



Alma Mater Studiorum
University of Bologna
Department of Psychology
"Renzo Canestrari"

Neuroscience of Consciousness

Francesco Di Gregorio, PhD

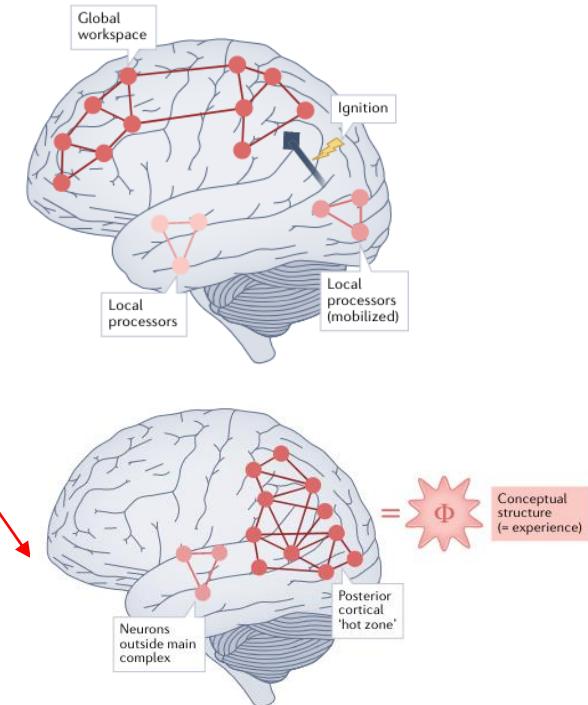
- ✉ francesco.digregori5@unibo.it
- 🌐 www.unibo.it/sitoweb/francesco.digregori5



Table 1 | A selection of theories of consciousness

Theory	Primary claim	Key refs
Higher-order theory (HOT)	Consciousness depends on meta-representations of lower-order mental states	31,46
Self-organizing meta-representational theory	Consciousness is the brain's (meta-representational) theory about itself	34,140
Attended intermediate representation theory	Consciousness depends on the attentional amplification of intermediate-level representations	141,142
Global workspace theories (GWTs)	Consciousness depends on ignition and broadcast within a neuronal global workspace where fronto-parietal cortical regions play a central, hub-like role	47–49
Integrated information theory (IIT)	Consciousness is identical to the cause–effect structure of a physical substrate that specifies a maximum of irreducible integrated information	57,59,60
Information closure theory	Consciousness depends on non-trivial information closure with respect to an environment at particular coarse-grained scales	143
Dynamic core theory	Consciousness depends on a functional cluster of neural activity combining high levels of dynamical integration and differentiation	144
Neural Darwinism	Consciousness depends on re-entrant interactions reflecting a history of value-dependent learning events shaped by selectionist principles	145,146
Local recurrence	Consciousness depends on local recurrent or re-entrant cortical processing and promotes learning	65,71
Predictive processing	Perception depends on predictive inference of the causes of sensory signals; provides a framework for systematically mapping neural mechanisms to aspects of consciousness	67,73,79
Neuro-representationalism	Consciousness depends on multilevel neurally encoded predictive representations	84
Active inference	Although views vary, in one version consciousness depends on temporally and counterfactually deep inference about self-generated actions	76; see also ⁹¹
Beast machine theory	Consciousness is grounded in allostatic control-oriented predictive inference	13,75,77; see also ⁹⁰
Neural subjective frame	Consciousness depends on neural maps of the bodily state providing a first-person perspective	24
Self comes to mind theory	Consciousness depends on interactions between homeostatic routines and multilevel interoceptive maps, with affect and feeling at the core	23,147
Attention schema theory	Consciousness depends on a neurally encoded model of the control of attention	148
Multiple drafts model	Consciousness depends on multiple (potentially inconsistent) representations rather than a single, unified representation that is available to a central system	149
Sensorimotor theory	Consciousness depends on mastery of the laws governing sensorimotor contingencies	85
Unlimited associative learning	Consciousness depends on a form of learning which enables an organism to link motivational value with stimuli or actions that are novel, compound and non-reflex inducing	150
Dendritic integration theory	Consciousness depends on integration of top-down and bottom-up signalling at a cellular level	151
Electromagnetic field theory	Consciousness is identical to physically integrated, and causally active, information encoded in the brain's global electromagnetic field	152
Orchestrated objective reduction	Consciousness depends on quantum computations within microtubules inside neurons	18

Our selection of theories includes those that are either neurobiological in nature or potentially expressible in neurobiological terms.

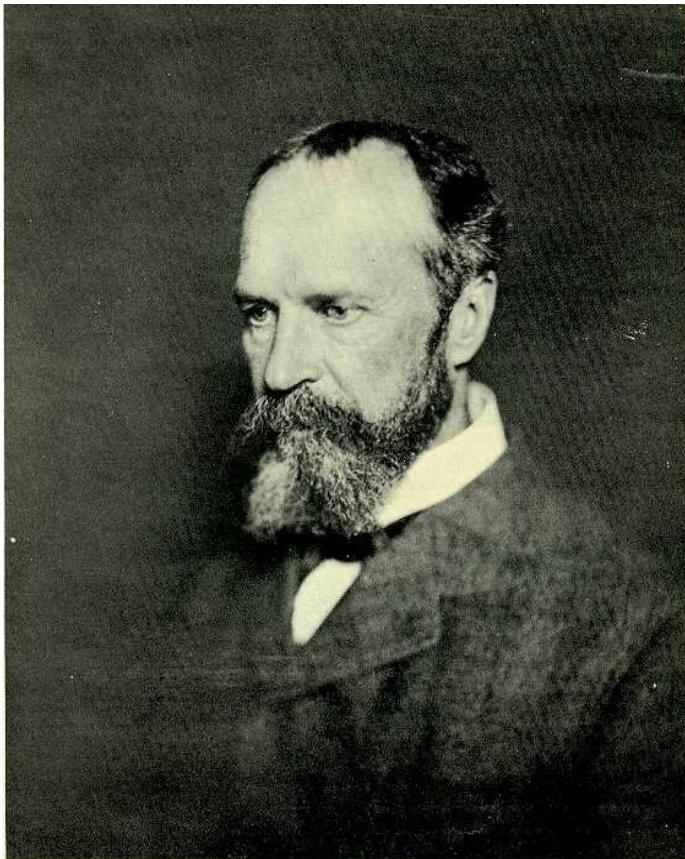


Theories of Consciousness and AI

Neuroscience of Consciousness

**What is
Consciousness?**

What is consciousness



“The study a posteriori of the *distribution* of consciousness shows it to be exactly such as we might expect in an organ added for the sake of steering a nervous system grown too complex to regulate itself.”

--- W. James, *The Principles of Psychology*, 1890, p. 141

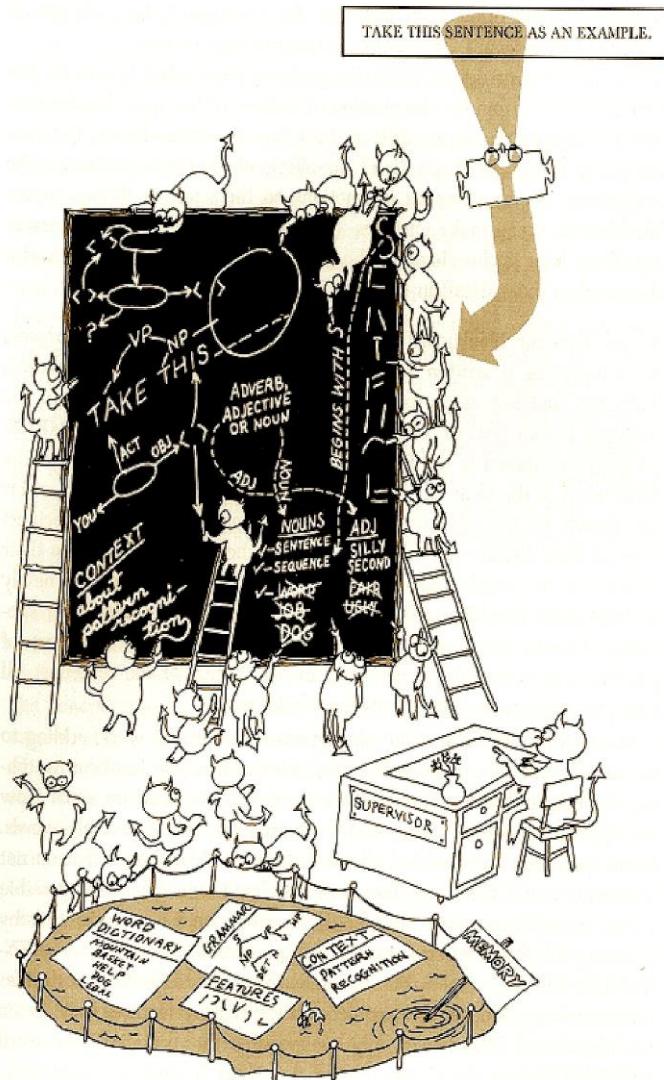
Distribution: Consciousness is a personal fact

Consciousness and Human cognition

Pandemonium

- A parallel distributed architecture for cognition.

FIGURE 7-7



- The “demons” are little processors that do specialized tasks.
- The blackboard is a place where they can all interact.
- The supervisor tries to keep them on task.
- The eyes scan one word at a time.
- Selective attention is a fundamental for consciousness

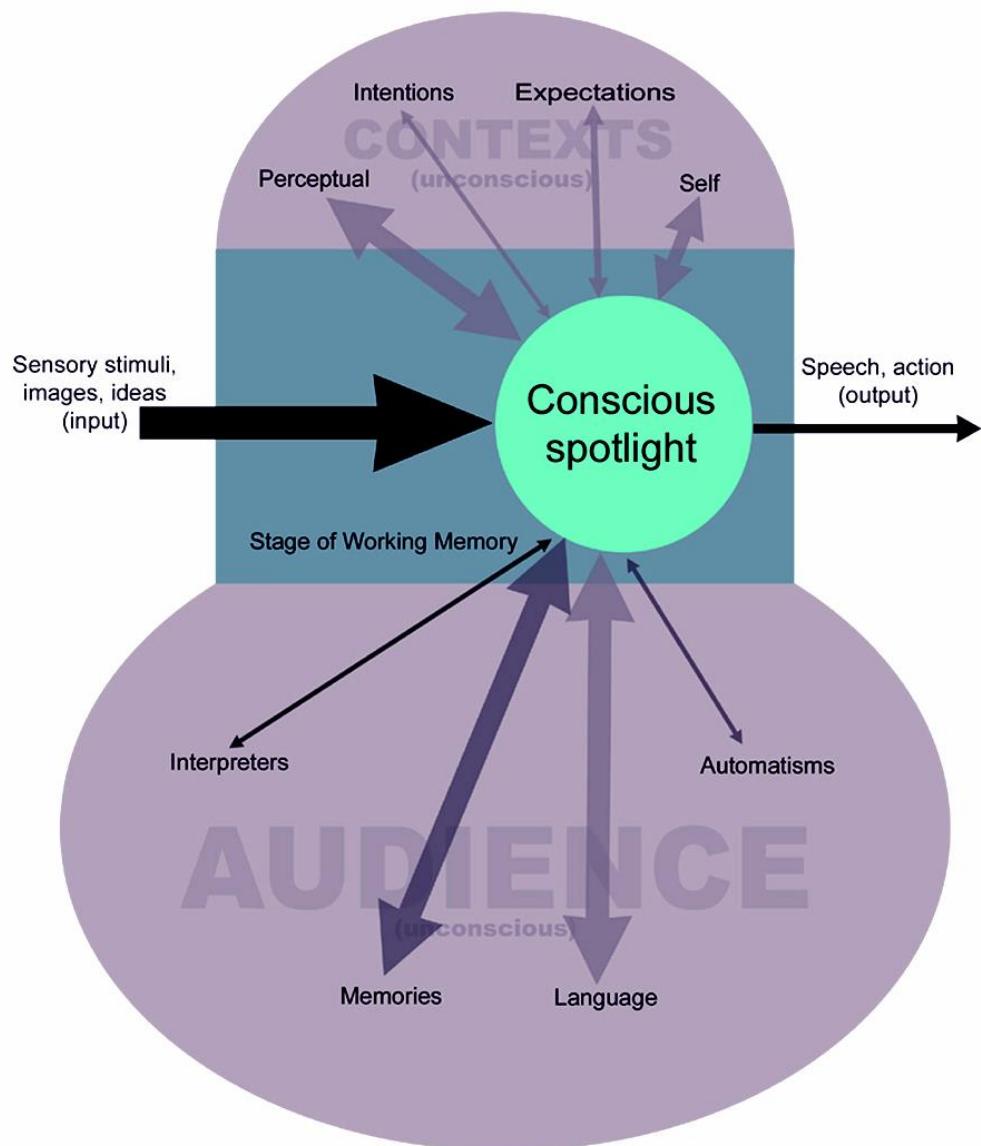
From Lindsay & Norman, Human Information Processing.

Awareness Test



copyright (c) 1999 Daniel J. Simons. All rights reserved.

