Cours 2025-2026:

Qu'est-ce que la conscience
et quels sont ses mécanismes cérébraux ?
What is consciousness, and what are its neuronal mechanisms?

Stanislas Dehaene
Chaire de Psychologie Cognitive Expérimentale

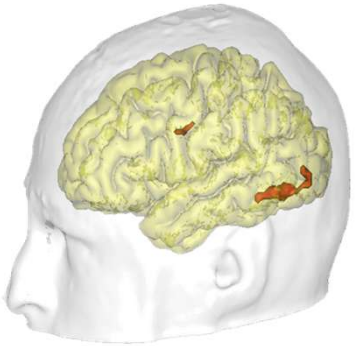
Cours n°1

Bref historique des recherches sur la conscience et modèle de l'espace de travail neuronal global

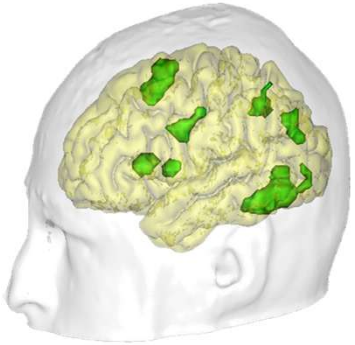
Historical outlook and main aspects of the global neuronal workspace model

28 years of Global Neuronal Workspace

Invisible words



Visible words




Lionel Naccache


Ghislaine Dehaene


Jean-Pierre Changeux


Lucie Charles


Sid Kouider


Darinka Trübutschek


Jérôme Sackur


Pieter Roeflsema


Bechir Jarraya


Jean-Rémi King

Sébastien Marti


Mariano Sigman

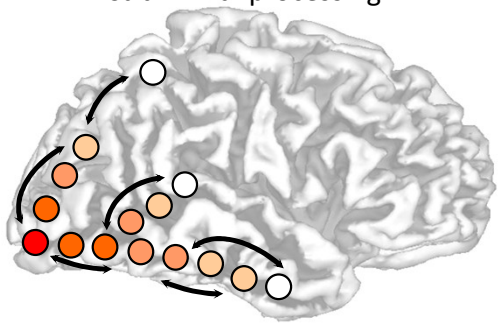

Claire Sergent


Aaron Schurger

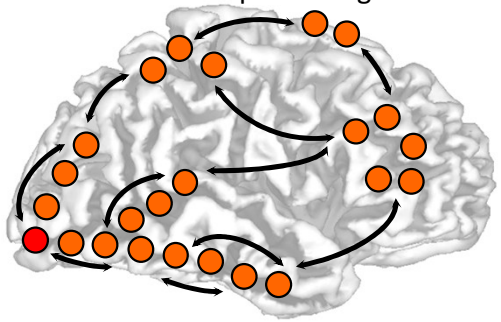

Jacobo Sitt


Moti Salti

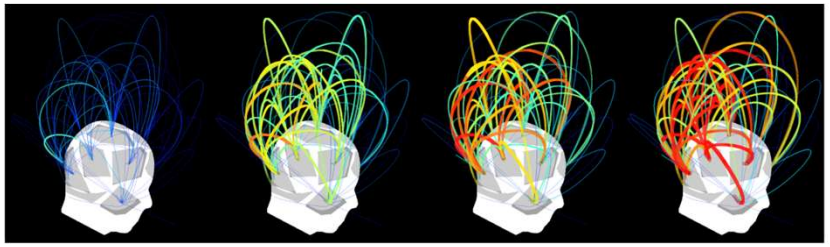
subliminal processing



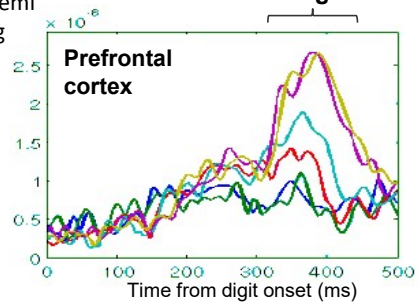
conscious processing



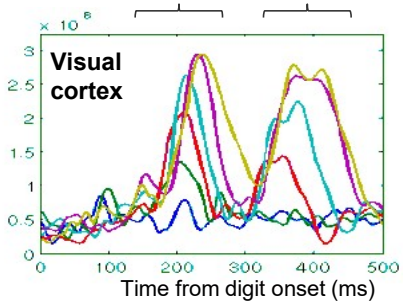
Symbolic mutual information



All-or-none ignition

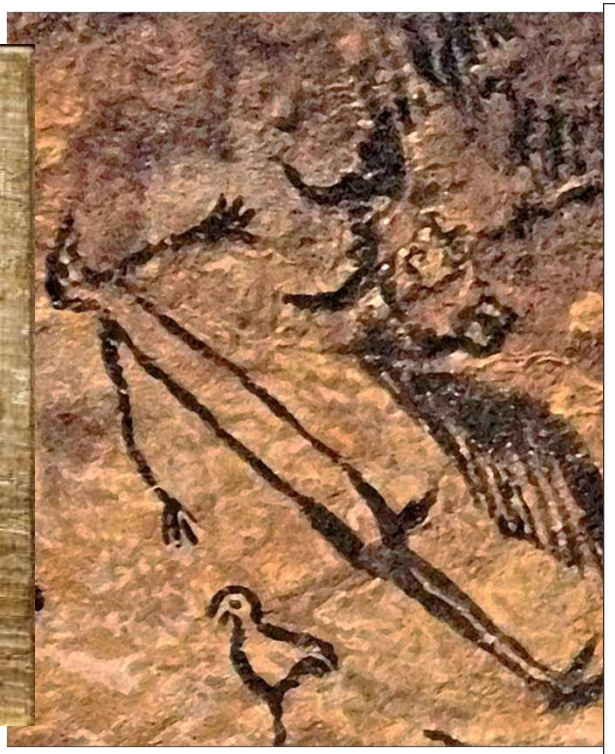


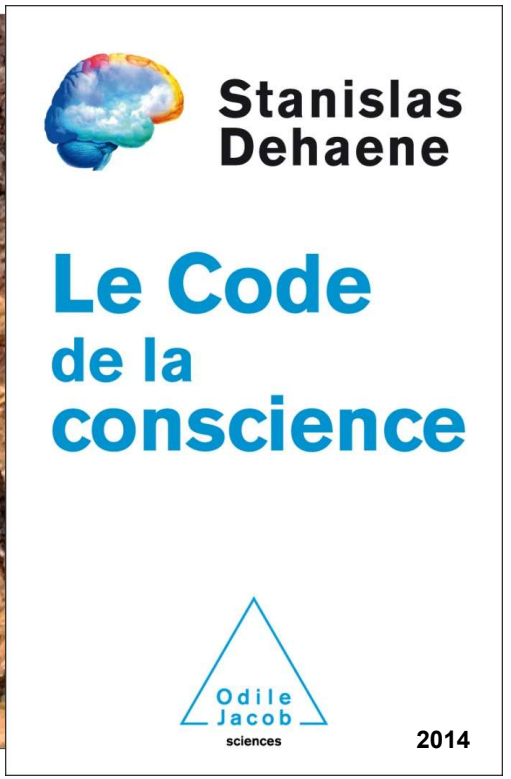
1. Linear accumulation **2. All-or-none ignition**



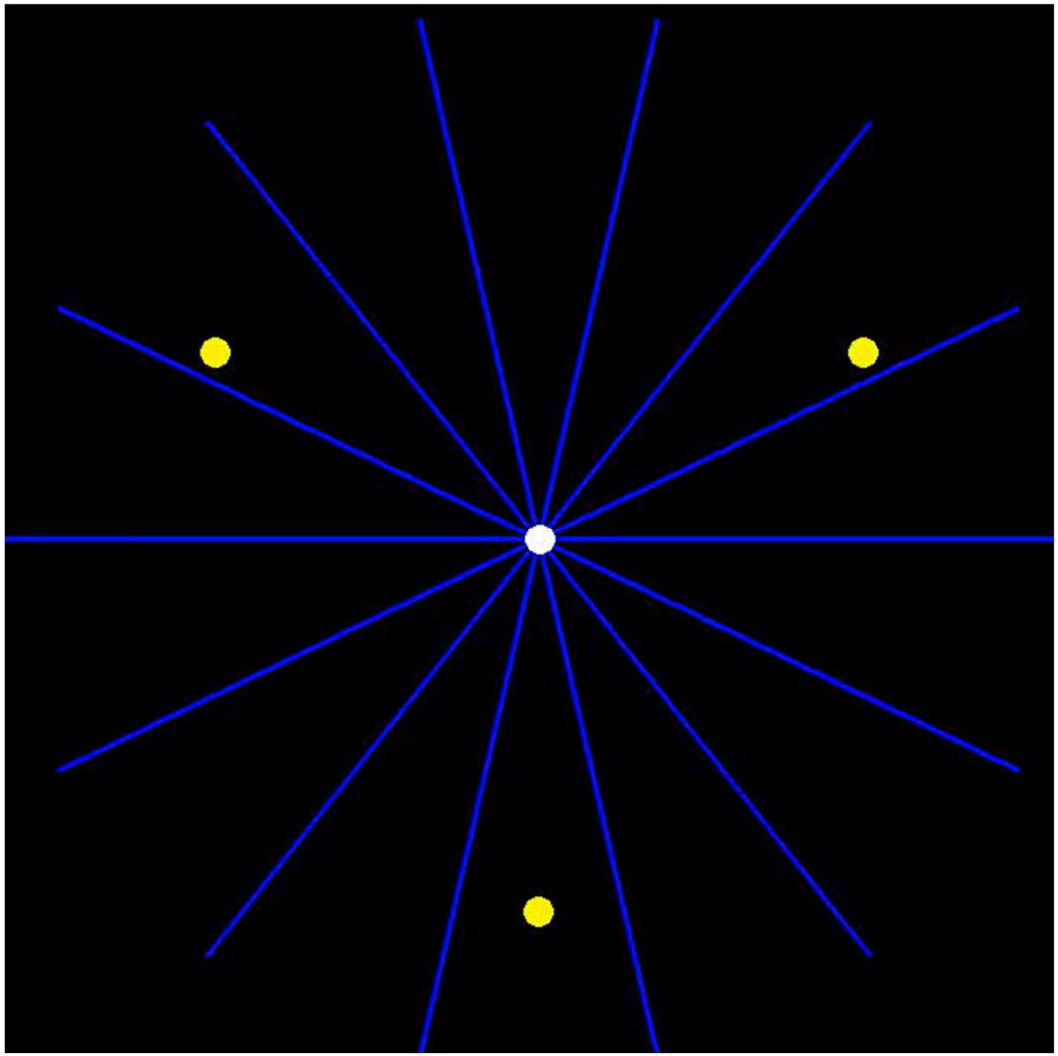
Is consciousness beyond the reach of science?

Consciousness is often seen as the **Holy Grail of neuroscience**, the hardest problem.
Isn't neuroscience limited to **objective** events?
Won't **subjective** conscious experience forever escape scientific investigation?
Our intuitions of the mind remain **dualist** : Doesn't the **soul** escape from the body, for instance during dreams?

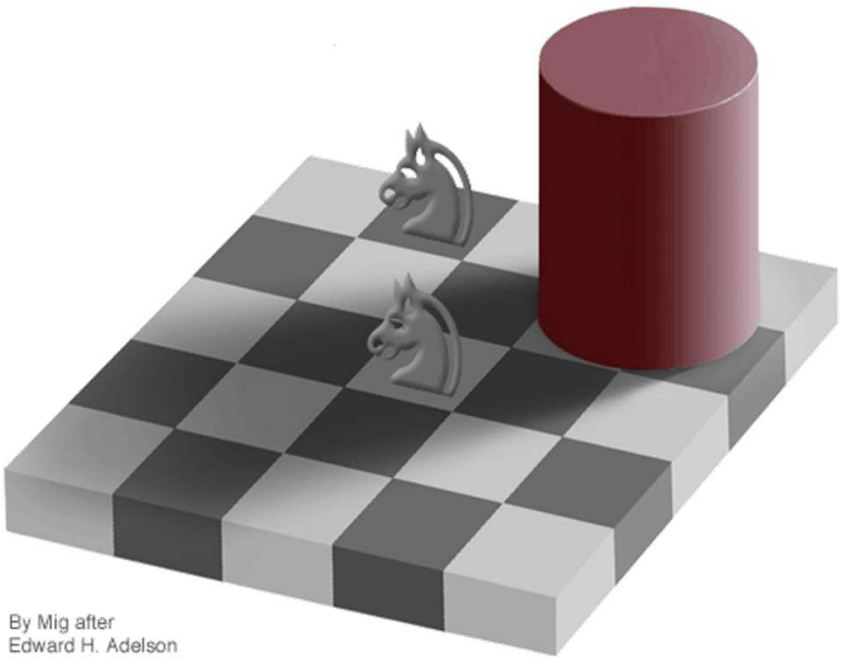




Visual illusions dissociate objective stimuli from subjective visual perception



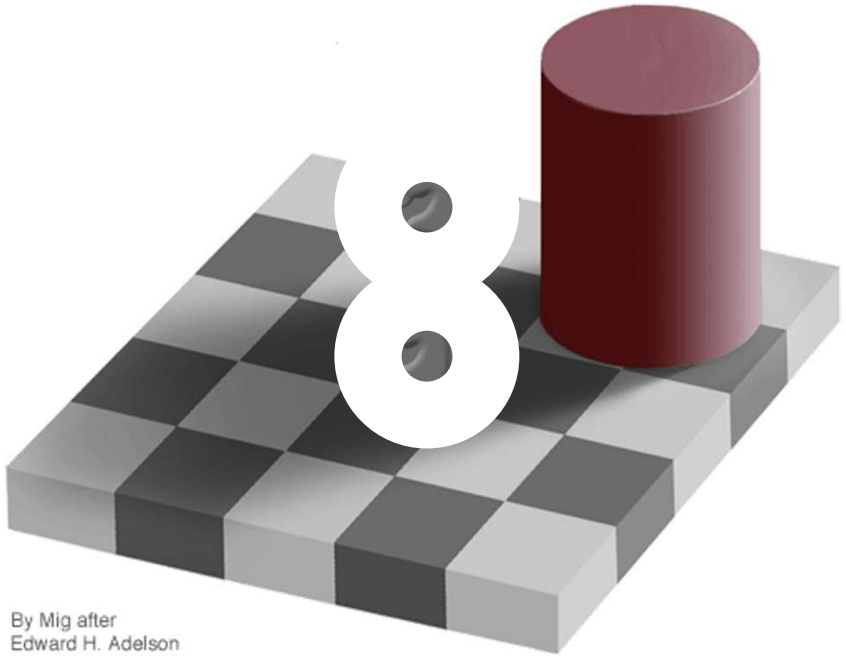
Visual illusions involve deep unconscious computations



