

Why Computers Don't Care

The Origins of Meaning in Efficient Computation

P. Read Montague

Baylor College of Medicine,
Houston, TX

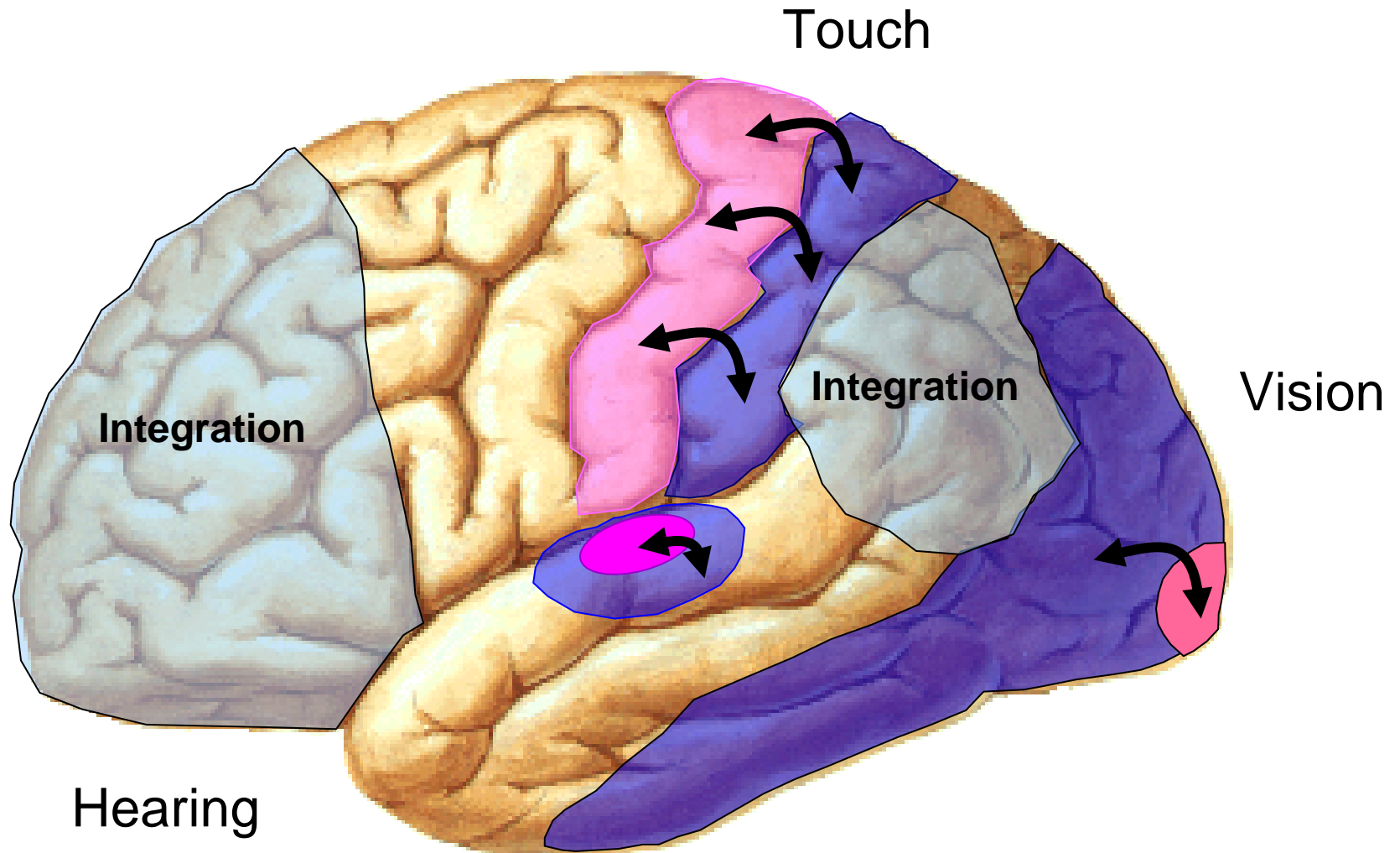
Part 1. Structural themes in the nervous system

Part 2. Energy, computation, and meaning

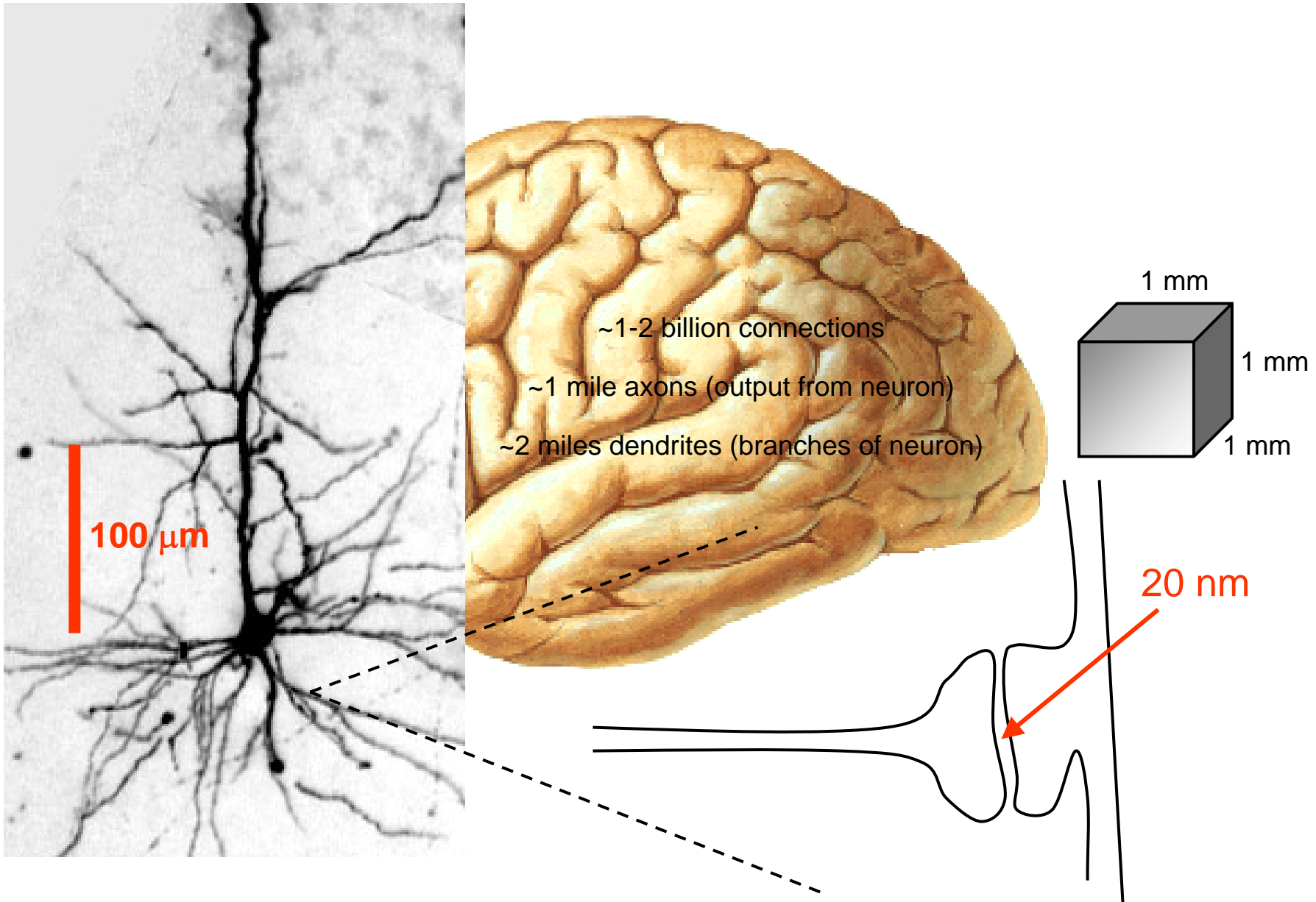
Part 3. Pursuing goals with valuation machinery

Why can ideas veto biological instincts for survival?

