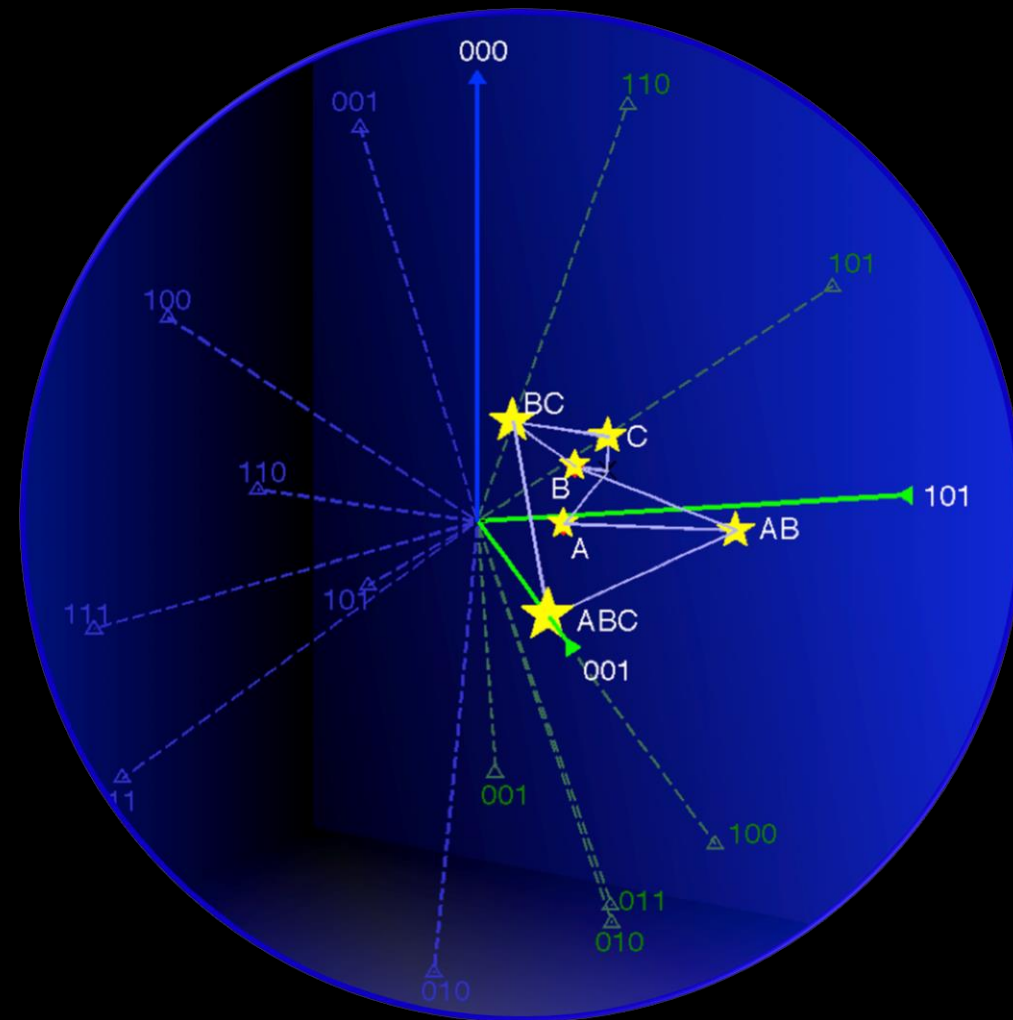


What is consciousness, and why did it evolve?

A view from within



Integrated Information Theory (IIT)

- **Starts from phenomenology**, not from behavioral or neural correlates
- Identifies the essential properties of every experience (**axioms**)
- Derives the requirements that physical systems must satisfy to account for them (**postulates**)
- Has **predictive, explanatory, and inferential** power

Tononi 2004, 2008, 2012

Balduzzi and Tononi, 2008, 2009

Oizumi, Albantakis and Tononi, 2014

Tononi and Koch, 2014

Intrinsic existence



Intrinsic existence



Composition



Intrinsic existence



Composition



Information



Intrinsic existence



Composition



Information



Integration



Intrinsic existence



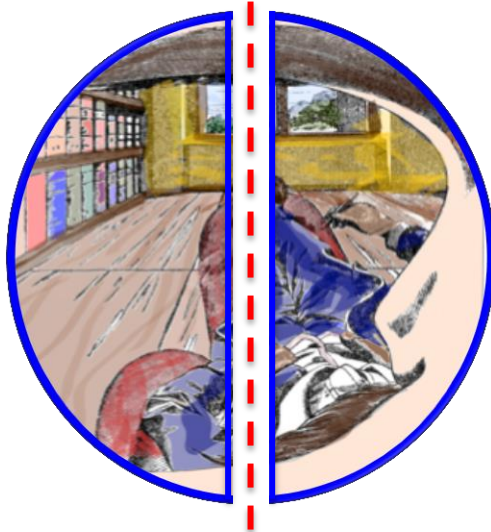
Composition



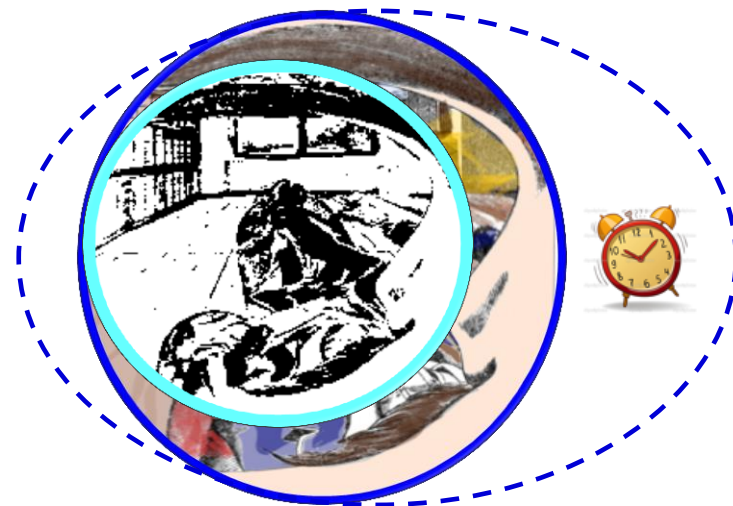
Information



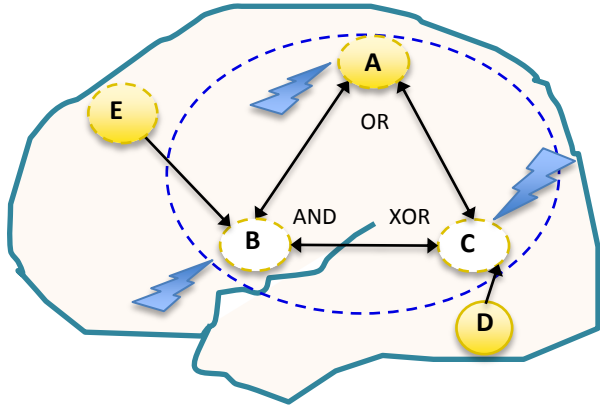
Integration



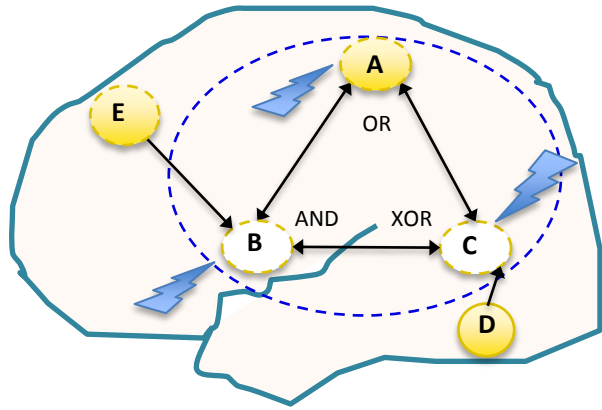
Exclusion



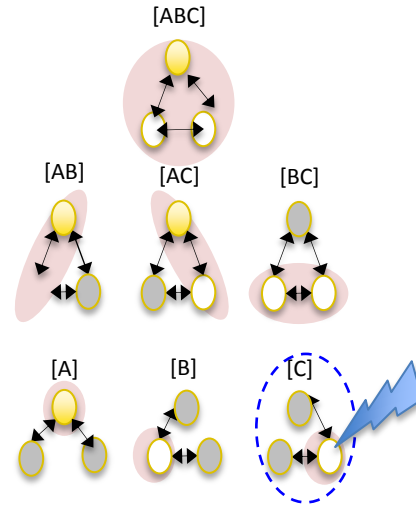
Intrinsic existence



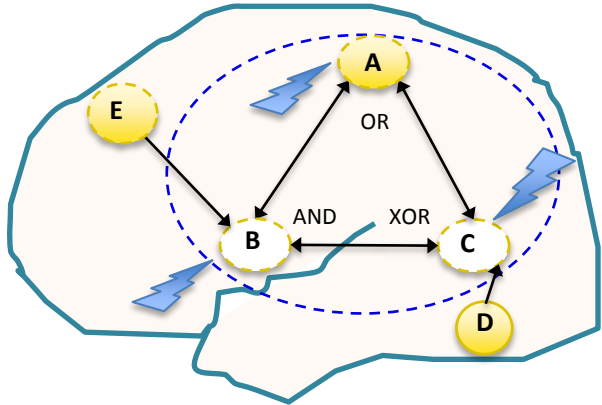
Intrinsic existence



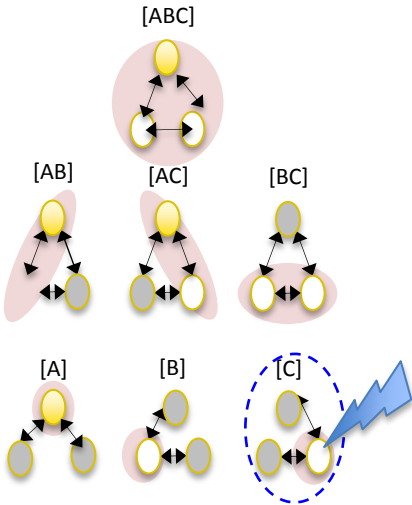
Composition



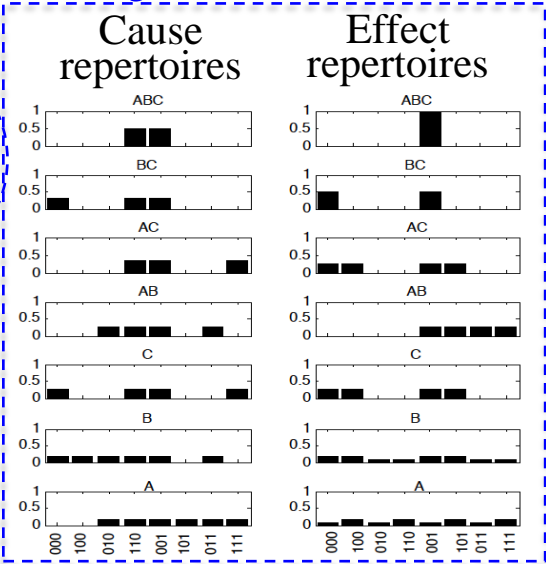
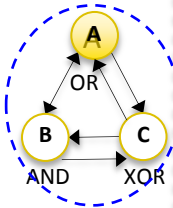
Intrinsic existence



Composition

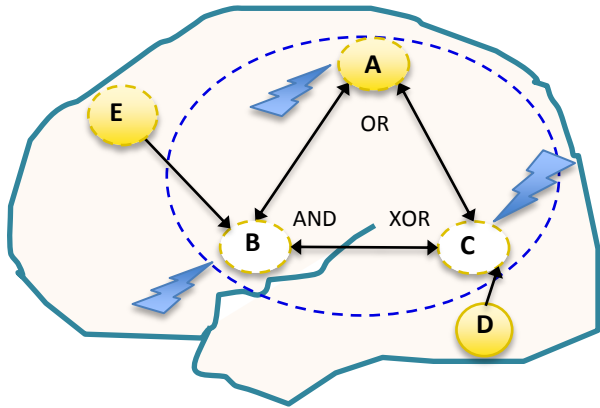


Information

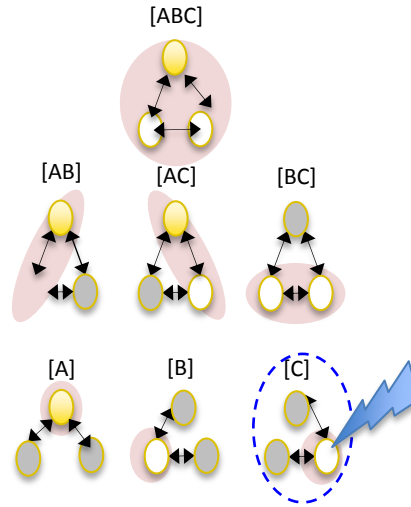


Cause-effect structure

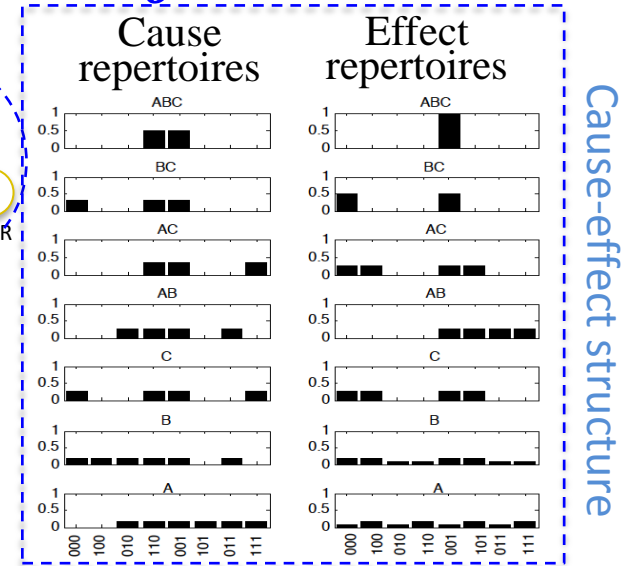
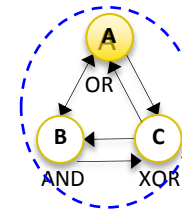
Intrinsic existence