A theater of consciousness - a useful theoretical metaphor



--- only the bright spot on stage is conscious (consciousness is very limited in capacity=selective attention)

--- sensory inputs compete for access to the conscious bright spot

--- the "stage" corresponds to Working Memory

--- all other features as unconscious, including long-term memory, the automatic processes of language, and the events going on backstage

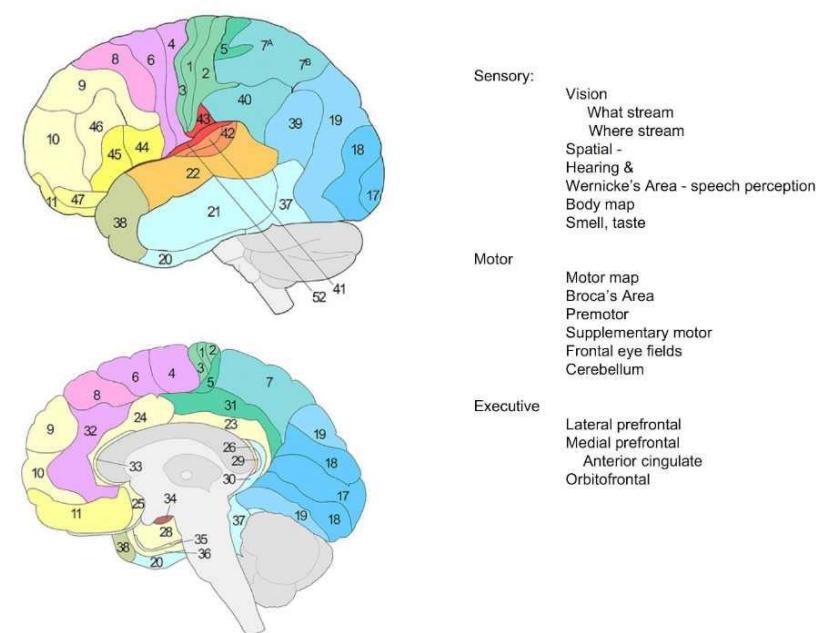
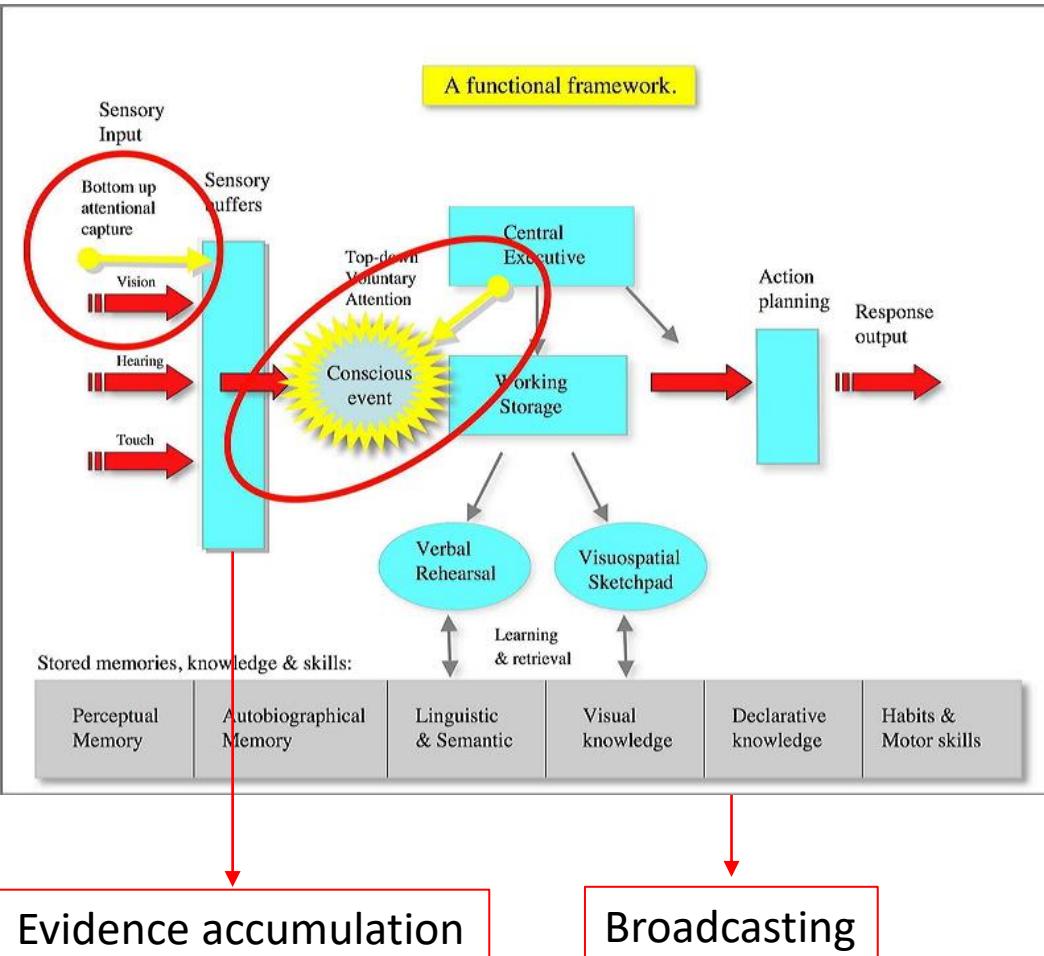
-- the theater metaphor has been turned into several testable models.

The Global Workspace Theory (GWT)

Applying GWT to cognitive functions.

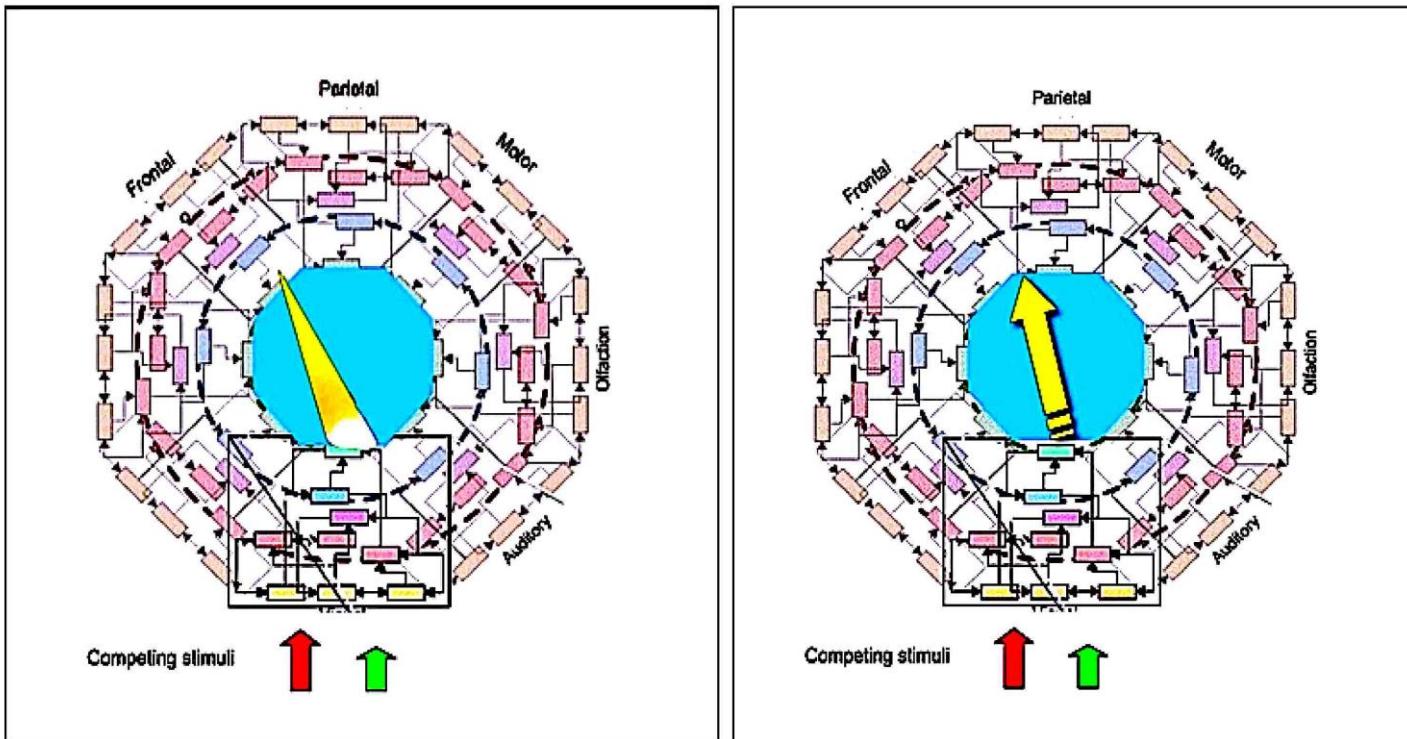
Baars & Gage, 2007.

- Brain correlates:



Two fundamental concepts in the GWT:

1. "Evidence accumulation (spotlight)"
2. "Broadcasting of accumulated evidence."



Spotlight - controlled by prefrontal regions.

Broadcasting arrow - from sensory cortex?

Binocular competition for consciousness - gamma resonance.

(Engel & Singer, 1995)

