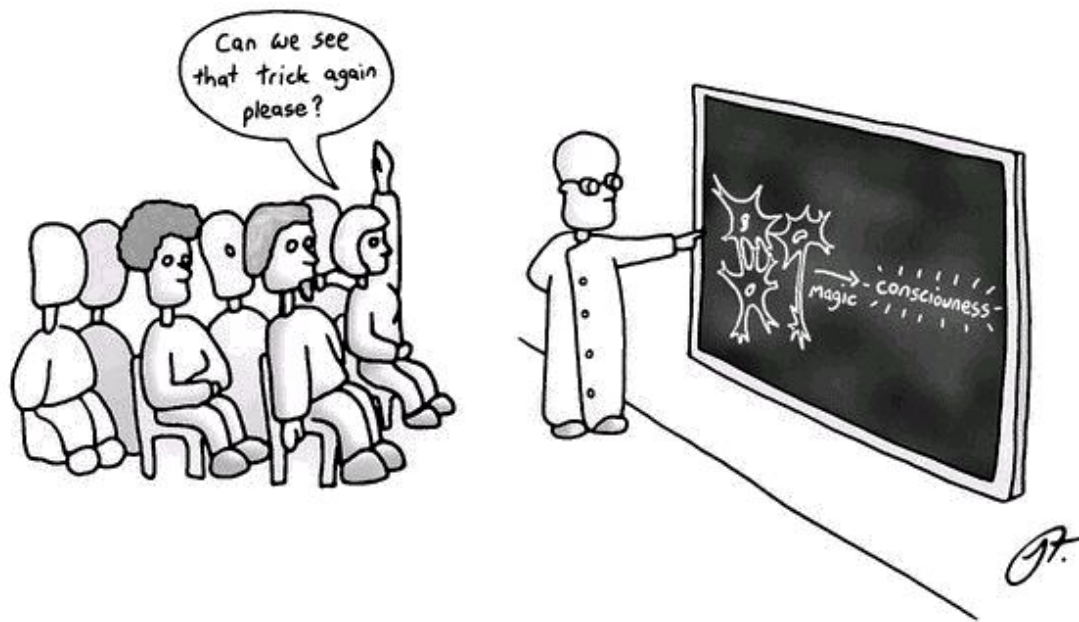
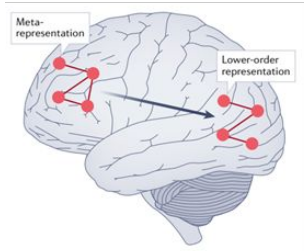


The unfolding argument: Why IIT and other causal structure theories cannot explain consciousness

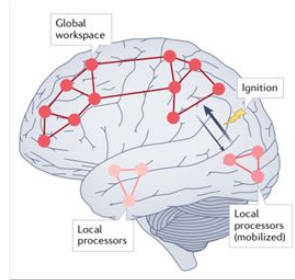
Adrien Doeriga, Aaron Schurgerb, Kathryn Hessf, Michael H. Herzoga

Presentation by Ksenia Soloveva
TReNDS Center MLBBQ
2/10/2023

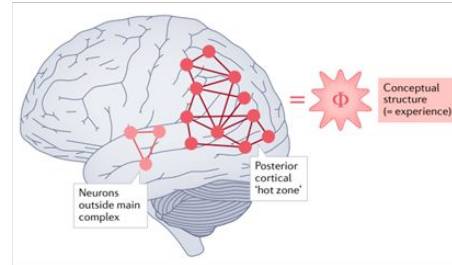




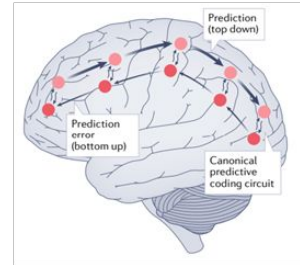
HOT



GWT



IIT

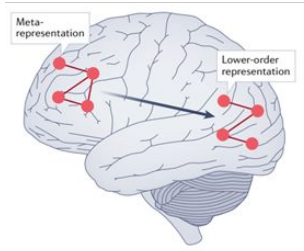


Re-entry/Predictive

Theories of consciousness

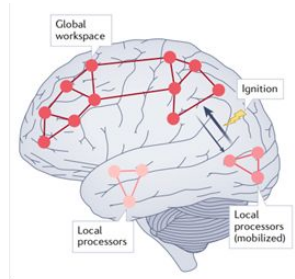
By Anil K. Seth and Tim Bayne

A mental state is
conscious =
it is the target of
another mental
representation



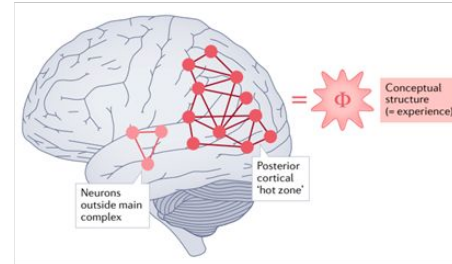
HOT

A mental process
conscious =
it has access to the
global workspace



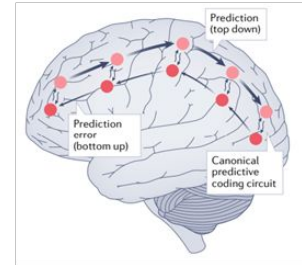
GWT

consciousness =
causal structure
(causal mechanism)
and its state



IIT

conscious perception
= top-down
(recurrent,
re-entrant) signaling



Re-entry/Predictive

Theories of consciousness

By Anil K. Seth and Tim Bayne

Theories of consciousness

Biological:

- how did it appear in evolution?
- What genes are associated with?
- How does it appear in ontogeny?
- How does it function?
- What biological structures of the body is associated with?
- What animals have?

Physical:

- What is its nature?
- Quantum or not?
- What laws describe?
- Deterministic or not?

Psychological:

- How is the human psyche built?
- How to change settings?
- how to switch states?
- how to develop a stable mind?

Economic:

- How to manage people's decisions?

Mathematical:

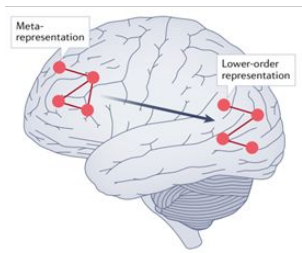
- How to define consciousness through mathematics?
- How is consciousness related to information?
- What mathematical structures are related to consciousness?

Philosophy: no theories, just new problems and questions!