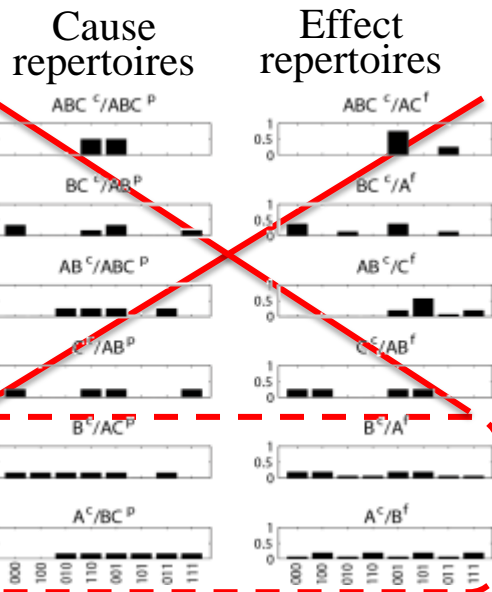
Composition



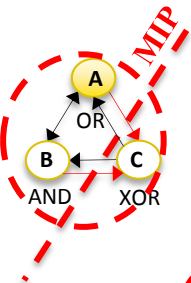
Information



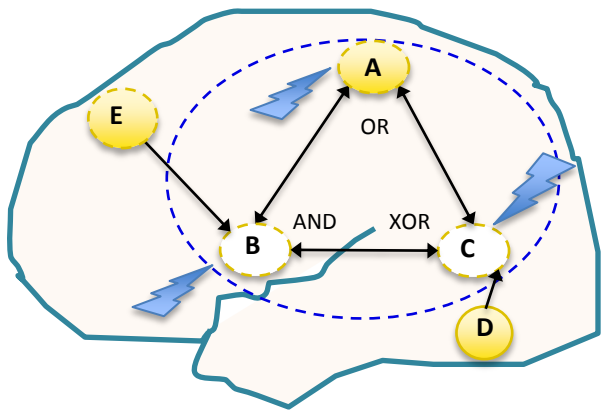
Integration



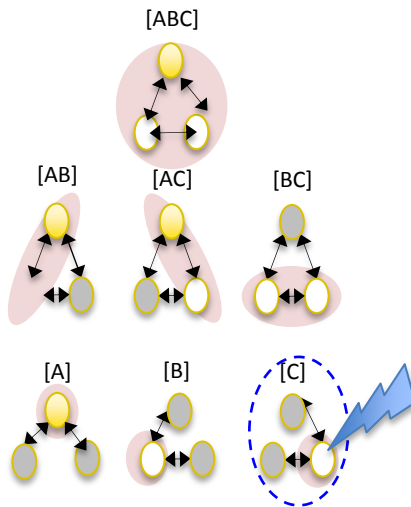
Cause-effect structure
of partitioned system



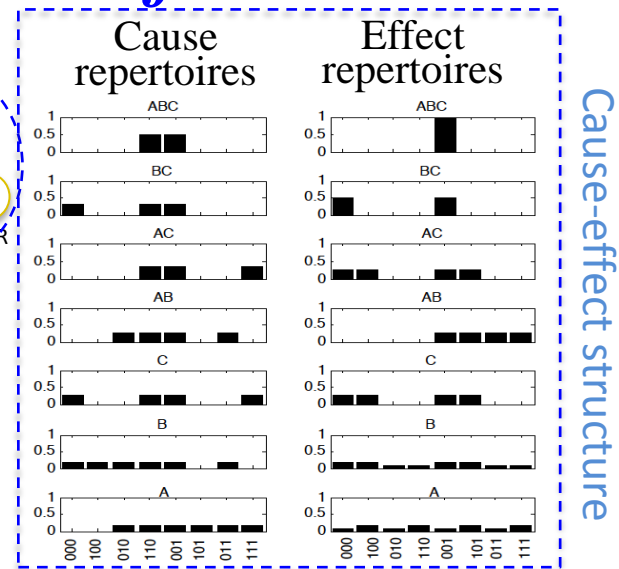
Intrinsic existence



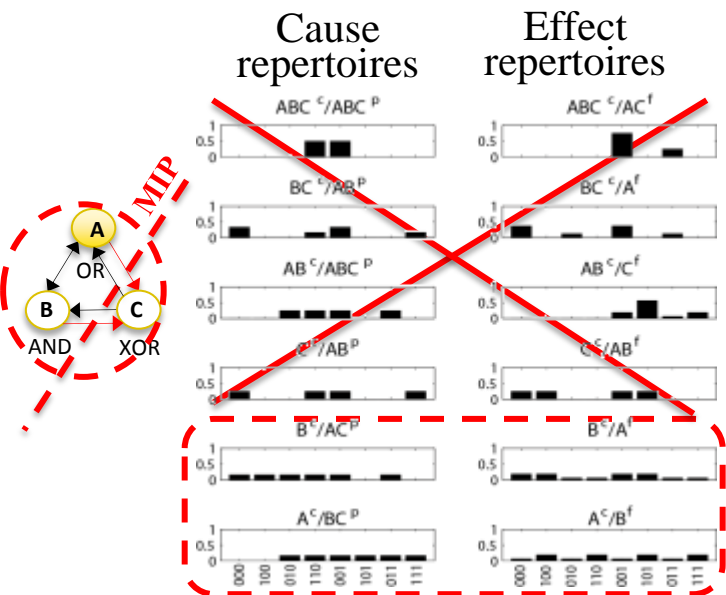
Composition



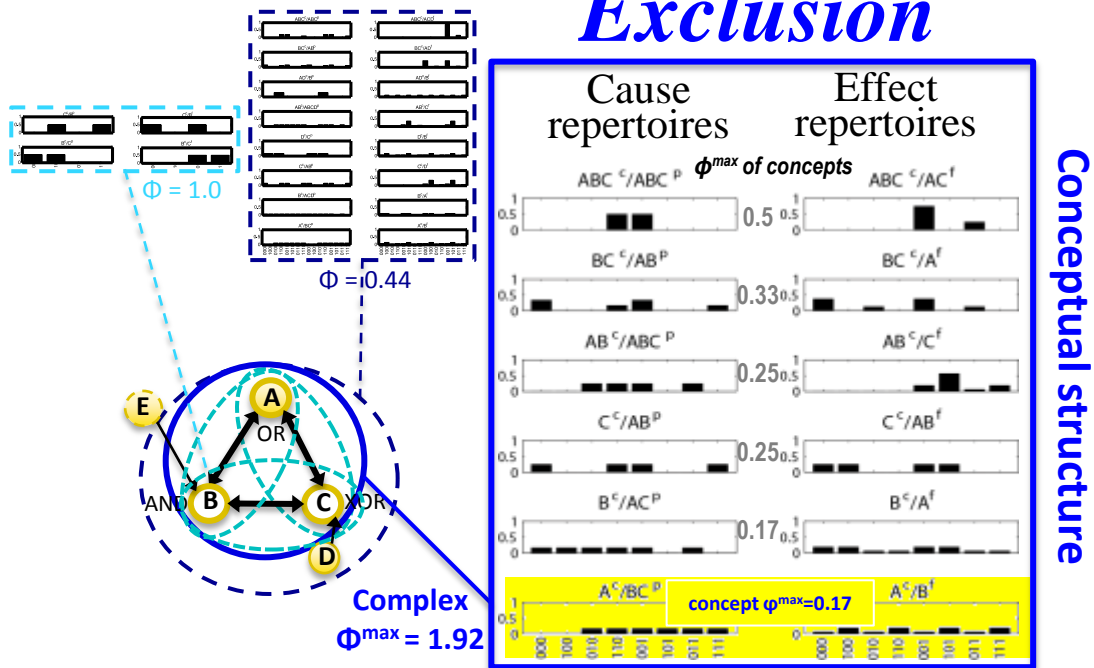
Information



Integration



Exclusion

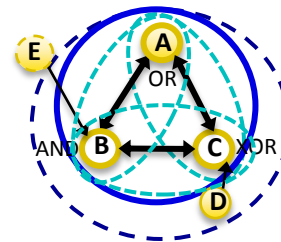
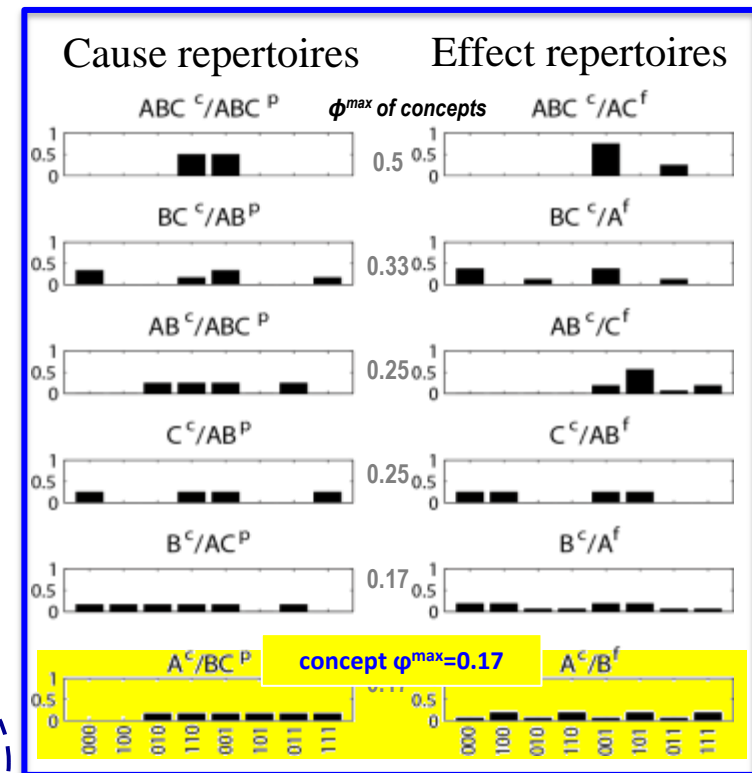


The central identity: An experience is a conceptual structure
a maximally irreducible cause-effect structure made of concepts (maximally irreducible cause-effect repertoires) specified by a complex of mechanisms in a state



≡

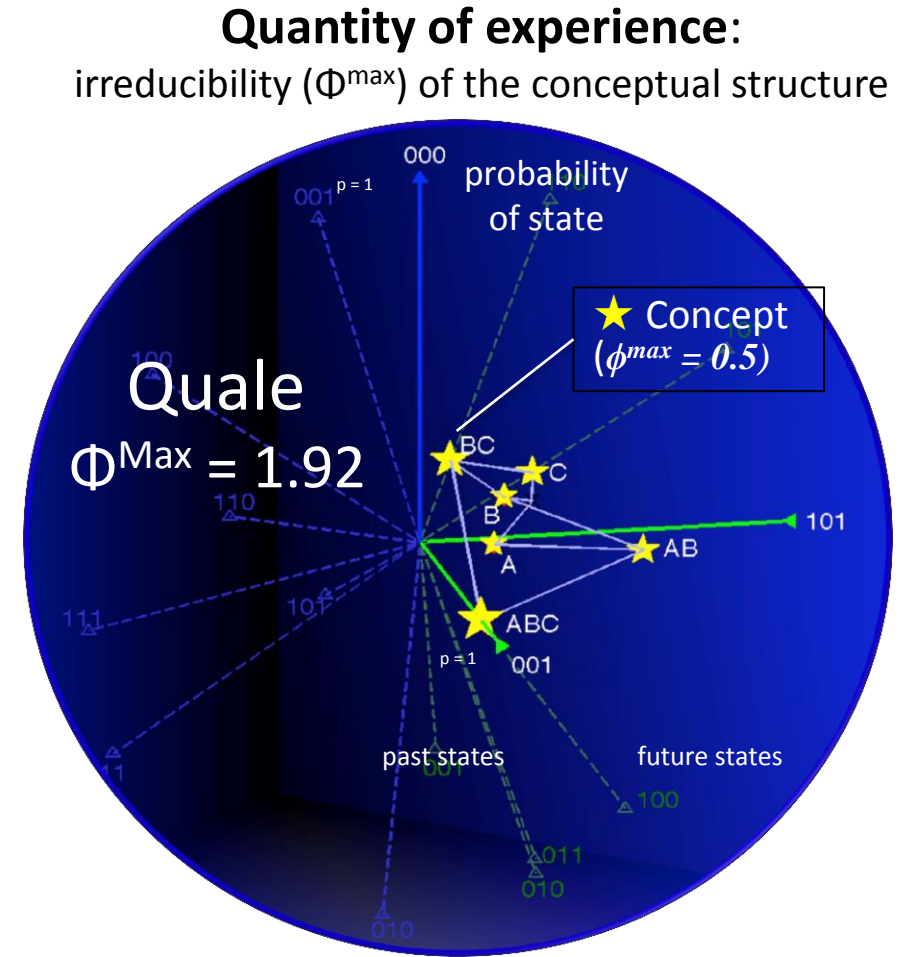
Conceptual structure Q



Complex
 $\phi^{\max} = 1.92$



The central identity: An experience is a conceptual structure
(a “form” Q in cause-effect space)





