- AI will provide key computational concepts
  - Algorithms
  - Representations
  - Theoretical results
- Machine learning methods will add
  - virtual "mental state" sensors
  - and more : info bottleneck codes, ICA, HMM's,…
- The final theory of brain function will *be* an AI program!

# Brain Sciences: Likely Impact on AI

- Won't soon reveal detailed brain algorithms
- Will reveal decomposition of cognitive tasks
  - substages in object recognition
  - location and orientation represented separately
- Organizational principles of the brain
  - Population codes
  - One learning mechanism, or many?
- New issues for AI
  - Forgetting, attention, motivation, habituation,…

U.S. Food and Drug Administration

# FDA Talk Paper

**FDA Talk Papers are prepared by the Press Office to guide FDA personnel in responding with consistency and accuracy to questions from the public on subjects of current interest. Talk Papers are subject to change as more information becomes available.**

T02-03
January 14, 2002

Sharon Snider: 301-827-6242
Consumer Inquiries: 888-INFO-FDA

## FDA APPROVES EXPANDED USE OF BRAIN IMPLANT FOR PARKINSON'S DISEASE

FDA today approved an expanded use of a brain implant to help control some symptoms of advanced Parkinson's disease.

The device, a deep brain stimulator, made by Medtronic, Inc., of Minneapolis, Minn., was initially approved by FDA in 1997 for use in one side of the brain to help control tremors on one side of the body.

Today, after review of additional studies conducted by the manufacturer, the agency approved the device, called the Activa Parkinson's Control System, for use in both sides of the brain to help reduce some of the other symptoms of advanced Parkinson's that cannot be adequately controlled with medication.

An estimated 1.5 million Americans have Parkinson's disease, which results in tremors, rigidity, postural instability, slowness and difficulty moving and, in some people, intellectual deterioration.

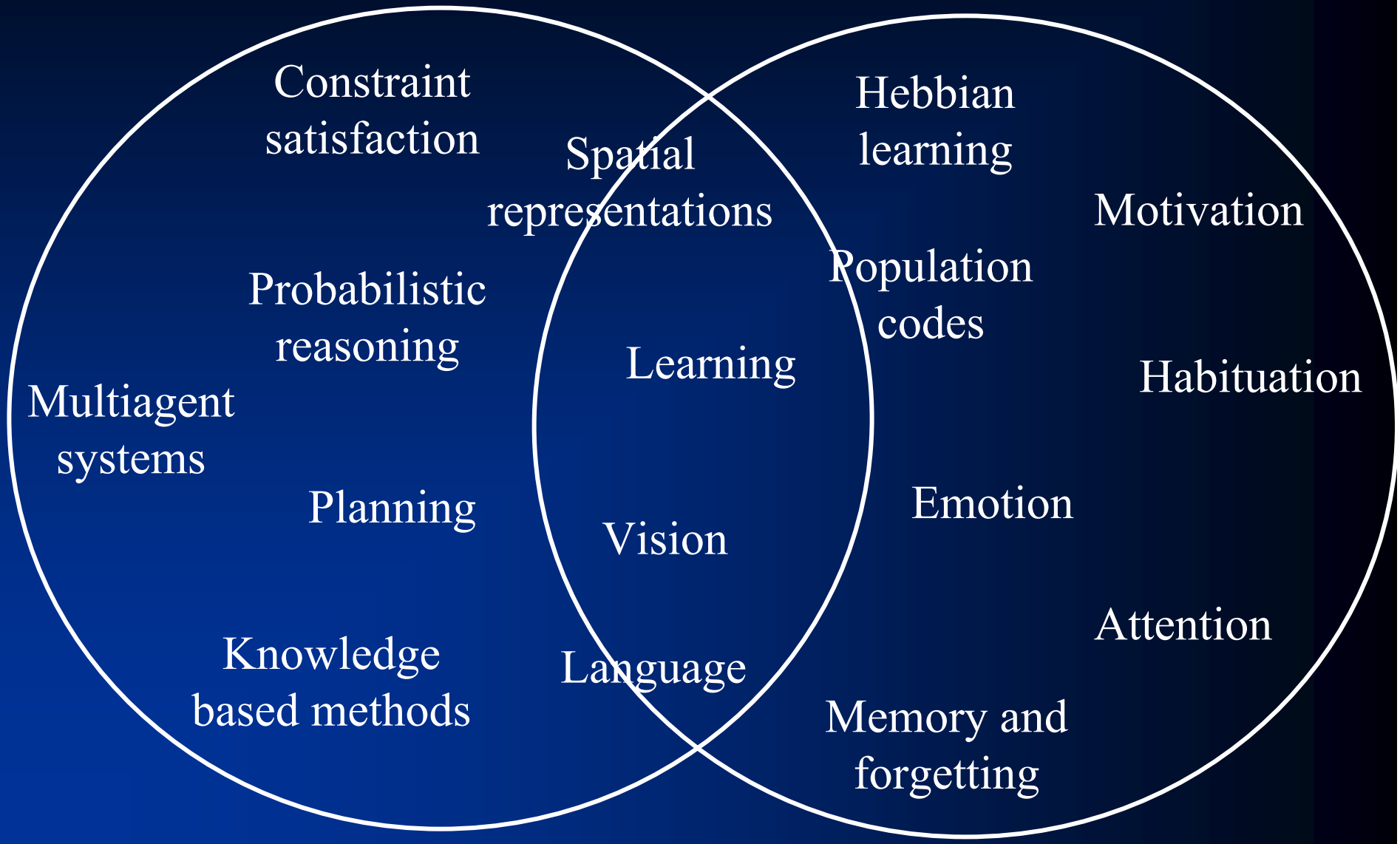
The Activa system consists of electrodes that are implanted into the brain and connected by leads (wires) under the skin to a pulse generator implanted in the abdomen or chest. The pulse generator sends a constant stream of tiny electrical pulses to the brain, blocking tremors. When the device is implanted in both sides of the brain, two separate systems are used.

To turn the stimulator on and off, the patient holds a magnet over the pulse generator. The generator must be replaced every three to five years, the life of the battery.

FDA based approval of the device on a clinical study conducted by Medtronic of the system's safety and effectiveness when implanted in both sides of the brain and on the recommendation of the Neurological Devices Panel of FDA's Medical Devices Advisory Committee.

Some 160 patients with advanced Parkinson's disease were enrolled at 18 medical centers in the United States, Canada, Australia and Europe. The device was implanted bilaterally in 134 patients. The implant procedures were done simultaneously or in stages. The patients were followed

# AI→ ←Brain Sciences



# Thank You!

- John Anderson
- Peter Bandettini
- Bruce Buchanan
- Pat Carpenter
- Rich Caruana
- Bill Eddy
- Marcel Just
- Jay McClelland
- Joan Mitchell
- Jack Mostow
- Stefan Niculescu
- Francisco Pereira
- Tomaso Poggio
- Yulin Qin
- Sebastian Thrun
- David Touretzky
- Manuela Veloso