Theme I: Put data near the processing capacity -
maps and integrative regions

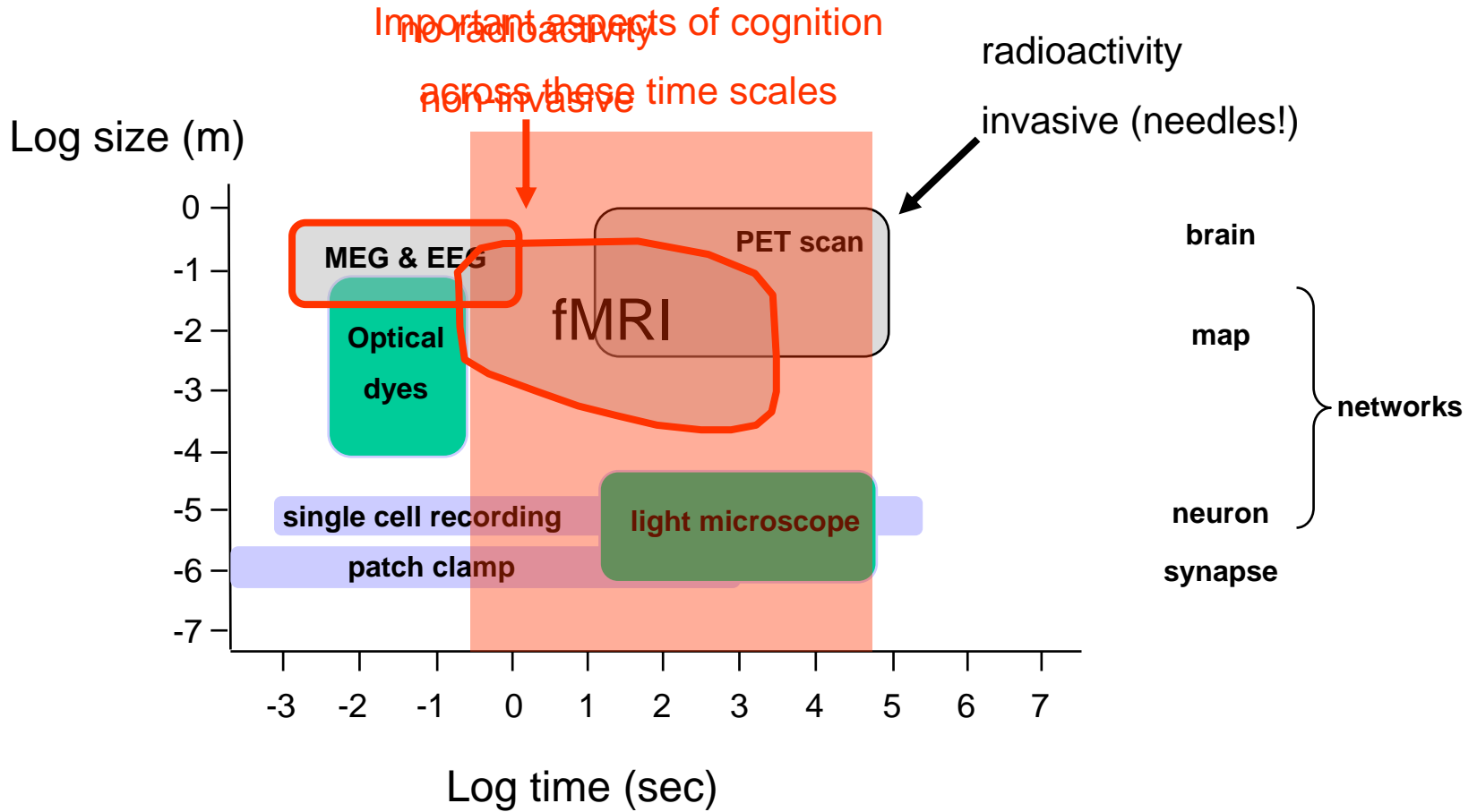


Theme II: Miniaturization, self-repair, re-configuration



Many physical probes of living neural tissue

Where does fMRI fit?



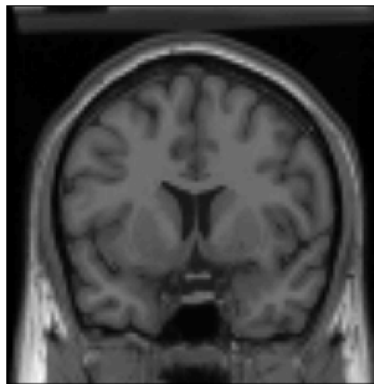
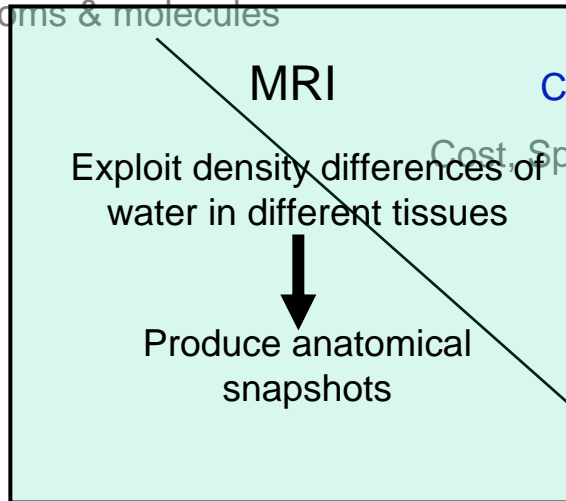
functional Magnetic Resonance Imaging (fMRI)



Four antecedents to modern fMRI

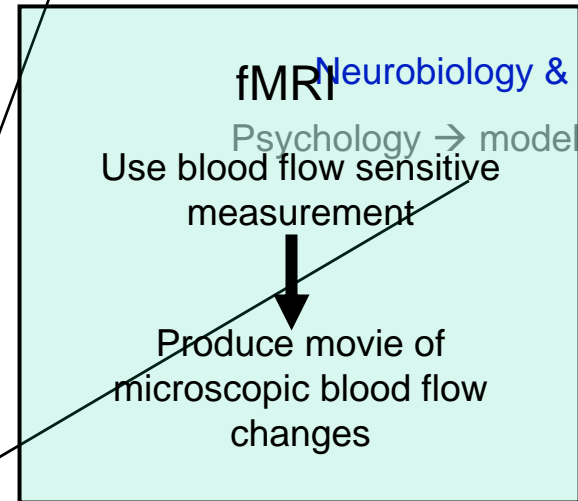
Physics and Chemistry

magnetic properties of
atoms & molecules



Physiology & Anatomy

Control of cerebral blood flow



Computing

Cost, Speed, Algorithms

MRI

Modern fMRI

Neurobiology & Models

Psychology → models of learning



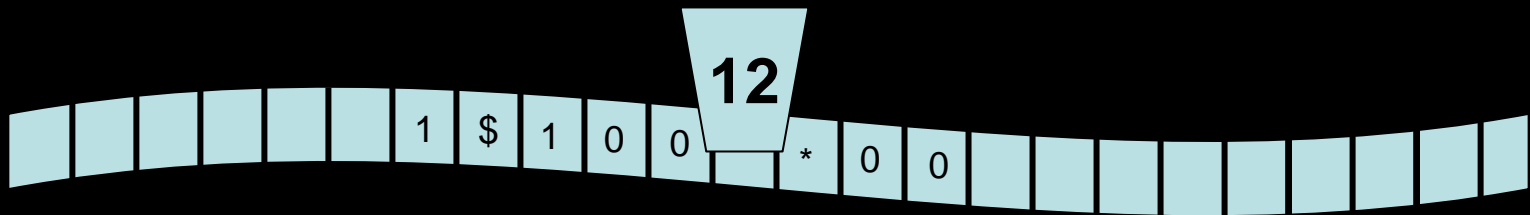
The neural myth

Despite the *slowness*, *noisiness*, and *imprecision* of neural computations...

...the brain - *usually through the 'miracle' of parallelism* - performs inexplicable feats of perception and action.



Alan Turing



Turing Machine

Computation

Computational Theory of Mind (CTOM)

Patterns of information processing

Hardware (physical interactions)

'Mind-like' stuff from 'Stuff-like' stuff

Turing - mind is not equivalent to our brain, but is equivalent to information processing supported by our brain.

Blurs distinction between device and algorithm

Computational Theory of Mind (CTOM)

Meaning is missing

Symbol list 1 → Symbol list 2

When does meaning get added back?

Recovering meaning

Two questions face every mobile creature

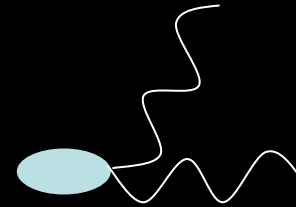
What is the value of available choices?