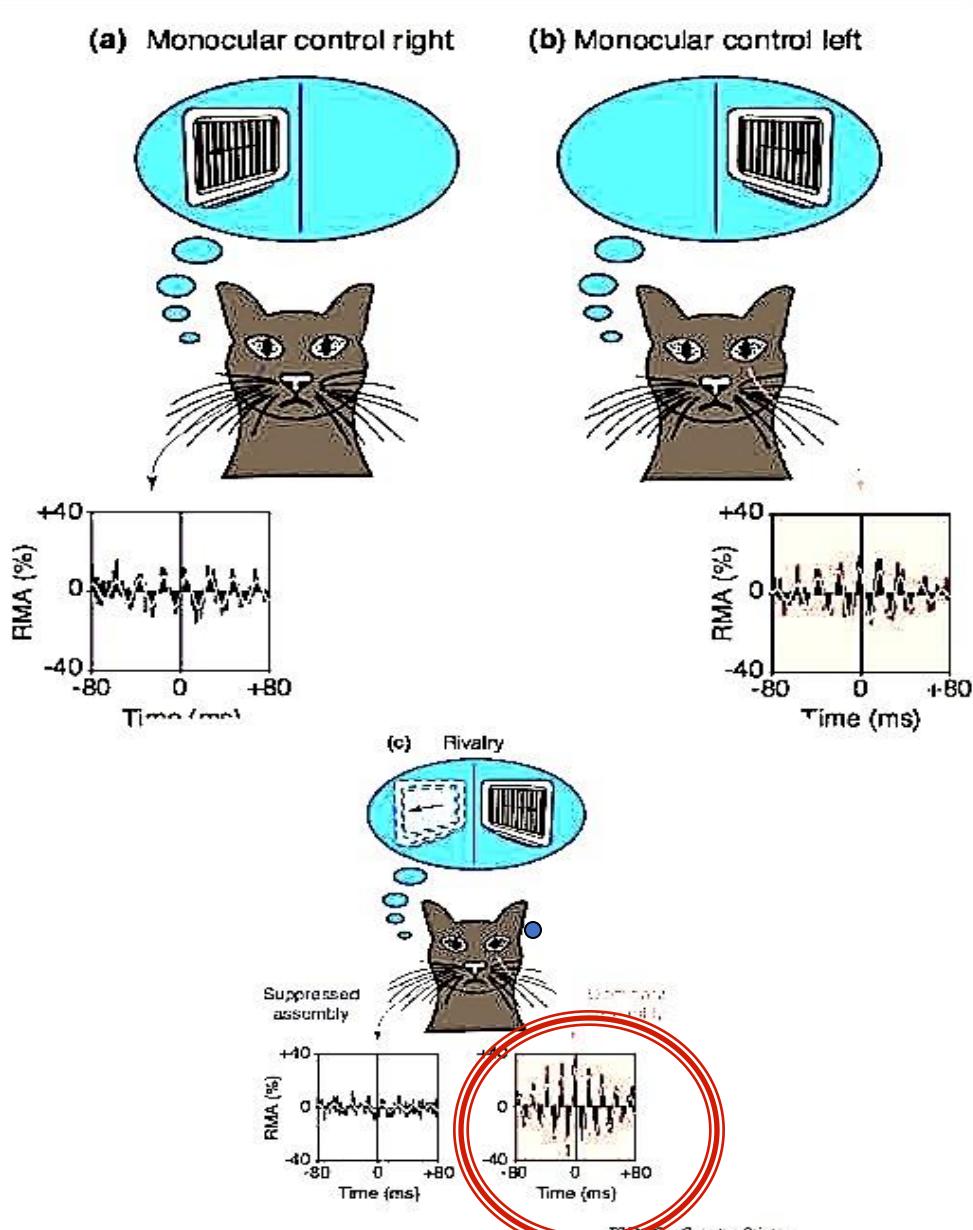
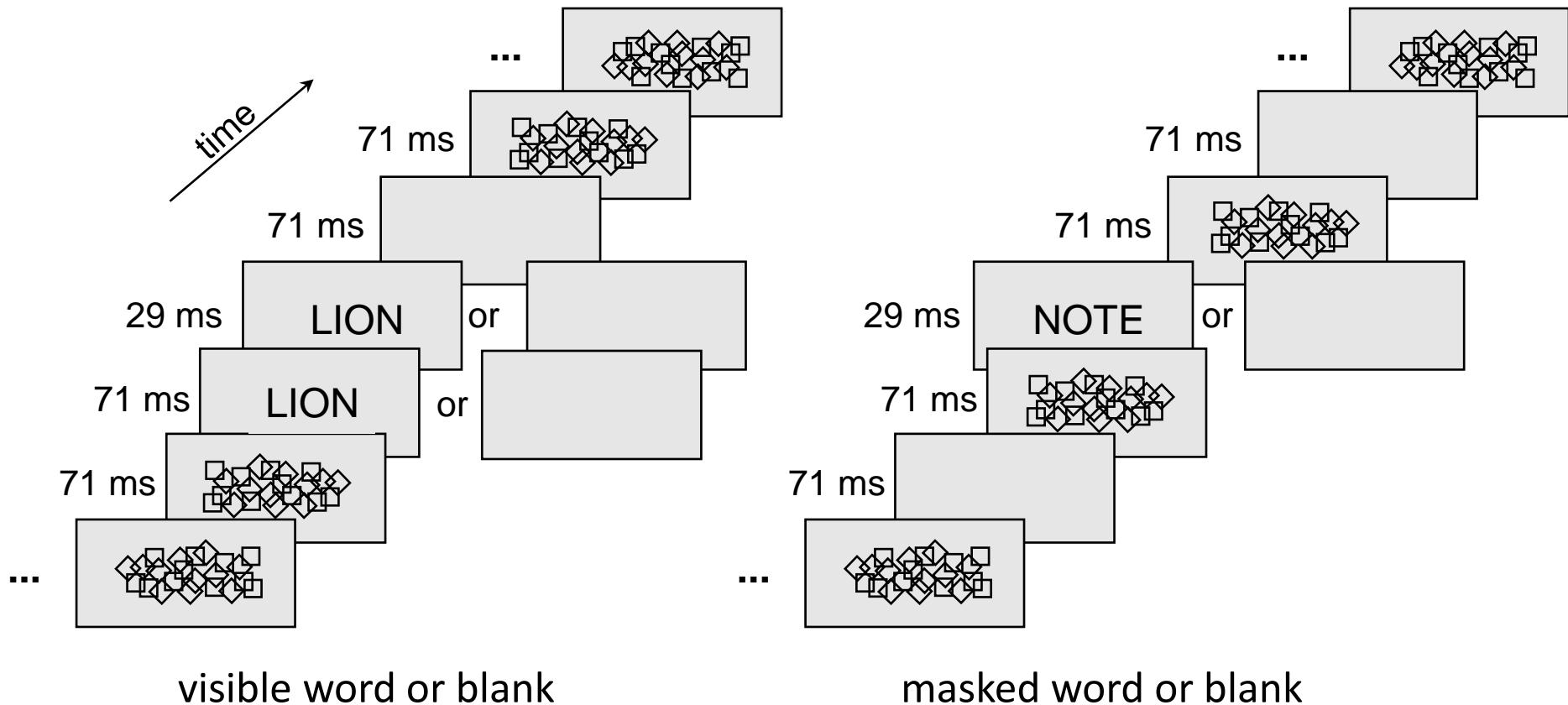
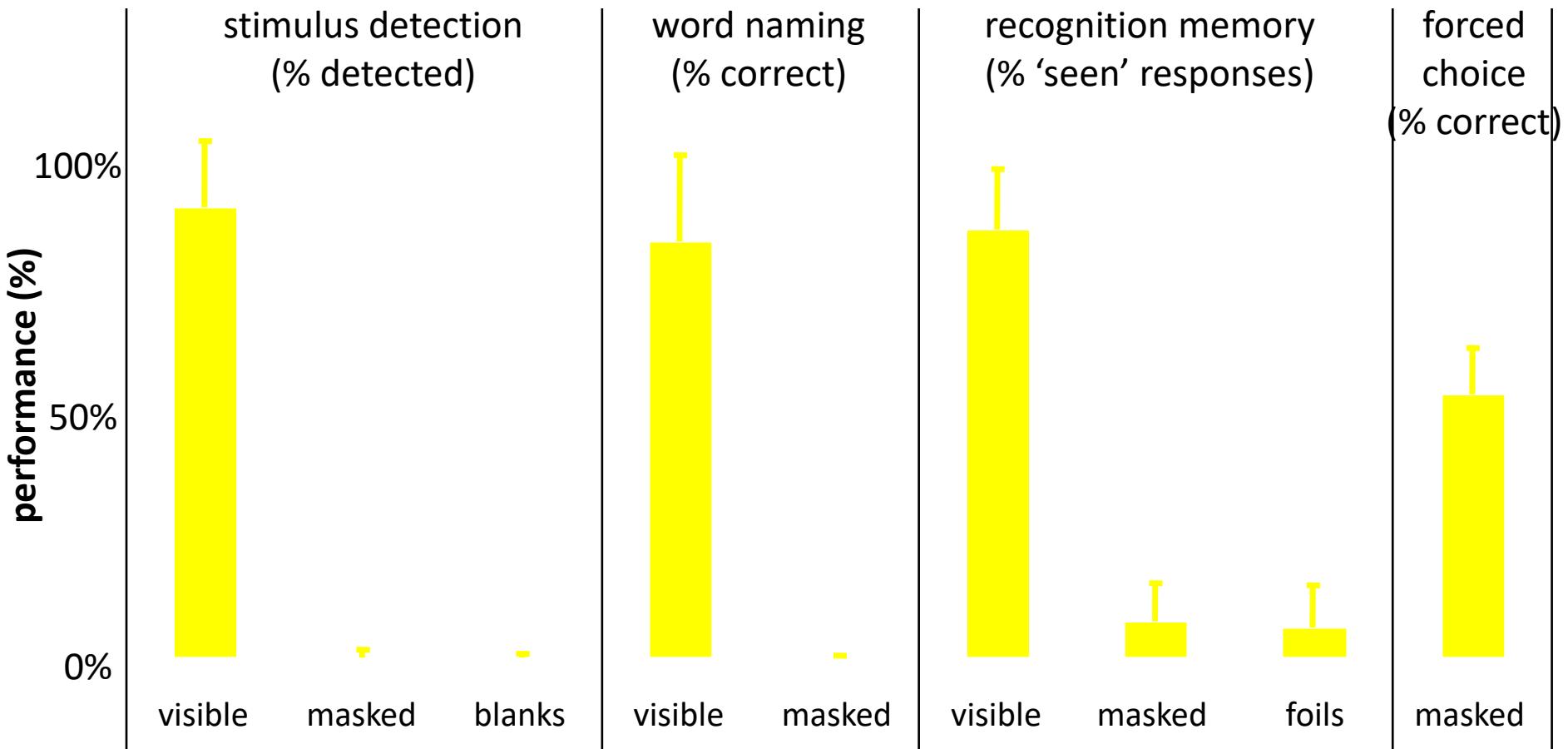


Fig. 2. Synchrony under conditions of binocular rivalry. Strabismic cats were used in these experiments because, in these animals, most cortical cells can be uniquely assigned to either the left or the right eye in terms of their ocular dominance. In front of the awake animal's head, two mirrors were mounted such that each eye viewed a separate computer monitor. Grating stimuli moving in different directions were presented on these monitors, resulting in perceptual rivalry between the two stimuli. (a,b) Cross-correlations are shown for two pairs of recordings driven by the right eye (blue plot) and left eye (red plot), respectively. Under a monocular control condition, both pairs of cells showed synchronized activity when their preferred eye was stimulated (as shown by clear peaks in the cross-correlograms). (c) Synchronization changed, compared with the monocular baseline, if both eyes were stimulated concurrently. Correlograms are shown for an epoch where the stimulus presented to the left eye was selected for perception. In this case, the cells driven through the left eye enhanced their correlation (dominant assembly, red plot), whereas the neurons that represent the suppressed stimulus (suppressed assembly, blue plot) decreased their temporal coupling. In epochs where the stimulus presented to the right eye dominates perception, the strength of the correlations reversed. The white continuous line superimposed on the correlograms represents a damped cosine function fitted to the data. RMA, relative modulation amplitude of the center peak in the correlogram, computed as the ratio of peak amplitude over offset of correlogram modulation. (Modified from Roit, 43.)

Brain Mechanisms of Conscious and Unconscious Reading



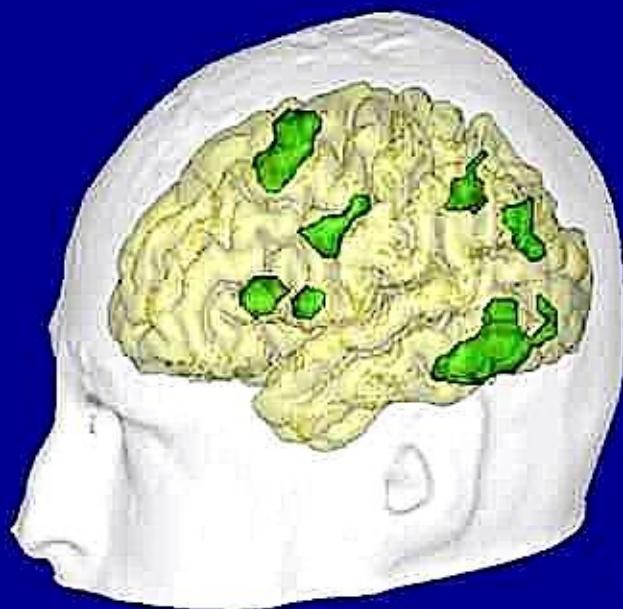
Behavioral Measures of Word Visibility



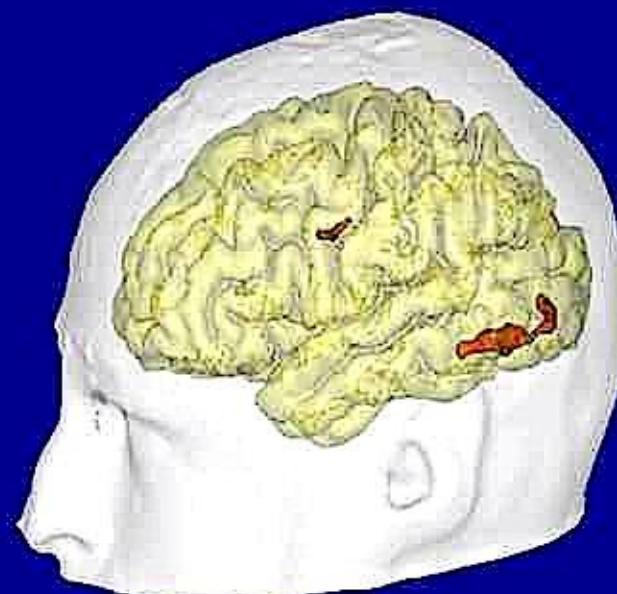
Experimental results:

Functional MRI

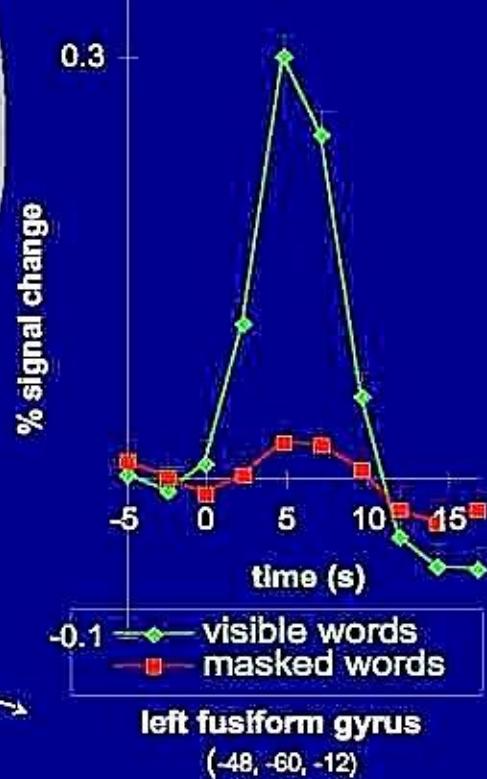
visible words
(conscious)



masked words
(unconscious)