By Mig after
Edward H. Adelson

Visual illusions involve deep unconscious computations

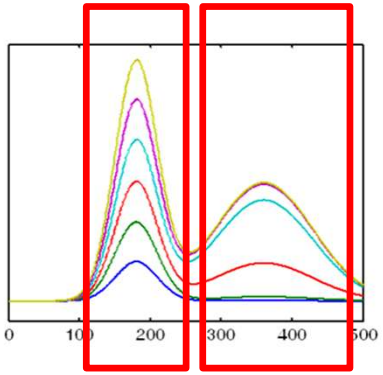
- Hermann von Helmholtz: The visual system operates by « unconscious inference »
- Conscious representations are not shallow percepts, but high-level representations that **integrate** incoming information into an **internal model of the world**

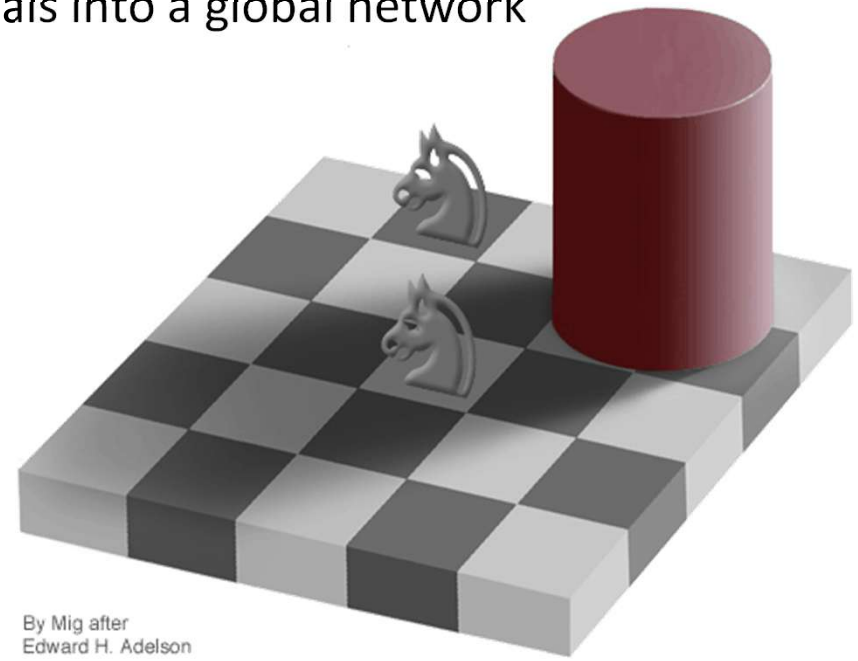


Neural implementation: Two successive waves

1. Early perceptual processing is **unconscious, parallel and continuous**

2. Conscious perception relates to a **late wave of non-linear ignition** that integrates incoming signals into a global network





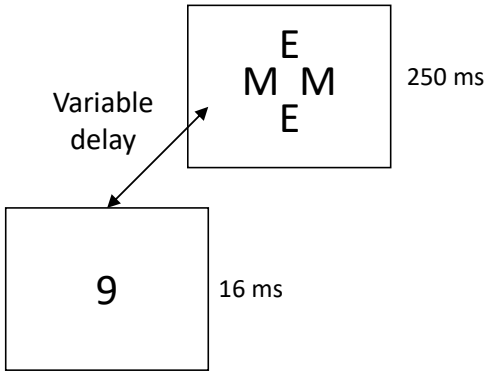
Key ingredients in the modern study of consciousness

Functional MRI

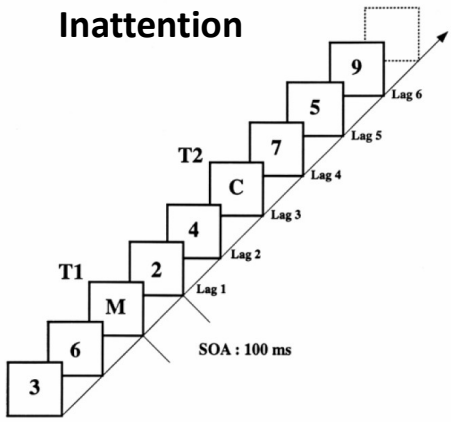
Event-related potentials

1. Minimal experimental paradigms to manipulate conscious perception

Masking

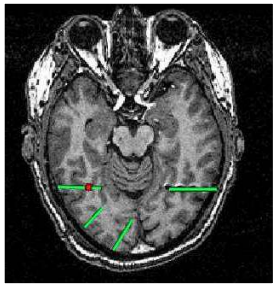


Inattention



Magnetoencephalography

Intracranial recordings



2. Precise delimitation of the polysemic term "**consciousness".**

This course will focus on **conscious access to a reportable content** (transitive meaning: consciousness of x).

3. Careful quantification of introspection

Introspection defines the very phenomenon that we want to study (subjectivity) and must be recorded.

We can present stimuli at threshold and **sort the trials by subjective reports.**

4. Objective brain imaging methods can then reveal the fate of conscious versus nonconscious information.

5. Guidance by a **theory of conscious access and reportability**

Several precursors

Many scientists demonstrated the depth of non-conscious processing. E.g.

- Larry Weiskrantz and blindsight
- Edoardo Bisiach and spatial neglect

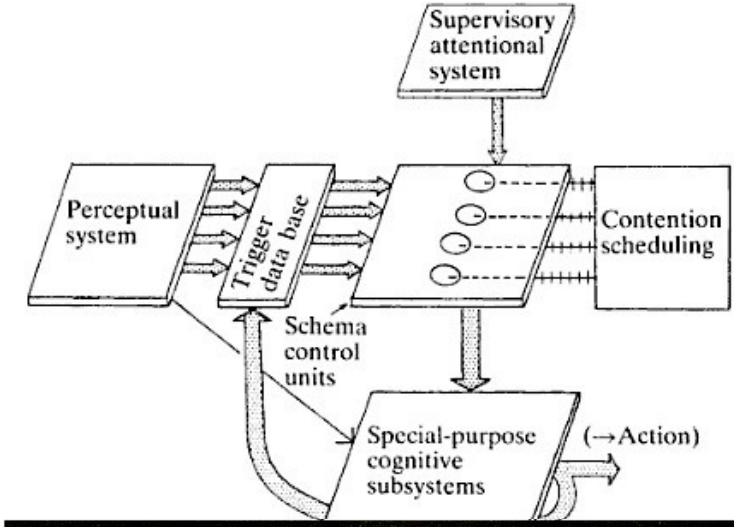
Others proposed models of the place of consciousness in cognitive architecture:

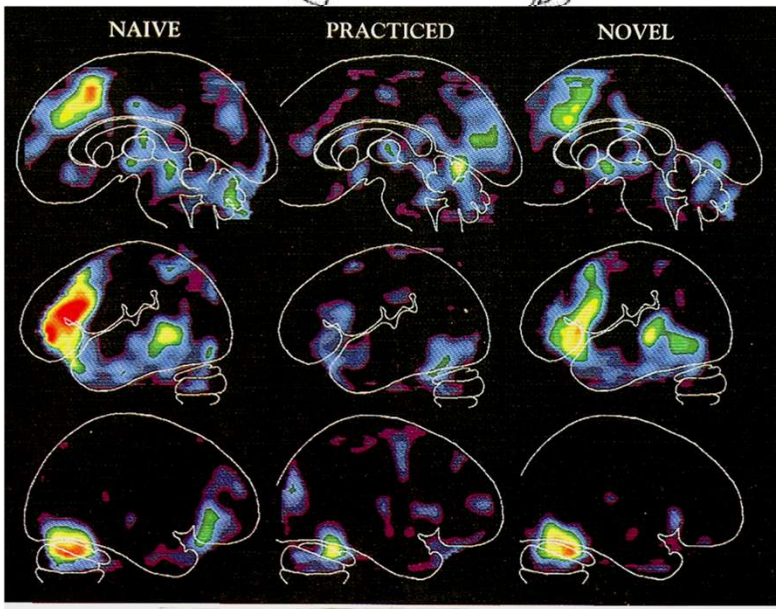
- William James (1890):
  - The need for a selective system to orient scarce resources. "consciousness is at all times **primarily a selecting agency**"
  - "Consciousness […] is only intense when nerve-processes are hesitant. In rapid, automatic, habitual action it sinks to a minimum."
- Tim Shallice (1978, Norman & Shallice, 1980): a "**supervisory attentional system**", **hierarchically higher** to "special-purpose cognitive subsystems", responsible for voluntary behavior through inhibitory control.
- Michael Posner : a distinction between automatic processes (unintentional, non-interfering with others, outside of consciousness) and controlled ones (Posner & Klein, 1973; Posner & Snyder, 1975)

Selective attention is the "gateway to consciousness"

With Marc Raichle: first brain-imaging studies showing a strong initial activation of **prefrontal and cingulate cortex**, disappearing after automatization.

- Crick & Koch : many successive models, raising the question of why visual processing is not necessarily accompanied by conscious perception.





M. Raichle, J. Fiez et al., *Cerebral Cortex*, 1994

The global workspace model

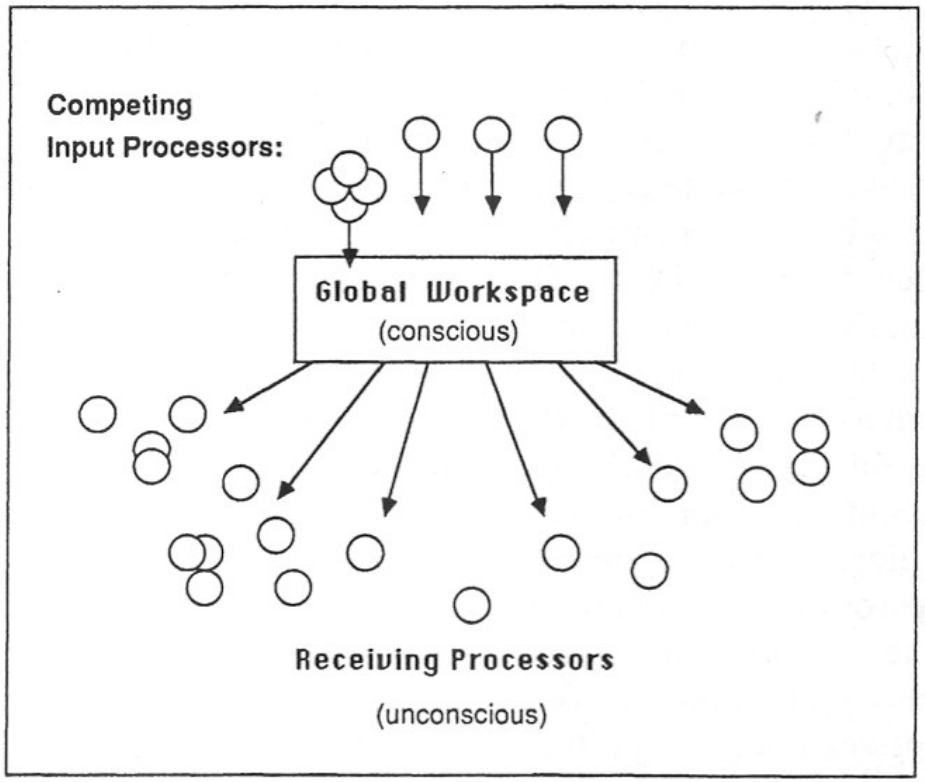
Bernard Baars, *A cognitive theory of consciousness*, 1989

Perception and cognition are supported by a great variety of **specialized processors** that operate in parallel (e.g. face recognition, spatial orientation…). Cf. Fodor's Modularity of mind.

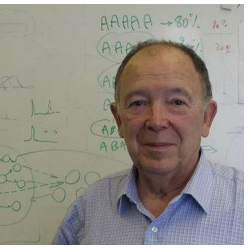
At any given moment, only a single **coalition of processors** sends its outputs to the **conscious global workspace**: a capacity-limited system that **broadcasts** information to all other processors.

Table 2.1. *Capabilities of comparable conscious and unconscious processes*

	Capabilities of conscious processes		Capabilities of unconscious processors
1	Computationally inefficient: High number of errors, low speed, and mutual interference between conscious computations	1	Highly efficient in their own tasks: Low number of errors, high speed, and little mutual interference
2	Great range of different contents over time; great ability to relate different conscious contents to each other; great ability to relate conscious events to their unconscious contexts	2	Each specialized processor has limited range over time; each one is relatively isolated and autonomous
3	Have internal consistency, seriality, and limited capacity	3	Diverse, can operate in parallel, and together have great capacity



Model 1. A global workspace in a distributed system.