Quality of experience:
“form” of the conceptual structure

From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0

Masafumi Oizumi^{1,2}, Larissa Albantakis¹, Giulio Tononi^{1*}

¹ Department of Psychiatry, University of Wisconsin, Madison, Wisconsin, United States of America, ² RIKEN Brain Science Institute, Wako-shi, Saitama, Japan

Integrated Information Theory (IIT)

- **Predictions**
- **Explanations**
- **Extrapolations**

IIT: some predictions

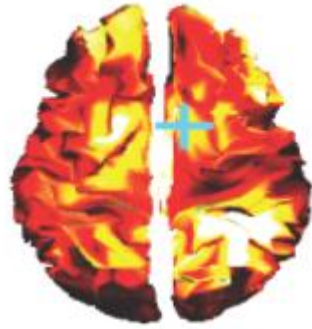
From theory to practice: Evaluating integrated information using TMS and hd-EEG



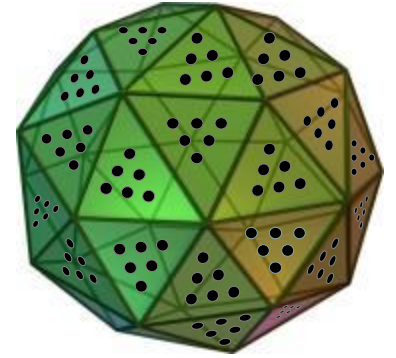
(from Massimini et al., Science TM, 2005)

Like consciousness, information integration is high in wake, breaks down in slow wave sleep, and returns during REM sleep

Wake



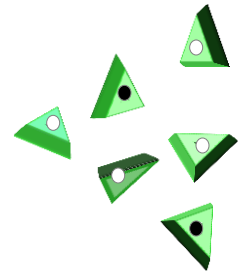
Highest inf. integration



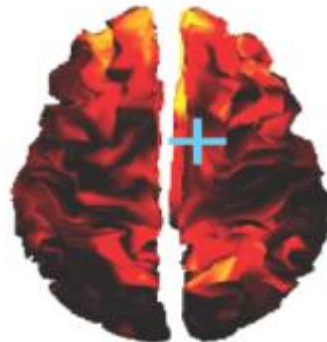
Early NREM Sleep



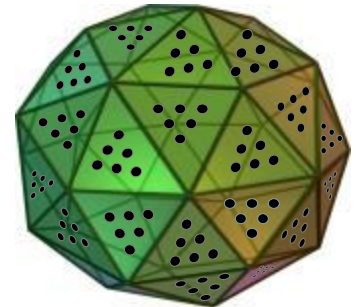
Low inf. integration



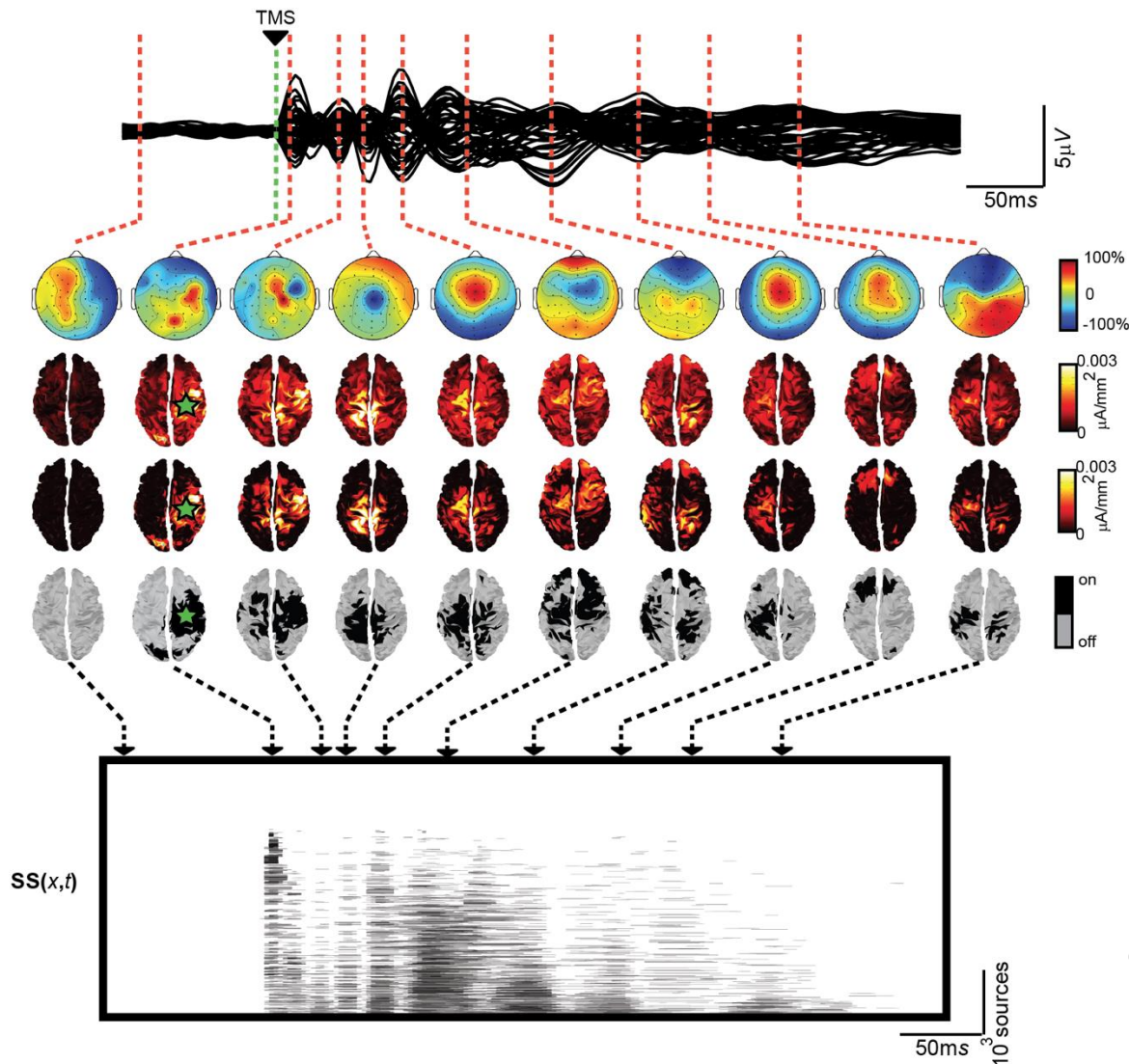
REM Sleep



High inf. integration



Towards a Consciousness – Meter: “zap and zip”



A. Time course of TMS-hdEEG responses

A. Voltage maps

B. Current sources

C. Significant sources (nonparametric)

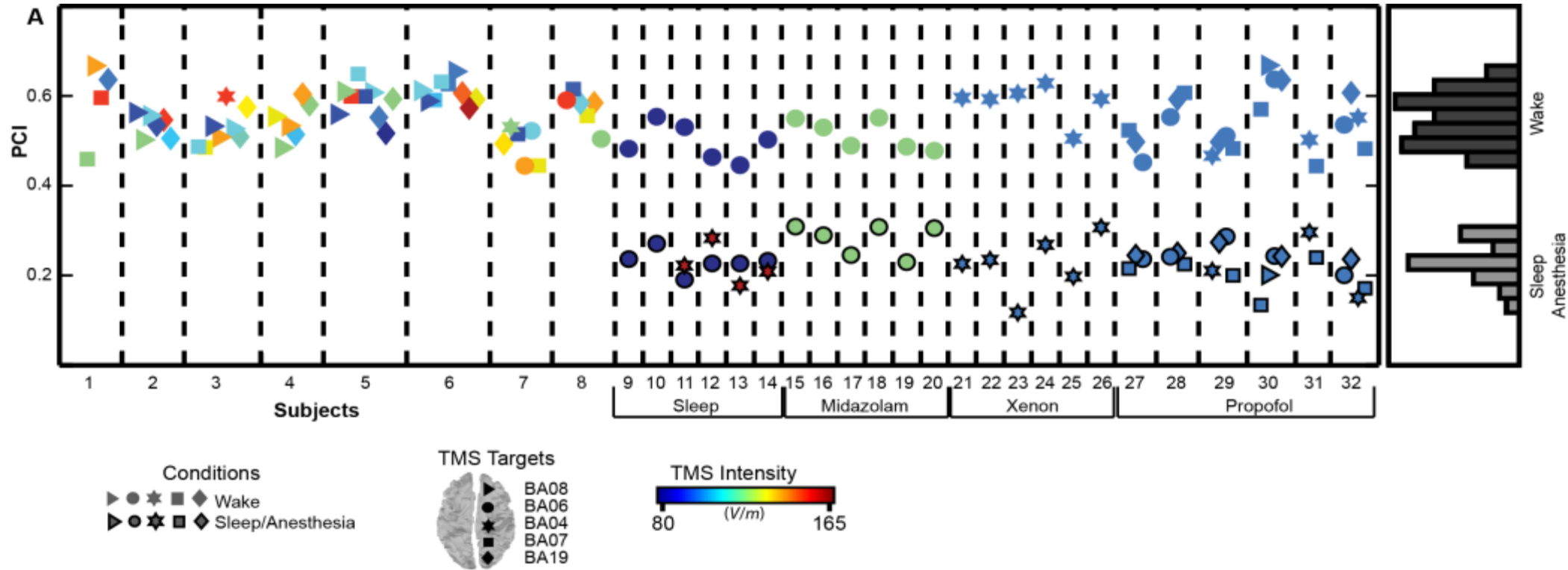
D. Binarized matrix

Perturbational Complexity Index (PCI),
a practical measure of information integration
using TMS (“zapping”)

computed using Lempel-Ziv encoding of hd-EEG
sources time series (“zipping”)

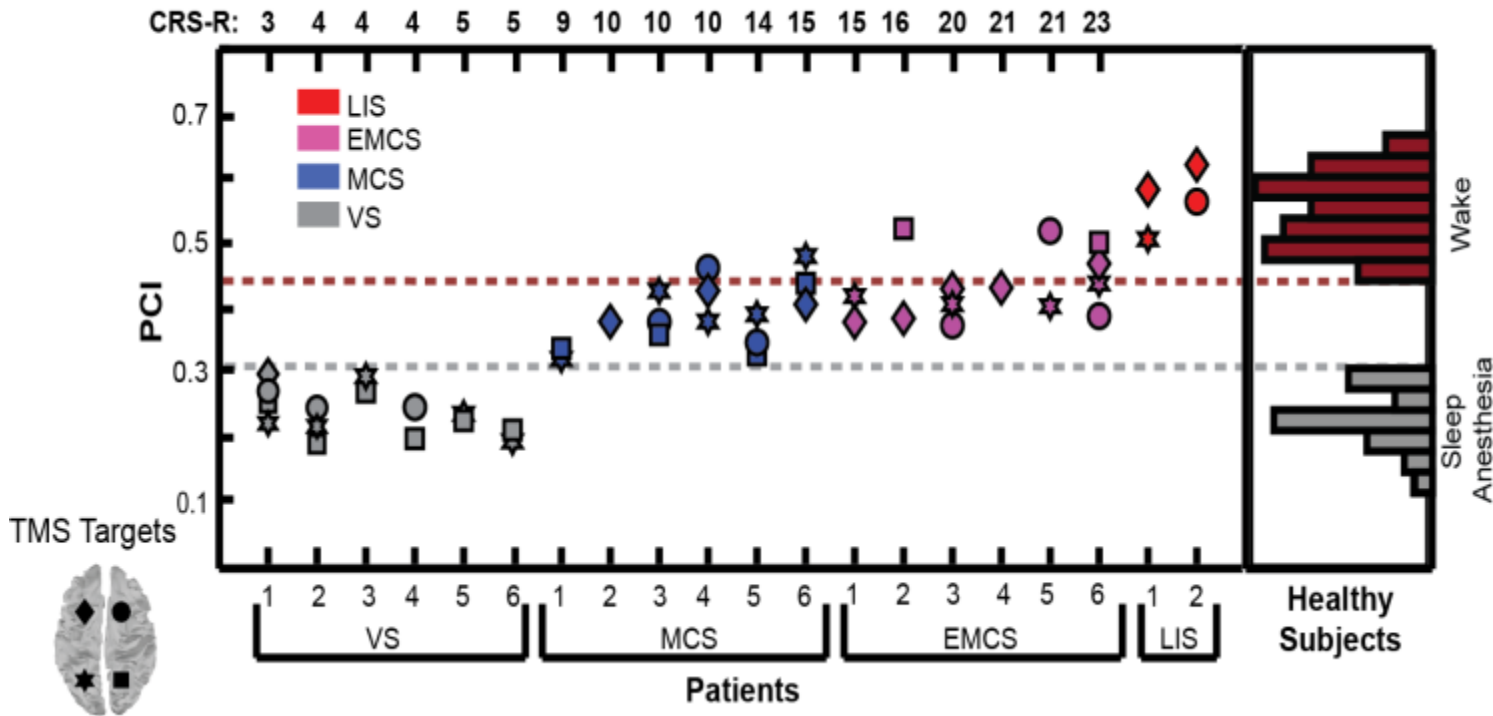
(Casali et al., Neuroimage 2010, Science TM, 2013)