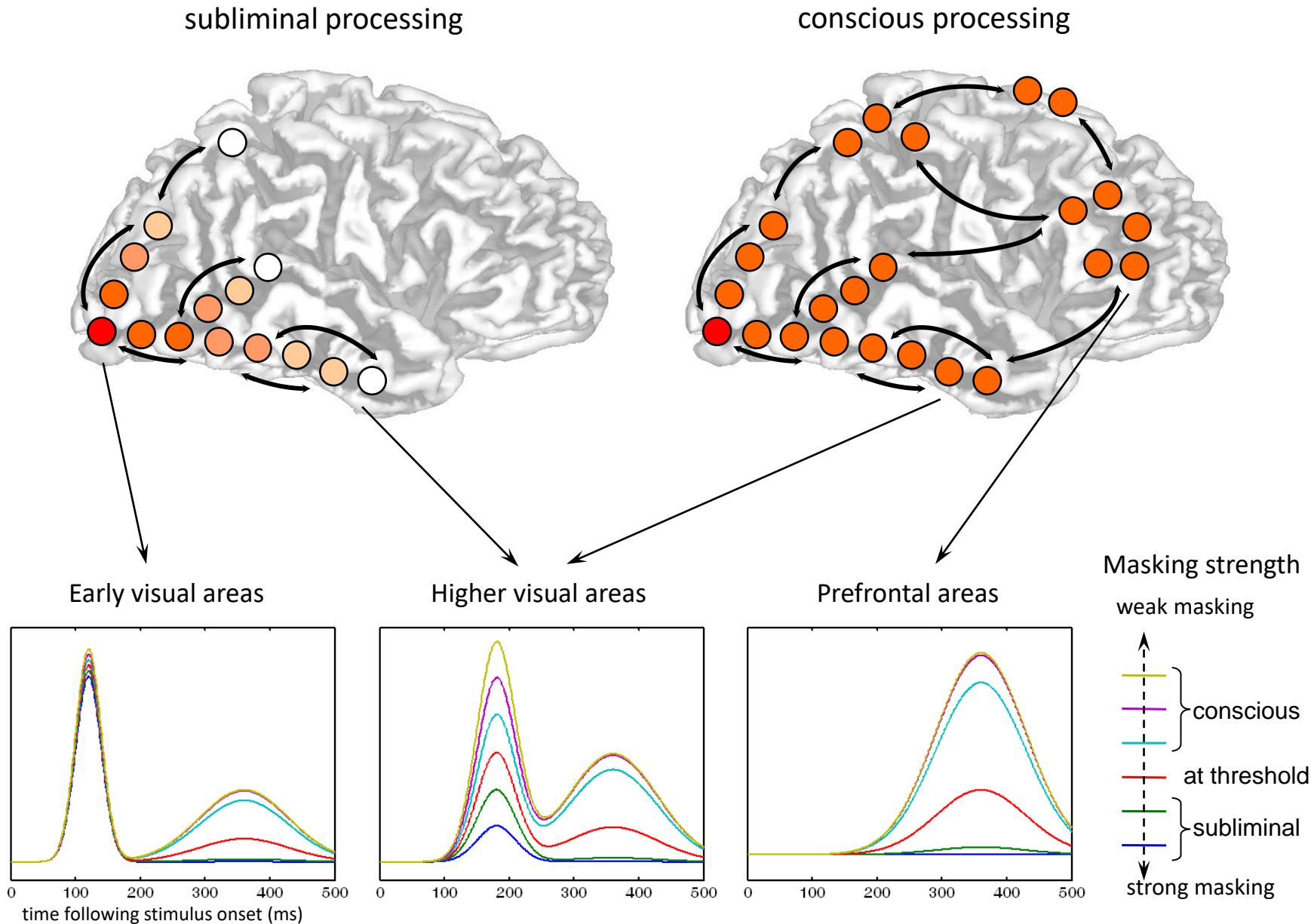


ERP's

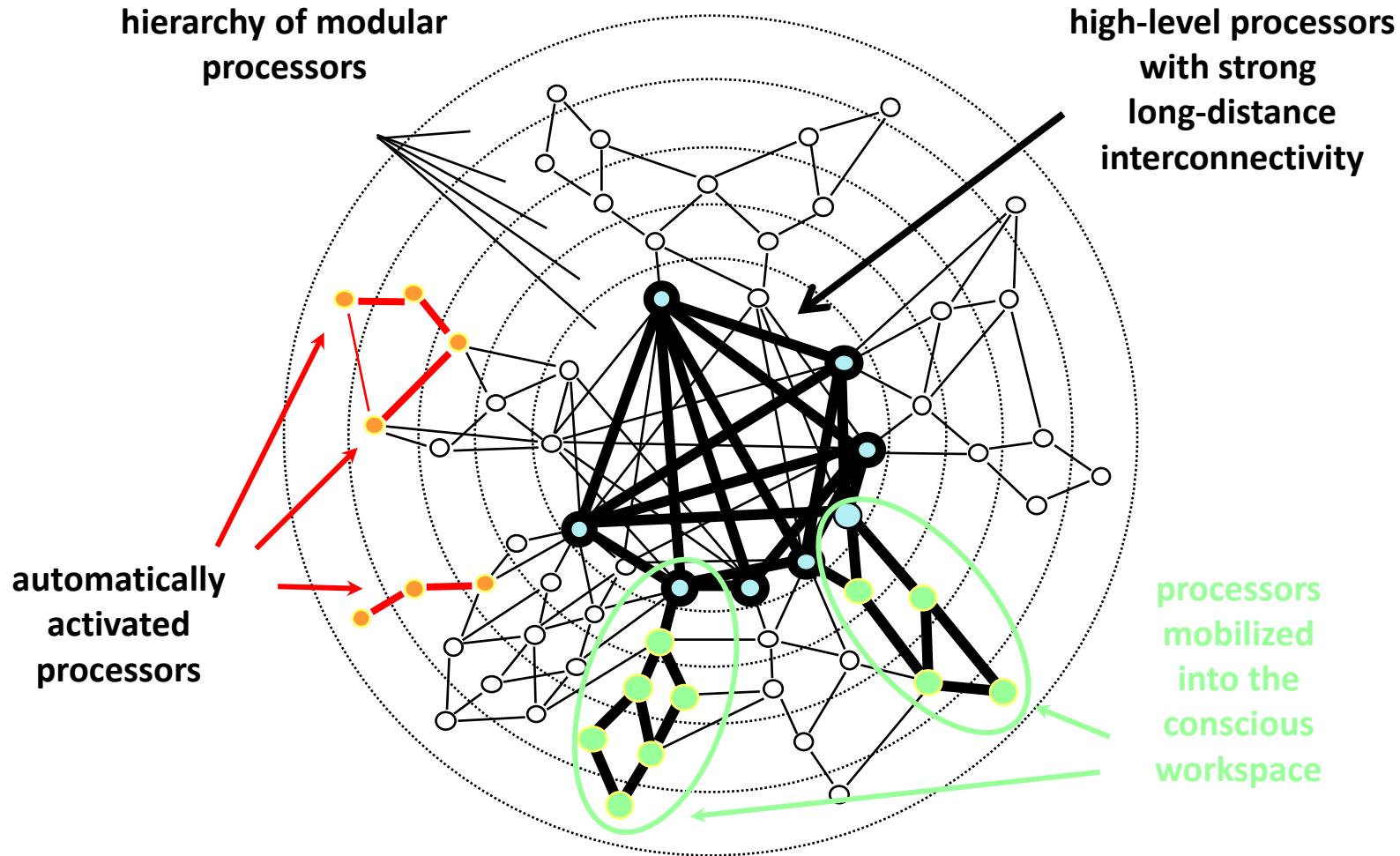


From Dehaene et al, 2001

The consciousness continuum and evidence accumulation

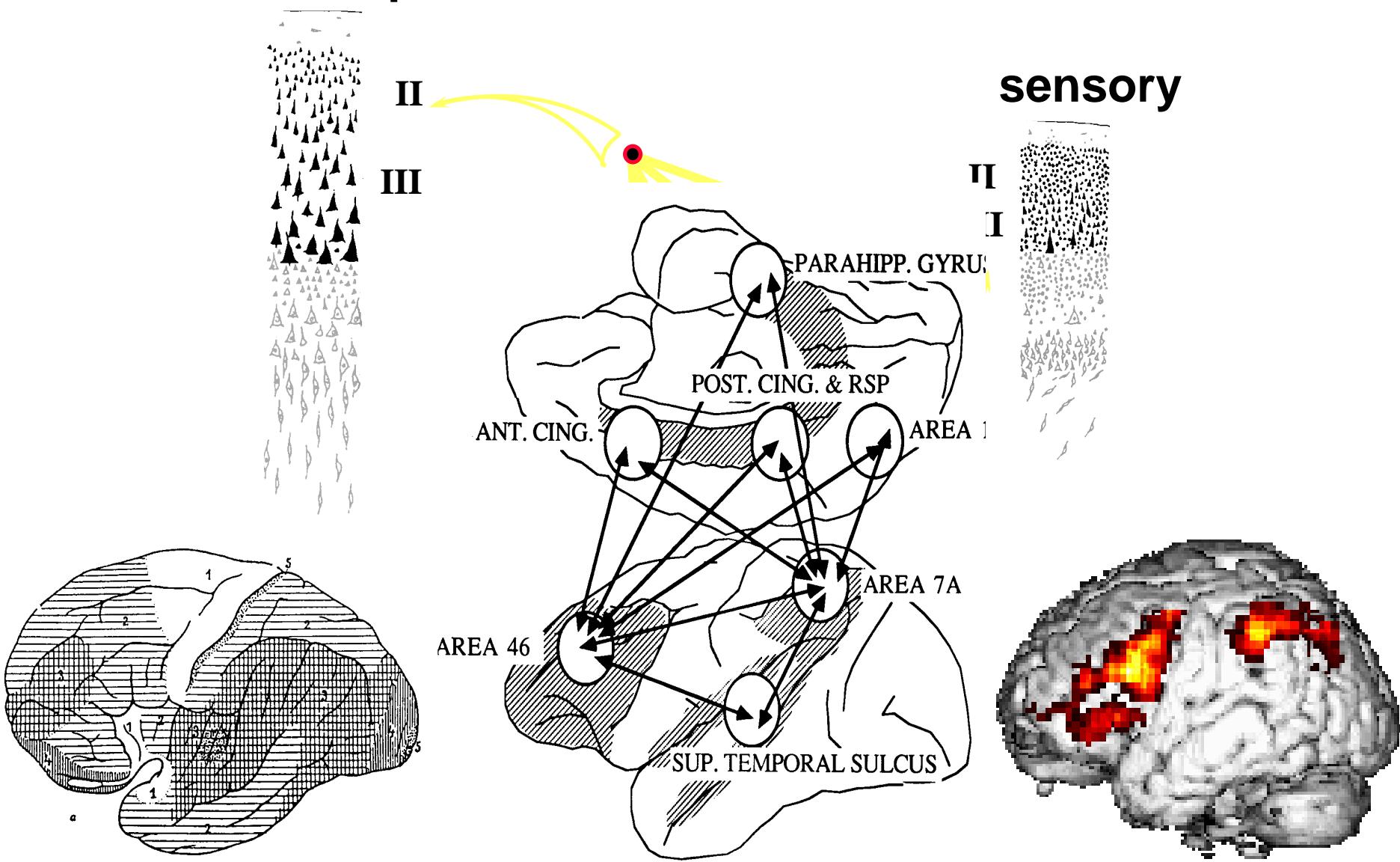


A schematic representation of the workspace model



Long-distance connectivity of Workspace Neurons: Putative role of layers II/III

Frontal / parietal



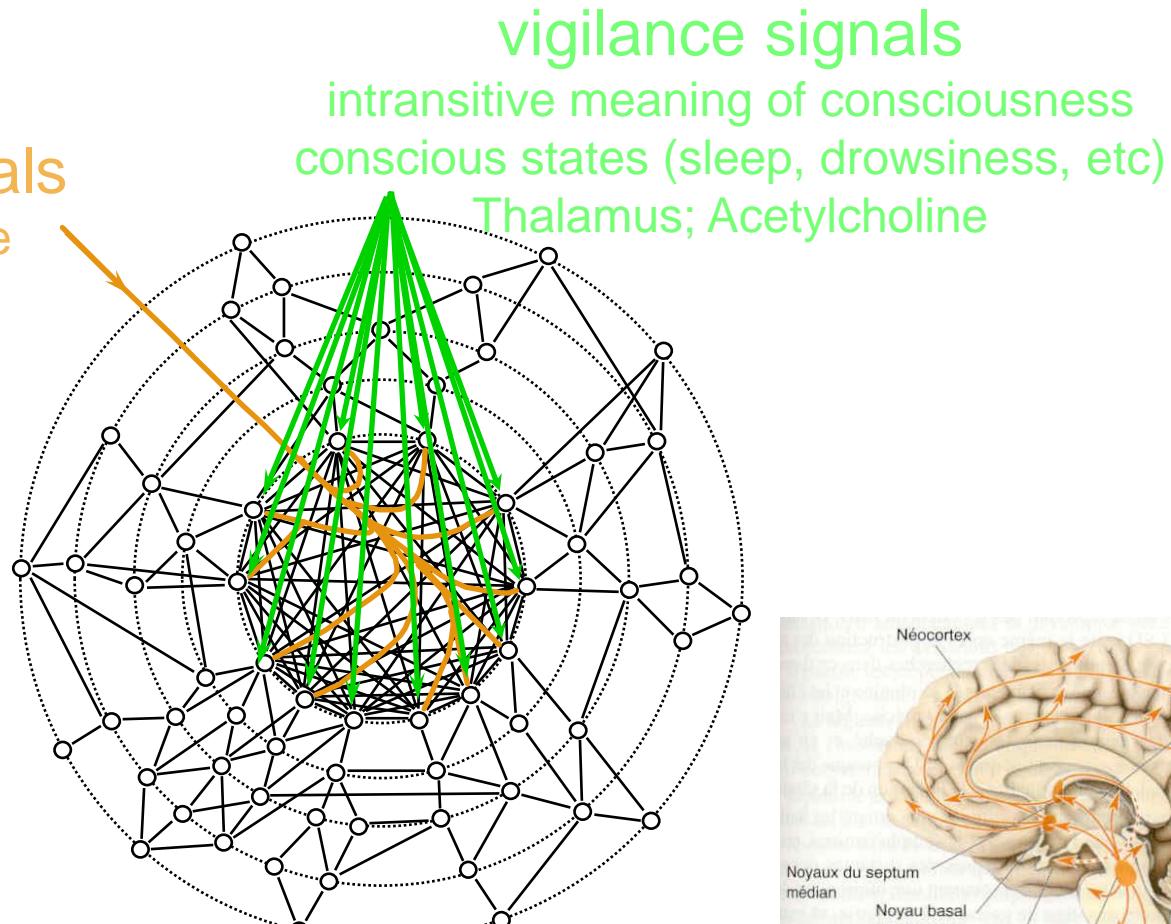
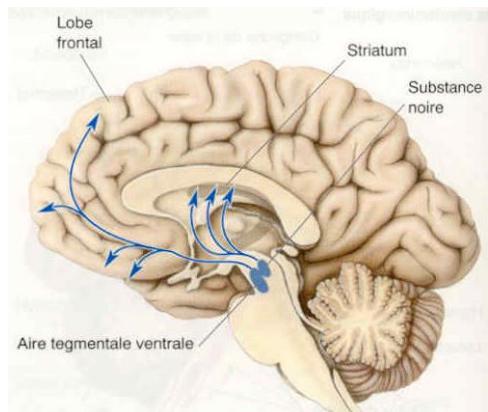
AUTONOMY OF WORKSPACE ACTIVITY