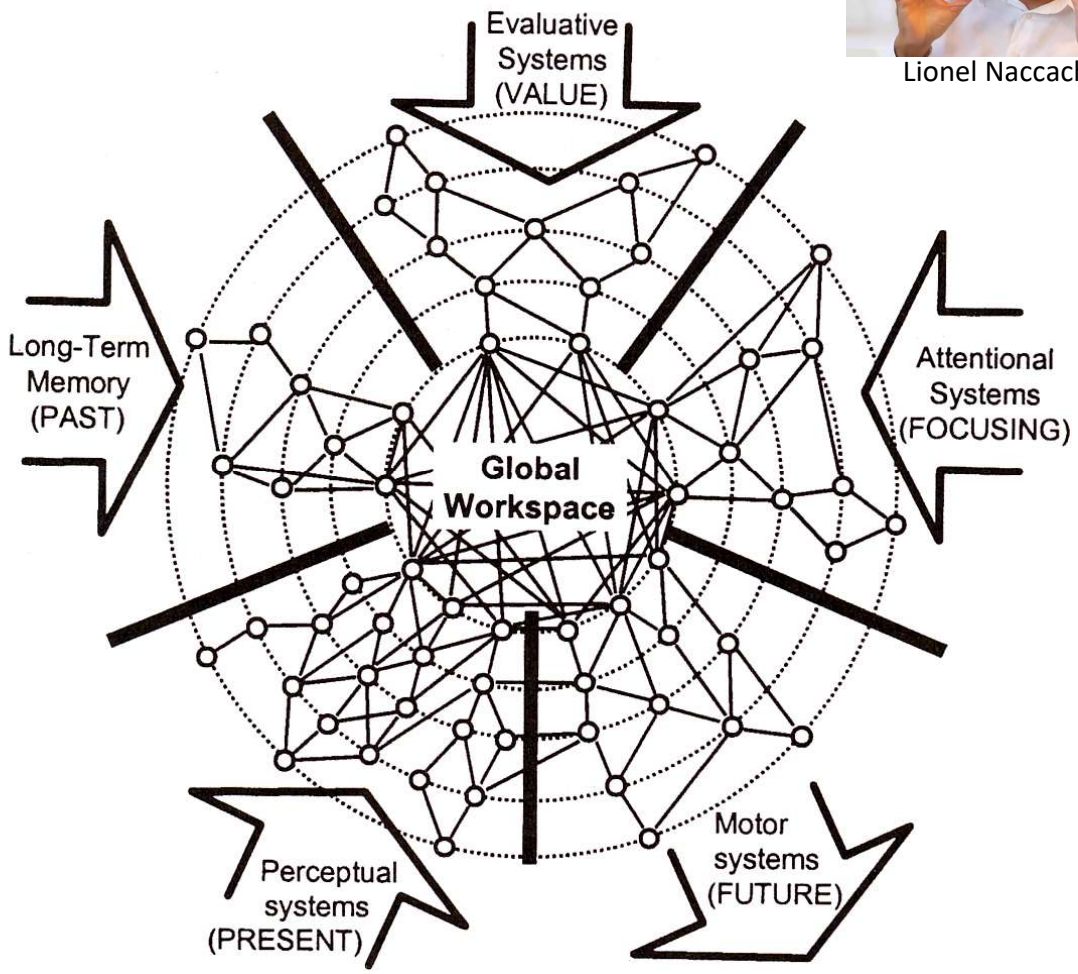
Jean-Pierre Changeux

The global neuronal workspace (GNW) hypothesis

Dehaene & Changeux, PNAS 1998 ; Dehaene & Naccache, Cognition 2001

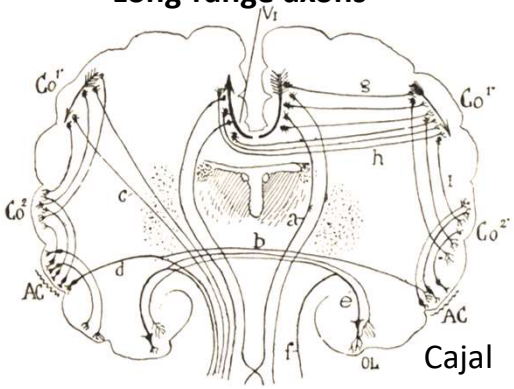
Lionel Naccache

- Virtually every automatized chain of cortical processors can **operate unconsciously**.
- What we subjectively experience as consciousness is the **global availability of information**
- Consciousness relates to the activity of a **global neuronal workspace** that evolved to **select** and **broadcast**, in a brain-wide manner, a relevant piece of information, allowing it to be reported.
- Conscious access corresponds to the "**ignition**" of a **subset** of workspace neurons, **distributed** in prefrontal and other associative cortices, whose topology (as a vector) defines the conscious contents, and that send **top-down signals** back to all processors.
- "This **global availability of information** through the workspace **is** what we **subjectively experience** as a conscious state."



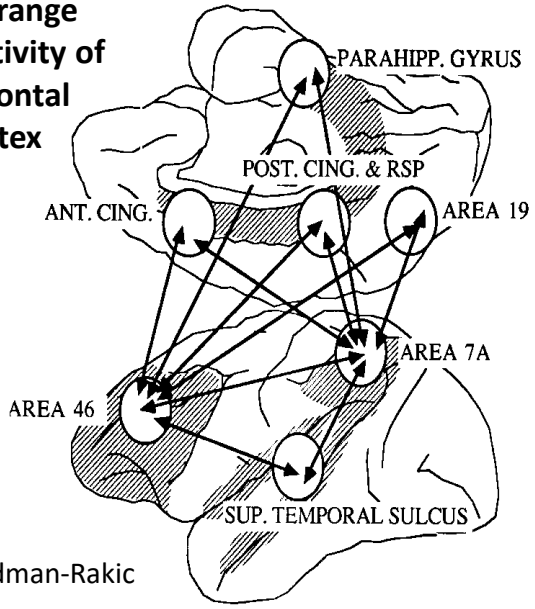
Consciousness rests on an evolved cortical long-distance communication network

Long-range axons



Cajal

Long-range connectivity of prefrontal cortex

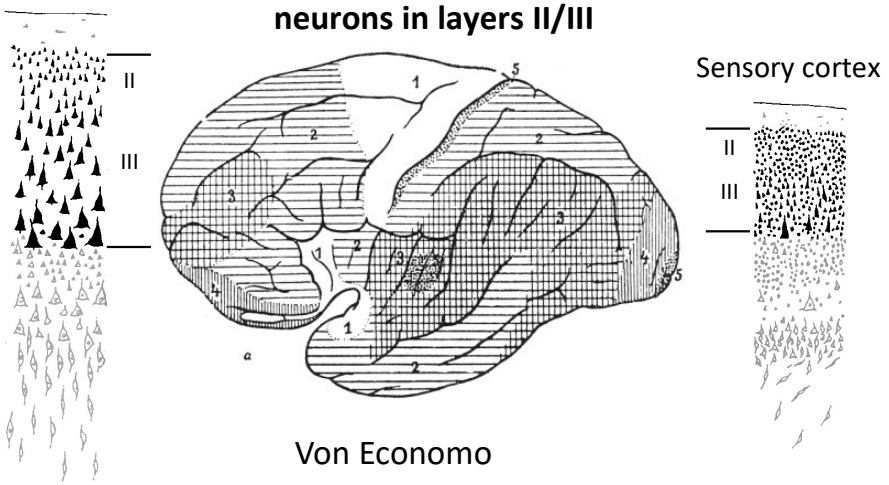


Goldman-Rakic

Prefrontal cortex

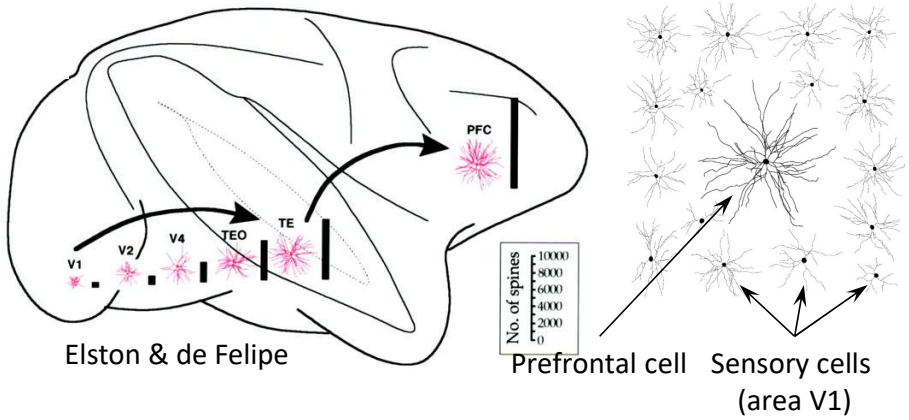
Large pyramidal neurons in layers II/III

Sensory cortex



Von Economo

Enlarged dendritic trees



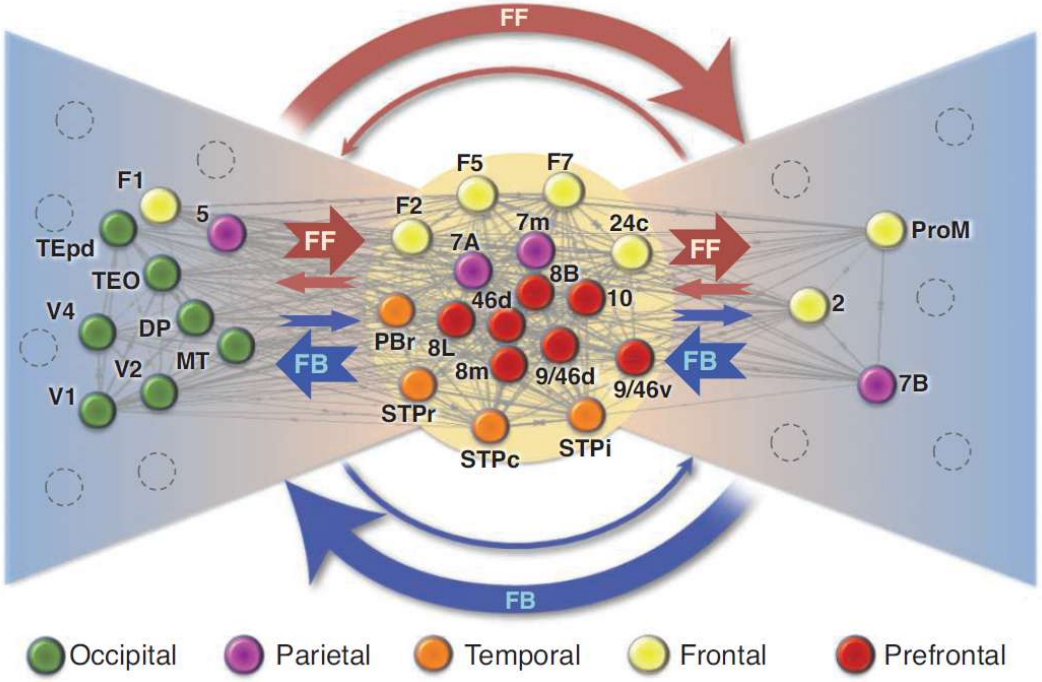
Elston & de Felipe

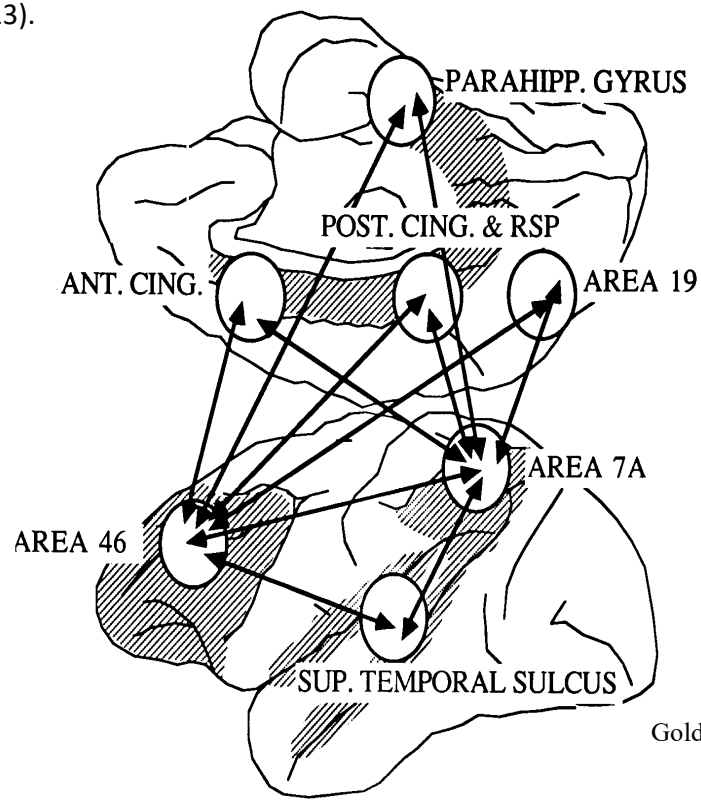
Prefrontal cell

Sensory cells (area V1)

An anatomical cortical “rich club” of highly interconnected areas

Markov, N. T., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., Toroczkai, Z., & Kennedy, H. (2013). Cortical high-density counterstream architectures. *Science (New York, N.Y.)*, *342*(6158), 1238406.