Common questions to answer:

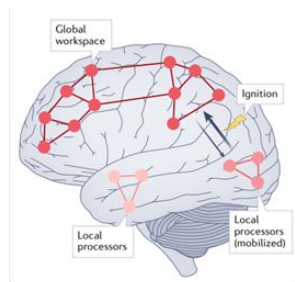
- 1) In what part of the brain is consciousness located?
- 2) How does activity in the brain become conscious?
- 3) What distinguishes different local and global states of consciousness?

A mental state is
conscious =
it is the target of
another mental
representation



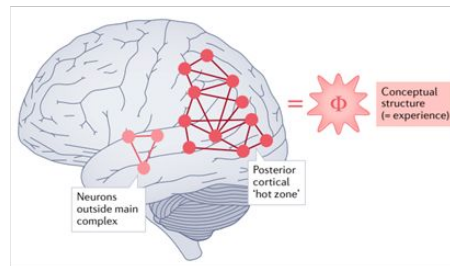
HOT

A mental process
conscious =
it has access to the
global workspace



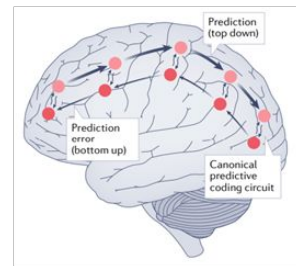
GWT

consciousness =
causal structure
(causal mechanism)
and its state



IIT

conscious perception
= top-down
(recurrent,
re-entrant) signaling



Re-entry/Predictive

Theories of consciousness

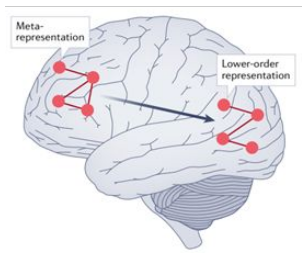
By Anil K. Seth and Tim Bayne

Common questions to answer:

- 1) In what part of the brain is consciousness located?
- 2) How does activity in the brain become conscious?
- 3) What distinguishes different local and global states of consciousness?

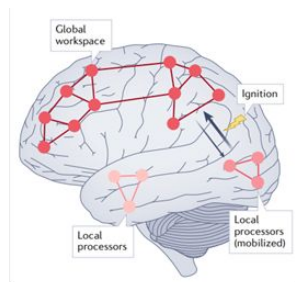
Find the odd one out among these four theories:

A mental state is
conscious =
it is the target of
another mental
representation



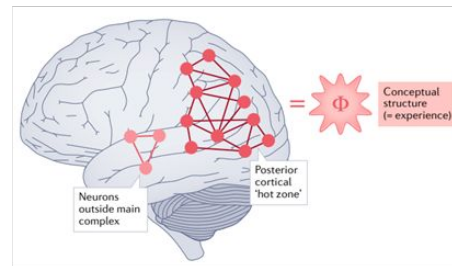
HOT

A mental process
conscious =
it has access to the
global workspace



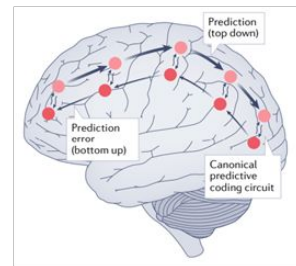
GWT

consciousness =
causal structure
(causal mechanism)
and its state



IIT

conscious perception
= top-down
(recurrent,
re-entrant) signaling



Re-entry/Predictive

Theories of consciousness

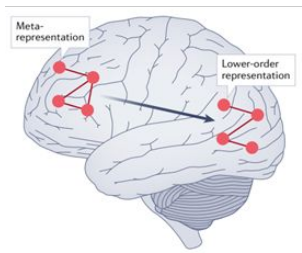
By Anil K. Seth and Tim Bayne

Common questions to answer:

- 1) In what part of the brain is consciousness located?
- 2) How does activity in the brain become conscious?
- 3) What distinguishes different local and global states of consciousness?

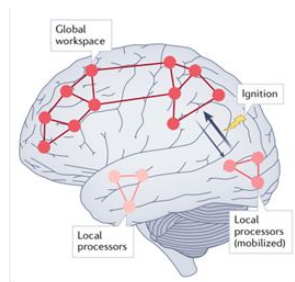
Find the odd one out among these four theories:

A mental state is
conscious =
it is the target of
another mental
representation



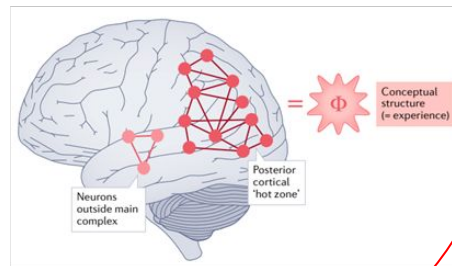
HOT

A mental process
conscious =
it has access to the
global workspace



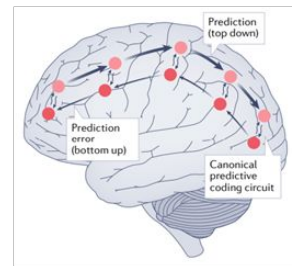
GWT

consciousness =
causal structure
(causal mechanism)
and its state



IIT

conscious perception
= top-down
(recurrent,
re-entrant) signaling



Re-entry/Predictive

Theories of consciousness

By Anil K. Seth and Tim Bayne

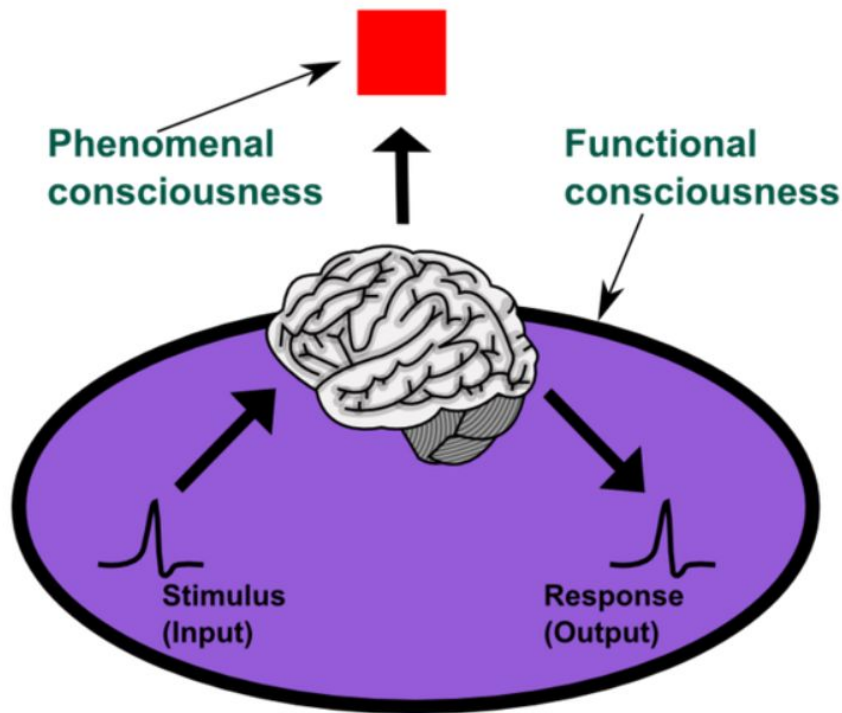
Hard problem

David Chalmers in the 1990s:

The easy problems = neural correlates of **functions and behaviors**

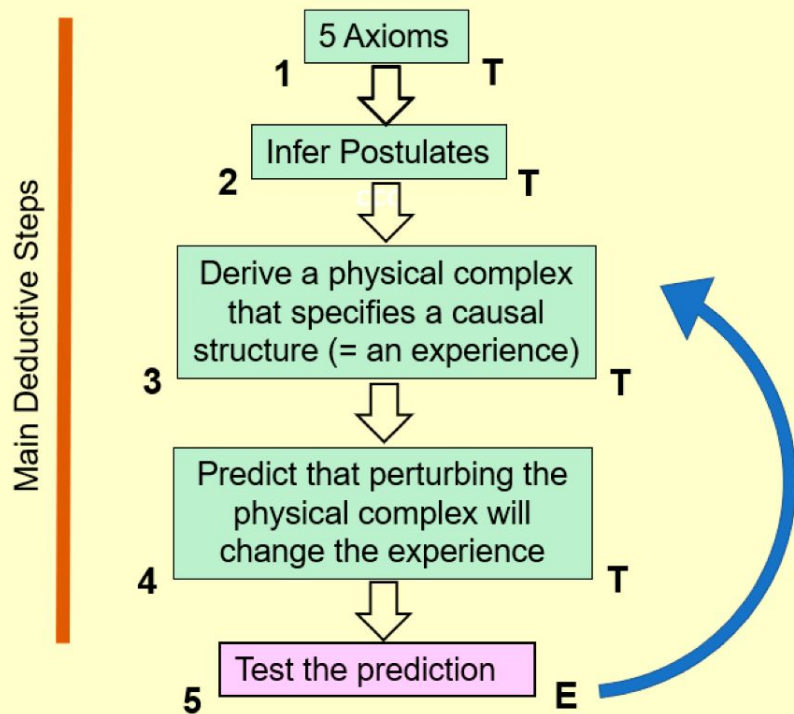
The hard problem = **subjective** experiences

This cannot be explained by the activity of neurons



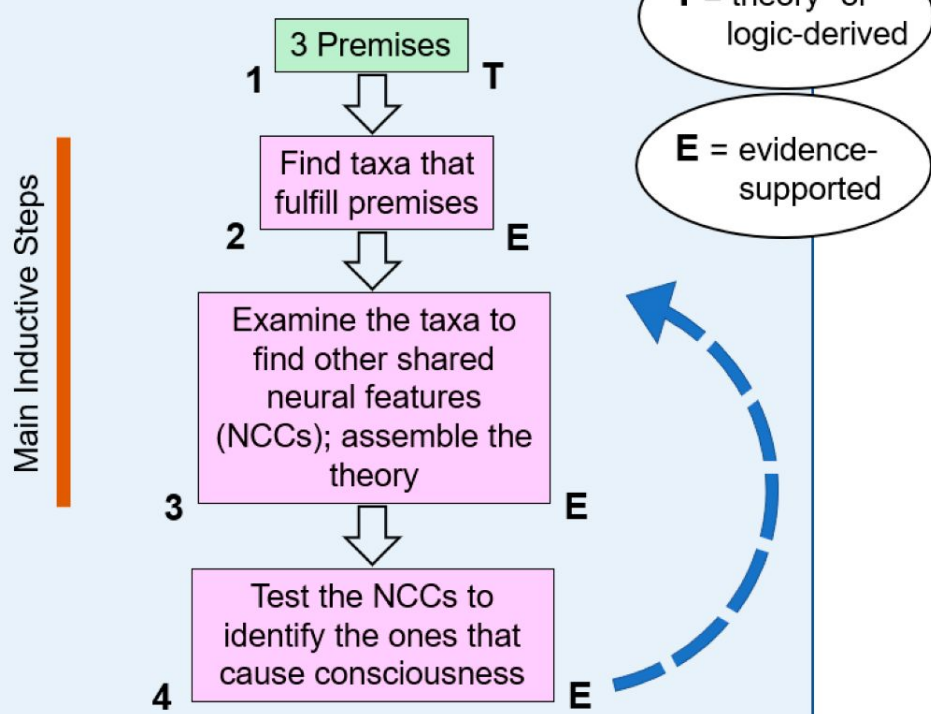
A. Integrated Information Theory

1. Starts from phenomenal properties of experience
2. Definition of consciousness: raw experience, as any interactions between the parts
3. Deductive inference: axioms to general theory to specific tests
4. Derivation:



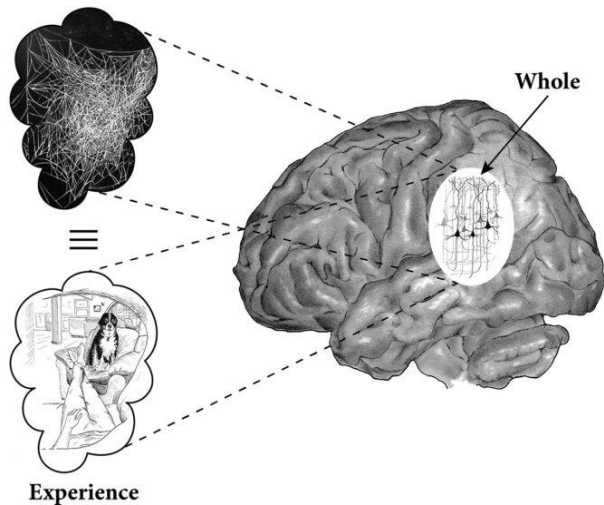
B. Neurobiological Naturalism

1. Seeks physical properties to explain experience
2. Definition of consciousness: raw experience
3. Inductive inference: specific observations to a general theory (then test it)
4. Derivation:

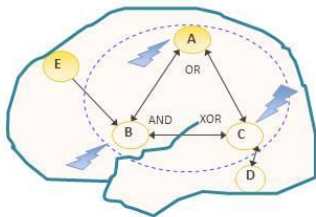


Integrated information theory

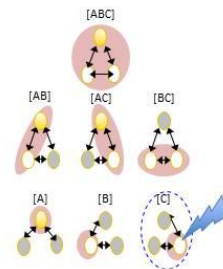
Cause-Effect Structure



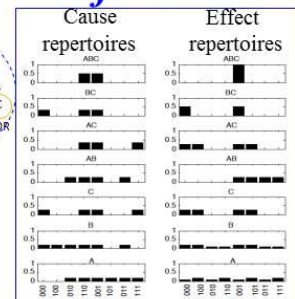
Intrinsic existence



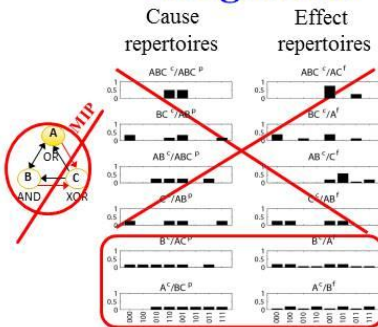
Composition



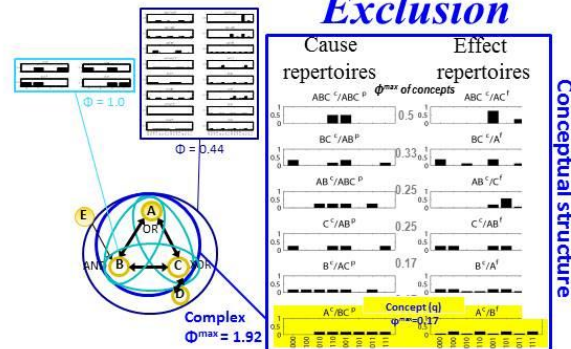
Information



Integration

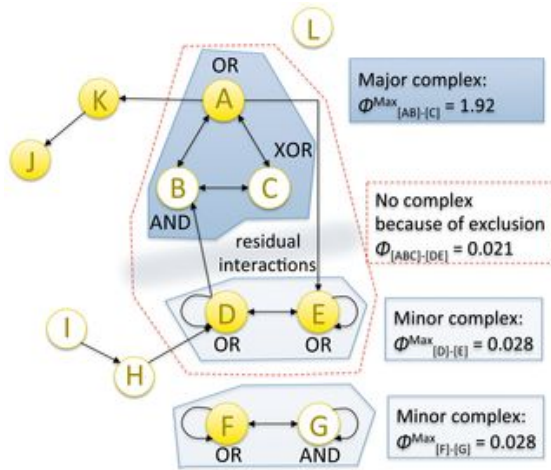


Exclusion

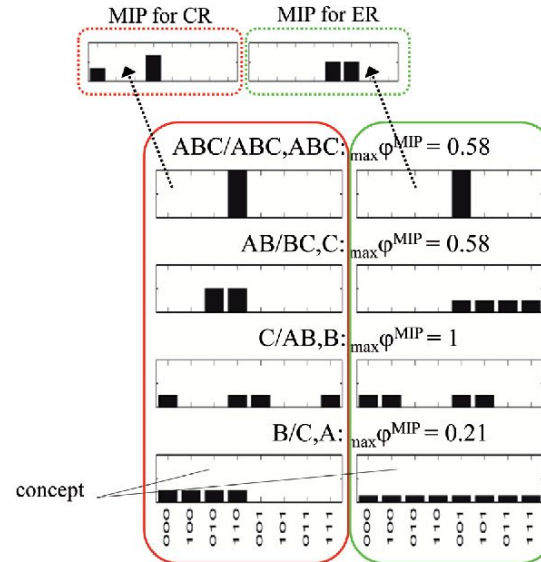
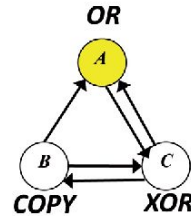


Integrated information theory

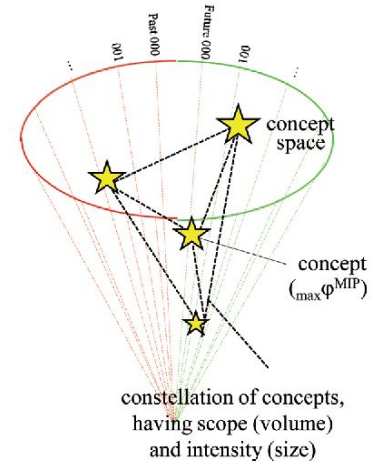
*A set of mechanisms
in their present state*



*The irreducible cause-effect
repertoires they generate (concepts)*

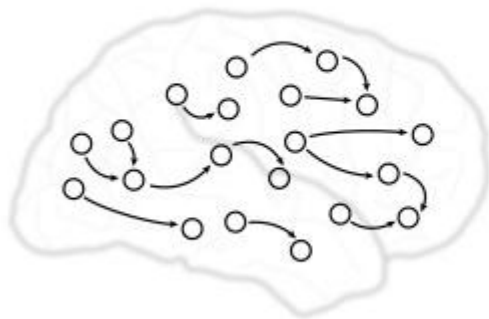


*The resulting
conceptual information structure
(a constellation of concepts
in concept space)*