Separating higher from lower levels of consciousness

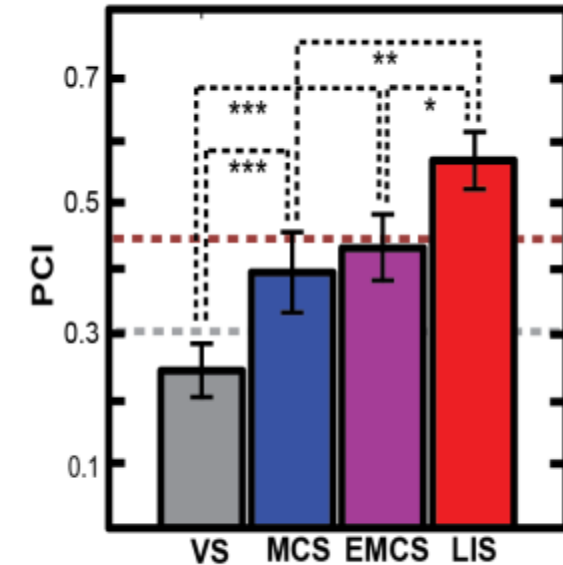


Separating higher from lower levels of consciousness

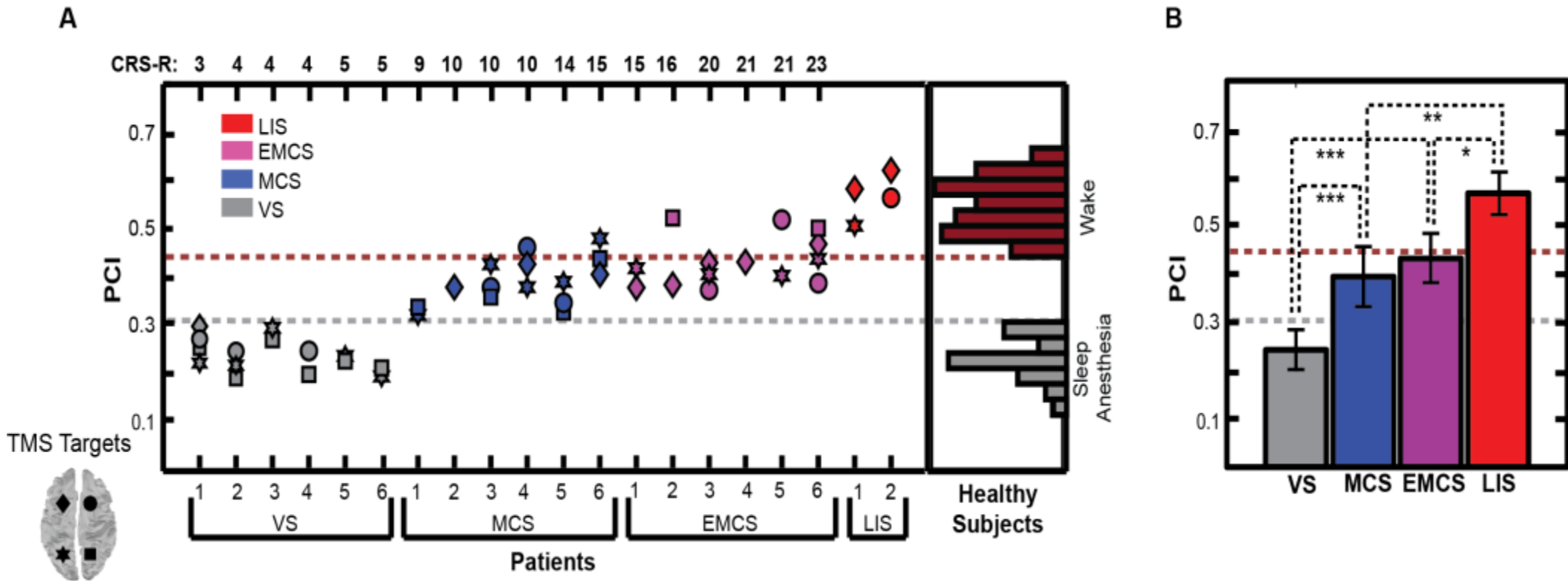
A



B



Separating higher from lower levels of consciousness



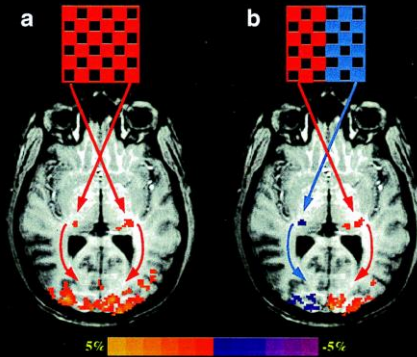
**At present, PCI is the only index that works (no false negatives)
in different conditions of loss of consciousness
and at the level of individual subjects**

Explanatory power

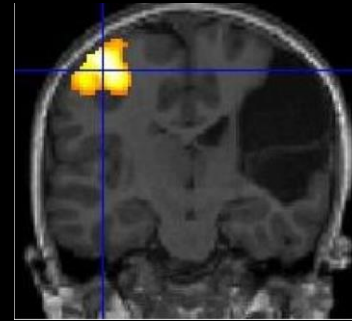
Why not the cerebellum?



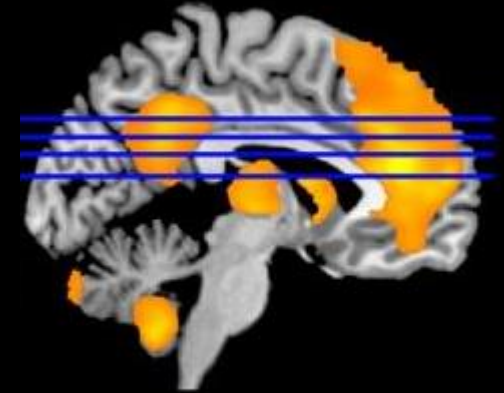
Why not afferent pathways?



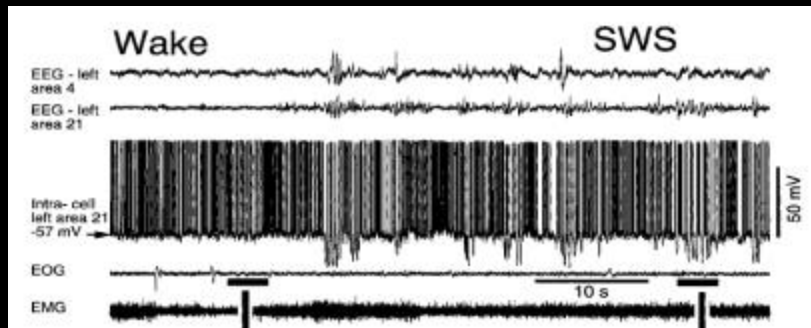
Why not efferent pathways?



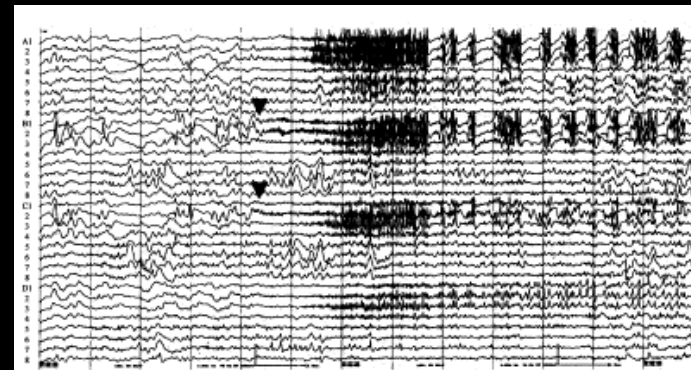
Why not cortico-subcortico-cortical loops?



Why not the cortex during deep sleep?



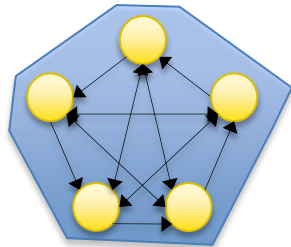
Why not the cortex during a seizure?



Explanatory power

Cortical system

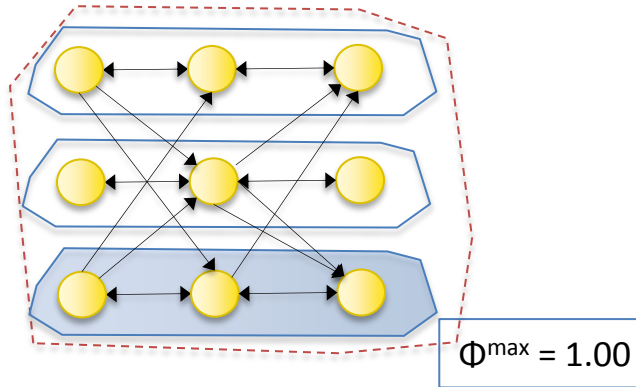
Inhomogeneous network,
functional specialization and
integration