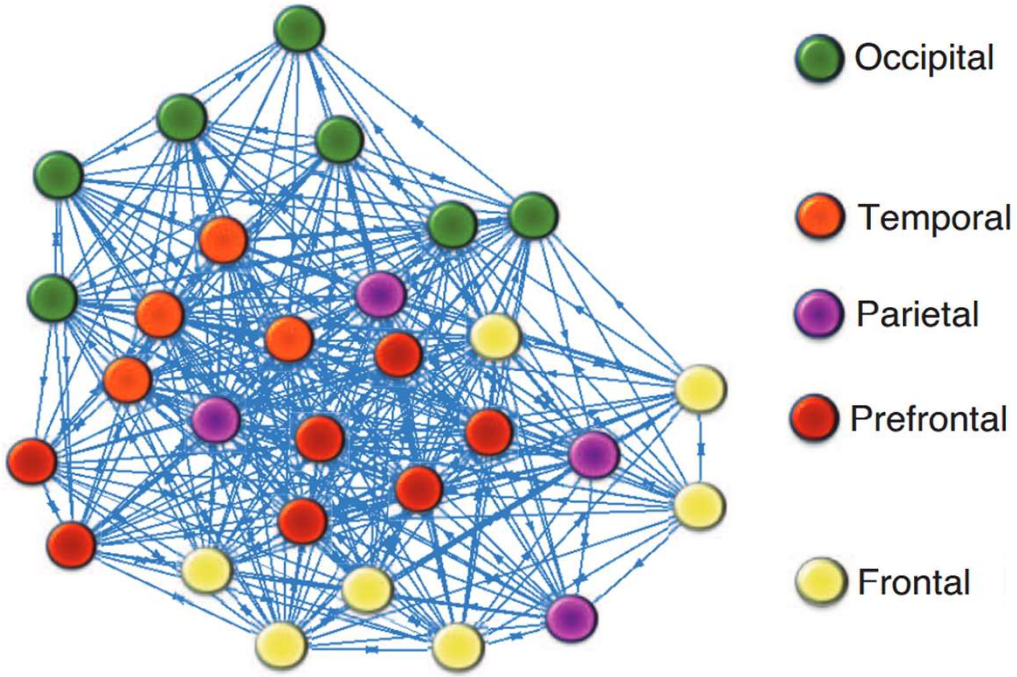


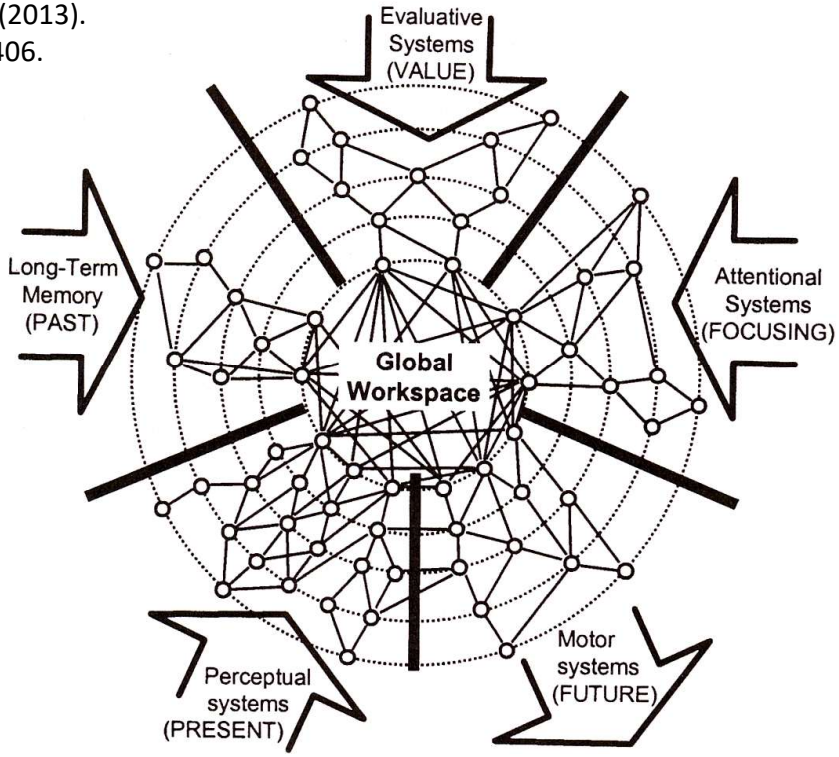
Goldman-Rakic

A systematic review of decades of tracer studies in the macaque brain.
The resulting graph exhibits a hierarchical structure of areas based on their feedforward and feedback connections.
Short-distance connections are the most dense, but there is also a sparse set of long-distance connections (66% of area pairs).
Identification of a very high density **core** (92% connected) of frontal, parietal and superior temporal areas, linked by bi-directional connections.

An anatomical cortical “rich club” of highly interconnected areas

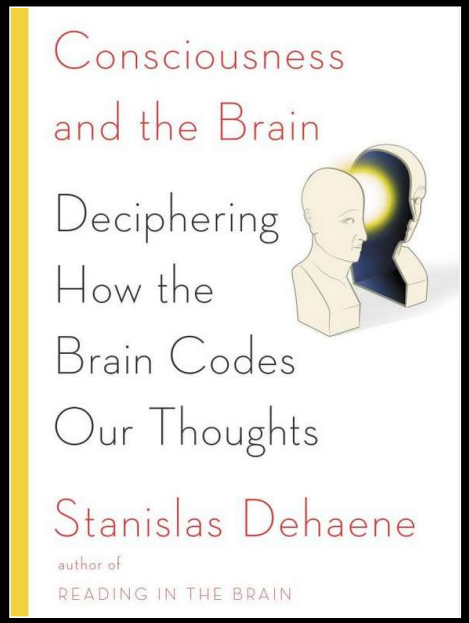
Markov, N. T., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., Toroczkai, Z., & Kennedy, H. (2013). Cortical high-density counterstream architectures. *Science (New York, N.Y.)*, *342*(6158), 1238406.





Another representation of the same connectivity matrix, with optimal placement of areas according to their connectivity. The core of prefrontal, parietal and temporal areas is clearly visible.

What is specific to human consciousness ?



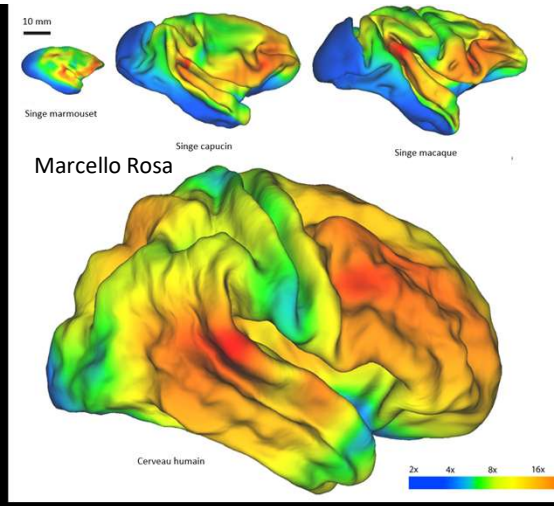
1. The basic mechanisms of conscious access are similar in human and non-human animals.

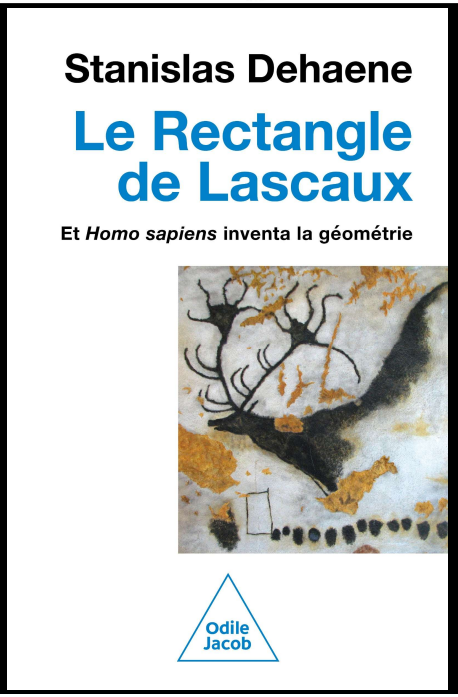
→ We and others have identified **signatures of consciousness,** shared between animals and humans, and which can now be used to detect residual consciousness in brain-lesioned patients.

→ In agreement with the **Global Neuronal Workspace framework**, the prefrontal cortex (PFC) contains a detailed neural code for the current conscious contents.

2. The **contents of consciousness are richer in humans**.

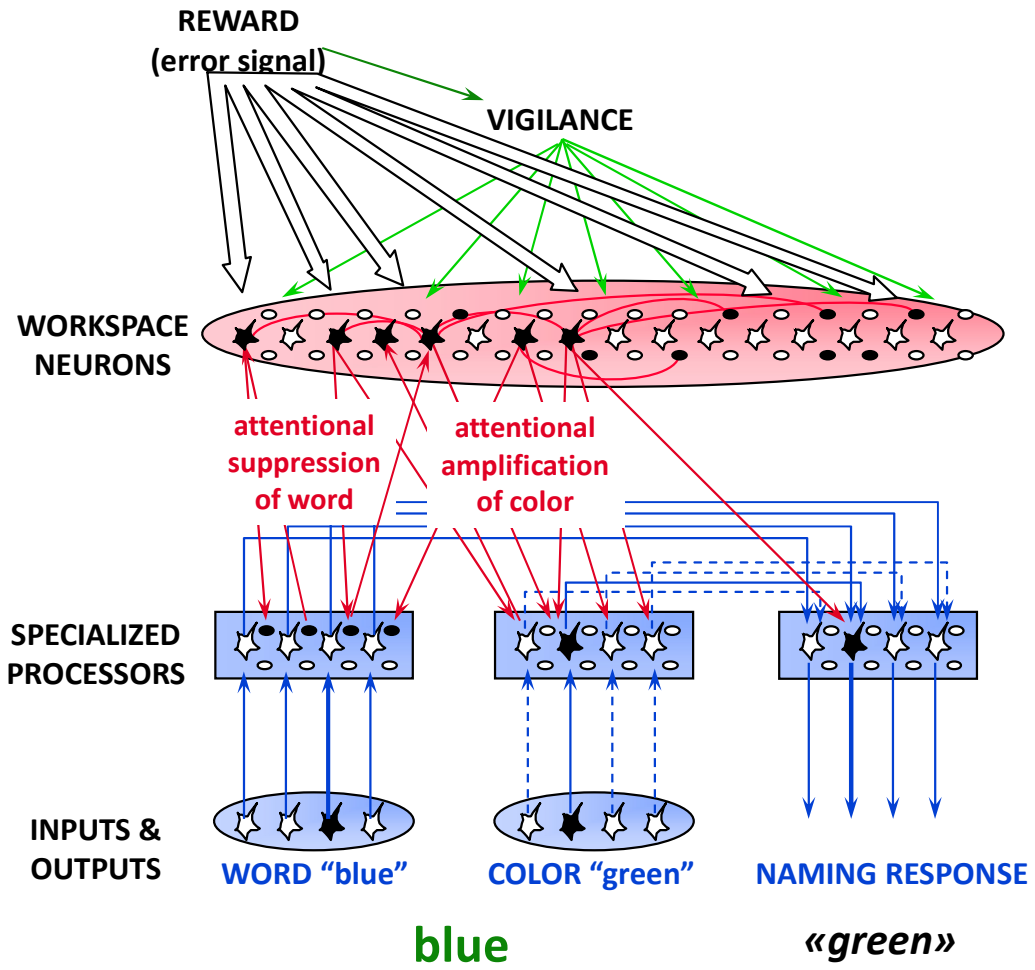
→ Only humans possess **symbols and recursive compositionality,** leading to unique competences for language, mathematics, music, and the conscious representation of ourselves and others.