What does each choice cost?

Organisms must assign meaning to the processes that comprise them.



run



tumble

They run on batteries and the future is uncertain.

Recharge or die

A principle for value-based computation

How is one computation assigned more or less meaning than another?

Organisms distribute their finite resources to the computations that sustain them.

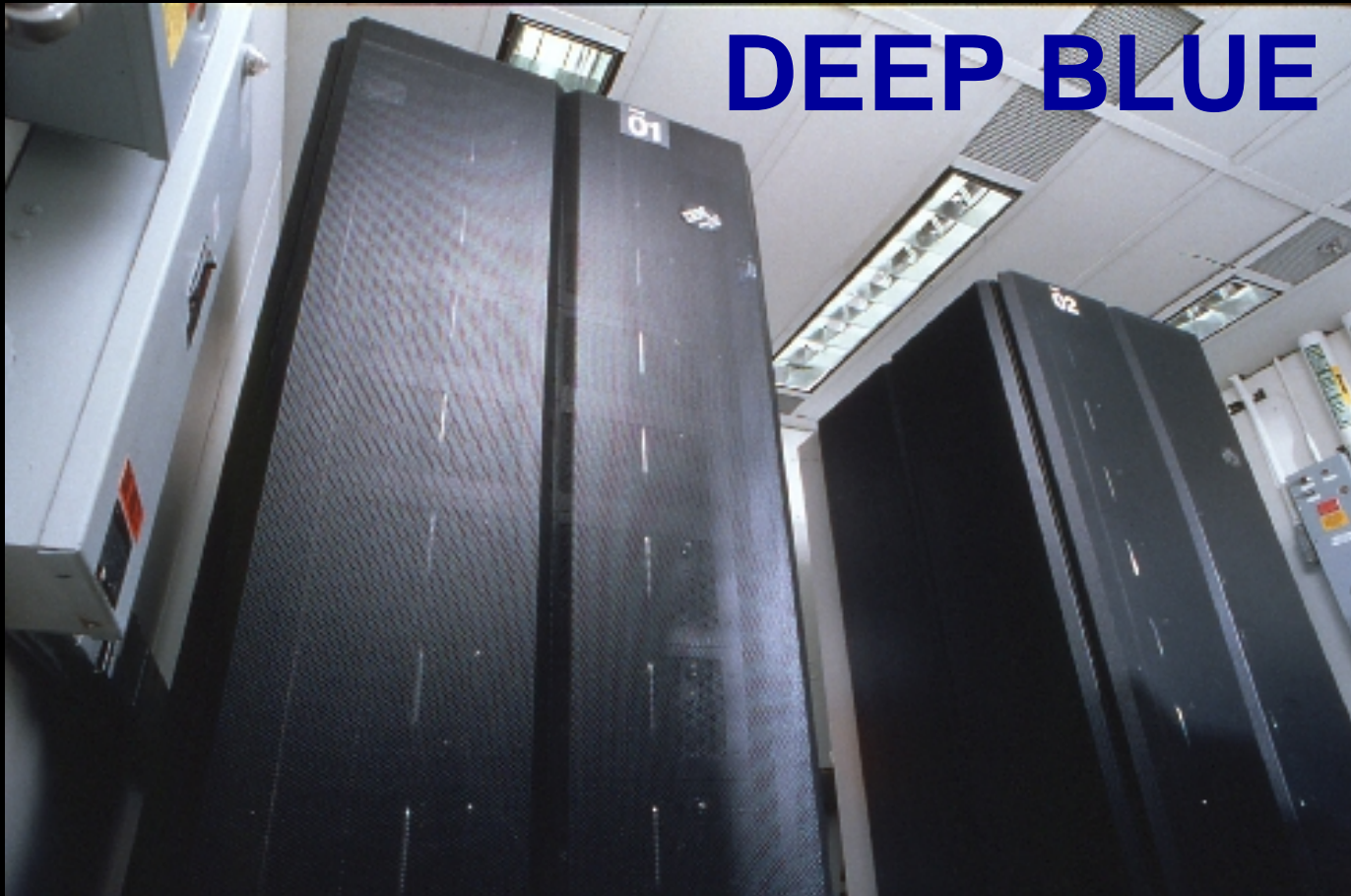


Their willingness to do so *should scale* with the expected long-term returns from the computation.

Where's the evidence for efficient computation in the brain?



DEEP BLUE



32 nodes

8 dedicated chess processors per node (total 256)

1.4 tons (lots of AC)

200,000,000 positions per second

Deep Blue was wildly inefficient

Kasparov stayed warm to the touch

Kasparov's brain is freakishly efficient

~100 watts →
brain uses ~20 watts

What about evidence from detailed neural properties?

Principles of efficient computation

Drain batteries slowly

neurons compute slowly

Save Space

Small parts, sparse connectivity

Save bandwidth

Stay off the lines, noisy codes

Have goals

Set and pursue goals

MYTH is now debunked

What happened?

Why are modern computers so wasteful?

Why is meaning missing from
modern model of computation?

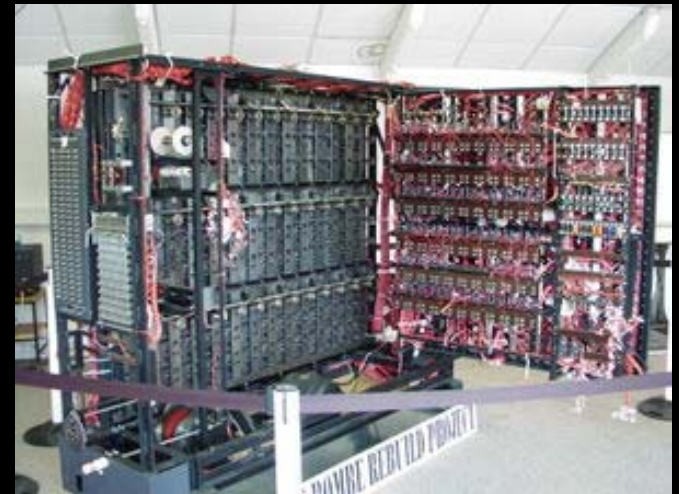
World War II happened



Bletchley Park Mansion



Enigma



Value-based Computation

Computer programs don't care (enough)

Flexible goals should guide their behavior (multi-level)

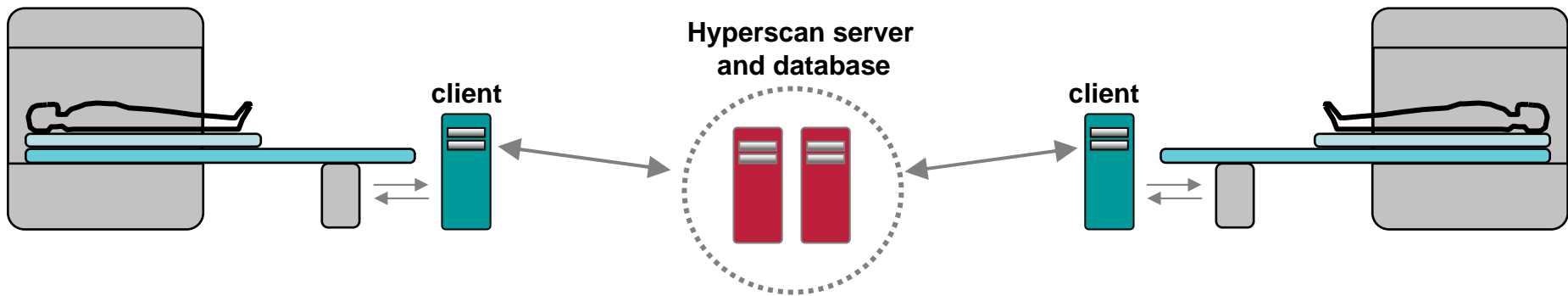
hyperscan fMRI (h-fMRI)

information:

www.hnl.bcm.tmc.edu/hyperscan

software download:

www.hnl.bcm.tmc.edu/nemo



Houston (BCM)

Houston (BCM)

Houston (BCM)

California (Caltech)

New Jersey (Princeton)

Georgia (Emory)

Germany (Uni-Ulm)

Hong Kong (HKUST)