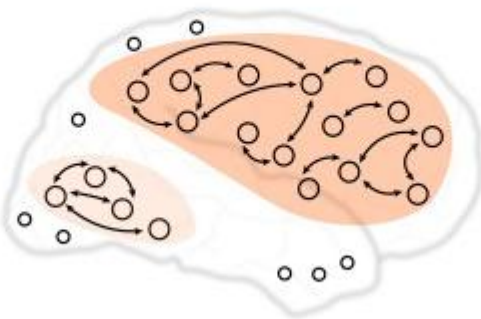


Integrated information theory

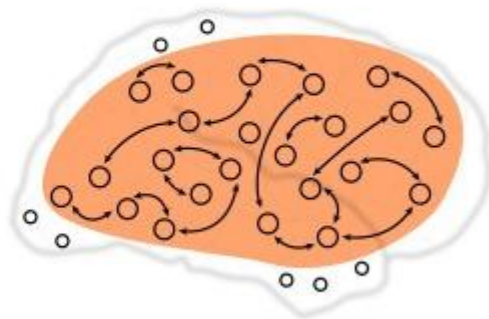
Unconscious integration



Minimally-conscious integration



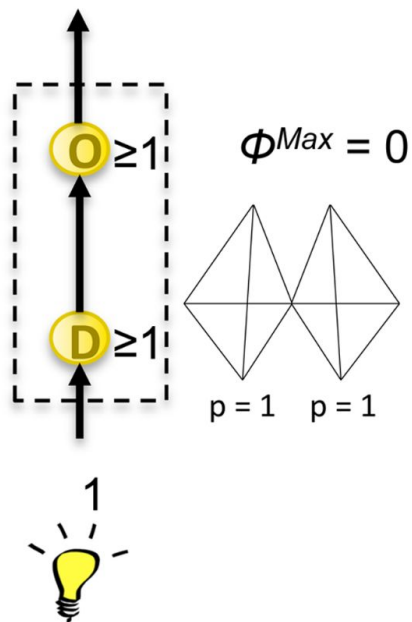
Conscious integration



TRENDS in Cognitive Sciences

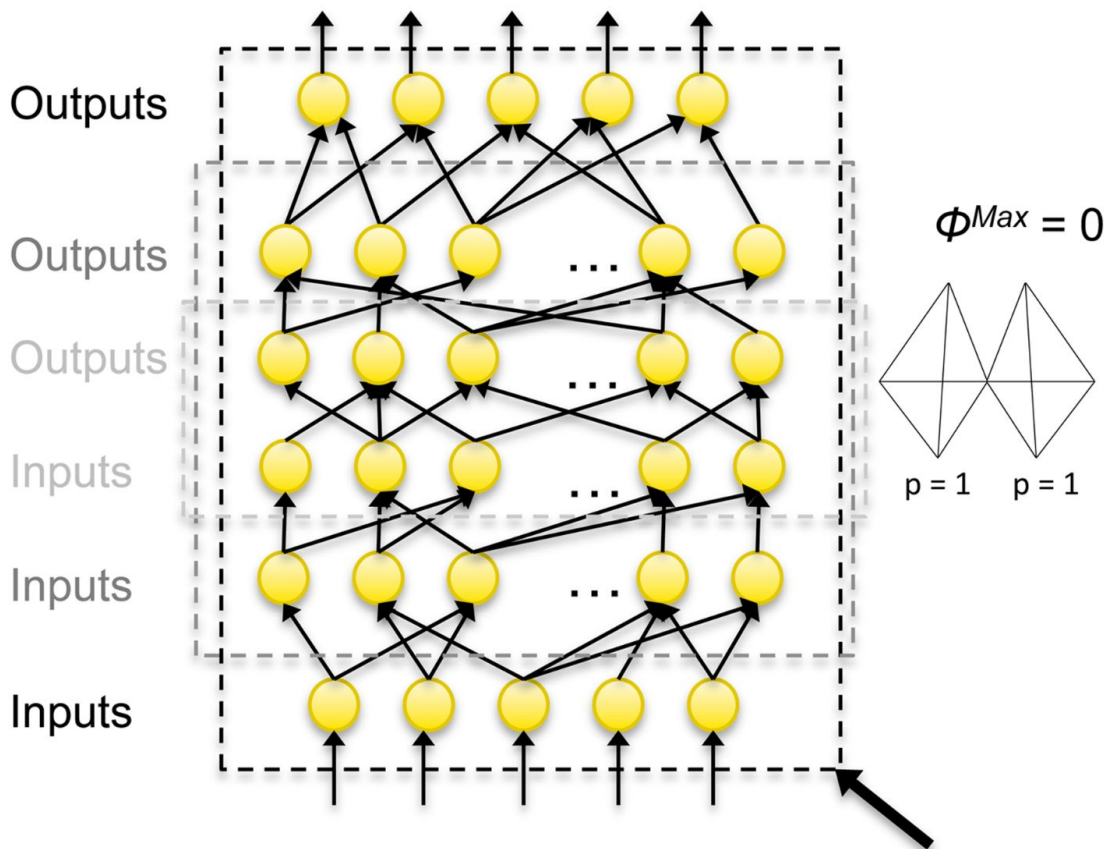
→
Increase of Φ

(A) Unconscious photodiode



(B)

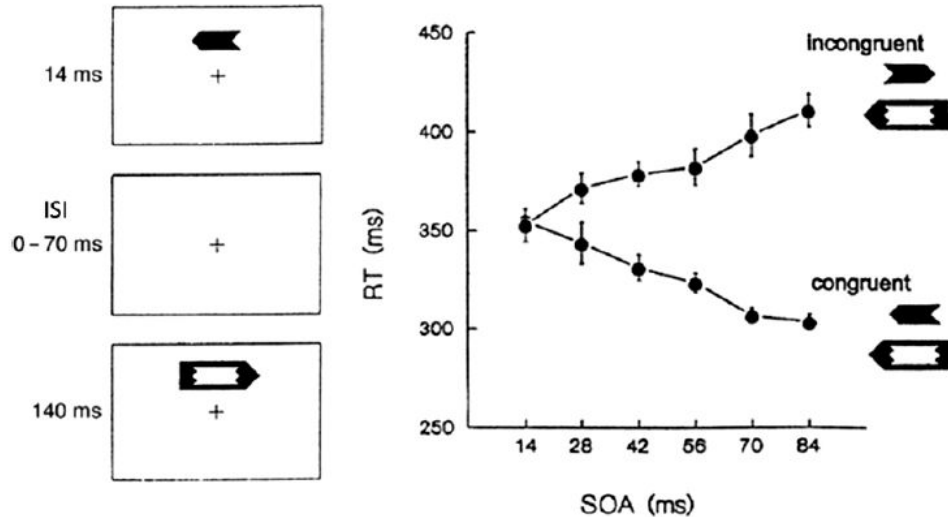
Feed-forward system



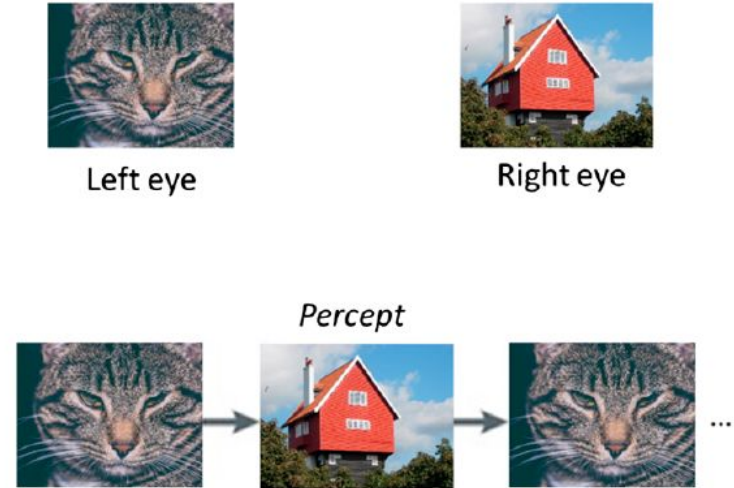
Extrinsic definition of a system

Examples of tests that are used to test theories of consciousness

a)