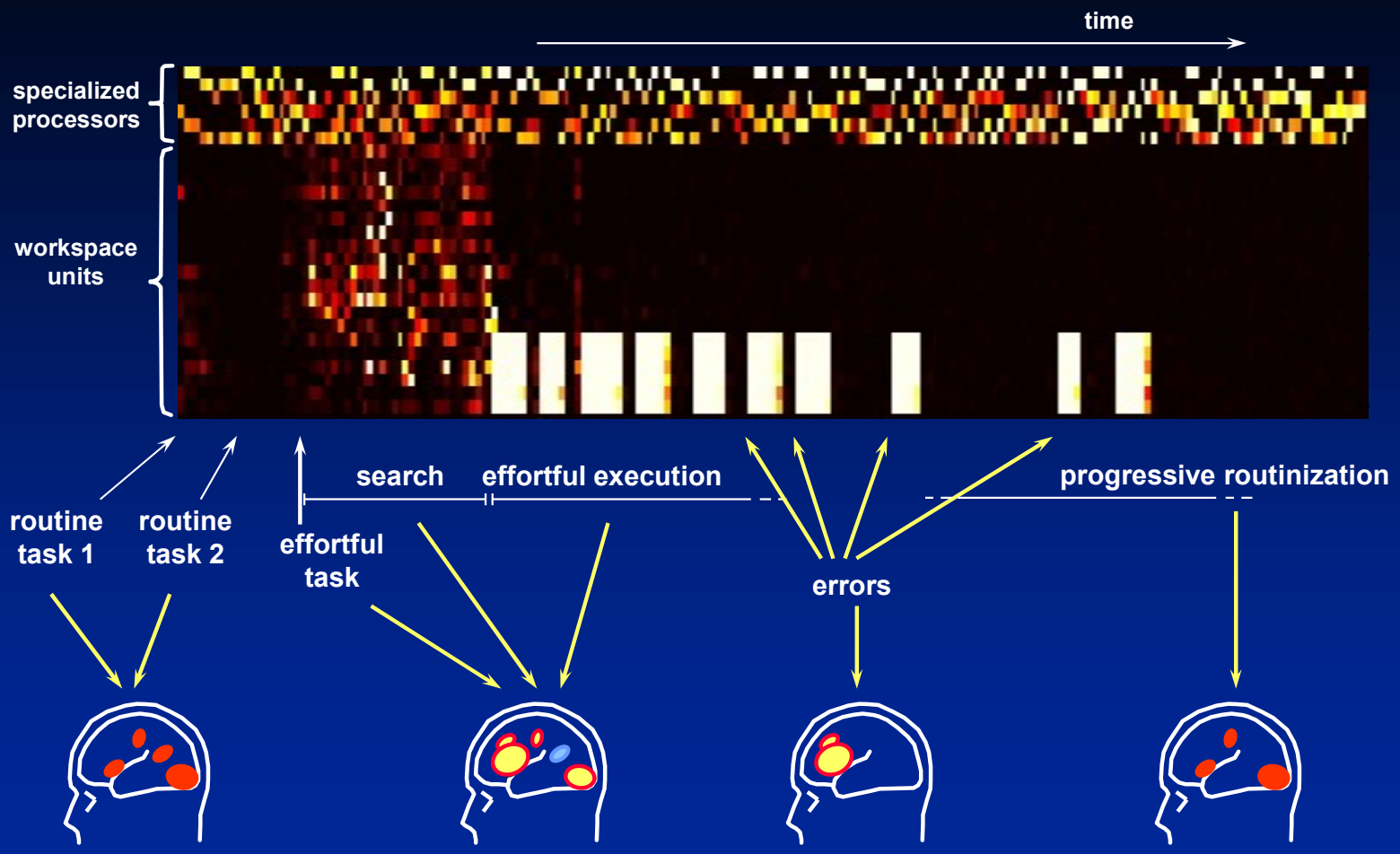


Elementary computer simulation of the neuronal workspace during the Stroop Test

Dehaene, Kerszberg & Changeux, *PNAS,* 1998



TEMPORAL DYNAMICS OF WORKSPACE ACTIVITY



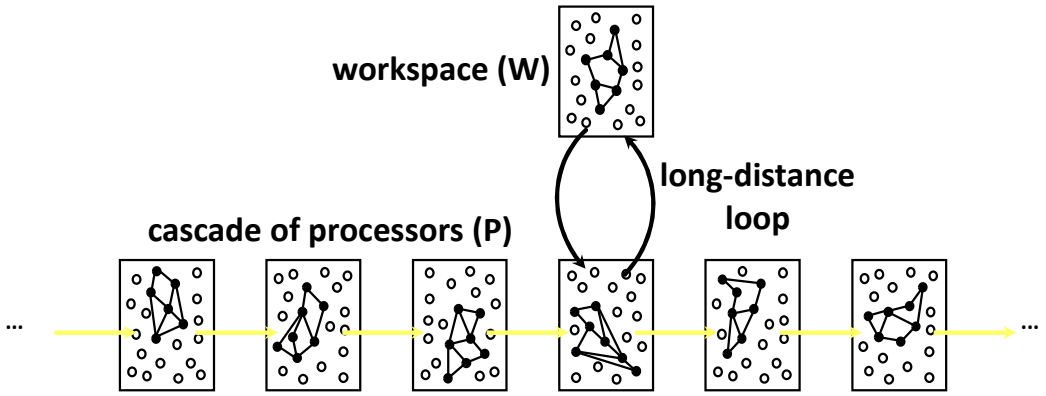
Dehaene, Kerszberg & Changeux, *PNAS,* 1998

First simulations of subliminal versus supraliminal processing

Dehaene & Naccache, *Cognition,* 2001

« ignition » :

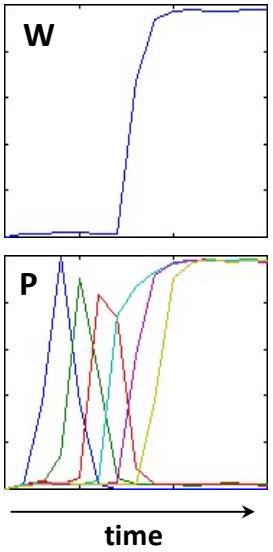
- The activation of workspace neurons is sustained for an arbitrary duration.
- The activity of the corresponding processors is also sustained, due to bidirectional bottom-up and top-down connectivity ("mobilization" of processors)



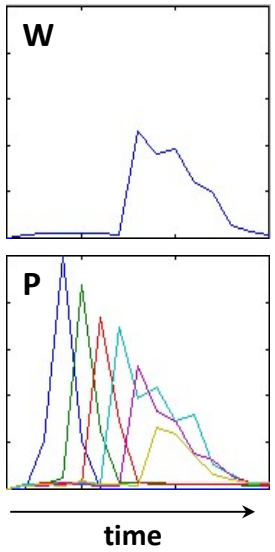
« failed ignition » :

A transient activity of workspace neurons can occur in the absence of consciousness.

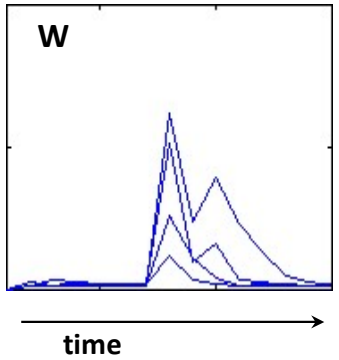
supra-threshold stimulus



sub-threshold stimulus



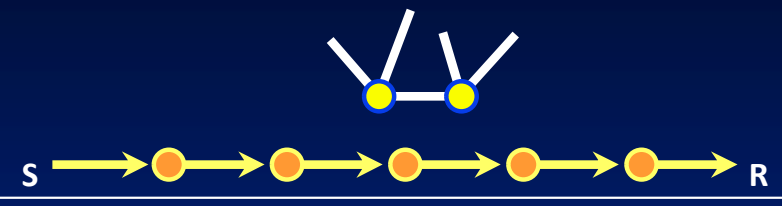
various levels of transient workspace activity



Prediction : Conscious intentions can influence unconscious processing

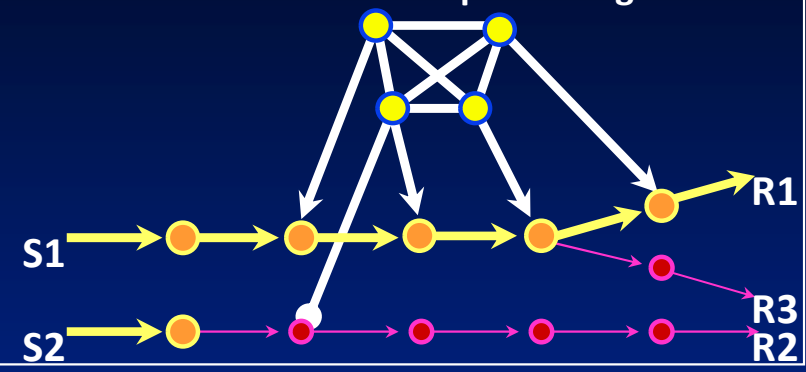
Dehaene & Naccache, *Cognition,* 2001

A. Non-conscious processing stream executing without conscious control

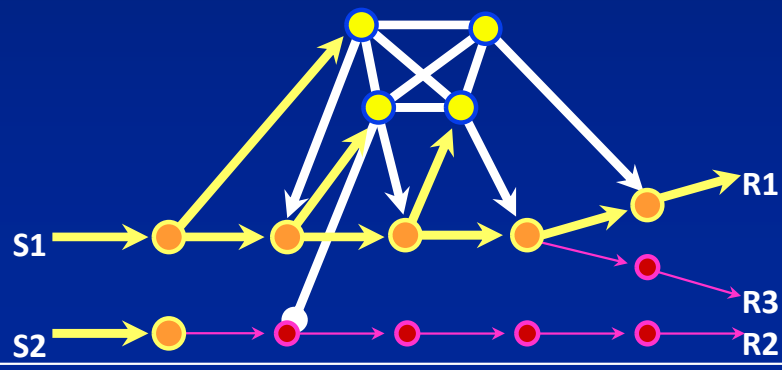


Many experimental proofs. E.g Blindsight ; or Dehaene et al., Nature 1998 : brain-imaging evidence for a complete sensori-motor chain.

B. Influence of a conscious context or instruction on unconscious processing



C. IMPOSSIBLE SITUATION: Top-down control by and to the same unconscious stimulus



New idea ! Tested by Naccache, Blandin & Dehaene (2002): even subliminal priming can be affected by top-down attention. Non-conscious processing is NOT synonymous with automatic and independent of attention.

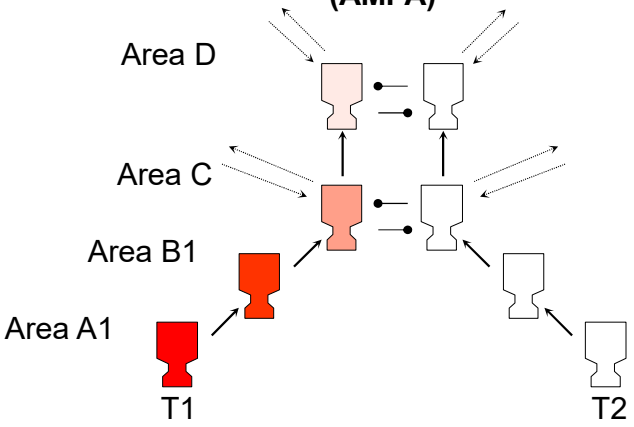
C is more controversial:

- Non-conscious stimuli can attract attention and affect the processing of other *conscious* stimuli
- Simon Van Gaal has shown that non-conscious stimuli can prime conflict adaptation, inhibitory control (no-go), task-switching…

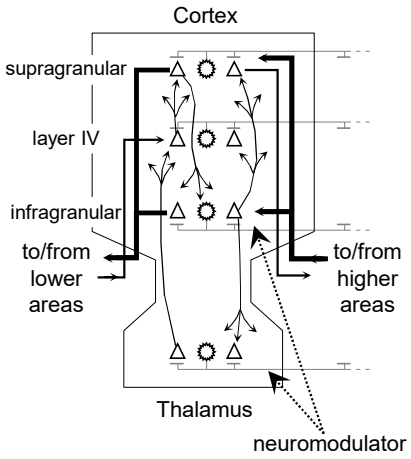
but these effects are minuscule and can be explained by a small transient PFC activation (failed ignition).

Detailed simulations of the global neuronal workspace using a semi-realistic network of spiking neurons

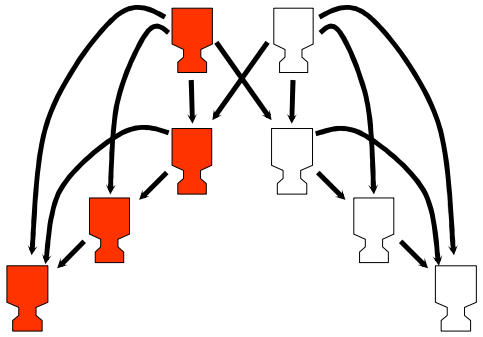
Feedforward connections (AMPA)



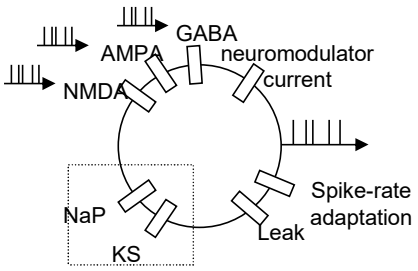
Thalamocortical column



Feedback connections (NMDA)



Spiking neurons



Optional cellular oscillator currents

Bottom-up propagation

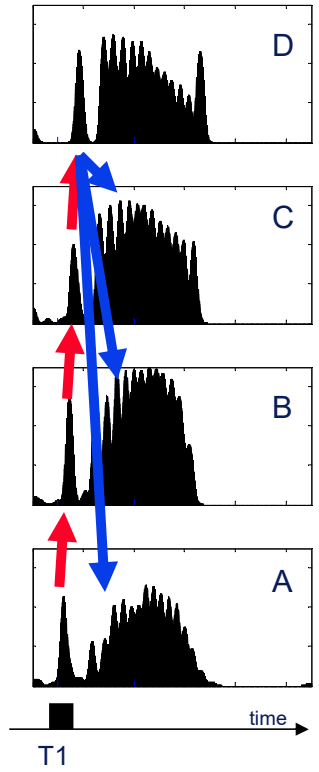
followed by

top-down amplification

More predictions :

- Feedforward activation increases **monotonically** with stimulus strength
- Late ignition is **all-or-none** (bimodal distribution).
- **Threshold** for ignition
- Link to evidence accumulation
- Long-distance **synchrony**
- Role of **NMDA** receptors

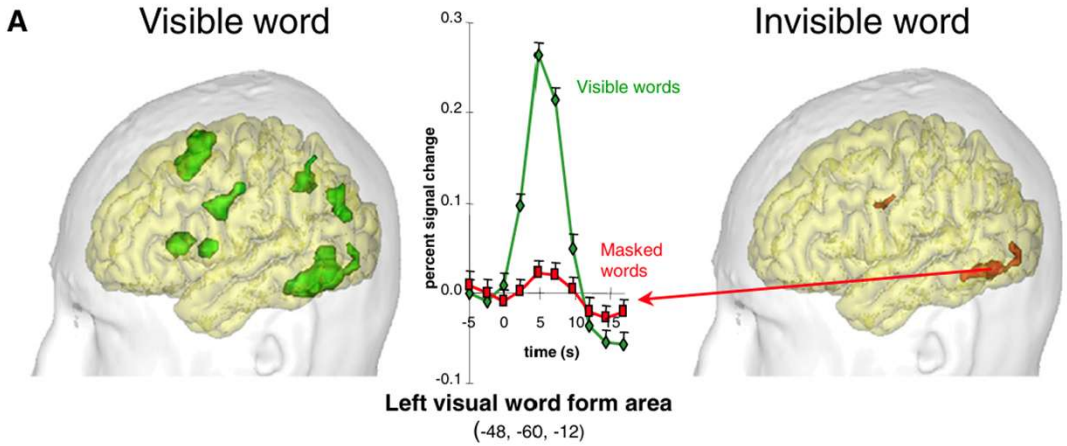
Instantaneous firing rate as a function of time in each of 4 hierarchical simulated areas