Goal-directed choice

Select goal

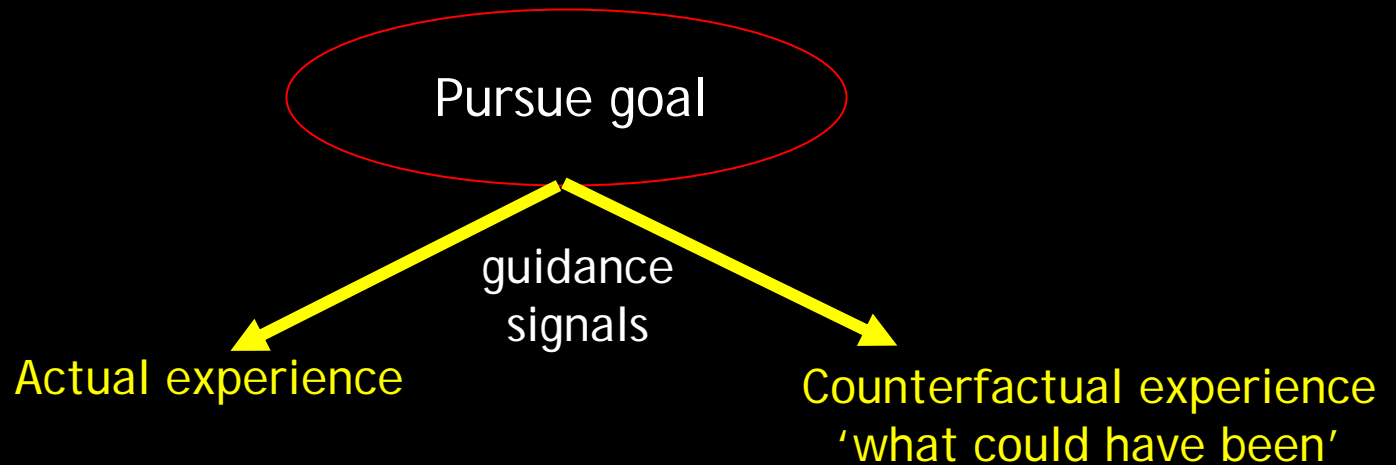
Sustain goal

Pursue goal

guidance
signals

Actual experience

Counterfactual experience
'what could have been'



Sharks don't go on hunger strikes



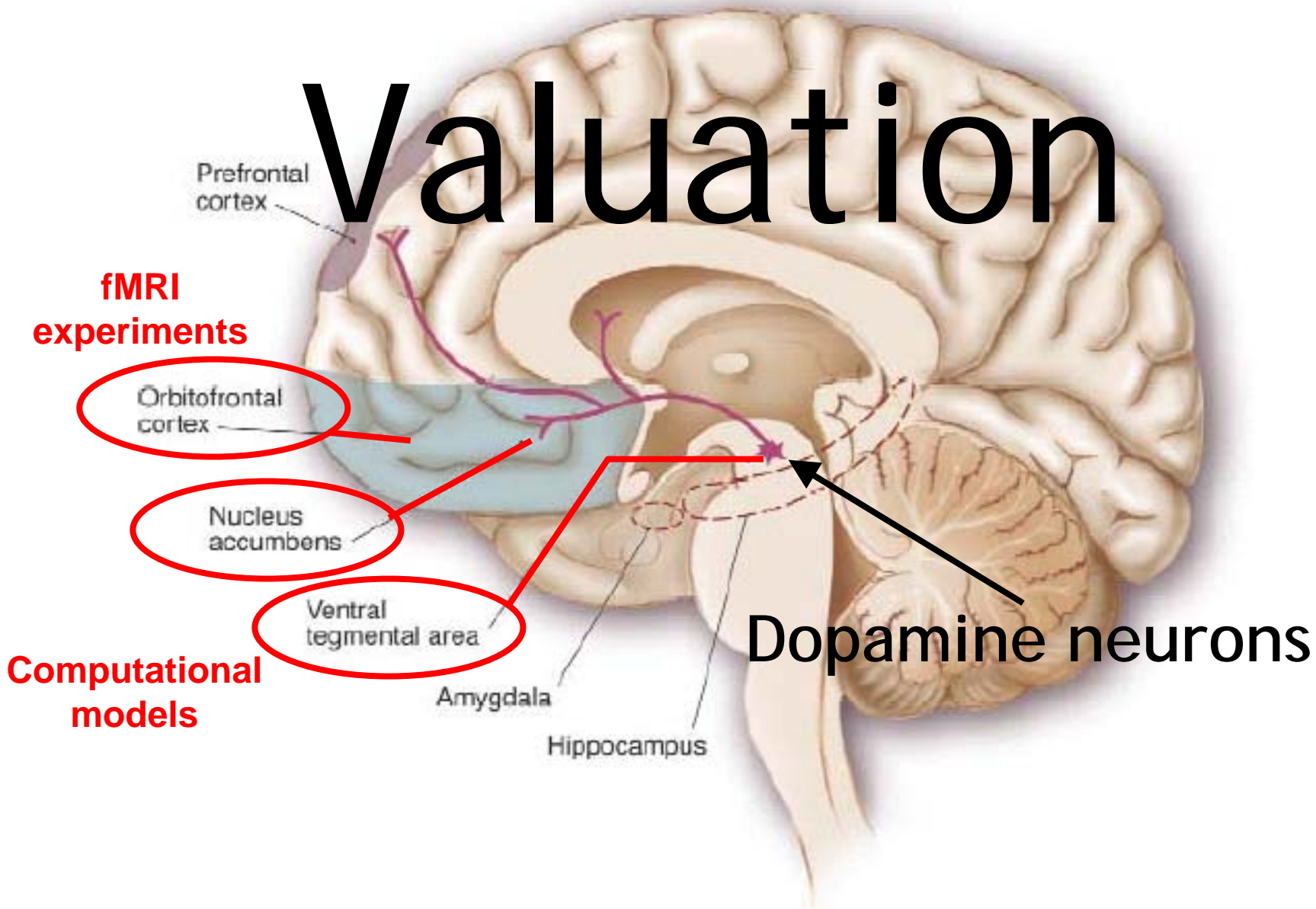
But humans do

Ideas gain behavioral potency
of primary rewards like food and water

How?

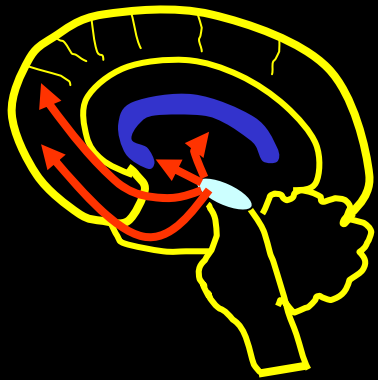
Goal pursuit requires guidance signals → dopamine

Valuation



Midbrain dopamine neurons

Pause, burst, and 'no change' responses represent reward prediction errors



burst
↑
—
↓
pause



naive



After learning



After learning
(catch trial)

Red Light becomes a proxy for the value of the juice

Midbrain dopamine neurons

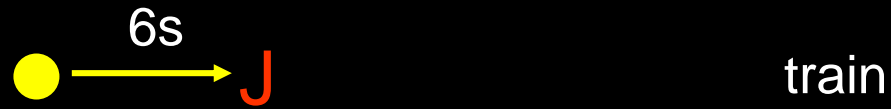
Constant emission of reward prediction errors

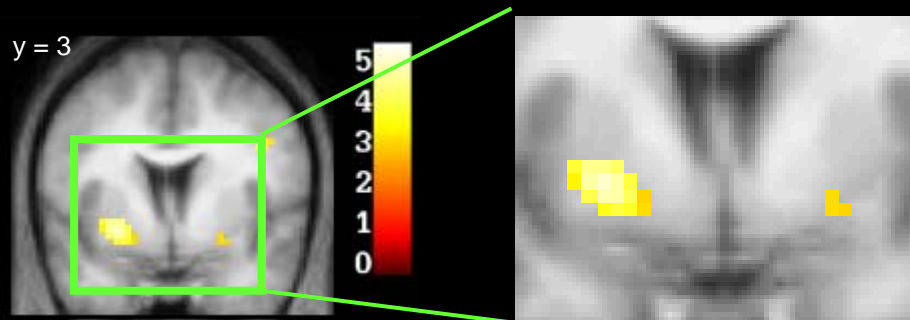
$$\text{ERROR SIGNAL} = \text{current reward} + \gamma \text{ next prediction} - \text{current prediction}$$
The equation is displayed in yellow text on a black background. A red rectangular box encloses the terms 'current reward + γ next prediction'. A red curved arrow points from the top right corner of this box to the 'current prediction' term. A yellow arrow points upwards from the text below to the 'current reward' term.