- Neurometabolism

evaluation signals

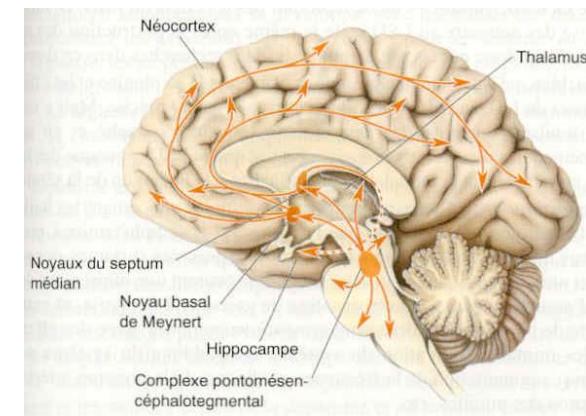
stimulus relevance

Dopamine



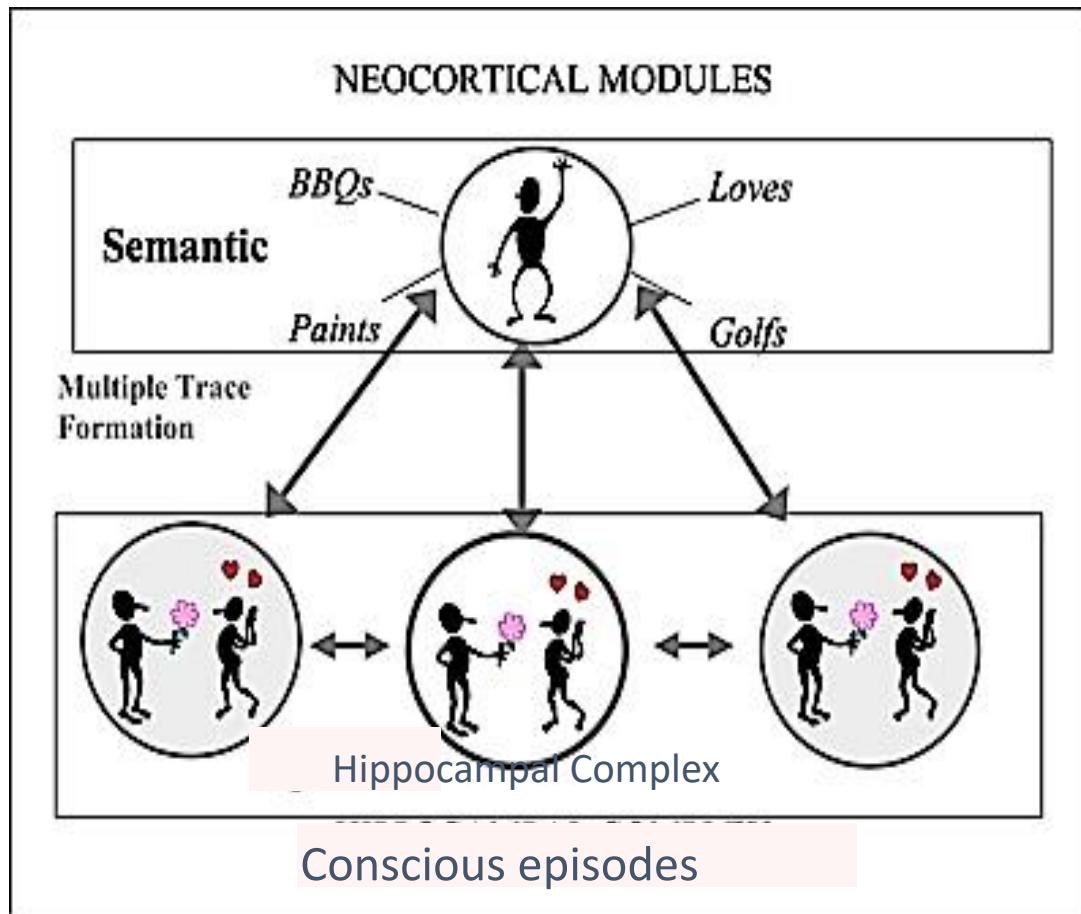
vigilance signals
intransitive meaning of consciousness
conscious states (sleep, drowsiness, etc)

Thalamus; Acetylcholine



Conscious input is *also* turned into longterm memory traces

--- via hippocampal-neocortical distribution.



Conscious visual input flows freely through the Hippocampal Complex) to be encoded in multiple distributed traces in neocortex.
(Traces are unconscious)

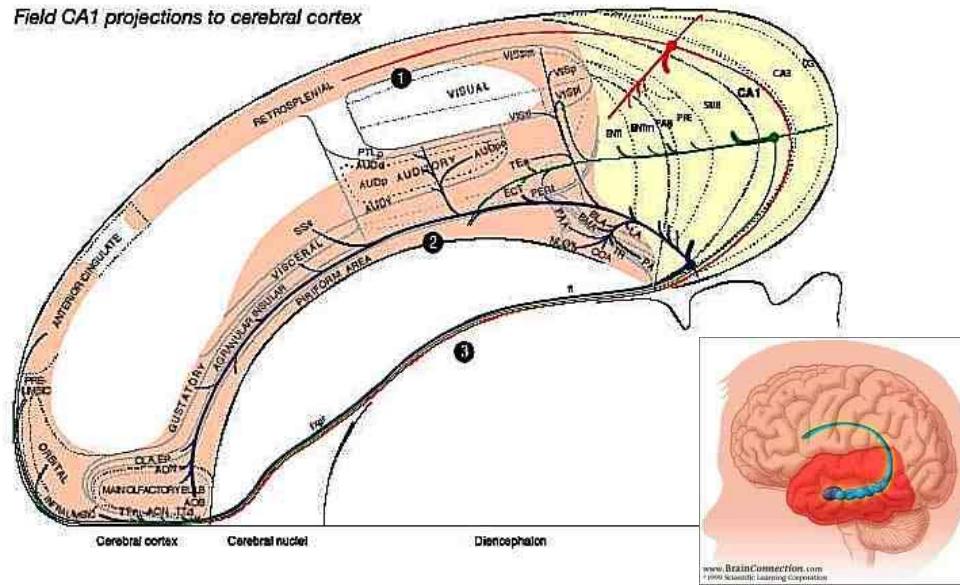
This allows neocortex to constantly *learn* and *update* itself with novel and significant information.

(Nadel & Moscovitch - Multiple Trace Theory. Figure from M. Moscovitch, personal comm.)

"Episodic memory" = memory for conscious episodes
HC = Hippocampal Complex, medial temporal lobe

Conscious input is also turned into longterm memory traces

--- via hippocampal-neocortical distribution.



Hippocampal connections to neocortex --- huge distribution.

This allows neocortex to constantly *learn* and *update* itself with novel and significant conscious information.

(Nadel & Moscovitch - Multiple Trace Theory. Figure from M. Moscovitch, personal comm.)

"Episodic memory" = memory for conscious episodes

The Hippocampal Complex includes neighboring regions in the medial temporal lobe.

Global Workspace Theory

- Bernard Baars (1988; Baars et al., 1998) argues that the function of consciousness is to broadcast information (accumulated evidence) to separate functional modules through-out the brain.
- His ‘global workspace’ is a central processor that contains the contents of consciousness (in long-term memory).
- Consciousness as a theatre

Global Workspace Theory

1. The modularity of mind

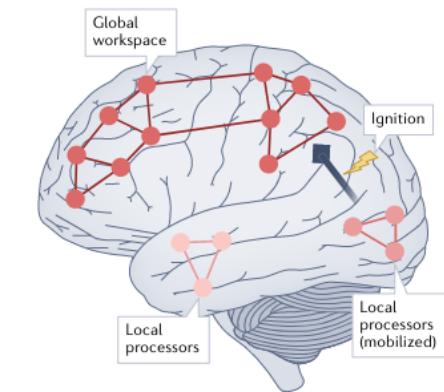
- A task, involving several mental operations, can proceed unconsciously whenever a set of adequately interconnected specialized processors is available to perform each of the required operations.

2. Attentional amplification and dynamic mobilization

- An information becomes conscious if the corresponding neural population is mobilized by top-down attentional amplification into a brain-scale state of coherent activity

3. The long-distance connectivity

- A distributed neural system with long-distance connectivity (the “conscious workspace”) can potentially interconnect multiple specialized processors in a coordinated manner (broadcasting).



Criticism

- Baar's theory does not address the problem of **phenomenal consciousness** (i.e., Chalmer's 'hard' problem).



David Chalmers

How does the physical brain give rise to
conscious experience?

Clinical Picture

Brain of a white-collar worker

Lionel Feuillet, Henry Dufour, Jean Pelletier

Lancet 2007; 370: 262

Department of Neurology
(L Feuillet MD, J Pelletier PhD),
and Department of
Neurosurgery (H Dufour PhD),
Faculté de Médecine de
Marseille, Université de la
Méditerranée, Assistance
Publique hôpitaux de
Marseille—Hôpital de la
Timone, Marseille, France

Correspondence to:
Dr Lionel Feuillet,
Department of Neurology,
Faculté de Médecine de Marseille,
Université de la Méditerranée,
Assistance Publique hôpitaux de
Marseille—Hôpital de la Timone,
Marseille, France
lionel.feuillet@mail.ap-hm.fr

A 44-year-old man presented with a 2-week history of mild left leg weakness. At the age of 6 months, he had undergone a ventriculoatrial shunt, because of postnatal hydrocephalus of unknown cause. When he was 14 years old, he developed ataxia and paresis of the left leg, which resolved entirely after shunt revision. His neurological development and medical history were otherwise normal. He was a married father of two children, and worked as a civil servant. On neuropsychological testing, he proved to have an intelligence quotient (IQ) of 75; his verbal IQ was 84, and his performance IQ 70. CT showed severe dilatation of the lateral ventricles (figure); MRI revealed massive enlargement of the lateral, third, and fourth ventricles, a very thin cortical mantle and a posterior fossa cyst. We diagnosed a non-communicating hydrocephalus, with probable stenosis of Magendie's foramen (figure). The leg weakness improved partly after neuro-endoscopic ventriculocisternostomy, but soon recurred; however, after a ventriculoperitoneal shunt was inserted, the findings on neurological examination became normal within a few weeks. The findings on neuropsychological testing and CT did not change.

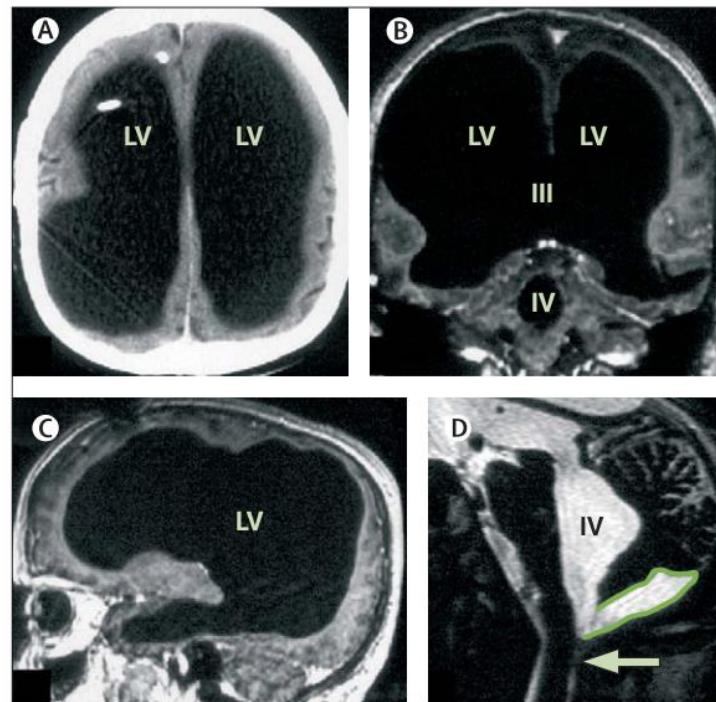


Figure: Massive ventricular enlargement, in a patient with normal social functioning

(A) CT; (B, C) T1-weighted MRI, with gadolinium contrast; (D) T2-weighted MRI.
LV=lateral ventricle. III=third ventricle. IV=fourth ventricle. Arrow=Magendie's foramen. The posterior fossa cyst is outlined in (D).



number of neurons?
particular region?
neural activity?