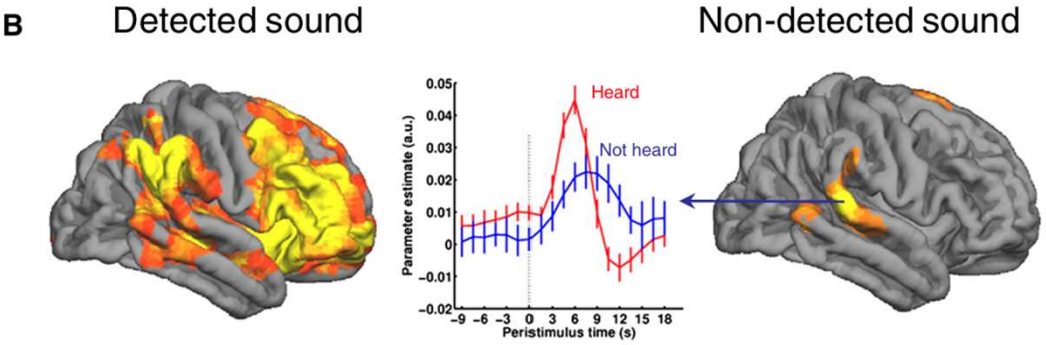
Dehaene, Sergent, & Changeux, *PNAS,* 2003

Many paradigms have observed global ignition (see Course 3 and the following)

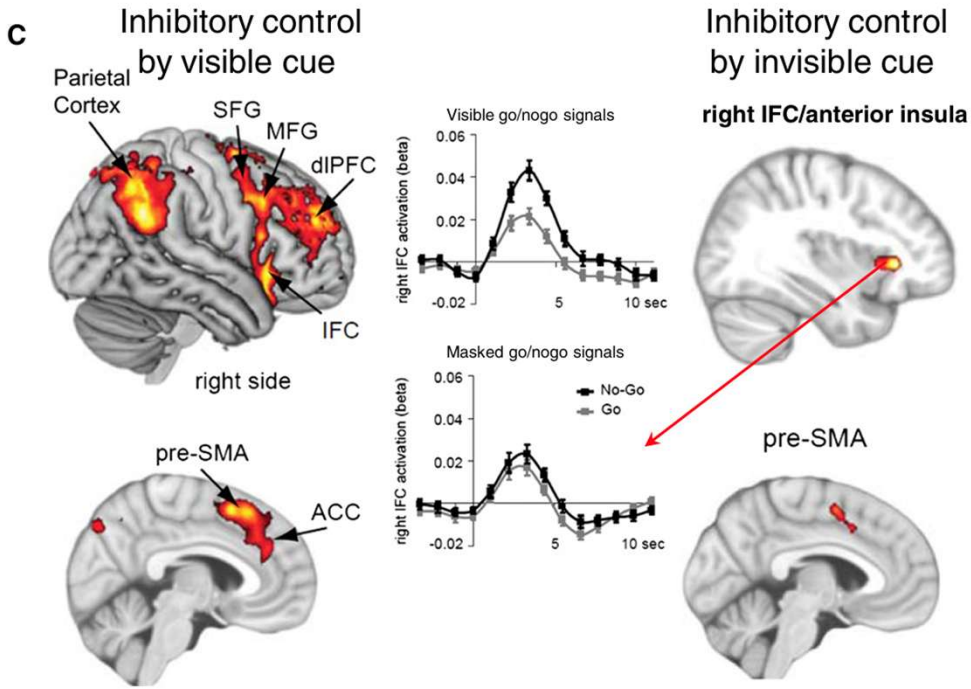
Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*(2), 200-227. https://doi.org/10.1016/j.neuron.2011.03.018



Dehaene et al., Nature Neuroscience 2001

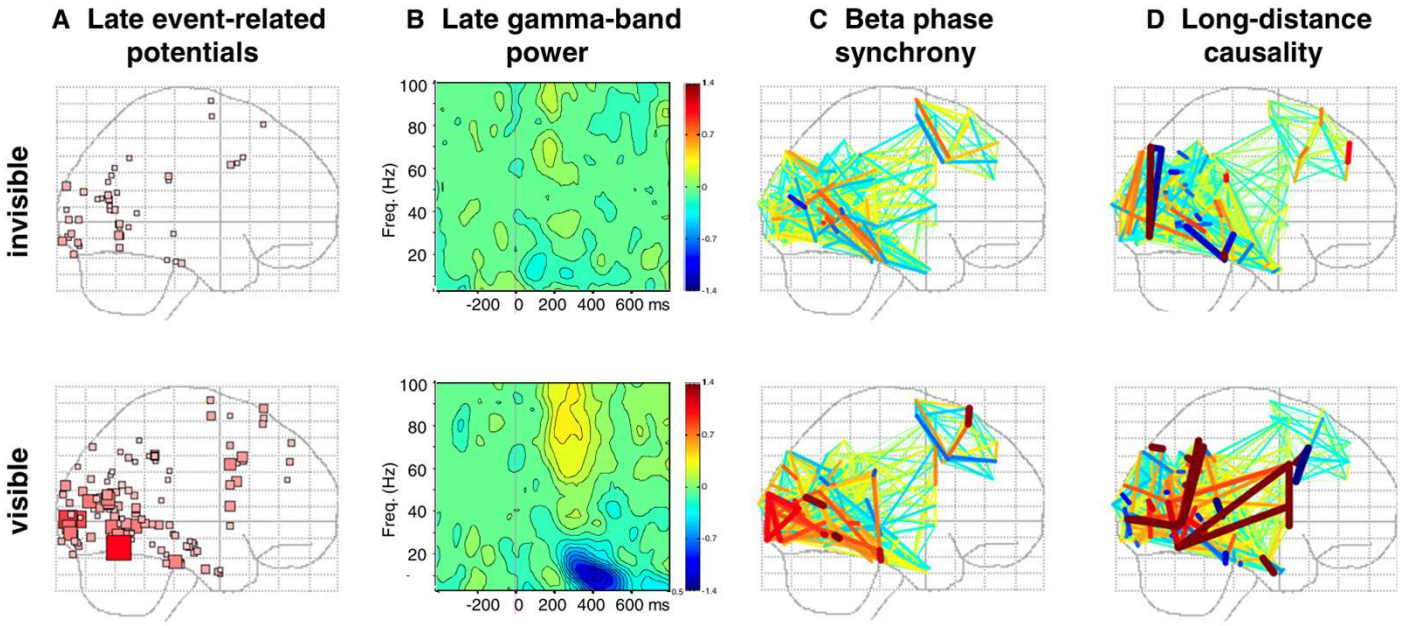


Sadaghiani, Hesselman & Kleinschmidt, J. Neuroscience 2009

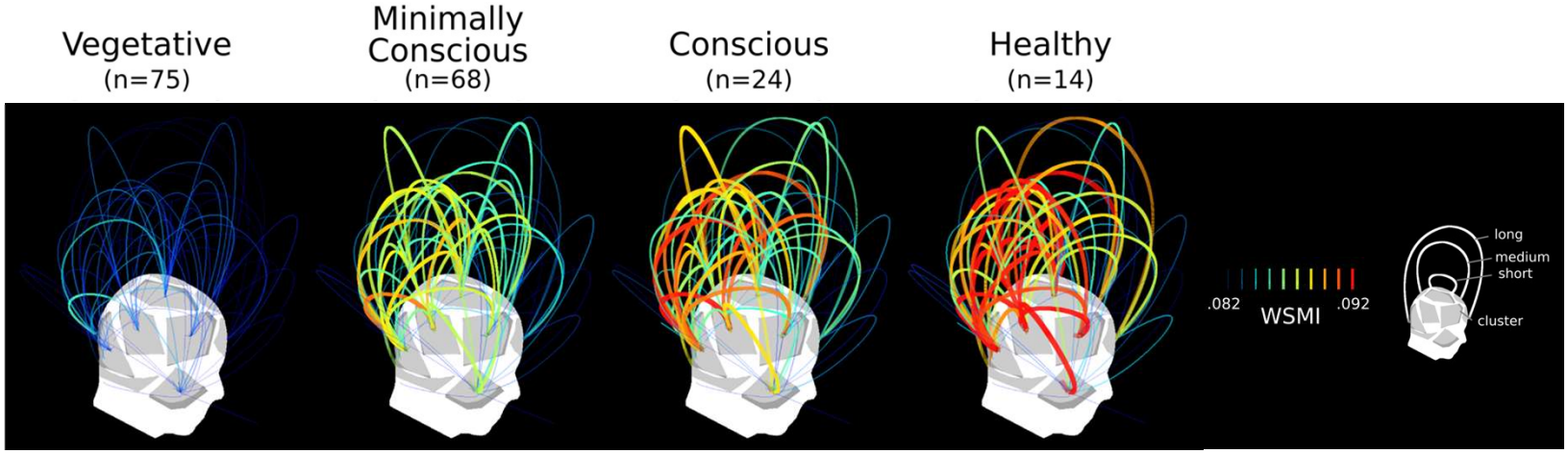


Van Gaal et al., J Cognitive Neuroscience 2011

Long-distance synchrony in the beta band is a reliable signature of consciousness (probably reflecting global broadcasting)



Gaillard, Dehaene et al., *PLOS Biology* 2009



Weighted symbolic mutual information : Jean-Rémi King, Jacobo Sitt, Stanislas Dehaene, Lionel Naccache, *Current Biology*, 2013

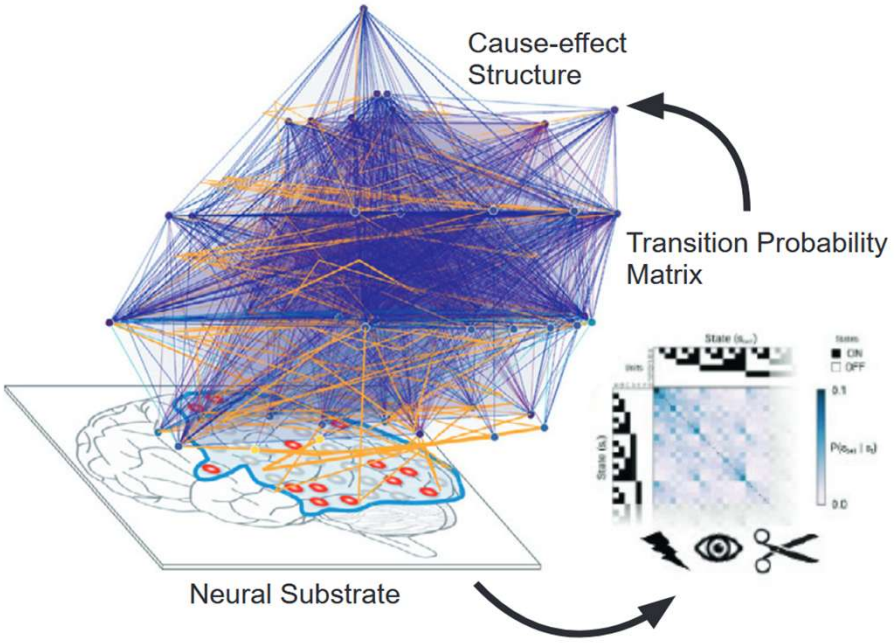
Convergence among theories of consciousness

Storm, J. F., Klink, P. C., Aru, J., Senn, W., Goebel, R., Pigorini, A., Avanzini, P., Vanduffel, W., Roelfsema, P. R., Massimini, M., Larkum, M. E., & Pennartz, C. M. A. (2024). An integrative, multiscale view on neural theories of consciousness. *Neuron*, *112*(10), 1531-1552. https://doi.org/10.1016/j.neuron.2024.02.004

Information integration theory (IIT)

Tononi et al. 2016

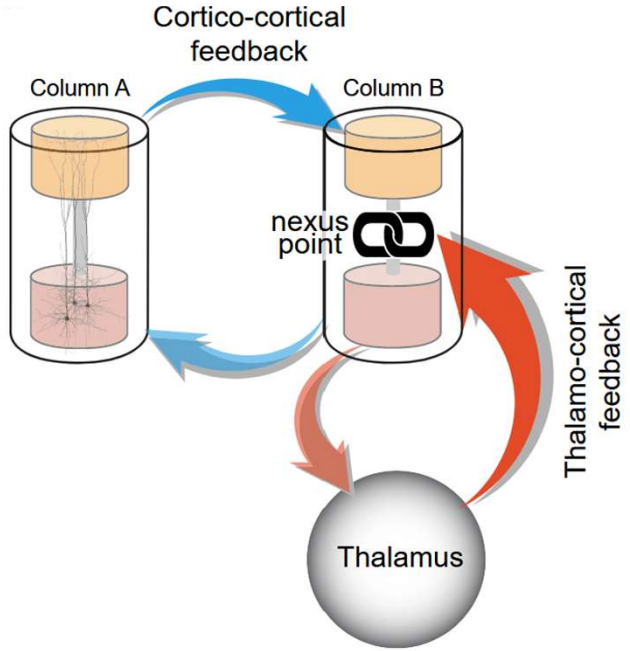
Consciousness depends on a global, highly integrated cause-effect structure



Dendritic integration theory (DIT)

Aru, Suzuki & Larkum 2020

Consciousness depends long-distance cortical-cortical feedback onto the apical dendrites of Layer 5 pyramidal cells, as well as thalamo-cortical feedback