$$\Phi^{\text{Max}} = 10.56$$

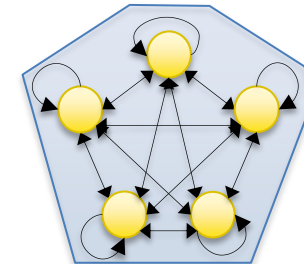
Cerebellum

Modular organization



Cortical system during deep sleep / anesthesia/ seizures

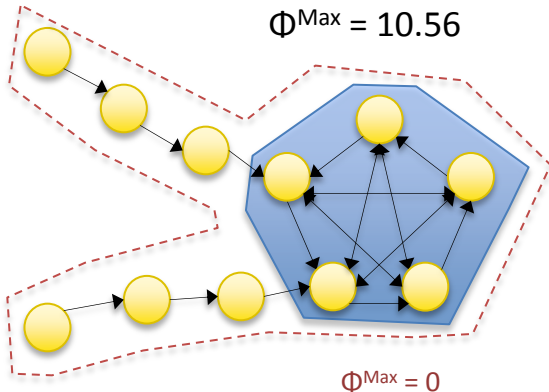
Homogeneous network



$$\Phi^{\text{Max}} = 0.003$$

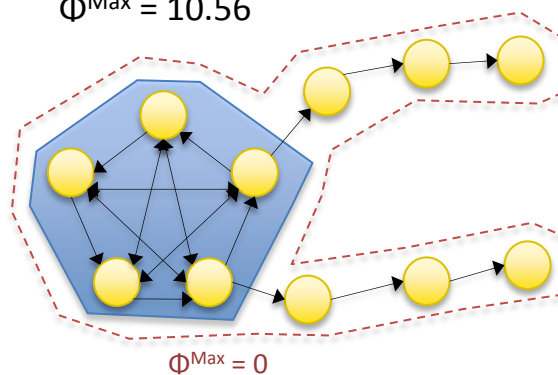
Afferent pathways

$$\Phi^{\text{Max}} = 10.56$$



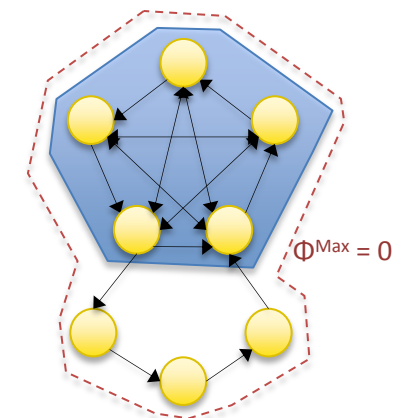
Efferent pathways

$$\Phi^{\text{Max}} = 10.56$$



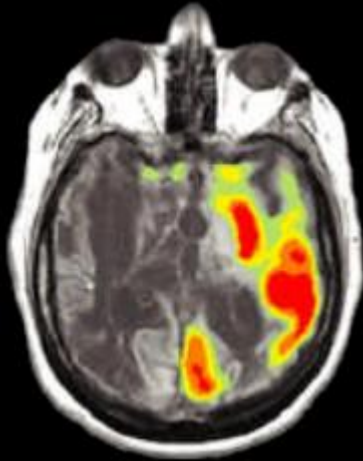
Cortico-subcortical loop

$$\Phi^{\text{Max}} = 10.56$$

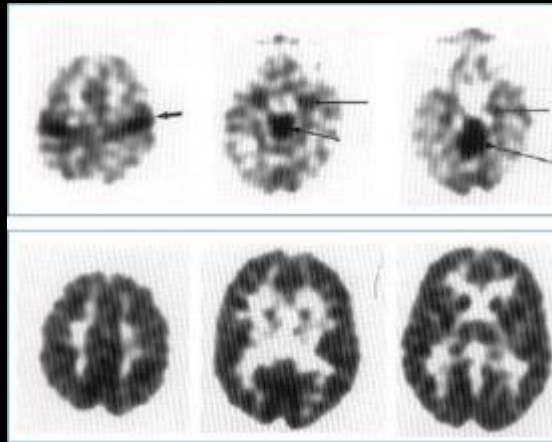


Inferential power

Brain “islands” in a vegetative subject



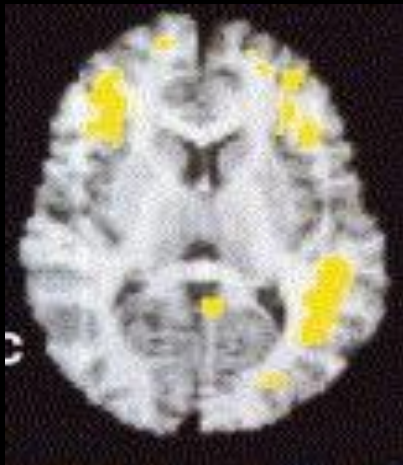
Newborn /
1 year old



Ketamine
anesthesia



Sleepwalking



Octopus

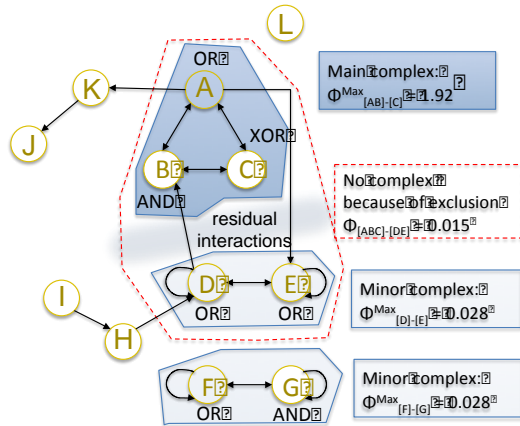


Apple Siri

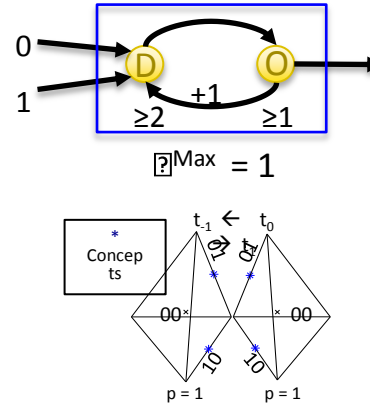


Inferential power

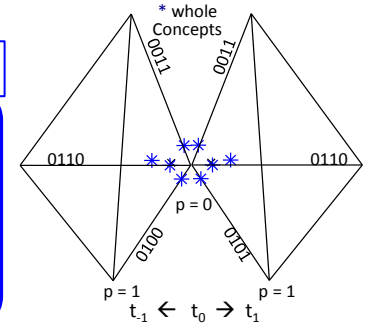
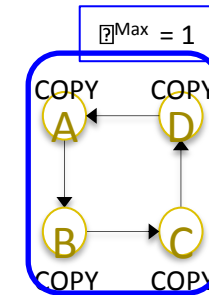
A systems can condense into major and minor complexes



Simple systems can be (minimally) conscious

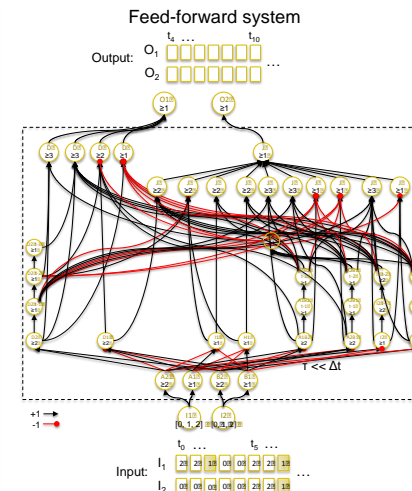
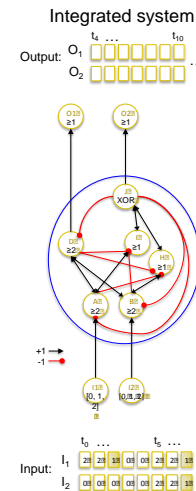
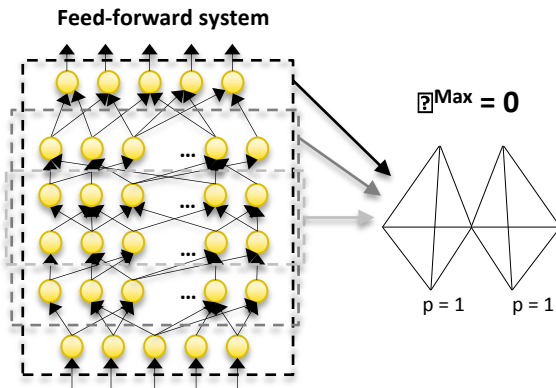


Inactive systems can be conscious



Systems can be functionally equivalent to conscious ones and yet be unconscious

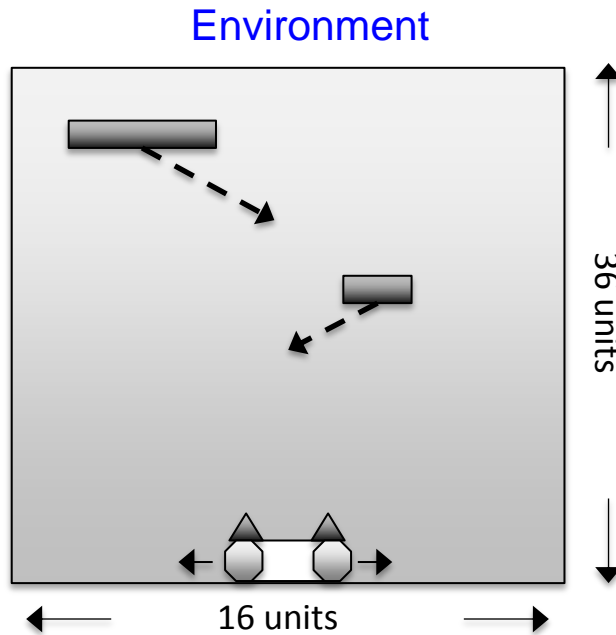
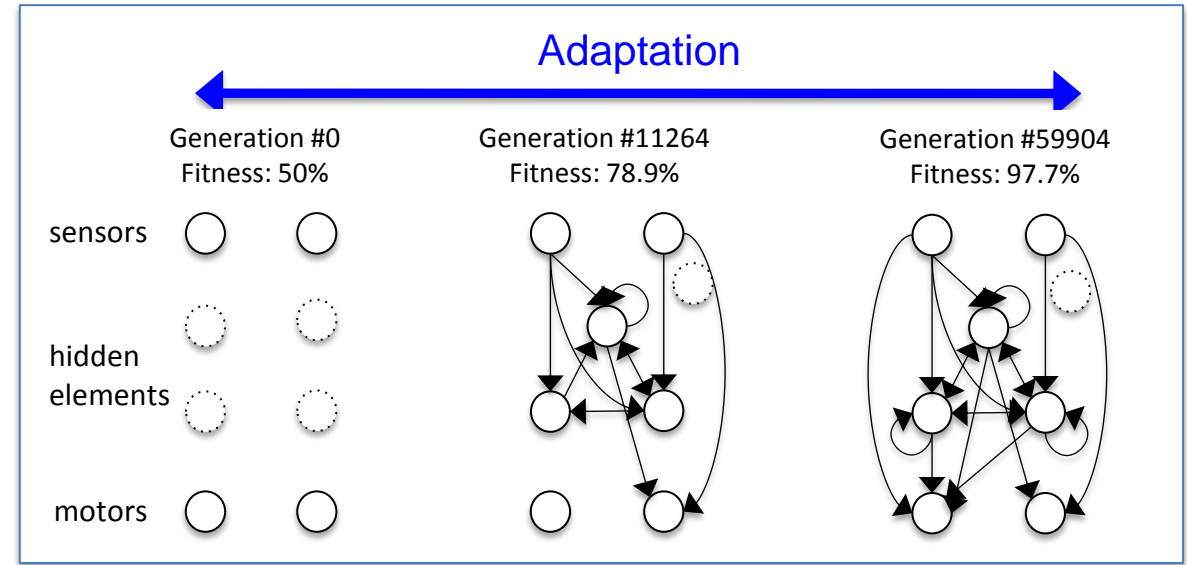
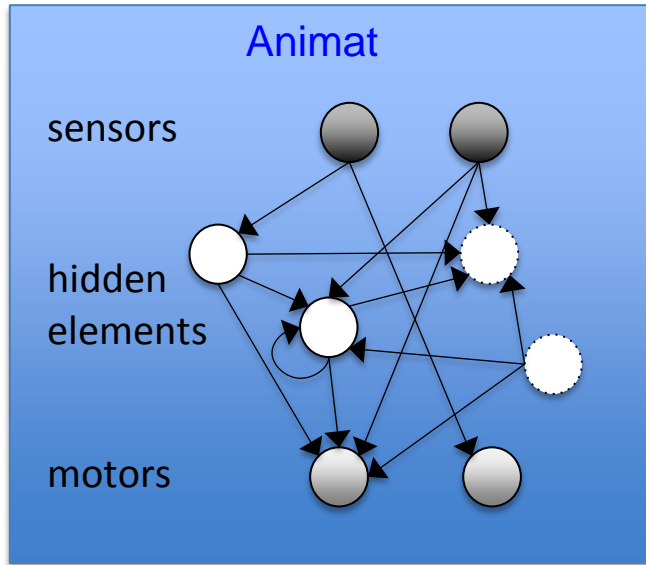
Complicated systems can be unconscious



Consciousness as integrated information:

Why did it evolve?

Does integrated information increase during adaptation in silico?

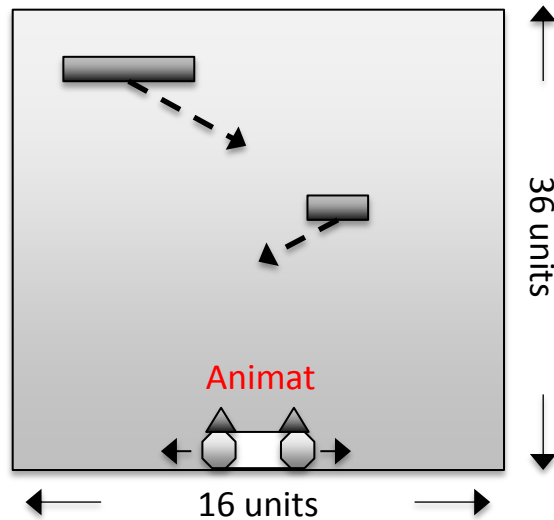


Selection & Mutation

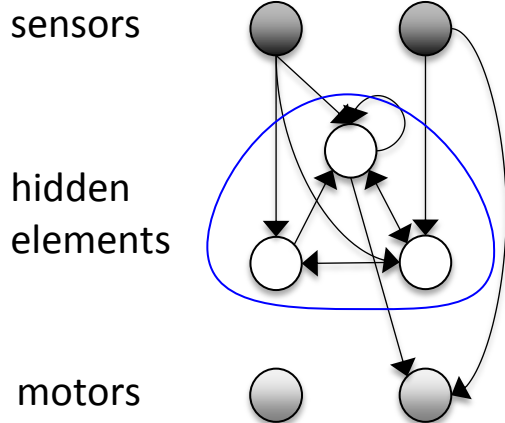
Fitness: % of successfully caught and avoided blocks (out of 128 trials)

Point mutations, deletions, and duplications in the genome after each selection

The more difficult the task, the higher integrated information ($\langle \Phi^{\max} \rangle$ and # of concepts) in the fittest animats

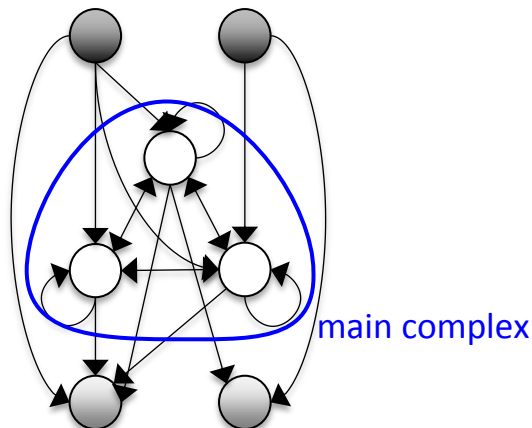


Generation #11264
Fitness: 78.9%



Main complex:
 $\langle \Phi^{\max} \rangle = 0.1674$
 $\langle \# \text{concepts} \rangle = 2.25$

Generation #59904
Fitness: 97.7%



Main complex:
 $\langle \Phi^{\max} \rangle = 1.13$
 $\langle \# \text{concepts} \rangle = 4.4$