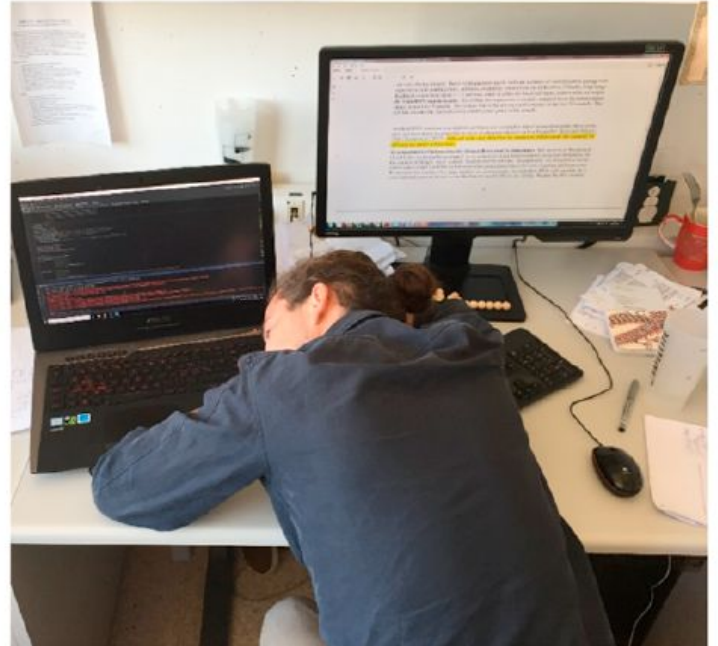
b)



d)

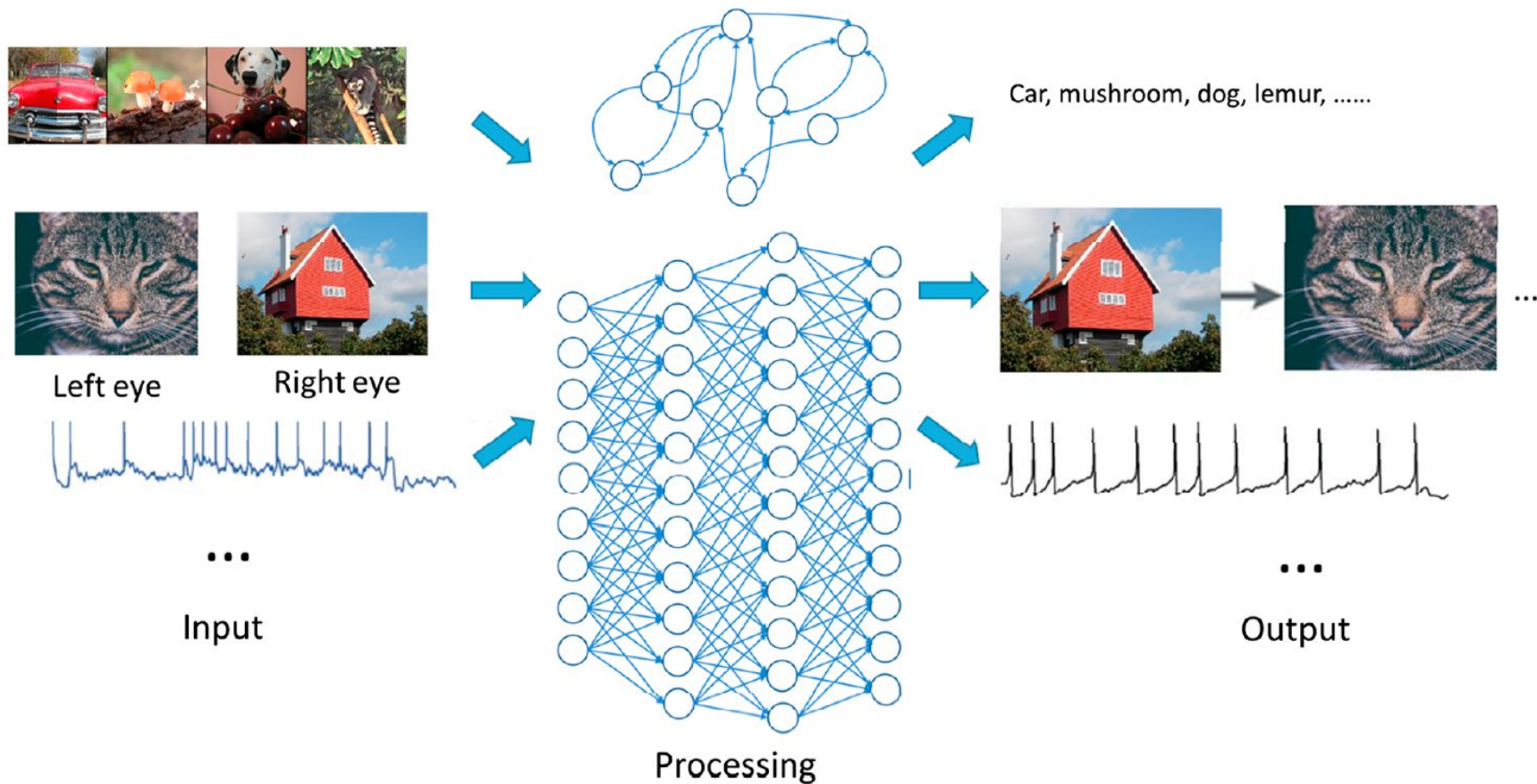


Awake human :
HIGH Φ



Human in dreamless sleep:
LOW Φ

Universal approximators & unfolding argument



The unfolding argument:

Feedforward neural networks are also universal function approximators (Fig. 2; Hornik, Stinchcombe, & White, 1989).