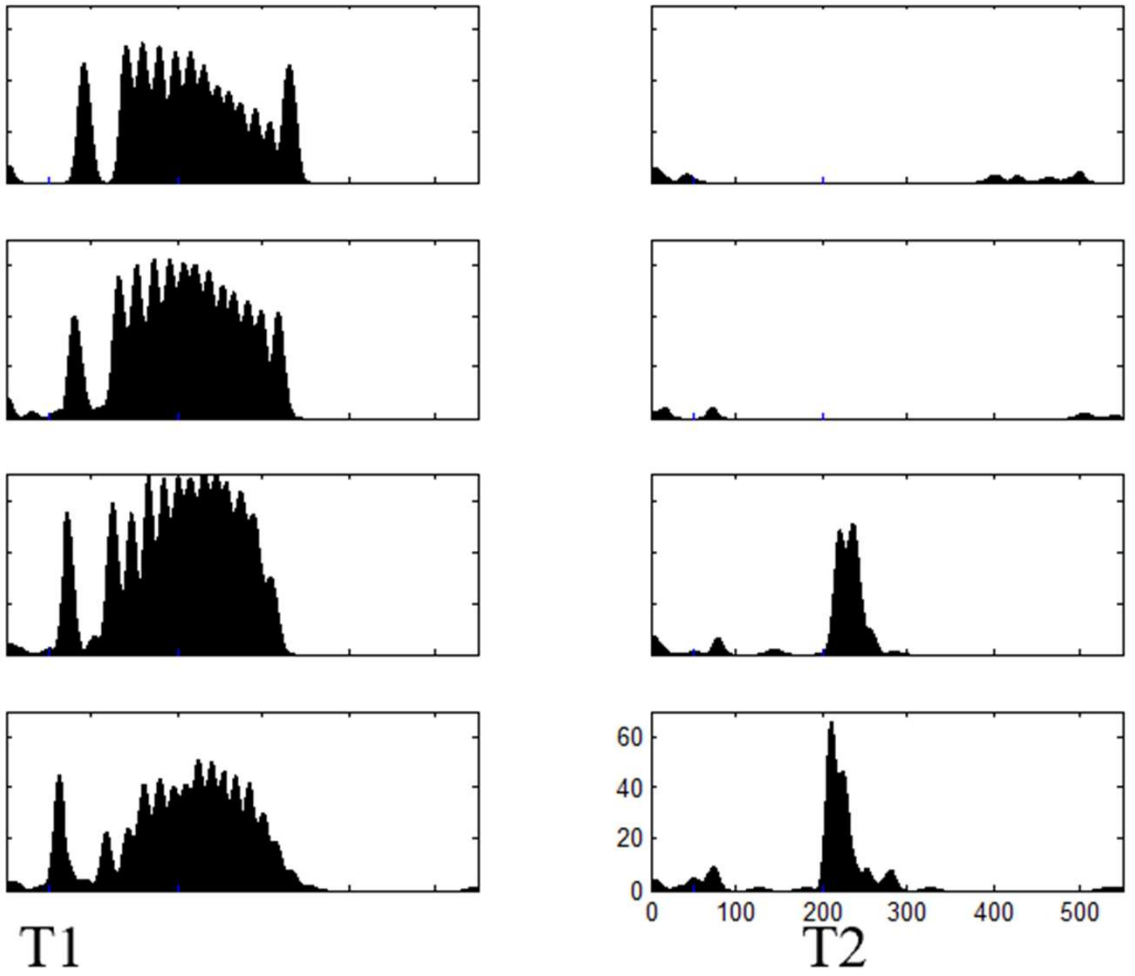


Simulations of another phenomenon: the attentional blink

When a stimulus T2 is briefly presented during the conscious ignition by a first stimulus T1, T2 may fail to be seen.
→ This phenomenon (a transient period of invisibilty) is known as the attentional blink.

Predictions :

- **Competition** for access to a central **capacity-limited system**.
- The blink should be **all-or-none**.
- The blink duration window should increase with T1 processing difficulty (Jolicoeur, JEPHPP 1999)
- The only necessary condition for the blink to occur should be that T1 has been consciously perceived (Nieuwenstein et al., J Vision 2009 varied just the conscious perception of a masked T1).

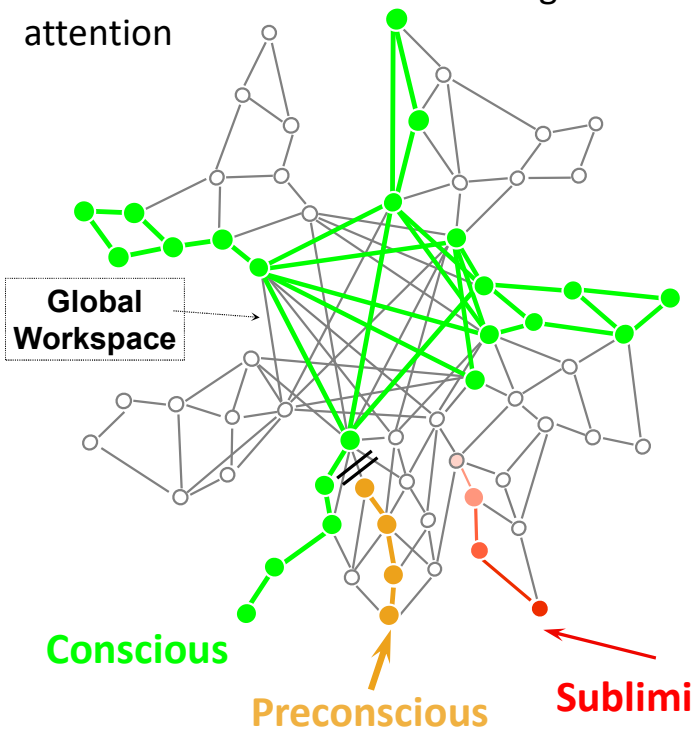


Dehaene, Sergent, & Changeux, *PNAS,* 2003

A taxonomy of conscious and non-conscious states

Dehaene, Changeux, Naccache, Sackur, & Sergent, TICS, 2006

A conjunction of two factors is needed for conscious access: stimulus strength and attention



Bottom-up stimulus strength	Top-down attention: absent	Top-down attention: present
weak or interrupted	**Subliminal** (unattended) - Very little activation - Activation is already weak in early extrastriate areas - Little or no priming - No reportability	**Subliminal** (attended) - Strong feedforward activation - Activation decreases with depth - Depth of processing depends on attention and task set - Activation can reach semantic level - Short-lived priming -No durable fronto-parietal activity - No reportability
sufficiently strong	**Preconscious** - Intense activation, yet confined to sensori-motor processors - Occipito-temporal loops and local synchrony - Priming at multiple levels -No reportability while attention is occupied elsewhere	**Conscious** - Orientation of top-down attention - Amplification of sensori-motor activity - Intense activation spreading to parieto-frontal network - Long-distance loops and global synchrony - Durable activation, maintained at will - Conscious reportability

This taxonomy remains controversial: Victor Lamme, Ned Block and others suggest that the « preconscious » category is actually a different form of **phenomenal** consciousness (without reportability) different from **access** consciousness.

An important aside: the relation between consciousness and attention

Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences*, *16*(8), 411-417. https://doi.org/10.1016/j.tics.2012.06.013

Attention is the name that we give to the system**s** that allow the brain to **select** one out of the many possible available contents or processes.

The GNW model predicts that

- attention can operate on non-conscious processors
- but that consciousness requires attentional selection (or **is** the ultimate selection!)

In other words :

- There is an entire pyramid of attentional selection processes, the last of which gates access to the conscious global neuronal workspace.
- attention is necessary, but not sufficient for conscious access.

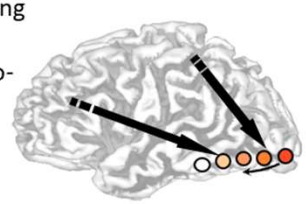
Cohen et al.'s review convincingly shows that "There are currently no stimuli that can be studied in laboratory settings [whose conscious perception] seems immune to attentional tasks that are designed to engage top-down, endogenous attention."

Subliminal (attended)

- Strong feedforward activation
- Activation decreases with depth
- Depth of processing depends on attention and task set
- Activation can reach semantic level
- Short-lived priming

-No durable fronto-parietal activity

- No reportability

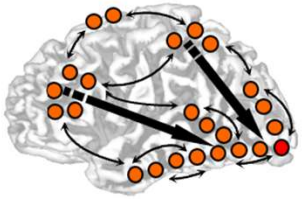


//

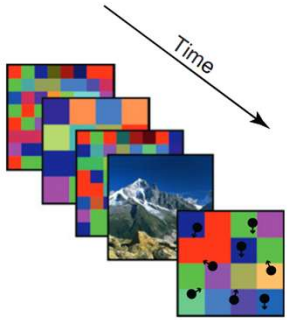
Conscious

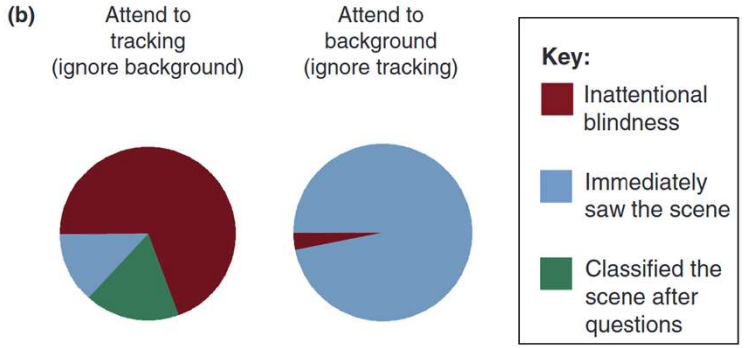
- Orientation of top-down attention
- Amplification of sensori-motor activity
- Intense activation spreading to parieto-frontal network
- Long-distance loops and global synchrony
- Durable activation, maintained at will

- Conscious reportability



Example: inattentional blindness. When focusing on dot tracking, participants fail to detect an image (and those who do probably switch away from the main task; PRP)





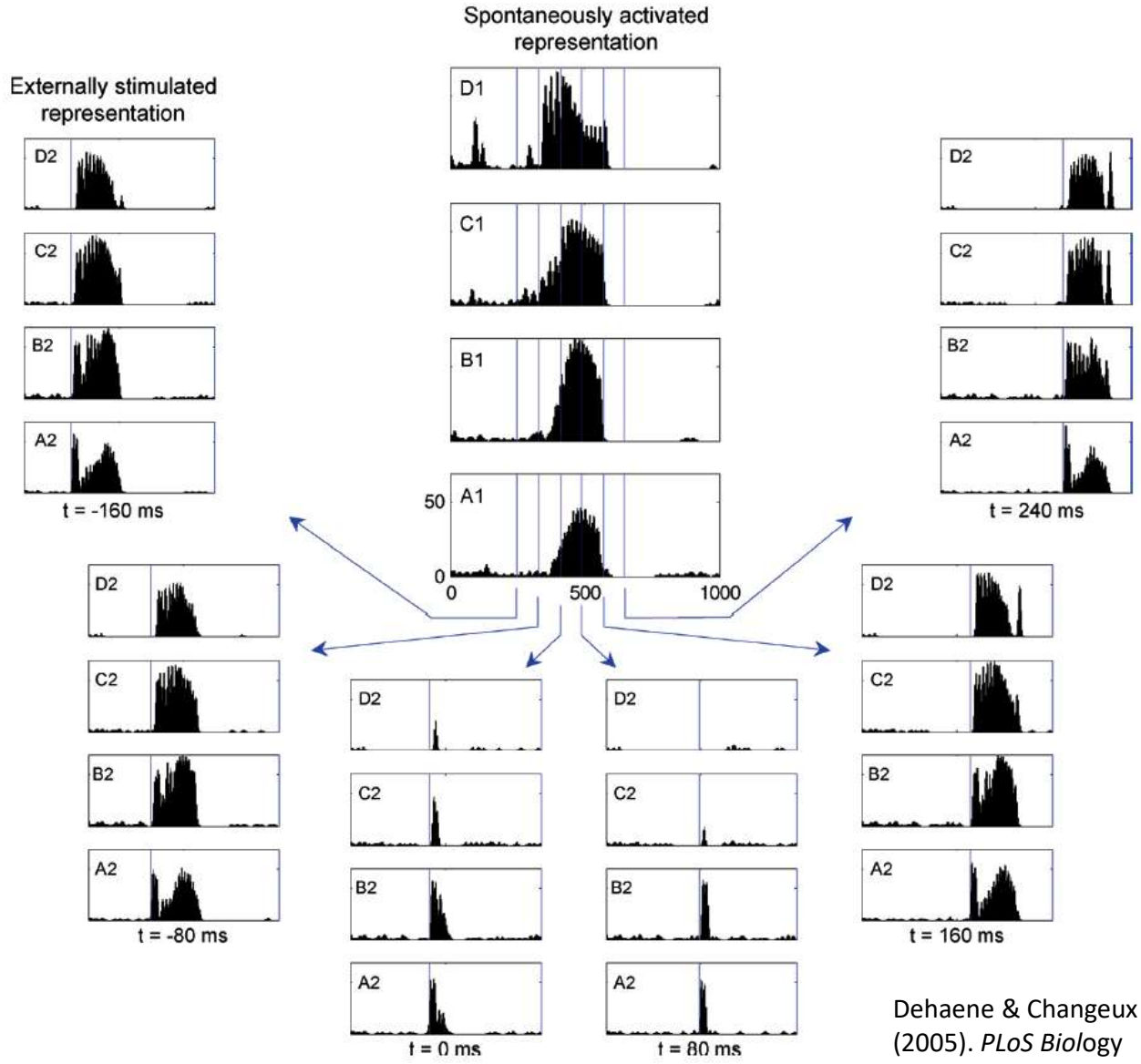
Spontaneous ignition and the flow of consciousness

Even in the absence of any input, **spontaneous ignition** may occur.