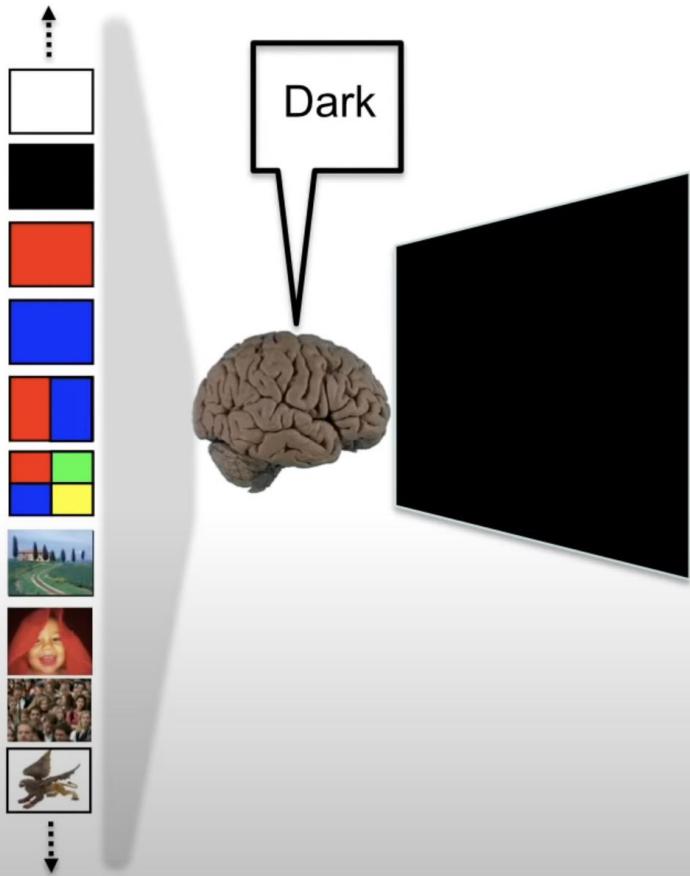
Larger number of neurons in the cerebellum than in the cortex

The brain of the white-collar worker

Epilepsy; Unresponsive Wakefulness Syndrome; MCS*

- Integrated information theory (IIT) approaches the problem of consciousness by starting not from the brain, but from experience itself.
- It does so by identifying the essential properties of experience (*axioms of phenomenal existence*) and inferring the required properties of its physical substrate (*postulates of physical existence*)

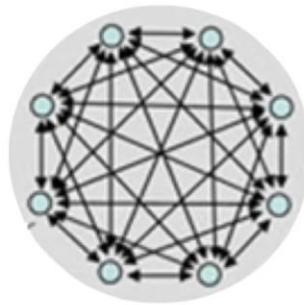
information



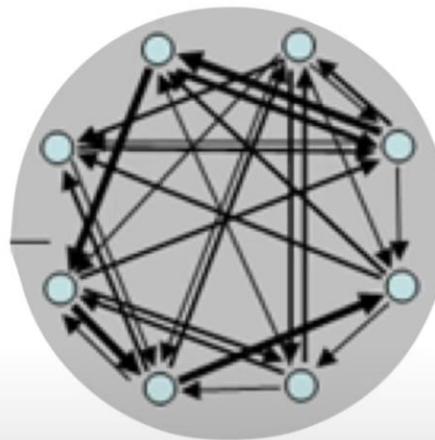
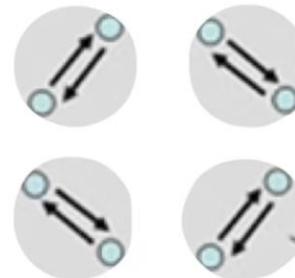
integration



no information



no integration



integration and information

Seth, Barrett, & Barnett (2011) *Phil Trans R Soc A*

Seth (2005) *Network*

Tononi (2008) *Biological Bulletin*

Tononi & Edelman (1998) *Science*



axioms
essential properties of every experience

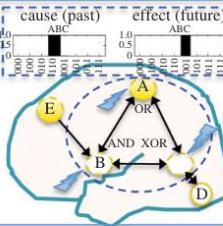


postulates
properties that physical systems (elements in a state)
must have to account for experience



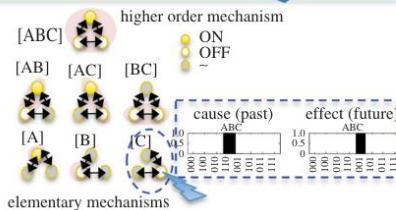
intrinsic existence
consciousness exists intrinsically:
each experience is real,
and it exists from its own
intrinsic perspective,
independent of
external observers (it is
intrinsically real)

to account for experience, a system of mechanisms
in a state must exist intrinsically. To exist, it must
have *cause-effect power*; to exist from its own
intrinsic perspective, independent of extrinsic
factors, it must have cause-effect power *upon itself*:
its present mechanisms and state must 'make a
difference' to the probability of some past and
future state of the system (its *cause-effect space*)



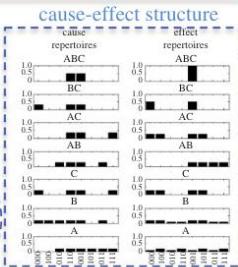
composition
consciousness is
structured: each
experience is
composed of
phenomenological
distinctions,
elementary or
higher-order,
which
exist within it

the system must be
structured: subsets of
system elements
(composed in various
combinations)
must have cause-effect
power upon the system



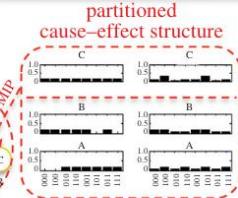
information
consciousness is
specific: each
experience is *the
particular way it is*
(it is composed of a
specific set of specific
phenomenological
distinctions), thereby
differing from other
possible experiences
(*differentiation*)

the system must specify a cause-effect structure
that is *the particular way it is*: a specific set of
specific cause-effect repertoires—thereby
differing in its specific way from other possible
structures (*differentiation*). A *cause-effect
repertoire* specifies the probability of all possible
causes and effects of a mechanism in a state. A
cause-effect structure is the set of
cause-effect repertoires specified
by all subsets of system elements and
expresses how the system gives an
actual form to the space of possibilities



integration
consciousness is
unified: each
experience is
irreducible to non-
interdependent
subsets of
phenomenal
distinctions

the cause-effect structure specified
by the system must be unified: it must be
intrinsically *irreducible* to that
specified by non-interdependent sub-
systems ($\Phi > 0$) across its weakest
(unidirectional) link:
MIP = minimum information partition



exclusion
consciousness is
definite, in
content and
spatio-temporal
grain: each
experience has
the set of
phenomenal
distinctions it has,
not less or more,
and flows at the
speed it does, not
faster or slower

the cause-effect structure specified
by the system must be definite:
specified over a *single* set of
elements—not less or
more—and spatio-
temporal grains—not
faster or slower;
this is a cause-effect
structure that is *maximally*
irreducible intrinsically (Φ_{\max}),
called *conceptual structure*,
made of maximally irreducible
cause-effect repertoires (*concepts*)

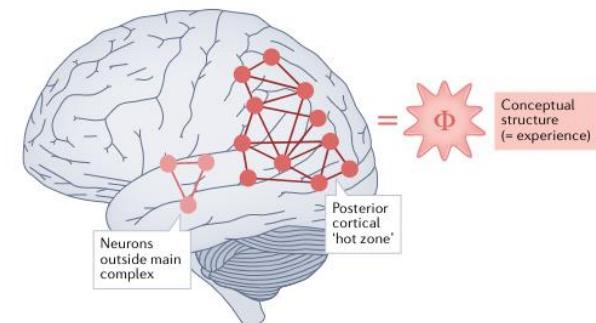
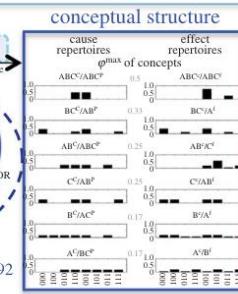


Figure 3. Axioms and postulates of integrated information theory (IIT). The illustration is a colourized version of Ernst Mach's 'View from the left eye' [84]. See also the mechanisms in figure 4.

Integrated Information theory

1. Intrinsic Existence

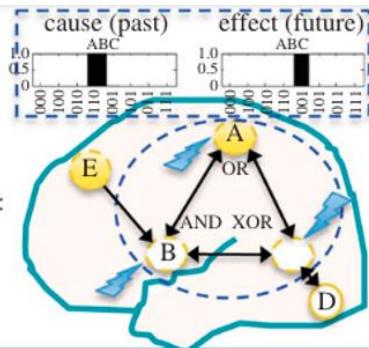
- Axiom: Consciousness exists intrinsically and it is personal: for example, I see my bedroom – is the only fact I can be immediately and absolutely sure of
- Postulate: Cause-effect power: simply mean a system that can be manipulated and after manipulation we see an effect upon him and not externally



intrinsic existence

consciousness exists intrinsically: each experience is real, and it exists from its own *intrinsic perspective*, independent of external observers (it is intrinsically real)

to account for experience, a system of mechanisms in a state must exist intrinsically. To exist, it must have *cause-effect power*; to exist from its own *intrinsic perspective*, independent of extrinsic factors, it must have cause-effect power *upon itself*: its present mechanisms and state must ‘make a difference’ to the probability of some past and future state of the system (its *cause-effect space*)



Integrated Information theory

2. Composition:

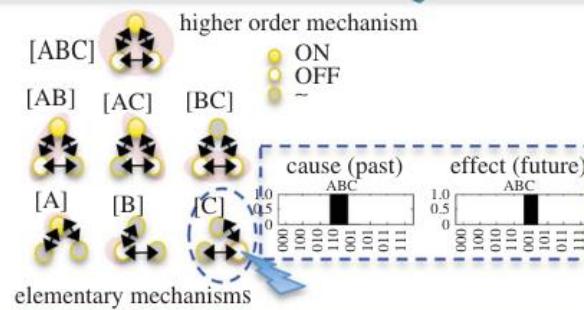
- Axiom: Consciousness is structured: each experience has internal structure, being composed of *phenomenal distinctions, bound together in various ways.* (richness of conscious content)
- Postulate: the system must be structured of subsets



consciousness is structured: each experience is composed of phenomenalological distinctions, elementary or higher-order, which *exist* within it

composition

the system must be structured: subsets of system elements (composed in various combinations) must have cause-effect power upon the system



Integrated Information theory

3. Information

- Axioms: Consciousness is specific. Each experience differs from other experiences
- Postulate: the system must have specific cause-effect repertoires (structure of the cause-effect is specific and different from other structures).

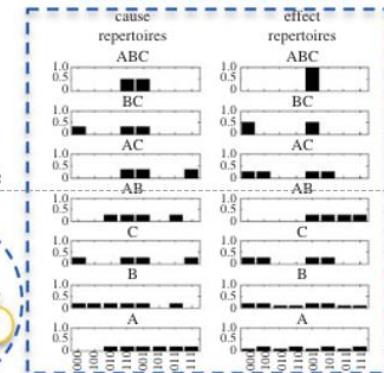


information

consciousness is specific: each experience is *the particular way it is* (it is composed of a specific set of specific phenomenological distinctions), thereby differing from other possible experiences (*differentiation*)

the system must specify a cause–effect structure that is *the particular way it is*: a specific set of specific cause–effect repertoires—thereby differing in its specific way from other possible structures (*differentiation*). A *cause–effect repertoire* specifies the probability of all possible causes and effects of a mechanism in a state. A *cause–effect structure* is the set of cause–effect repertoires specified by all subsets of system elements and expresses how the system gives an *actual form* to the space of possibilities

cause–effect structure



Integrated Information theory

4. Integration:

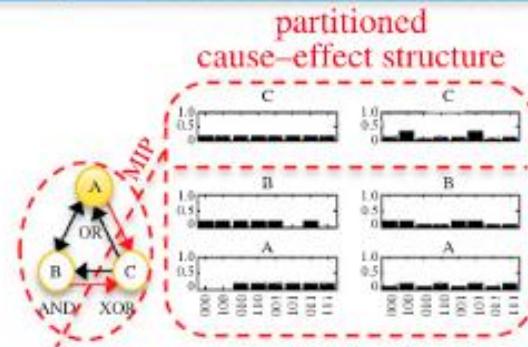
- Axiom: Consciousness is unitary: each experience is irreducible to non-interdependent components.



consciousness is unified: each experience is *irreducible* to non-interdependent subsets of phenomenal distinctions

integration

the cause–effect structure specified by the system must be unified: it must be intrinsically *irreducible* to that specified by non-interdependent subsystems ($\Phi > 0$) across its weakest (unidirectional) link:
MIP = minimum information partition



Integrated Information theory

4. Integration:

- Postulate: Can be calculated by the phi, which reflects to what extent a system can generate more information than the sum of his parts

OPEN ACCESS Freely available online

PLOS COMPUTATIONAL BIOLOGY

Practical Measures of Integrated Information for Time-Series Data

Adam B. Barrett*, Anil K. Seth

Sackler Centre for Consciousness Science and School of Informatics, University of Sussex, Brighton, United Kingdom

Abstract

A recent measure of ‘Integrated information’, Φ_{int} , quantifies the extent to which a system generates more information than the sum of its parts as it transitions between states, possibly reflecting levels of consciousness generated by neural systems. However, Φ_{int} is defined only for discrete Markov systems, which are unusual in biology; as a result, Φ_{int} can rarely be measured in practice. Here, we describe two new measures, Φ_1 and Φ_{int} , that overcome these limitations and are easy to apply to time-series data. We use simulations to demonstrate the in-practice applicability of our measures, and to explore their properties. Our results provide new opportunities for examining information integration in real and model systems and carry implications for relations between integrated information, consciousness, and other neurocognitive processes. However, our findings pose challenges for theories that ascribe physical meaning to the measured quantities.