"I want to solve Fermat's last theorem"

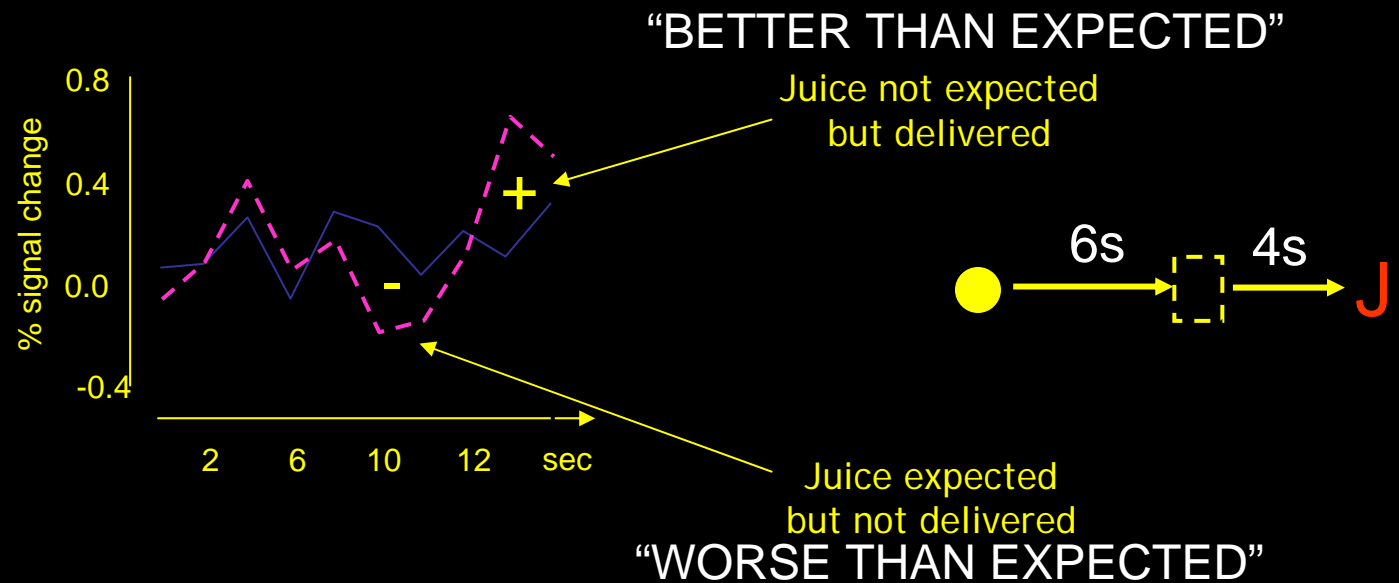
Neural evidence suggests that system can be re-deployed
for abstractly defined rewards → ideas

Induce reward prediction error in
a passive conditioning task → measure brain response



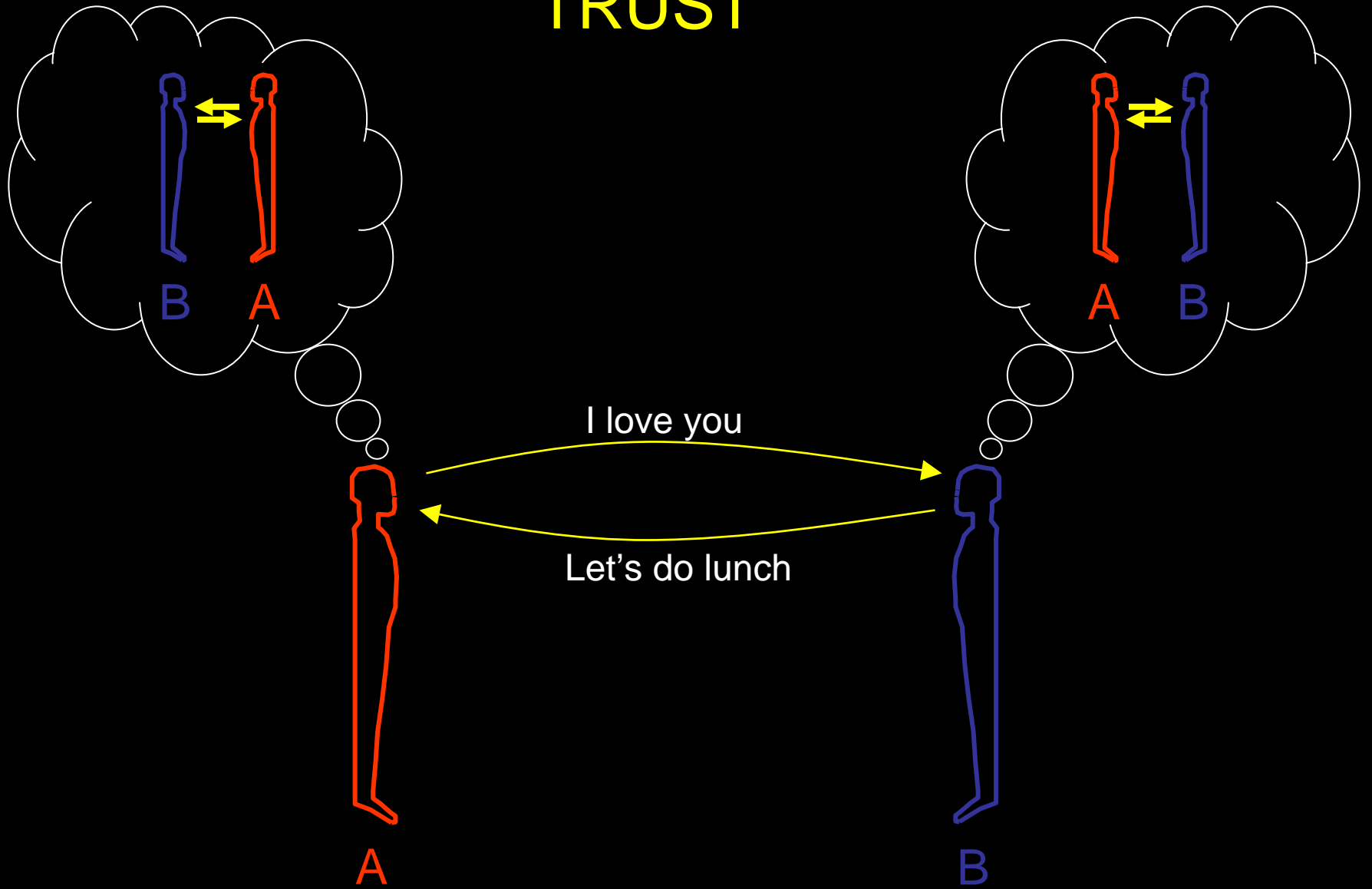


Hemodynamic response in ventral putamen



What about something more abstract
like trust?