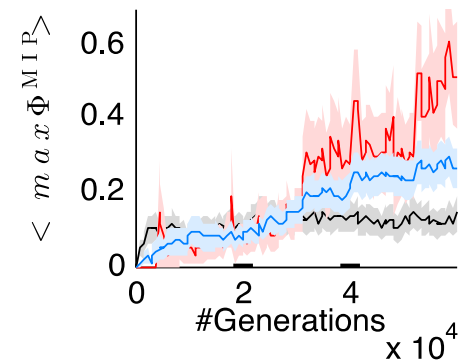
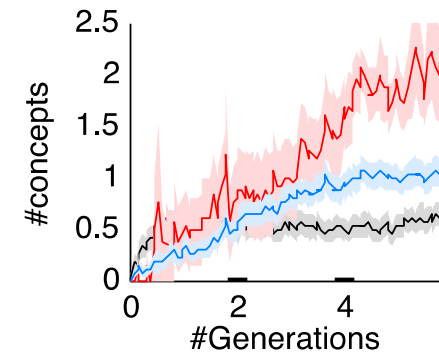
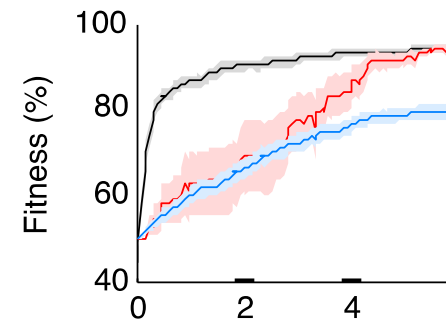
Task 1 (easy)

Catch 1
Avoid 3

Task 2 (difficult)

Catch 3 6
Avoid 4 5

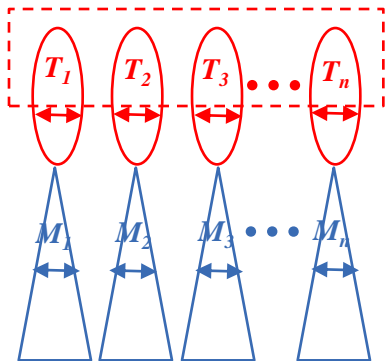
— Task 1 — Task 2 — Task 2 (7 fittest)



*Integrating information is potentially valuable
in environments with complex causal structures*

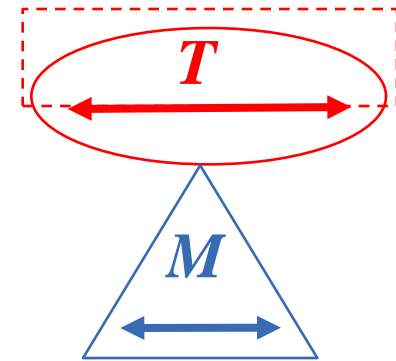


LOW Φ



- *Adaptive task can be subdivided into ~independent sub-tasks*
- *Each task domain small, few alternatives to discriminate*
- *Not context-sensitive*

HIGH Φ



- *Adaptive task cannot be subdivided into ~independent sub-tasks*
- *Task domain large, many alternatives to discriminate*
- *Context-sensitive*

*Measuring how information integration within a system
“resonates” with the causal structure of its environment*

- *Matching* is the distance D (weighted by Φ^{Max}) between the conceptual structures (Q) specified when the system is exposed to *World* and to *Noise*:

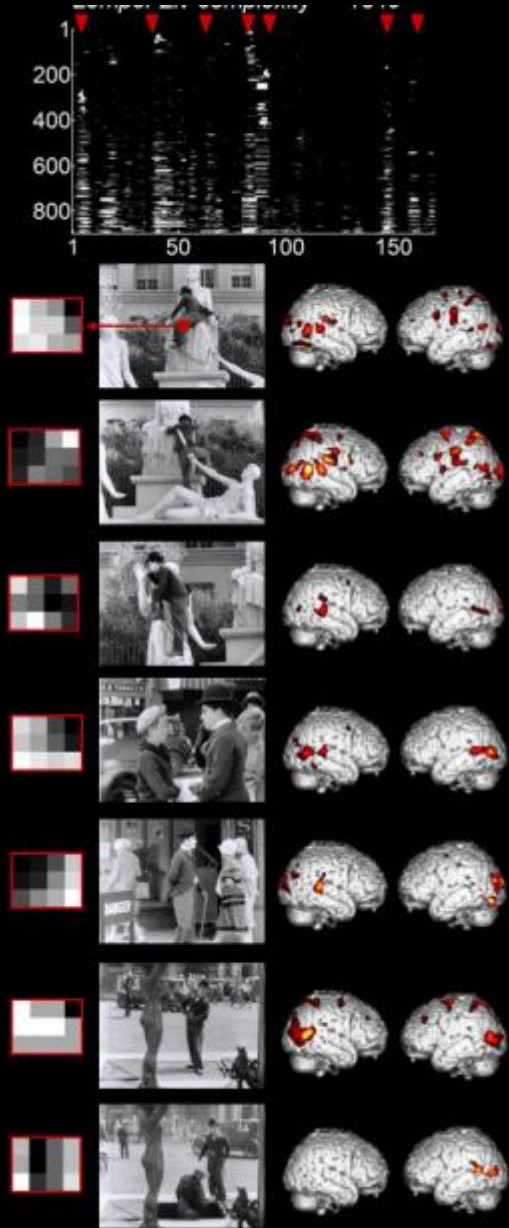
$$\text{Matching (M)} = D[Q(\text{World}) \parallel Q(\text{Noise})]$$

- *Matching* is high if a system’s responses to *World* (but not *Noise*) are both highly differentiated (many different conceptual structures) and integrated (high Φ^{Max})
- *Matching* increases in animats together with their value of Φ^{Max}
- *Matching* can be increased by infomax, prediction error minimization, and wake/sleep potentiation/down-selection
- *Matching* is a general measure of the “resonance” between the causal structure of a system and that of its environment

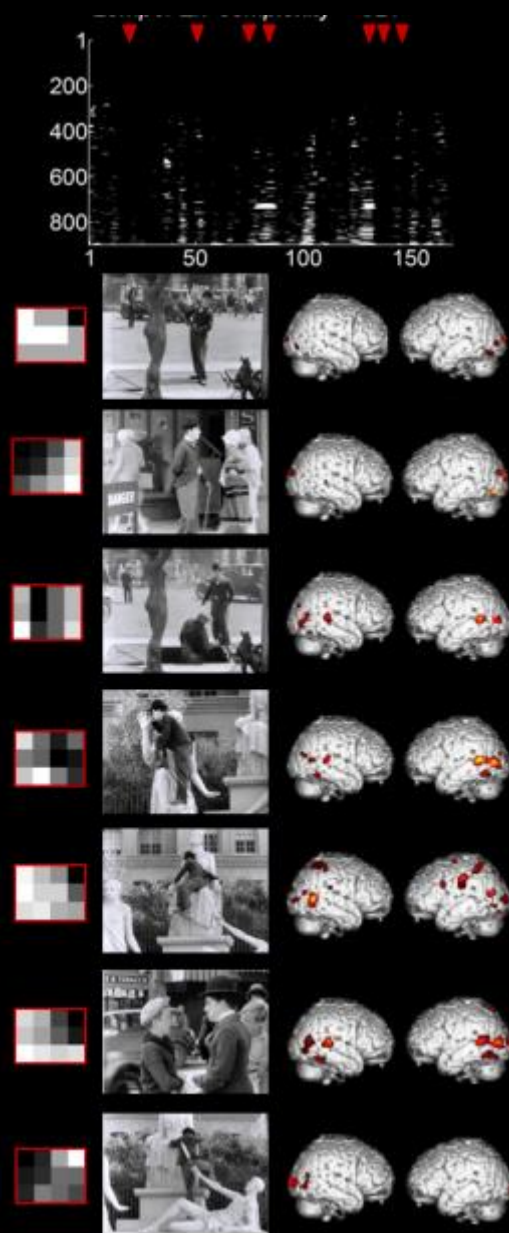
Tononi et al., 1996; 2012; Nere et al., 2013; Hashmi et al., 2013;
Boly et al., submitted; Gomez et al., in preparation

Assessing matching: differentiation of brain responses with stimulus set meaningfulness

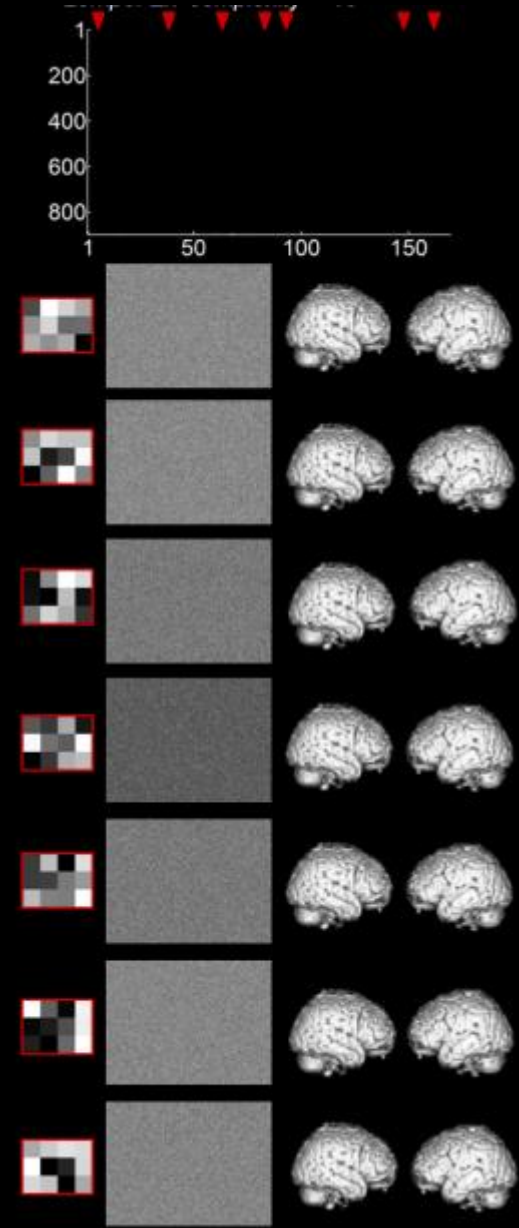
Movie: LZC = 1549



Scrambled: LZC = 921



TV noise: LZC = 10



Giulio Tononi: Disclosures

Consultant: Philips-Respironics Endowed Chair: Respironics

Acknowledgements

Internal collaborators

Marcello Massimini
Larissa Albantakis *Andy Nere*
David Balduzzi *Yuval Nir*
Melanie Boly *Masafumi Oizumi*
Chiara Cirelli *Umberto Olcese*
Daniela Dentico *Ben Shababo*
Fabio Ferrarelli *Francesca Siclari*
Olivia Gosseries
Atif Hashmi
Erik Hoel
Matteo Mainetti

External collaborators

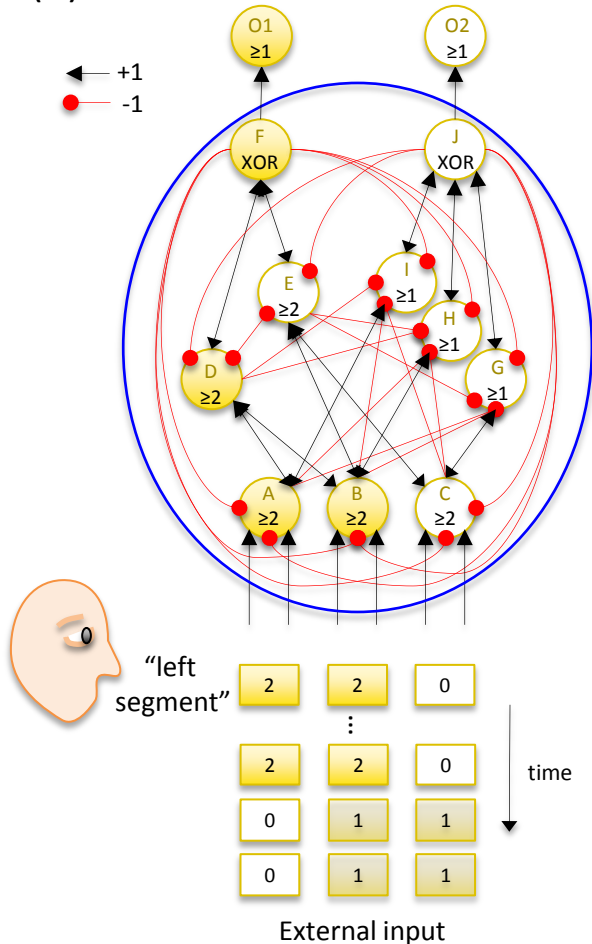
Chris Adami
Mike Alkire
Ruth Benca
Richie Davidson
Tony Hudetz
Christof Koch
Steven Laureys
Randy McIntosh
Bob Pearce
Brad Postle
Olaf Sporns
Barry Van Veen

Support

*NIH Director's Pioneer
Award, NIMH, NINDS,
DARPA, Paul Allen
Foundation, McDonnell
Foundation,
Mind Science Foundation
University of Wisconsin*

A complex can have ports in and ports out from and to the environment, but its qualia are 'solipsistic' (self-generated, self-referential, holistic)

(A)



(B)

Concept Order

#Elements in the concept

Elementary

(1st order) concepts:

$\varphi(A=1) = 0.125$ $\varphi(B=1) = 0.077$
 $\varphi(C=0) = 0.033$ $\varphi(D=1) = 0.186$
 $\varphi(E=0) = 0.077$ $\varphi(F=1) = 0.186$
 $\varphi(G=0) = 0.044$ $\varphi(H=0) = 0.025$
 $\varphi(I=0) = 0.044$ $\varphi(J=0) = 0.186$

2nd order concepts:

$\varphi(AB=11) = 0.107$
 $\varphi(BC=10) = 0.106$
 $\varphi(DE=10) = 0.417$
 $\varphi(DI=10) = 0.0625$
 ...