Citation: Barrett AB, Seth AK (2011) Practical Measures of Integrated Information for Time-Series Data. PLoS Comput Biol 7(7): e1001952. doi:10.1371/journal.pcbi.1001952

Editor: Olaf Sporns, Indiana University, United States of America

Received June 20, 2010; Accepted December 6, 2010; Published January 30, 2011

Copyright: © 2011 Barrett, Seth. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

unding: AB is supported by EPSRC Leadership Fellowship EP/G007543/1, which also supports the work of AKB. Support is also gratefully acknowledged from the Dr. Maxine and Theresa Sackler Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: adam.barrett@sussex.ac.uk



Phil. Trans. R. Soc. A (2011) 369, 3748–3767
doi:10.1098/rsta.2011.0079

Causal density and integrated information as measures of conscious level

BY ANIL K. SETH*, ADAM B. BARRETT AND LIONEL BARNETT

Sackler Centre for Consciousness Science, and School of Informatics,
University of Sussex, Brighton BN1 9QJ, UK

$$\text{CD}(\mathbf{X}) =: \frac{1}{n(n-1)} \sum_{i \neq j} \mathcal{F}_{X_i \rightarrow X_j | X_{\{ij\}}}$$

$$\tilde{\varphi}[X; \tau, \{M^1, M^2\}] = : \sum_{k=1}^2 H(M_{t-\tau}^k | M_t^k) - H(X_{t-\tau} | X_t)$$

Seth (2005) Network

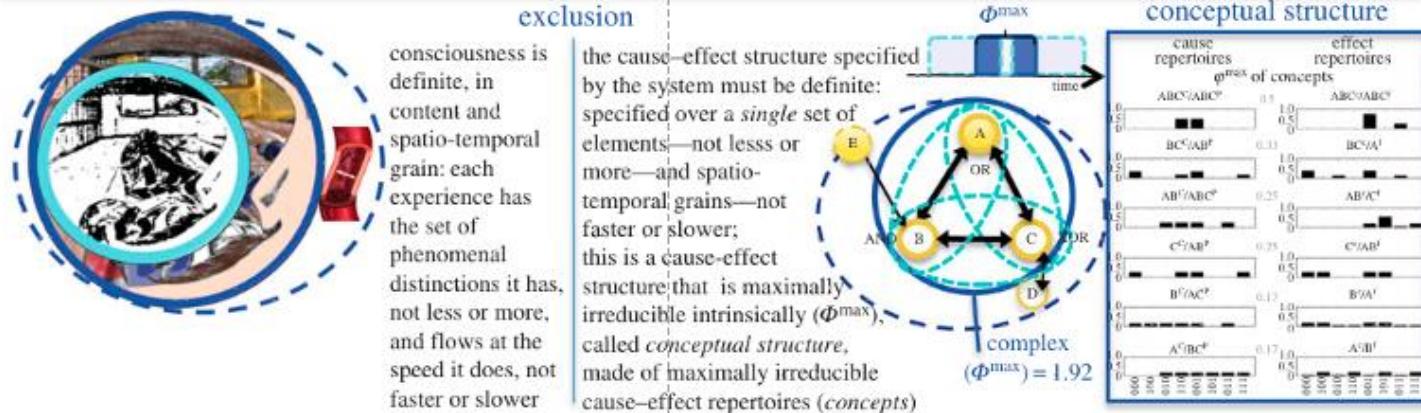
Seth, Barrett, & Barnett (2011) *Phil Trans R Soc A*

Barrett & Seth (2011) *PLoS Comp Biol*

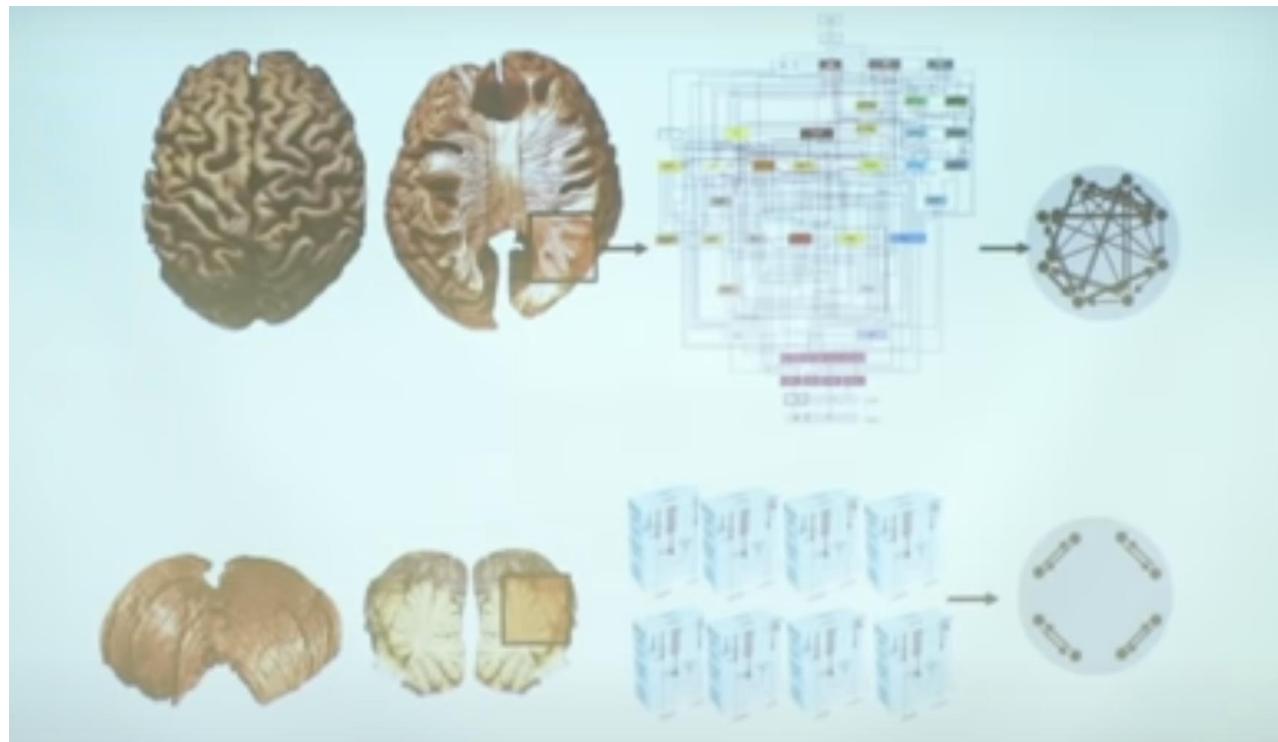
Integrated Information theory

5. Exclusion:

- Axiom: Consciousness is definite, thus my experience has just the content it has.
- Postulate: The system must exclude other cause-effect repertoires.

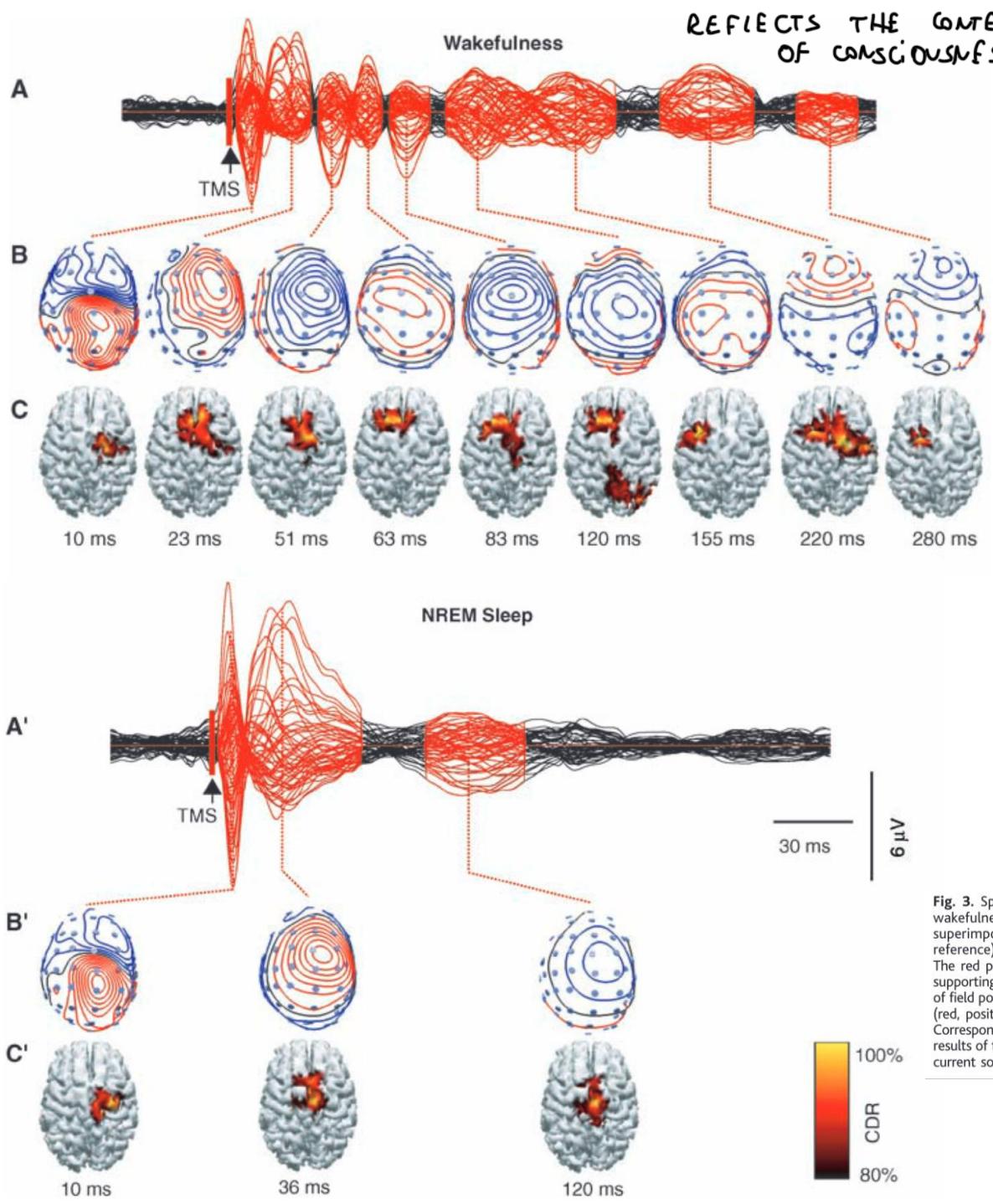


Computer simulations and brain structure show that systems that specify conceptual structures of high ϕ^{\max} must be effectively interconnected (integrated) and have a large Repertory of differentiate states (information)



THE MAGNETIC FIELDS PERTURBS THE ELECTRICAL ACTIVITY OF THE BRAIN

This prediction has been addressed using transcranial magnetic stimulation (TMS) in combination with high-density electroencephalography (EEG) in subjects who were alternately awake and conscious, asleep and virtually unconscious



The PCI
 You can treat the continuos
activity as a photo or a series
of photo and compress it as
the shortest file size that you
need to regenerate the photo,
this is a measure of
complexity or in other words
how much information I need
to describe this pattern and
you give a number to that
(PCI) perturbation complexity
index, which is in principle the
beginning of a consciousness
meter and reflect the
complexity of the processed
information.

Fig. 3. Spatiotemporal dynamics of scalp voltages and cortical currents evoked by TMS during wakefulness and sleep. (A and A') Averaged TMS-evoked potentials recorded at all electrodes, superimposed in a butterfly diagram (black traces; the horizontal red line indicates the average reference), for the same subject as in Figs. 1 and 2. The time of TMS is marked by a vertical red bar. The red portions of the traces indicate the times at which TMS induced a significant response (see supporting online material for calculation details). Source modeling was performed at the local maxima of field power within periods of significant activity. (B and B') Three-dimensional contour voltage maps (red, positive; blue, negative; step = 0.6 μ V for wakefulness and 1 μ V for NREM sleep). (C and C') Corresponding current density distributions plotted on the cortical surface. At each time point, the results of the L2 Norm (see methods) were auto-scaled and thresholded at 80% to highlight maximum current sources (CDR, current density reconstruction).