TRUST



Trust

Requires risk – no trust needed without risk

Most basic form is about model-building

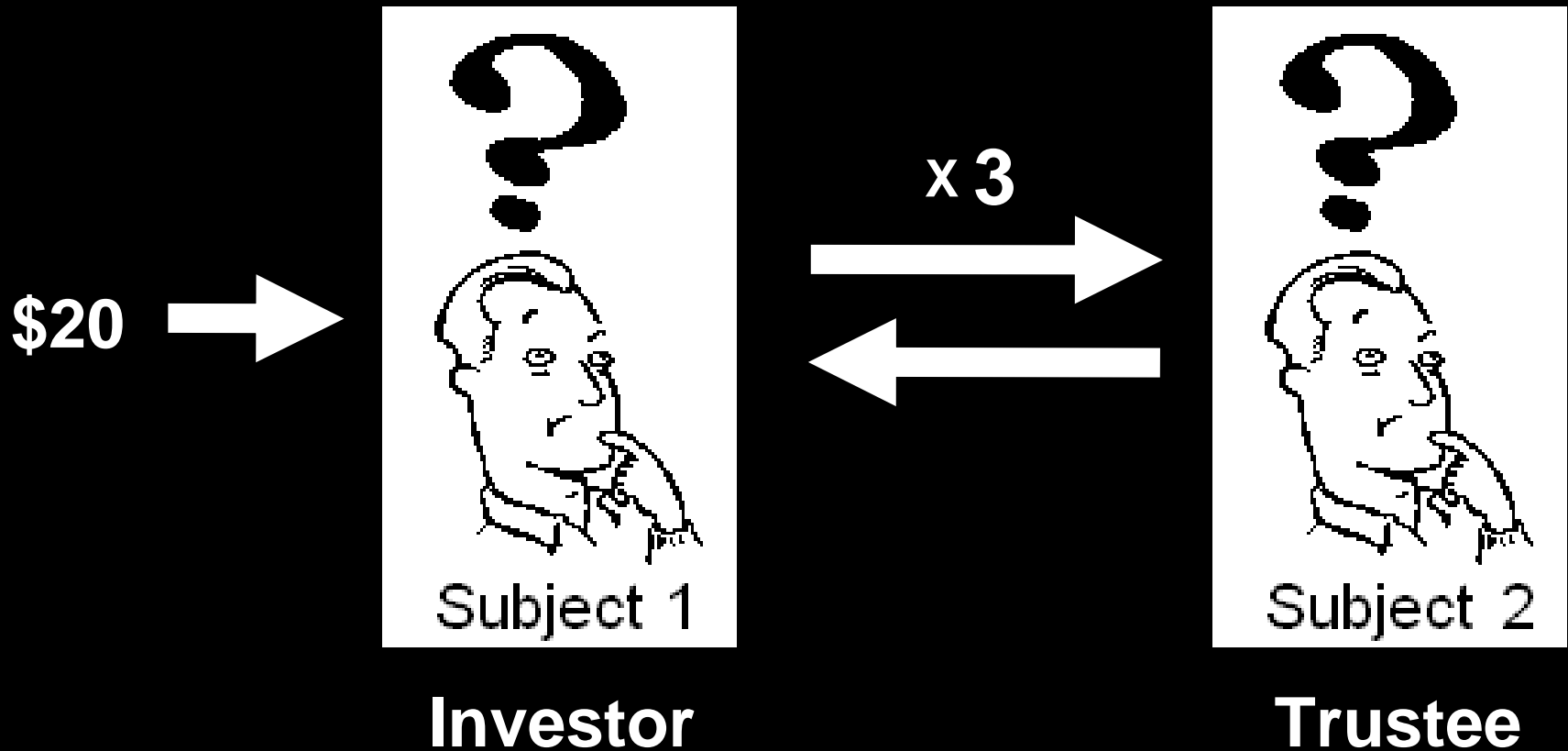
Cooperation

How can we study trust in a social exchange?

Use simple economic game between two players

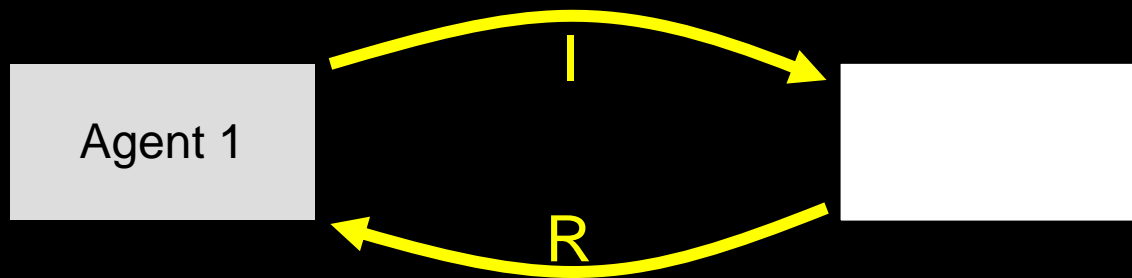
Synchronously image both interacting brains

A dynamic version of the Trust game (10 rounds)



Simplifying and quantifying Trust

Trust is the amount of money a sender sends to a receiver without external enforcement.



What is Fair?

Investor



Trustee



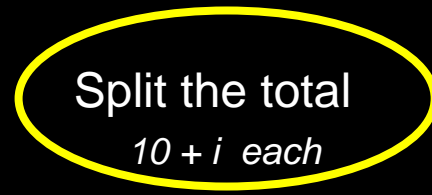
Split 'loaf'



Split the profit



Split the total
10 + i each



collective ownership? common goods?

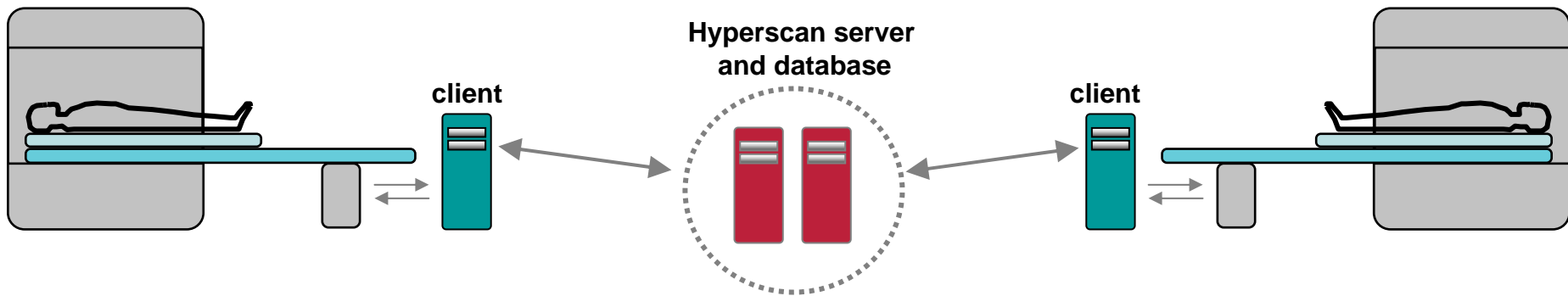
Hyperscanning: Scan all brains during the interaction

information:

www.hnl.bcm.tmc.edu/hyperscan

software download:

www.hnl.bcm.tmc.edu/nemo



Houston (BCM)

Houston (BCM)

Houston (BCM)

California (Caltech)

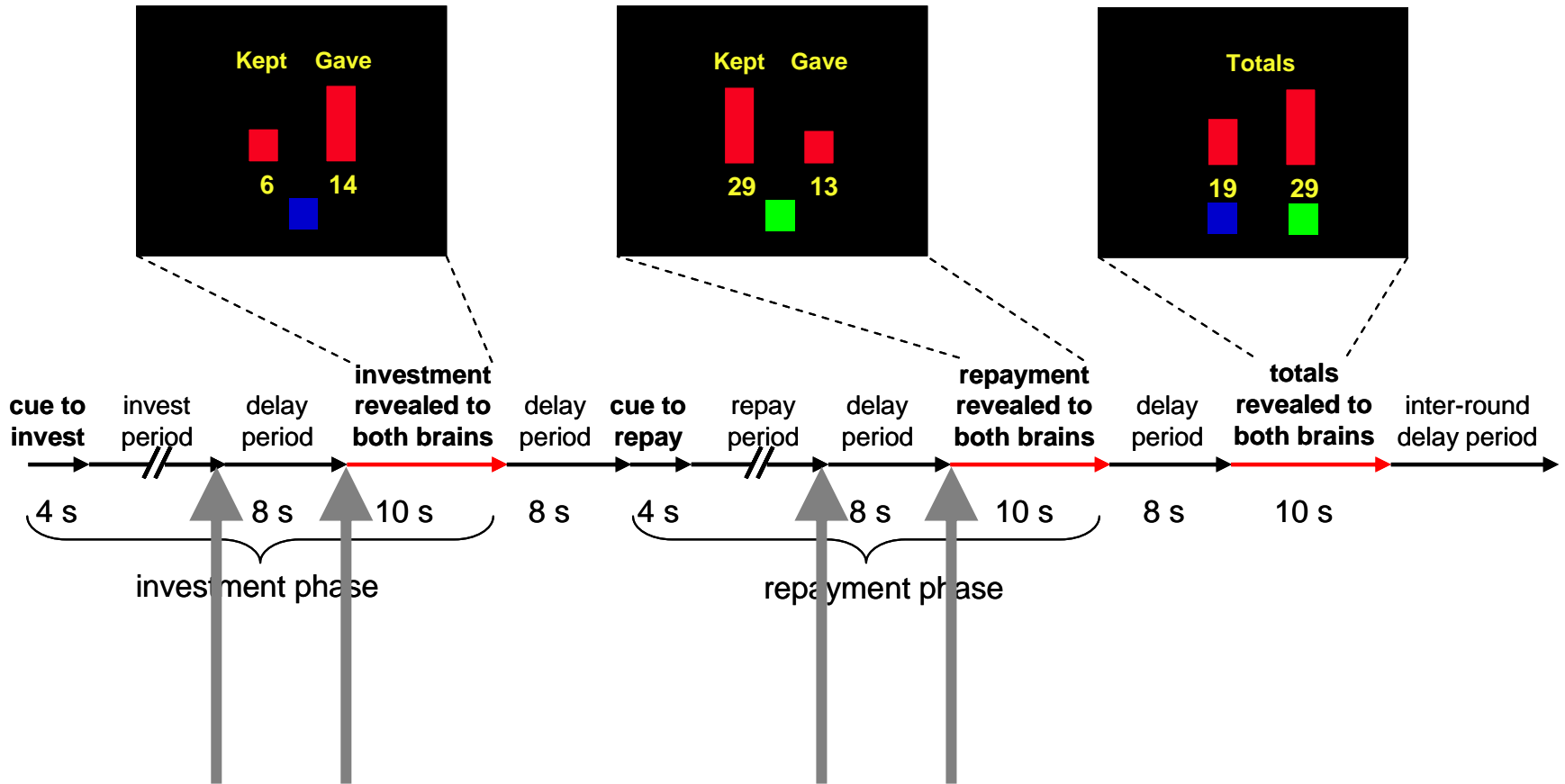
New Jersey (Princeton)