Recurrent neural networks are universal function approximators (Fig. 2; Schäfer & Zimmermann, 2006).

Therefore, for any recurrent network with a given input-output behaviour, there are corresponding feedforward networks with the same characteristics (although feedforward networks often need many more neurons than their recurrent counterparts).

Anything that can be done by recurrent networks can also be done in a feedforward manner.



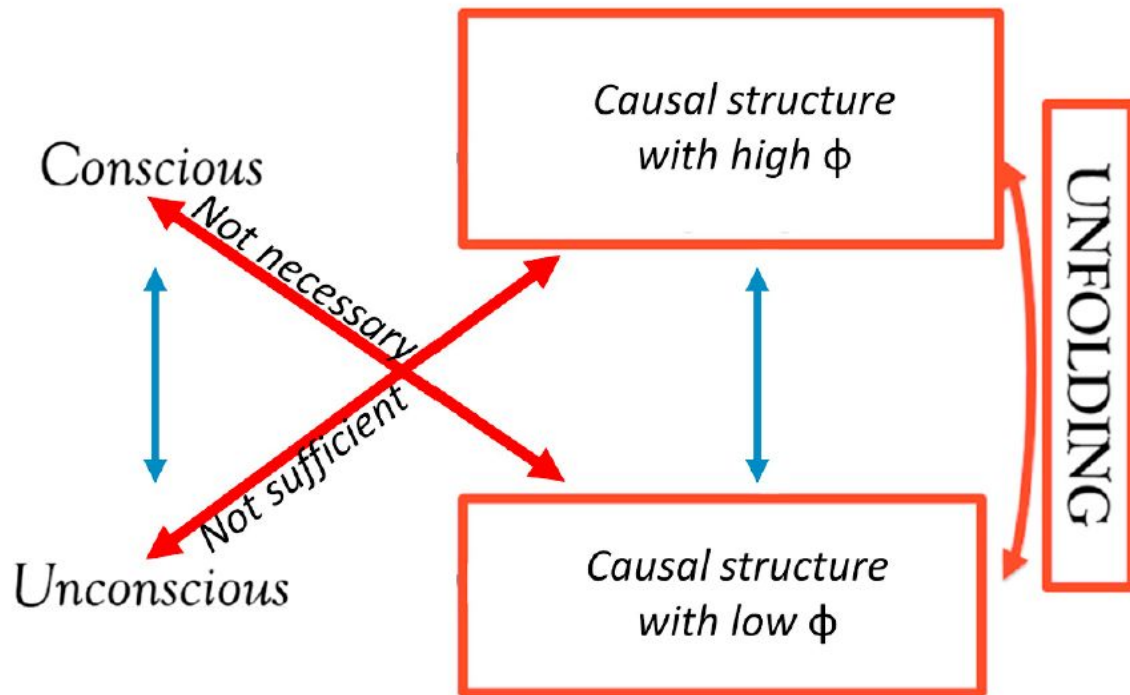
Wakeful

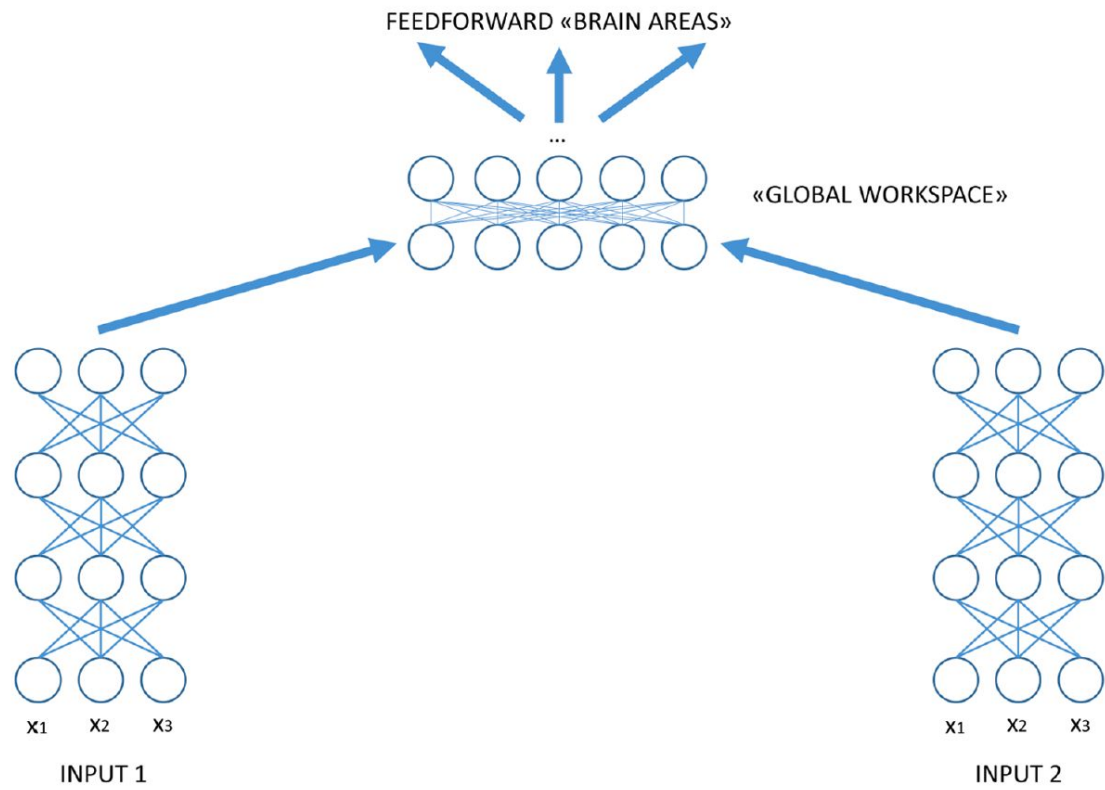
(Not masked)