3rd and higher order concepts:

$\varphi(ABC=110) = 0.063$
 $\varphi(GHI=000) = 0.095$
 $\varphi(DEF=101) = 0.188$
 $\varphi(ADF=111) = 0.063$
 ...
 $\varphi(ABDF=1111) = 0.063$
 ...
 $\varphi(ABCDEF=110101) = 0.050$
 ...

(C)

Concept truth value

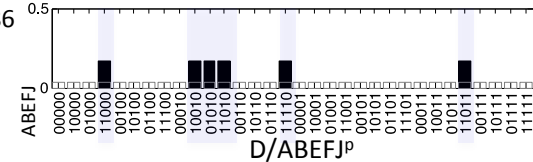
("on" or "off")

D: "on"

$\varphi(D=1) = 0.186$

Purview:

D/ABEFJ^p, A^f

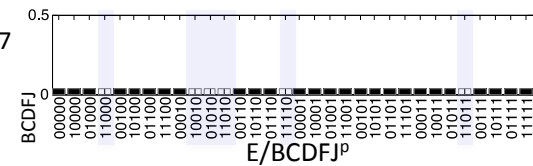


E: "off"

$\varphi(E=0) = 0.077$

Purview:

E/BCDFJ^p, BD^f



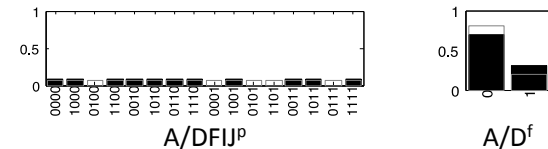
(D)

A: "on"

$\varphi(A=1) = 0.125$

Purview:

A/DFIJ^p, D^f



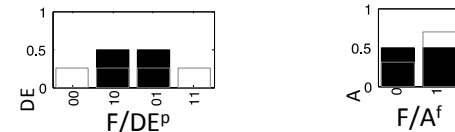
(E)

F: "on"

$\varphi(F=1) = 0.186$

Purview:

F/DE^p, A^f

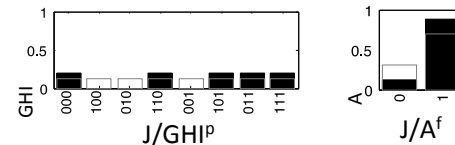


J: "off"

$\varphi(J=0) = 0.186$

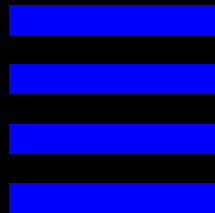
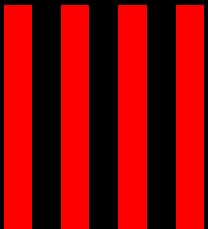
Purview:

J/GHI^p, A^f

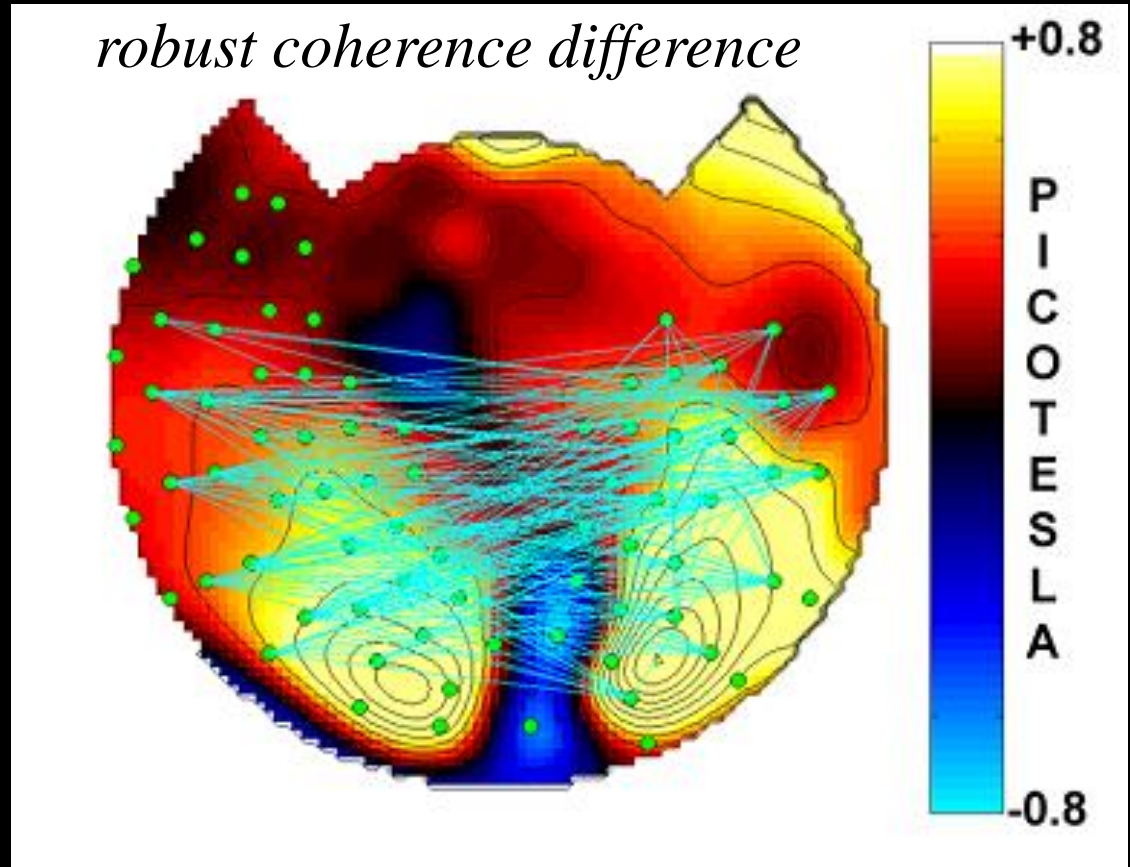


From BCC (behavioral correlates of consciousness) to NCC (neural correlates of consciousness)

BCC



NCC



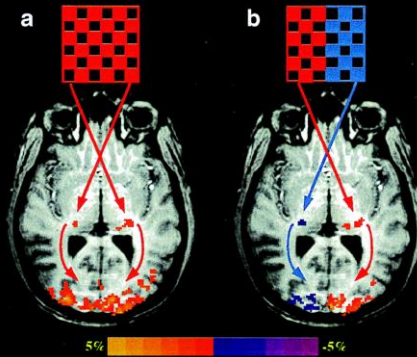
Increased MEG power and cortico-cortical long-range coherence during perceptual dominance in binocular rivalry (Tononi et al., PNAS, 1998)

Explanatory power

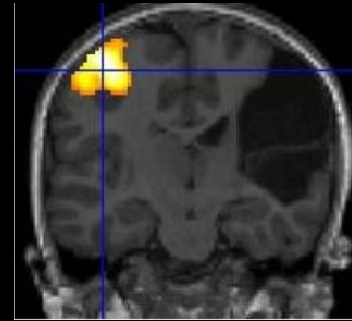
Why not the cerebellum?



Why not afferent pathways?



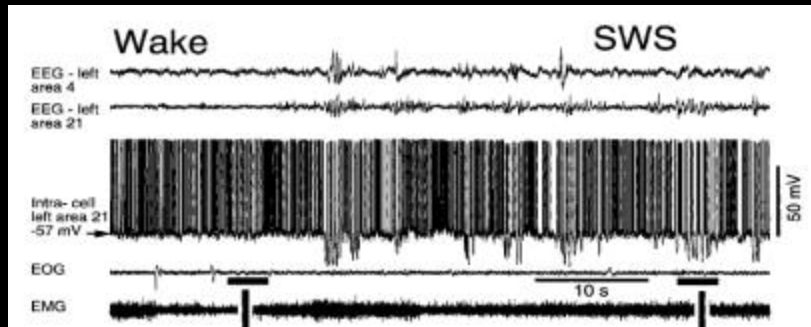
Why not efferent pathways?



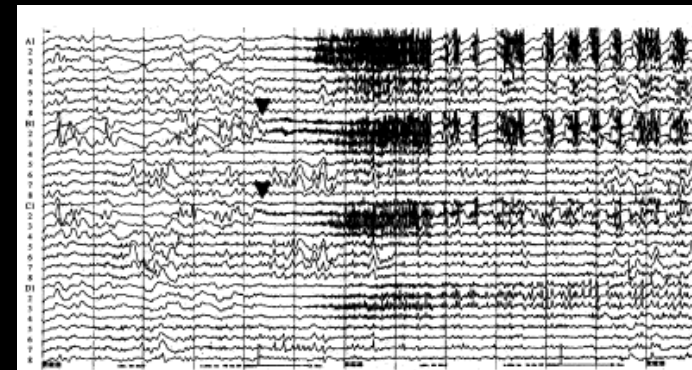
Why not cortico-subcortico-cortical loops?



Why not the cortex during deep sleep?

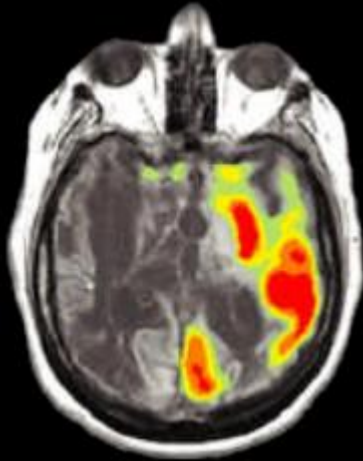


Why not the cortex during a seizure?

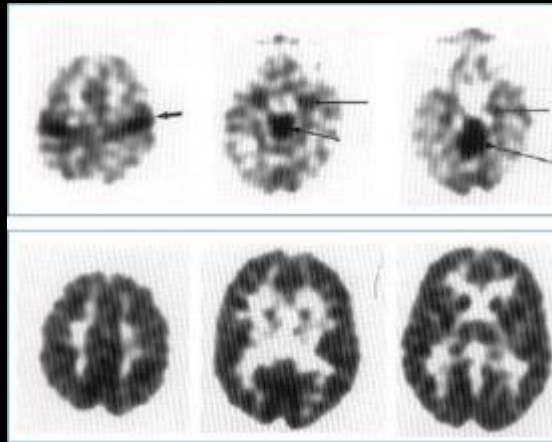


Inferential power

Brain “islands” in a vegetative subject



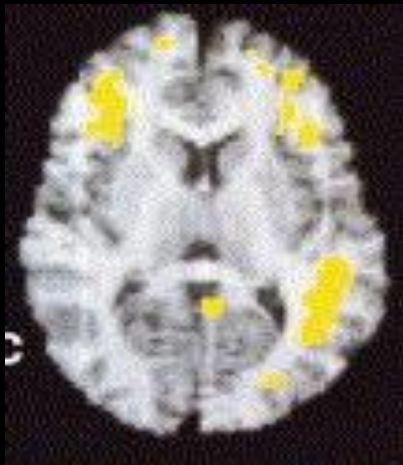
Newborn /
1 year old



Ketamine
anesthesia



Sleepwalking



Octopus



Apple Siri



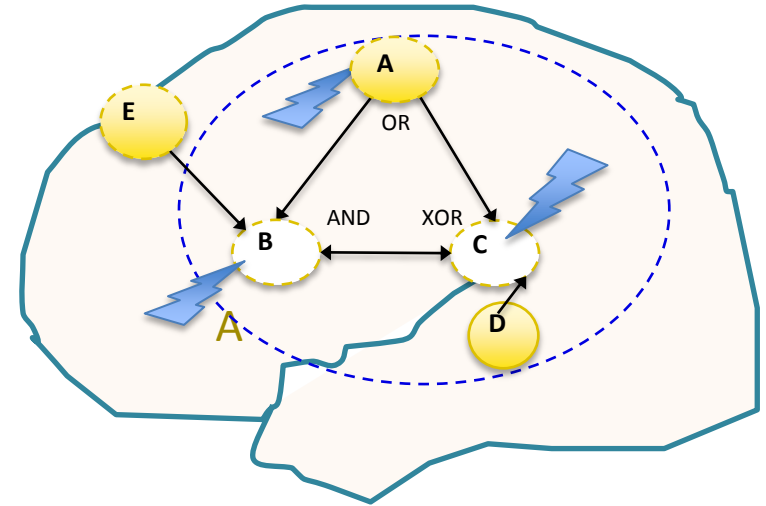
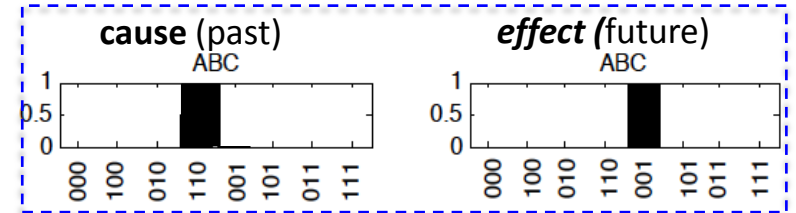
Axioms and Postulates