Georgia (Emory)

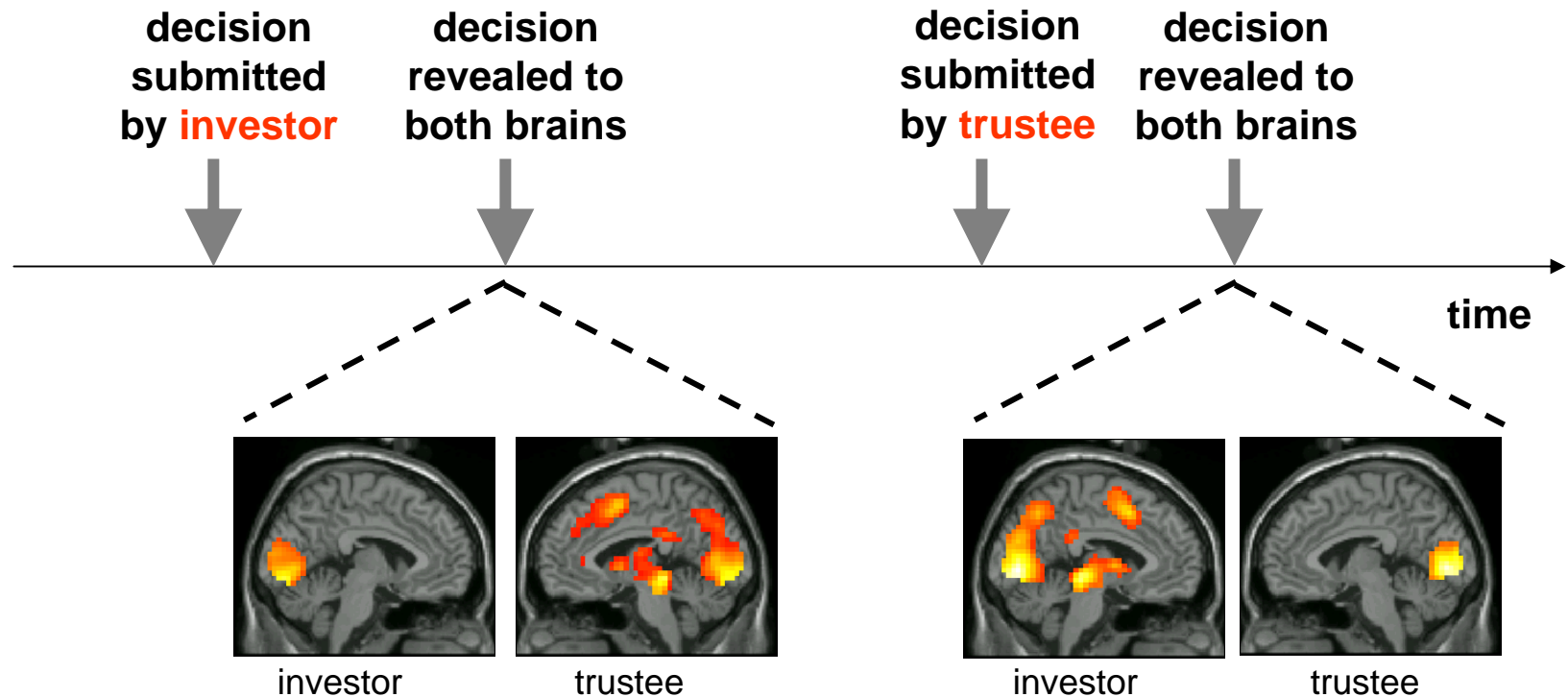
Germany (Uni-Ulm)

Hong Kong (HKUST)

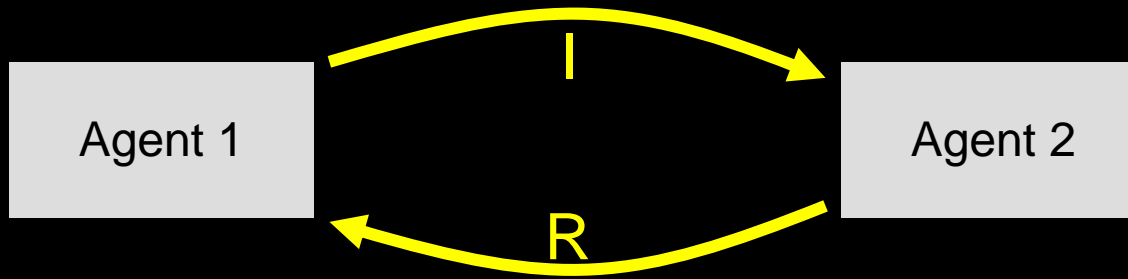
Structure of a round



What we generally see on this hyperscanned trust exchange



What is the behavioral signal that most strongly influences changes in trust (money sent) ?