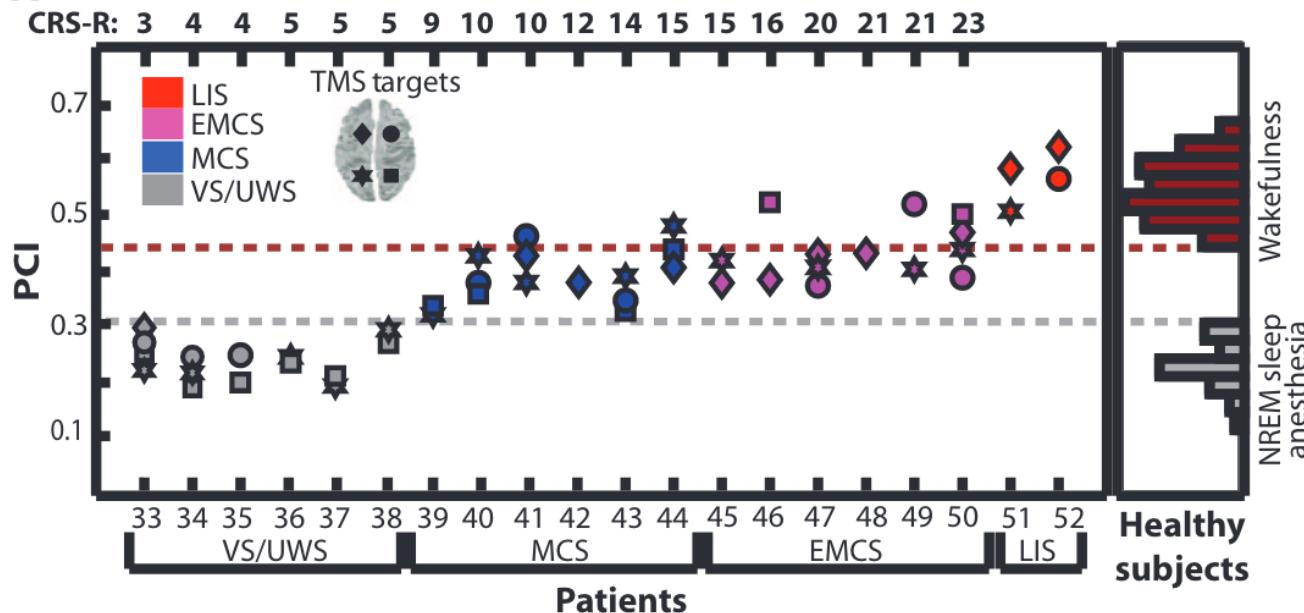
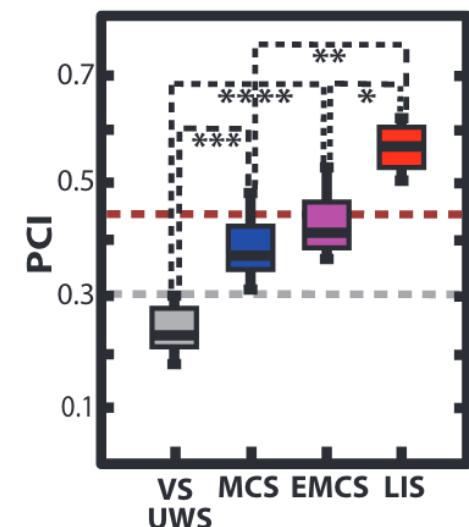
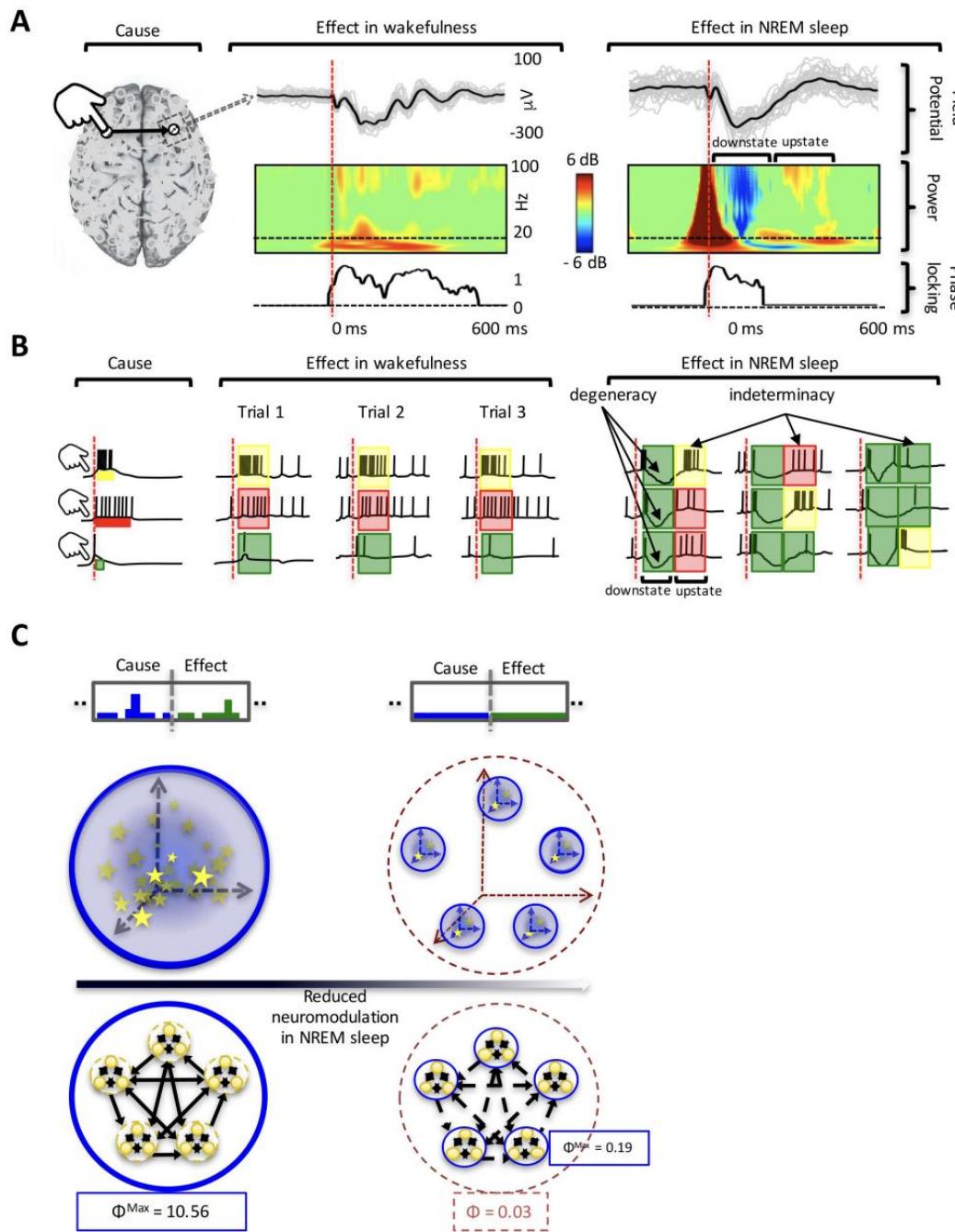
A**B**

Fig. 5. PCI discriminates the level of consciousness in brain-injured patients. (A) PCI values for 48 TMS sessions collected from 20 severely brain-injured patients (TMS was targeted to both left and right BA08 and BA07, as indicated at top left). Right: Distribution of PCI values from healthy individuals. **(B)** Box plots for PCI in brain-injured patients with the statistical significance between pairs of conditions (LMM: $*P = 0.002$, $**P = 0.0001$, $***P = 2 \times 10^{-5}$, $****P = 8 \times 10^{-7}$). Gray and red dashed lines in (A) and (B) represent the maximum complexity observed during unconsciousness ($PCI = 0.31$) and the minimum complexity observed during alert wakefulness ($PCI = 0.44$) in healthy subjects, respectively.

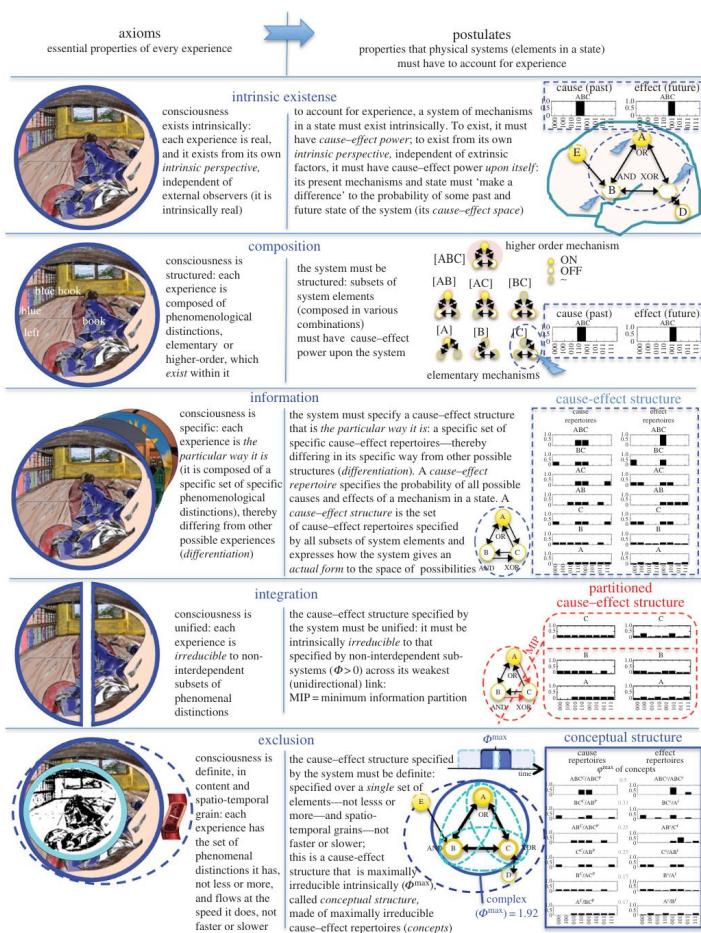
SUPPLEMENTARY INFORMATION

In format provided by Tononi *et al.* (JULY 2016)



Integrated Information theory

(IIT) approaches the problem of consciousness by starting from phenomenology
(axioms and postulates)



The PCI and clinical Applications in disorders of consciousness

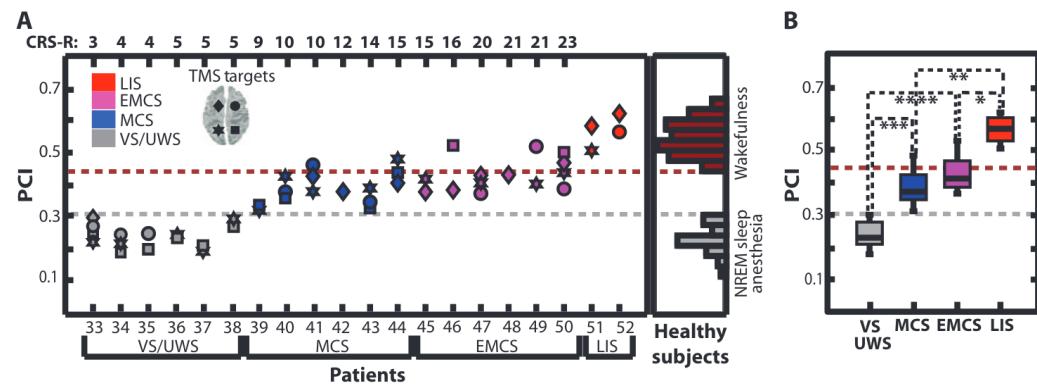


Fig. 5. PCI discriminates the level of consciousness in brain-injured patients. (A) PCI values for 48 TMS sessions collected from 20 severely brain-injured patients (TMS was targeted to both left and right BA08 and BA07, as indicated at top left). Right: Distribution of PCI values from healthy individuals. (B) Box plots for PCI in brain-injured patients with the statistical significance between pairs of conditions (LMM: *P = 0.002, **P = 0.0001, ***P = 2 × 10⁻⁵, ****P = 8 × 10⁻⁷). Gray and red dashed lines in (A) and (B) represent the maximum complexity observed during unconsciousness (PCI = 0.31) and the minimum complexity observed during alert wakefulness (PCI = 0.44) in healthy subjects, respectively.

Figure 3. Axioms and postulates of integrated information theory (IIT). The illustration is a coloured version of Ernst Mach's 'View from the left eye' [84]. See also the mechanisms in figure 4.

Important Issues in Consciousness Studies and AI

- Can computers ever duplicate human consciousness?
- Is consciousness really necessary in cognition?
 - **The Touring Test**
 - **The Chinese room**

CAN A
COMPUTER
TALK LIKE A
HUMAN?



会！流利地

Yes! fluently



Does the brain work like a computer?

- Over the last 40 years cognitive psychology has used computers as an analogy for the human brain.
- General-purpose computers have three main features:
 - (1) Input and output devices that allow the user to communicate with the computer.
 - (2) A memory system that permits the storage of information.
 - (3) A central processor that controls the major functions of the computer.

THE ONLY WAY TO REALLY ASSES CONSCIOUSNESS IS TO ASK

WE DON'T REALLY KNOW
WHAT IT IS

WE CAN'T SAY
AN ANIMAL IS
UNCONSCIOUS

Does the brain work like a computer?

- In principle a computer can be programmed to duplicate the principle cognitive functions of perception, memory, and problem-solving.
- The construction of computer programs that simulate human mental functions is called **artificial intelligence**.

Does the brain work like a computer?

Requirements

- Computer simulations require **formal modelling** of cognitive functions.
- Establishing what steps are necessary for a computer to simulate a cognitive ability may give insight into the kinds of process the brain must perform.
- Critics argue that computers may perform tasks in an entirely different way to the human brain.

Does the brain work like a computer?

What about consciousness?

- Strong A.I. position states that a computer which exactly duplicates the functions of the brain would by definition become conscious.
- Critics argue that true consciousness can never be achieved by an artificial non-organic system.

Is Consciousness Really Necessary?

Philosophers use the concept of Cognitive **zombies**.



MAYBE, SINCE ZOMBIES DON'T EXIST, IT COULD MEAN THAT IT'S NOT POSSIBLE TO WORK IN A CERTAIN WAY AND BE UNCONSCIOUS

- 'Zombies' are hypothetical beings that possess exactly the same cognitive processes as we do, but without conscious experience.
- Conscious experience does not appear to be an **inevitable** consequence of cognitive and neural processing.
- The majority of cognitive models do not feature a functional role for phenomenal consciousness.
- Some argue that phenomenal consciousness is an **epiphenomenon** of neural and cognitive processing (i.e., a by-product that plays no functional role in the system).



MARY'S ROOM

www

A PHILOSOPHICAL THOUGHT EXPERIMENT



QUALIA → see better

«Consciousness is a fascinating but elusive phenomenon: it is impossible to specify what is, what it does, or why it evolved. Nothing worth reading has been written on it»

Sutherland (1989) the international dictionary of Psychology