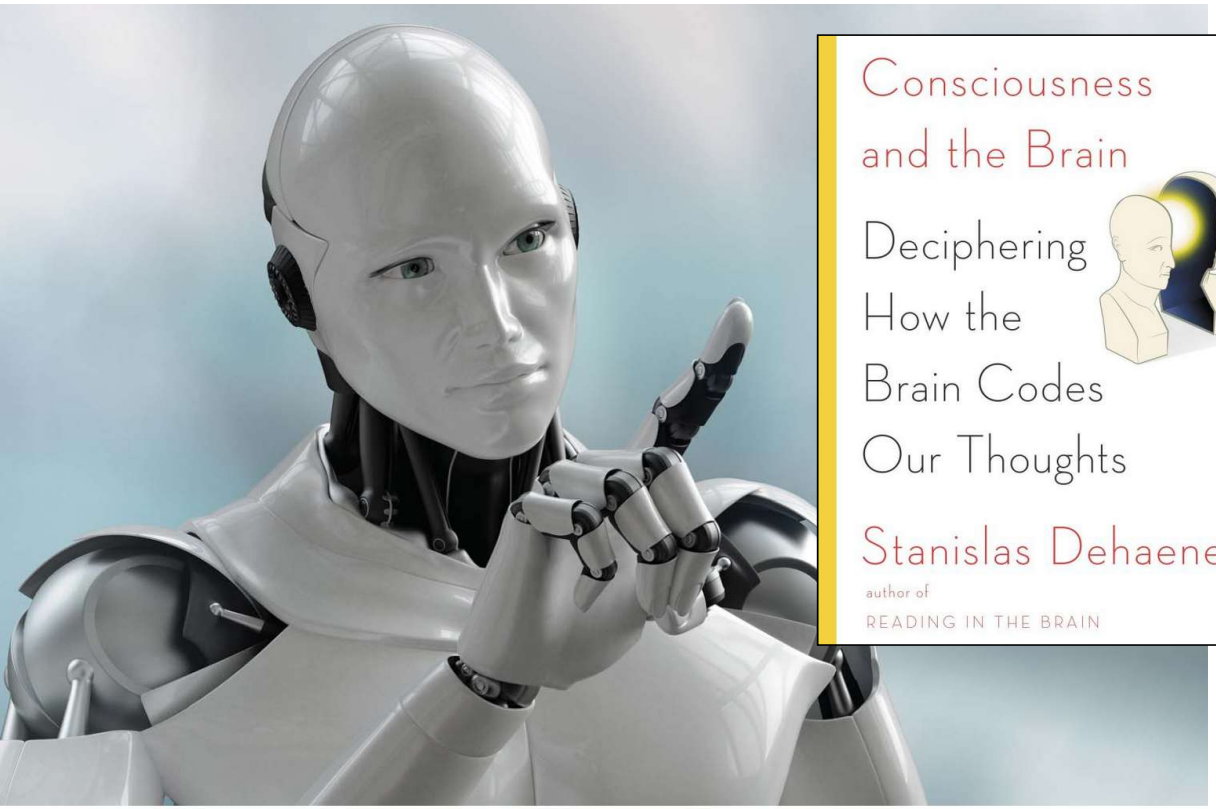
→ William James's « stream of consciousness » : a ceaseless succession of internal neural states.

Predictions from the simulations:

- Spontaneous ignition may cause **inattentional blindness**.
- Spontaneous activity depends on an **ascending vigilance signal** (modeling e.g. acetylcholine or norepinephrine signals)
- Spontaneous activity in the conscious resting state is characterized by **a dynamic diversity of long-distance brain states**.
- The loss of this diversity is a signature of non-consciousness (tested by Barttfeld et al. 2015, then Demertzi et al. 2019).
- The appearance of spontaneous activity corresponds to a **Hopf bifurcation**



Dehaene & Changeux (2005). *PLoS Biology*



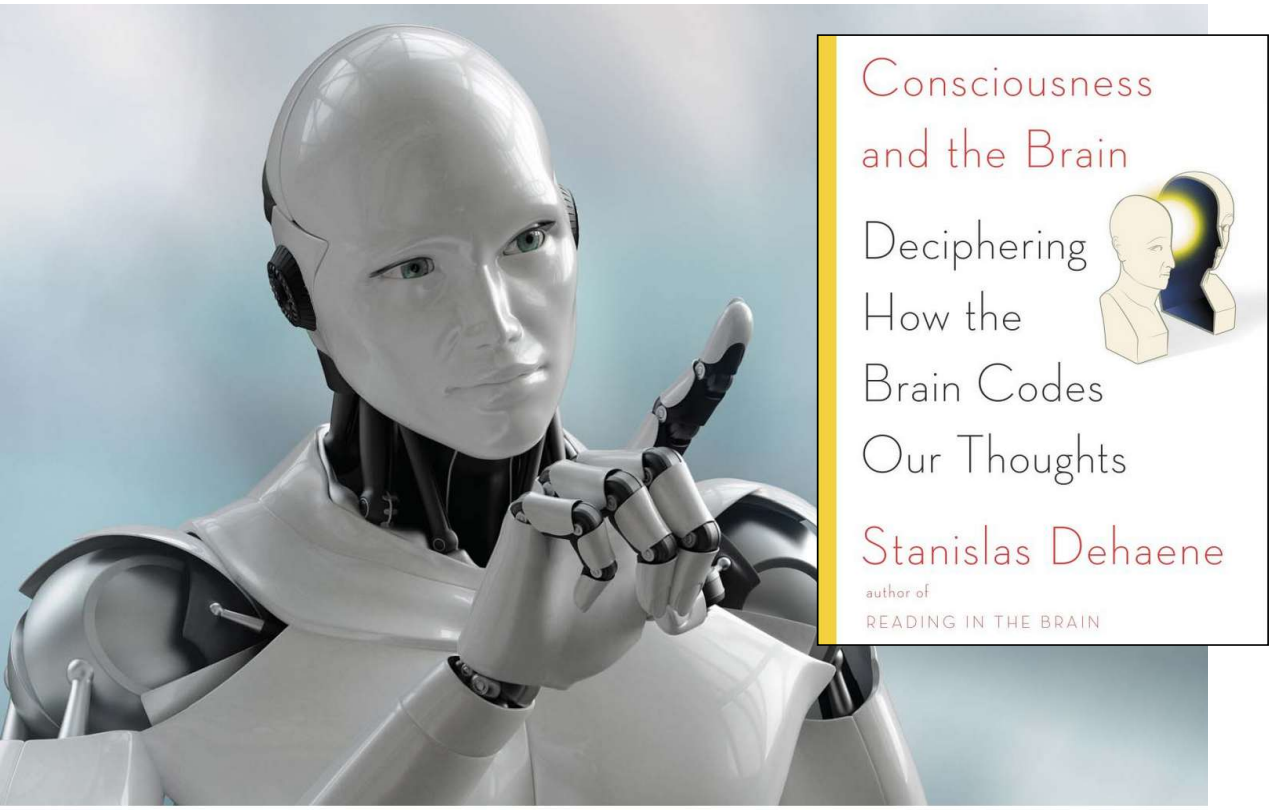
REVIEW

What is consciousness, and could machines have it?

Stanislas Dehaene,[1,2]* Hakwan Lau,[3,4] Sid Kouider[5]

Dehaene *et al.*, *Science* **358**, 486–492 (2017) 27 October 2017

Example of the low-fuel warning :
Imagine that you are driving when you suddenly realize that the fuel-tank light is on.
What makes you, a complex assembly of neurons, aware of the light?
And what makes the car, a sophisticated piece of electronics and engineering, unaware of it?
What would it take for the car to be endowed with a consciousness similar to our own?



REVIEW

What is consciousness, and could machines have it?

Stanislas Dehaene,[1,2]* Hakwan Lau,[3,4] Sid Kouider[5]

Dehaene *et al.*, *Science* **358**, 486–492 (2017) 27 October 2017

Consciousness is a **computational property** associated with a certain **type of information processing**, which could therefore be implemented in machines.

- Level C0. **Unconscious, modular computations**

- Level C1. **Conscious access**:
  a selected piece of information is amplified, and its results are broadcasted globally

→ a **global and flexible routing** system.

- Level C2. **Conscious self-representation**:
  the system learns to represent itself (e.g. it knows that it knows X, it knows that it has made an error, etc.)

→ a **self-monitoring** system

Machine consciousness: what is missing?

C1. Global availability : "Having information X in mind"
A global workspace for information sharing between modules

Current machines rely on modular, special-purpose programs and databases.

One program may « know » a piece of information while the others ignore it.

In the human brain, this is characteristic of **unconscious processing.**

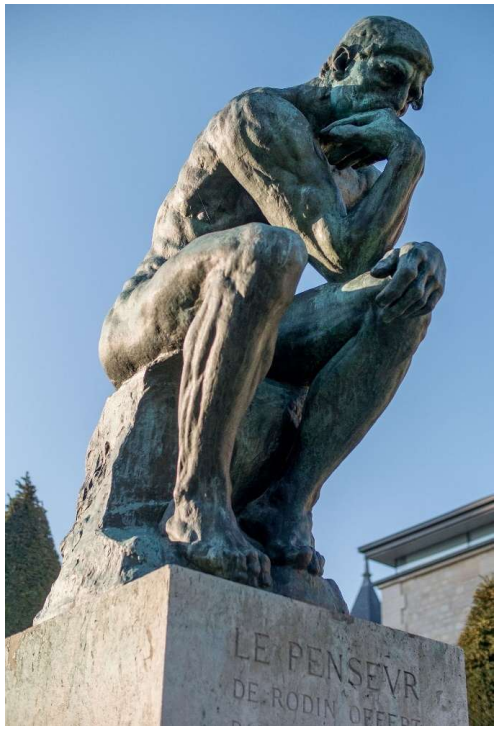
Global neuronal workspace hypothesis:
Consciousness evolved to **break this modularity,** and therefore to **make relevant information available** to the whole system for decision making

→ A **shared "workspace" or "blackboard" for flexible information sharing**

Machine consciousness: what is missing?

C2. Self monitoring and introspection

A repertoire of self-knowledge



A self-conscious machine must have a repertoire of self-knowledge, i.e. know

- **what modules** it possesses
- what **computations** each of them can perform
- **how fast** they can operate
- how likely they are to be **correct**

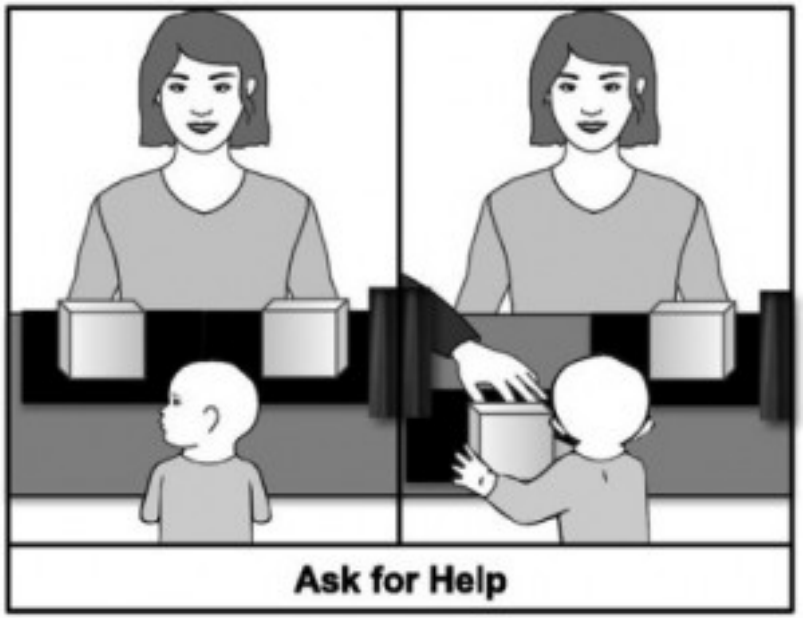
This knowledge should be learned and constantly updated.

→ "**Know thyself**"

Machine consciousness: what is missing?

C2. Self monitoring and introspection

Confidence and "knowing that you don't know"



Metacognition in 20 month old babies
(Goupil et al, with Kouider, Science 2016)

A conscious system knows when it is wrong or uncertain.

Even monkeys and preverbal infants know when they don't know (e.g. turn to their mother for help).

A machine could be equipped with

- **Bayesian programs** that do not just give an answer, but also compute the probability of being correct.
- **An error-detection system** that monitors contradictions and detects when the current behavior is likely to be wrong
- A **coupling between error detection and search**, such that the system looks for alternative ways to get an answer.

Relation to another theory of consciousness : Attention Schema Theory (AST)

Graziano, M. S., Guterstam, A., Bio, B. J., & Wilterson, A. I. (2019). Toward a standard model of consciousness : Reconciling the attention schema, global workspace, higher-order thought, and illusionist theories. *Cognitive neuropsychology*, 1-18.
Graziano, M. S. A. (2022). A conceptual framework for consciousness. *PNAS*, *119*(18), e2116933119. https://doi.org/10.1073/pnas.2116933119

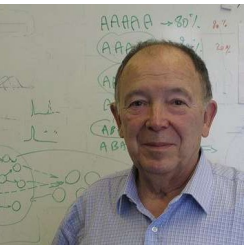
Proposal : subjective awareness is **the brain's simplified, internal model of its own attention process**.

Attention is a genuine set of unconscious brain processes that select and amplify incoming sensory signals.

To efficiently control this complex process, the brain constructs a simplified **Attention Schema** – an internal model that represents the current state and dynamics of attention (similar to the body schema).

AST proposes that **what we consciously experience is the *content* of this attention schema.**
Since every brain model is inaccurate, we may infer that consciousness has an intangible non-physical quality, whereas it is in fact just an imperfect description of the physical process of attention.

→ Similar to “theory of mind” but for our own minds.
→ Evolved for (1) better self-control ; (2) better understanding of other minds and cooperation
→ Similar to the GNW’s « internal repertoire of self-knowledge » : the prefrontal cortex must have a model, not just of its attention, but also of its knowledge states (e.g. “I can’t speak Chinese”).



Jean-Pierre Changeux

The global neuronal workspace (GNW) hypothesis
Summary of the main predictions

Baars, 1989; Dehaene & Changeux, 1998, 2003, 2005; Dehaene & Naccache, 2001

Lionel Naccache

- **Unconscious processing** can be very deep (course 2)
- **All-or-none ignition**, particularly in prefrontal cortex, whenever a novel content gains access to consciousness.

GNW activity (including PFC neurons) should contain a **complete neural code** for any conscious contents

Even in the absence of explicit report (no-report paradigms) (course 3)

- **Sustained activity :** Consciousness is the gateway to working memory (course 4)
- **Central collision and temporal decoupling** :

Only a single mental object can be consciously processed at a time. The other has to wait. Conscious access can be drastically delayed relative to stimulus onset, and need not be sustained throughout a long stimulus. (course 5)

- **Spontaneous ignition** : the global workspace is constantly traversed by a series of ever-changing global patterns. Clinical applications (course 6).