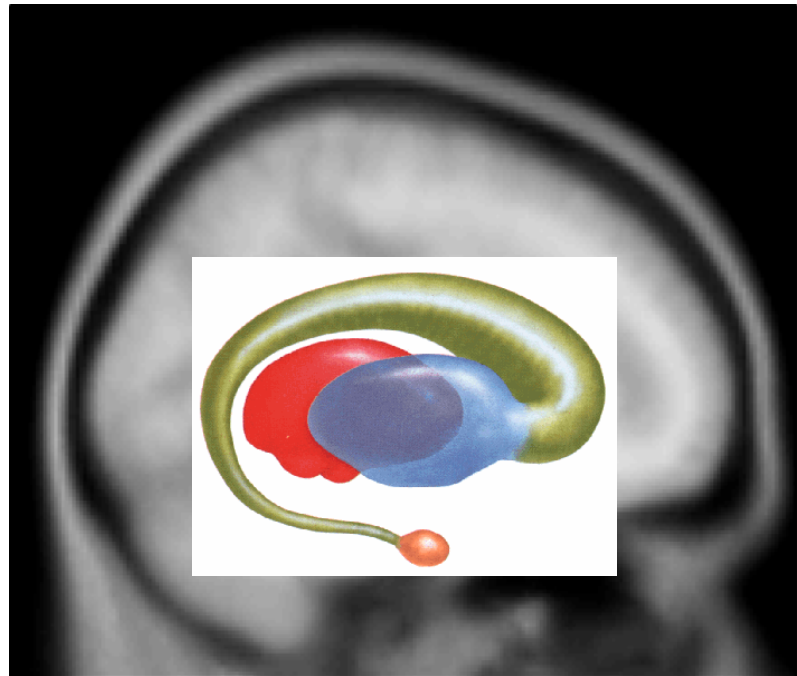


Reciprocity = TIT-FOR-TAT

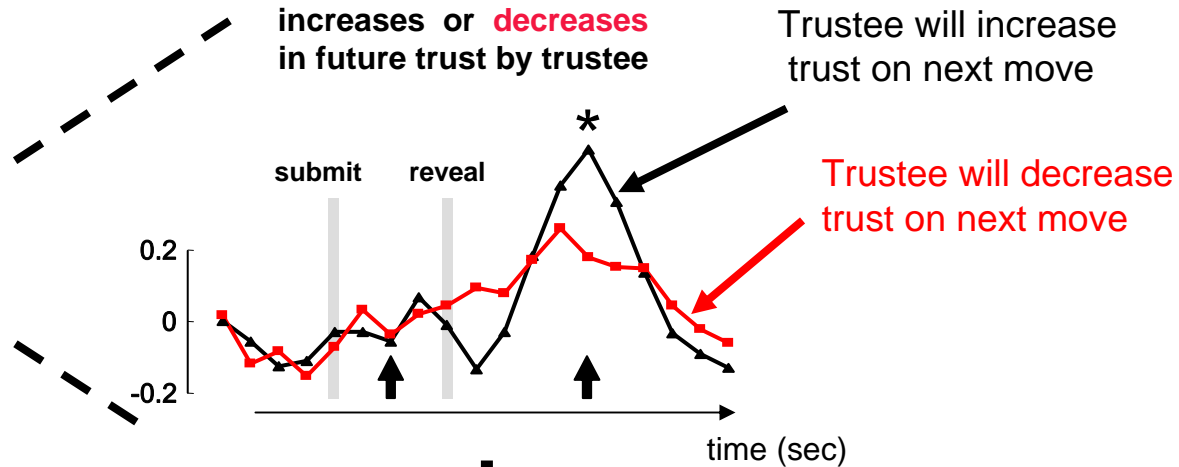
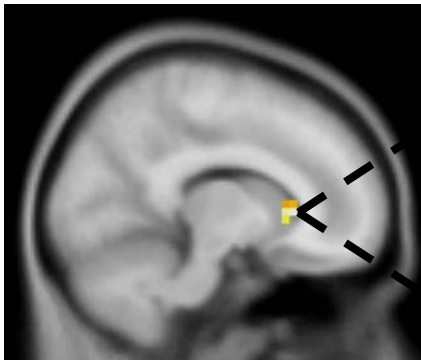
Trustee brain

Activity related specifically to benevolent reciprocity

Caudate nucleus – action choice, reward processing, ...

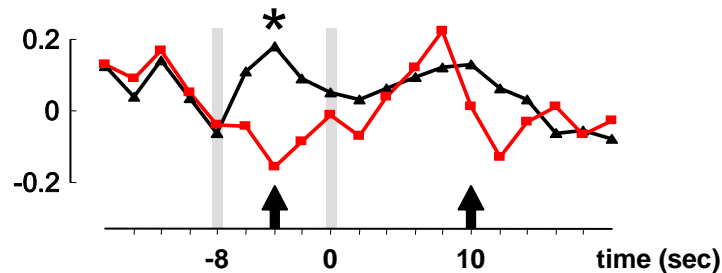


Trustee 'intention to increase trust' shifts with reputation building

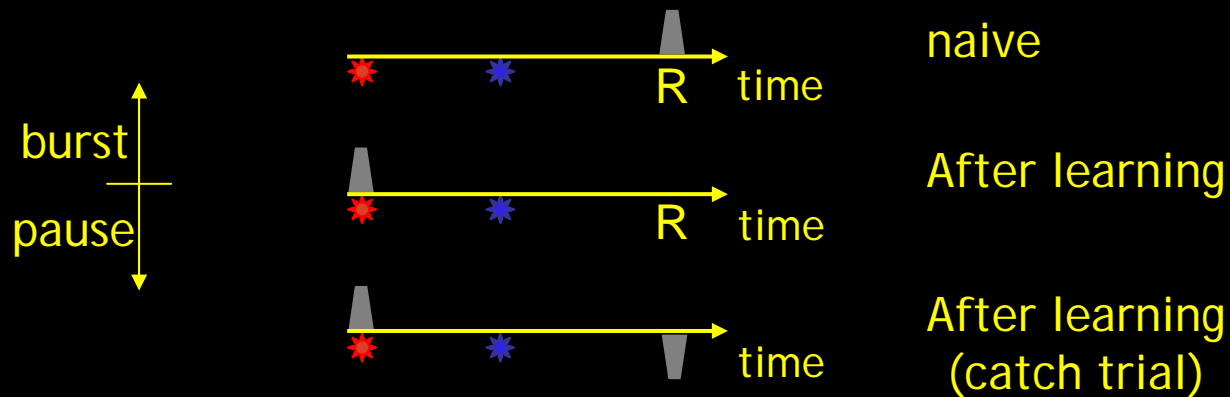


Reputation develops

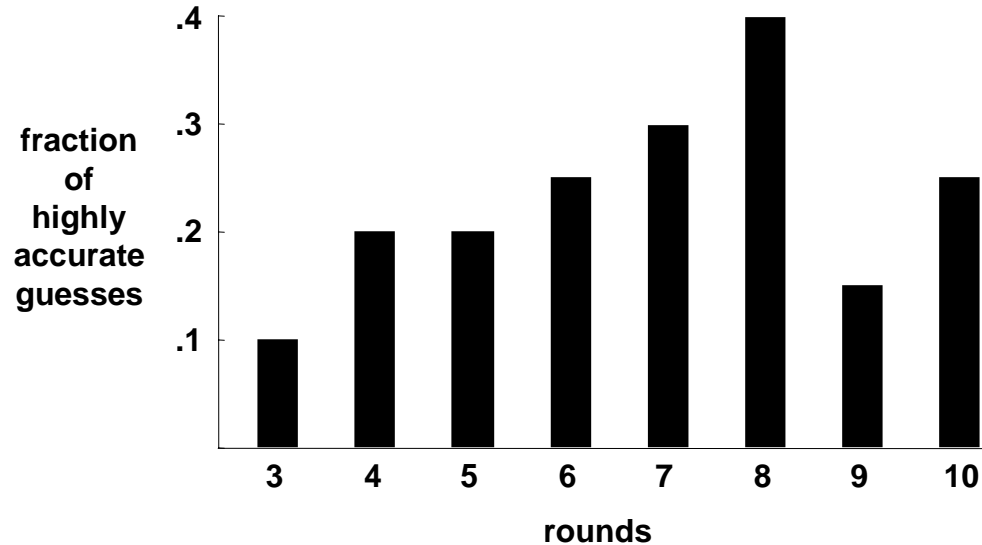
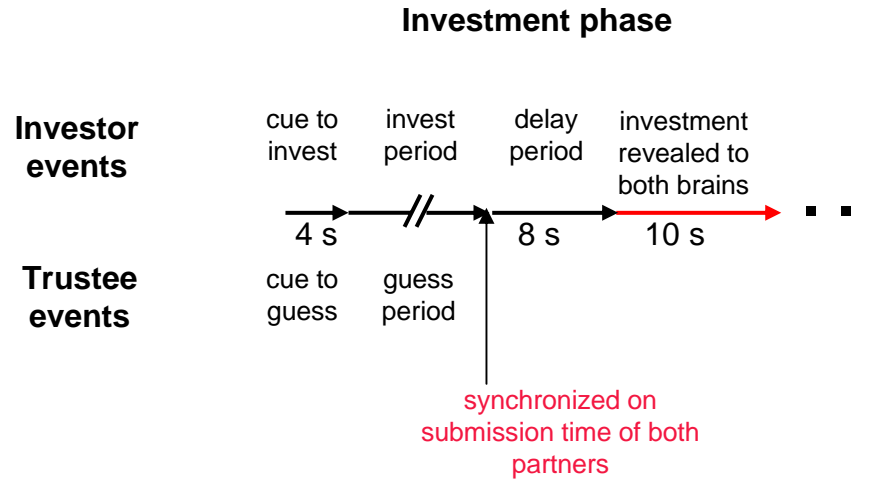
Signal is now anticipating the outcome



Temporal shift resembles value transfer in reward learning experiments



How do we know a reputation is forming?



Summary

Finite resources force distribution decisions on life forms

Biological computations always had to assign meaning

Efficient computational systems must have goals

Goal pursuing machinery identified and modeled

Ideas gain behavioral potency of food and sex

Re-deployment of reward harvesting machinery