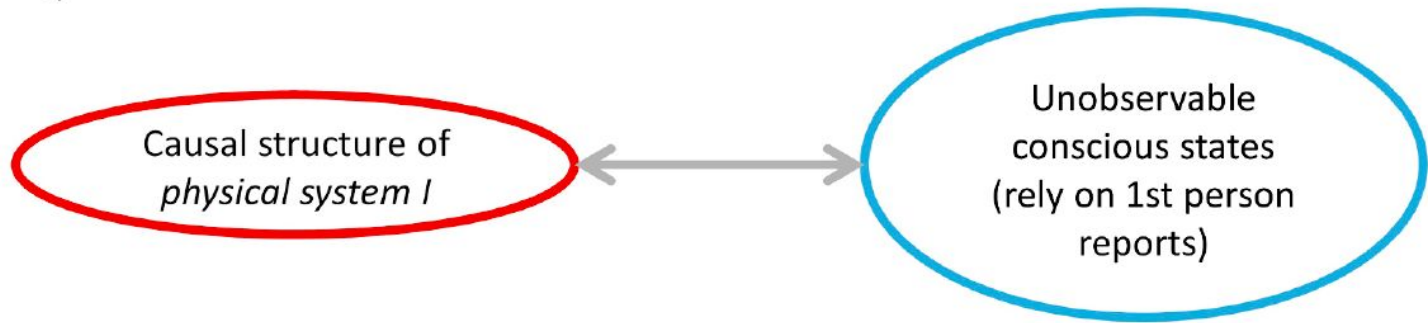
Asleep

(Masked)

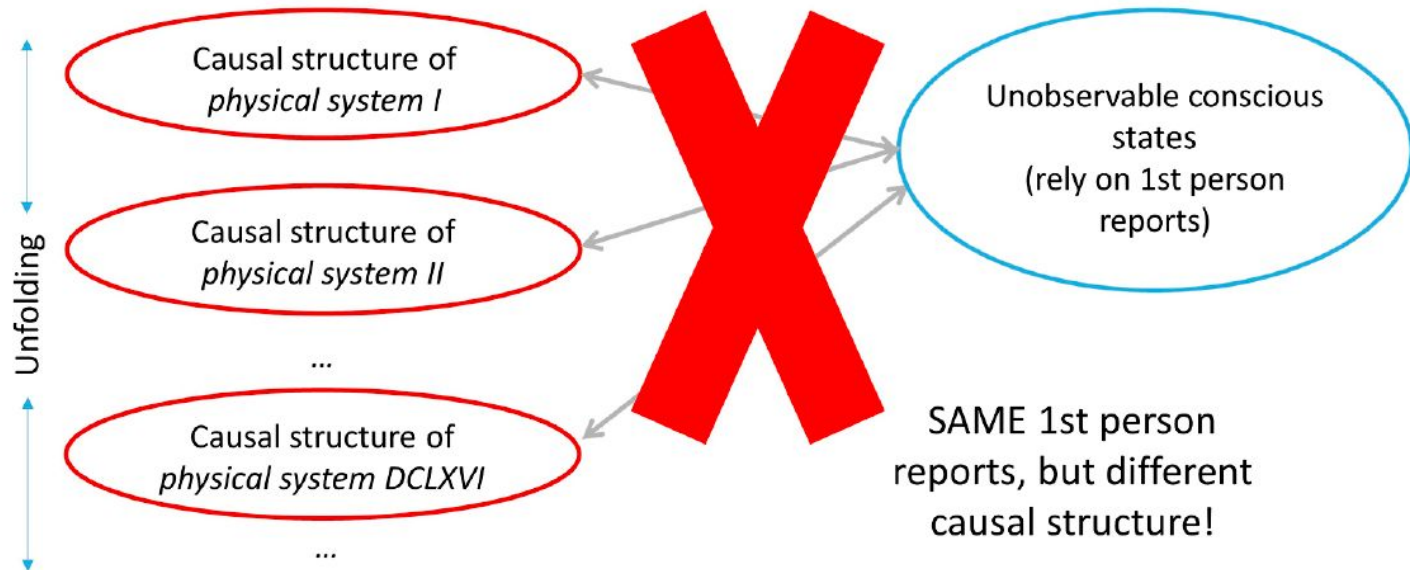




a)



b)



To summarize the unfolding argument, the conclusion follows from four premises.

(P1): In science we rely on physical measurements (based on subjective reports about consciousness).

(P2): For any recurrent system with a given input-output function, there exist feedforward systems with the same input-output function (and vice-versa).

(P3): Two systems that have identical input-output functions cannot be distinguished by any experiment that relies on a physical measurement (other than a measurement of brain activity itself or of other internal workings of the system).

(P4): We cannot use measures of brain activity as a-priori indicators of consciousness, because the brain basis of consciousness is what we are trying to understand in the first place.

(C): Therefore, EITHER causal structure theories are falsified (if they accept that unfolded, feedforward networks can be conscious),
OR causal structure theories are outside the realm of scientific inquiry (if they maintain that unfolded feedforward networks are not conscious despite being empirically indistinguishable from functionally equivalent recurrent networks).

Theories of consciousness

Biological:

- how did it appear in evolution?
- What genes are associated with?
- How does it appear in ontogeny?
- How does it function?
- What biological structures of the body is associated with?
- What animals have?

Physical:

- What is its nature?
- Quantum or not?
- What laws describe?
- Deterministic or not?

Psychological:

- How is the human psyche built?
- How to change settings?
- how to switch states?
- how to develop a stable mind?

Economic:

- How to manage people's decisions?

Mathematical:

- How to define consciousness through mathematics?
- How is consciousness related to information?
- What mathematical structures are related to consciousness?

Philosophy: no theories, just new problems and questions!

Why do we need
theories of consciousness at all?

The authors of theories
tell:

medical ethics
ethical treatment of
animals
ethical handling of AI
correct judgments

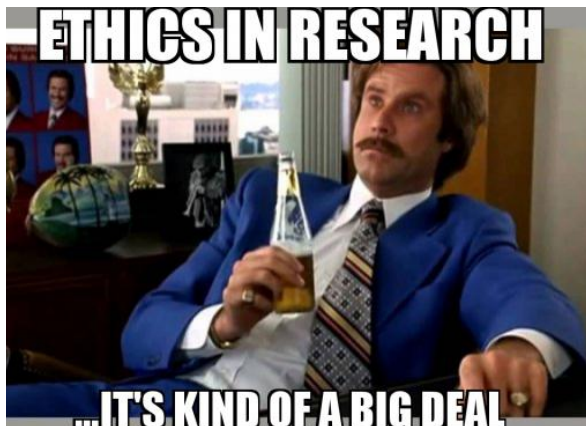
Writers and journalists
tell:



Why do scientists
actually create theories
of consciousness:

Fame
Fun
Grants

Motivation for naive students



Mind control
Upload
Conscious AI