- Intrinsic existence
- Composition
- Information
- Integration
- Exclusion

Intrinsic existence

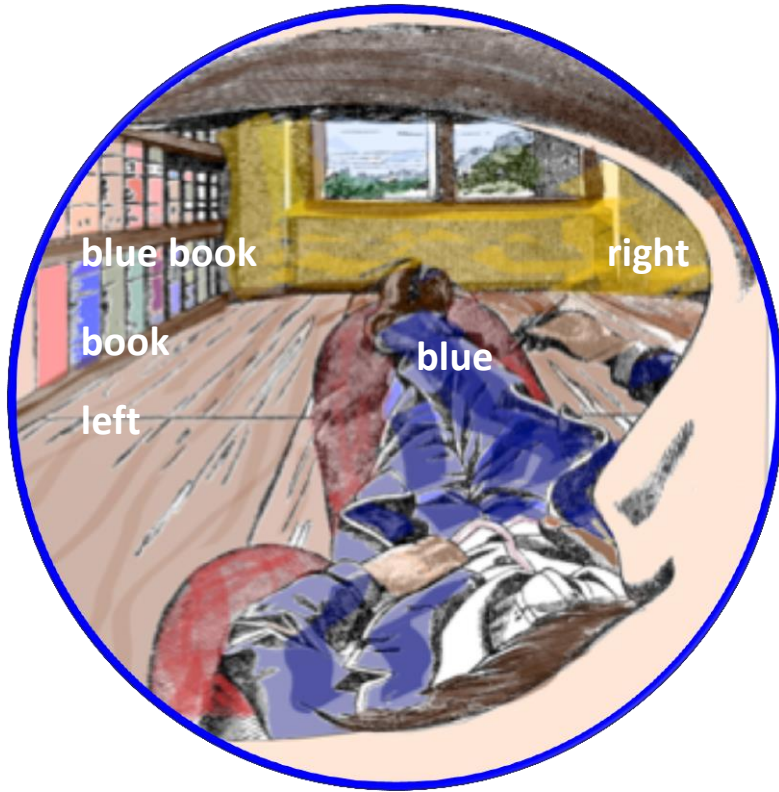


Axiom

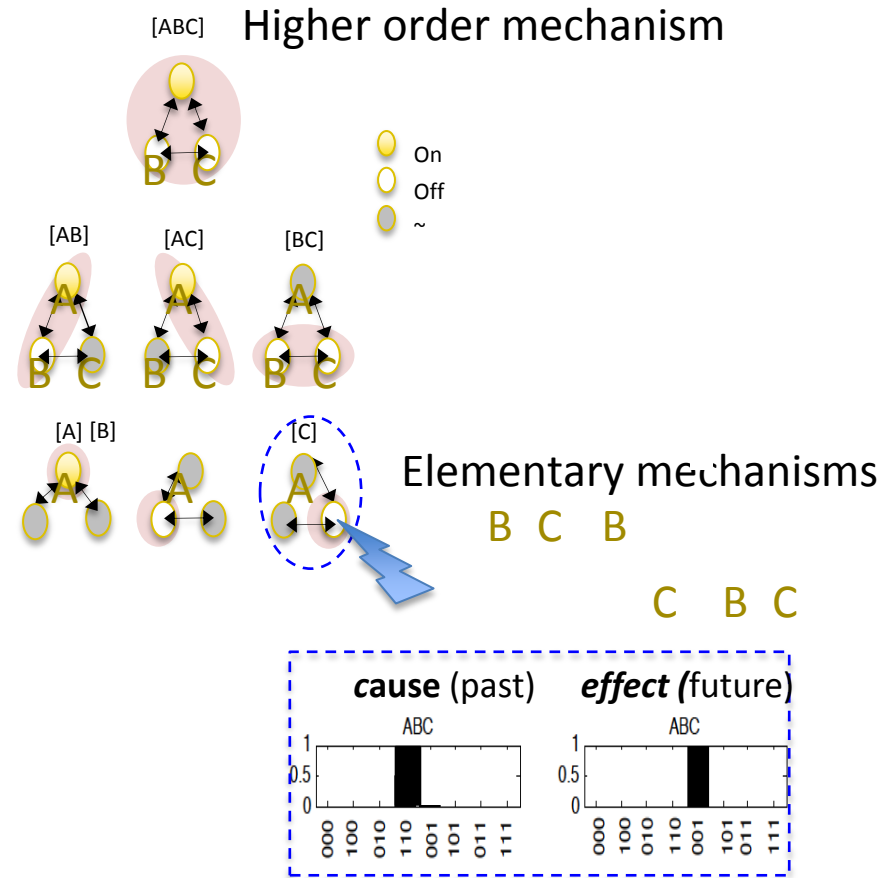


Postulate

Composition



Axiom

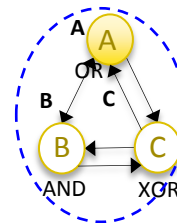


Postulate

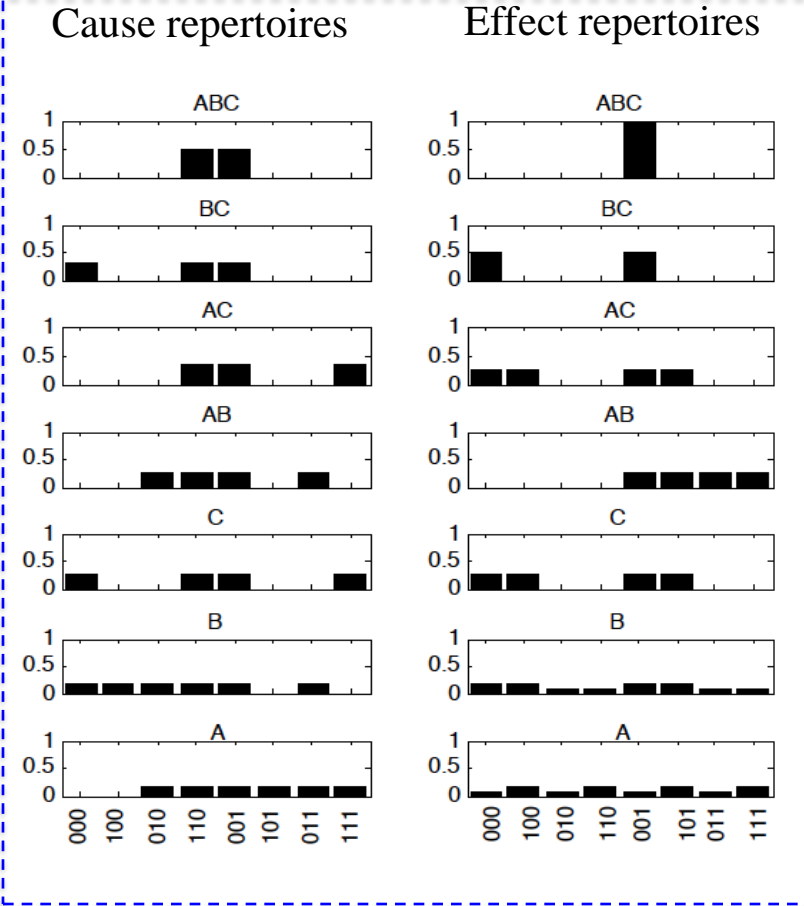
Information



Axiom



Cause-effect structure



Postulate

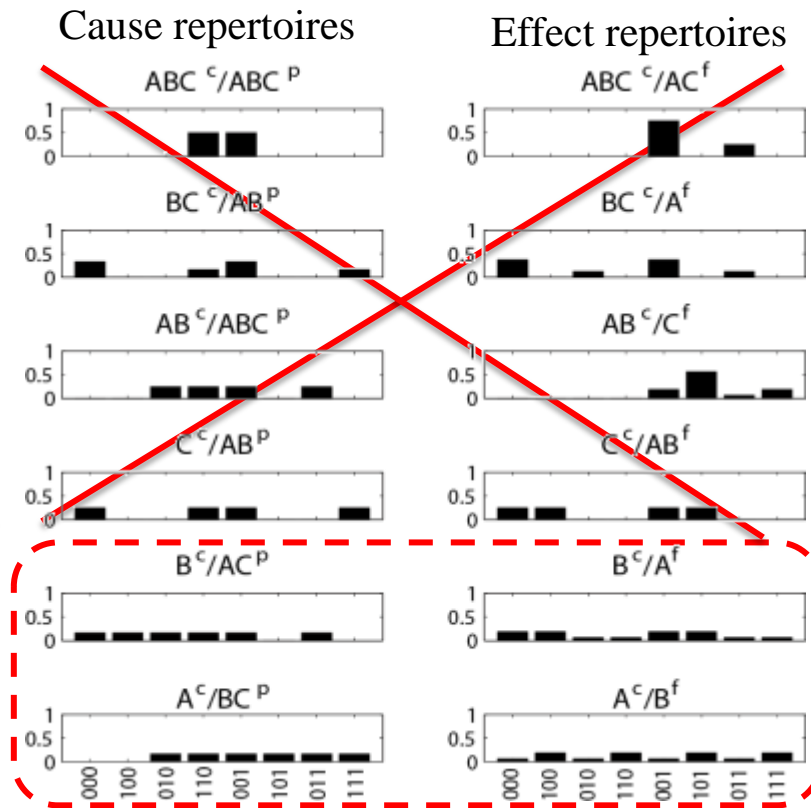
Integration



Axiom

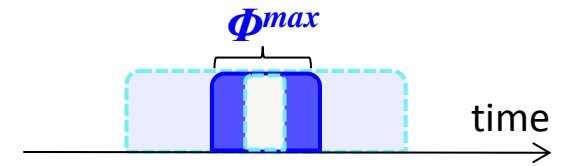


**Cause-effect structure
of partitioned system**

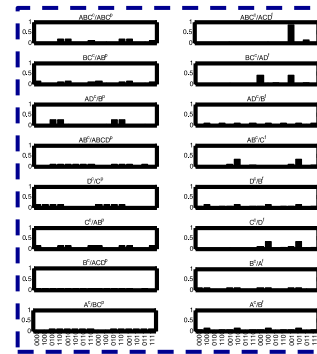
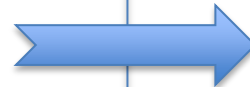
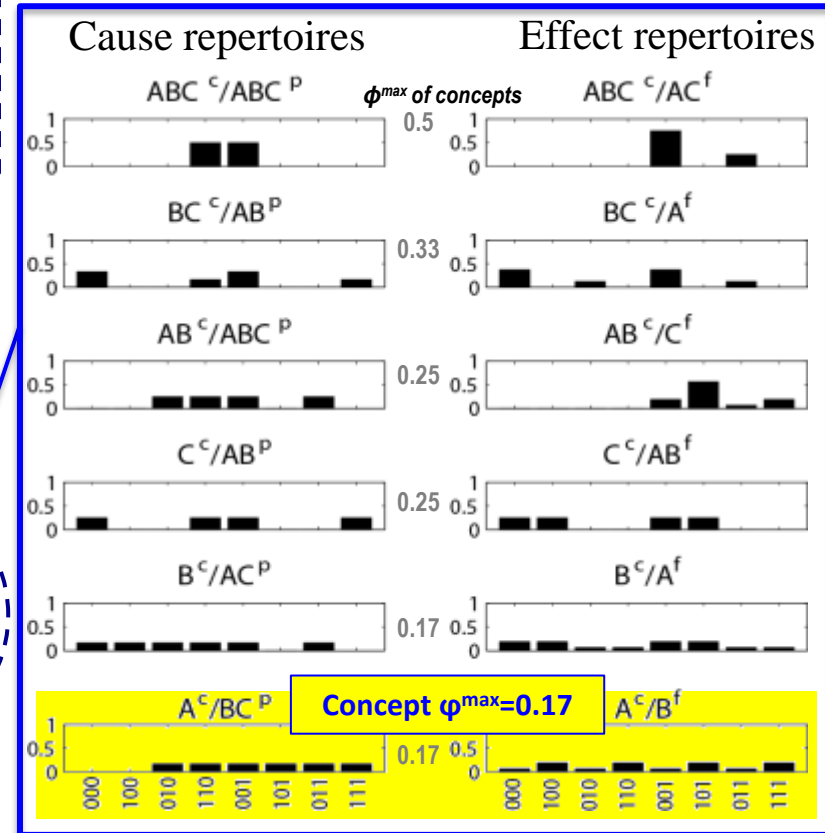


Postulate

Exclusion

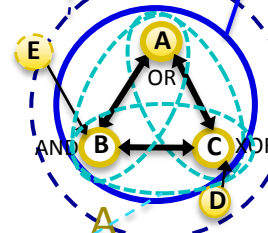


Conceptual structure



$\Phi=0.44$

Complex
 $\Phi^{max}=1.92$



$\Phi=1.0$

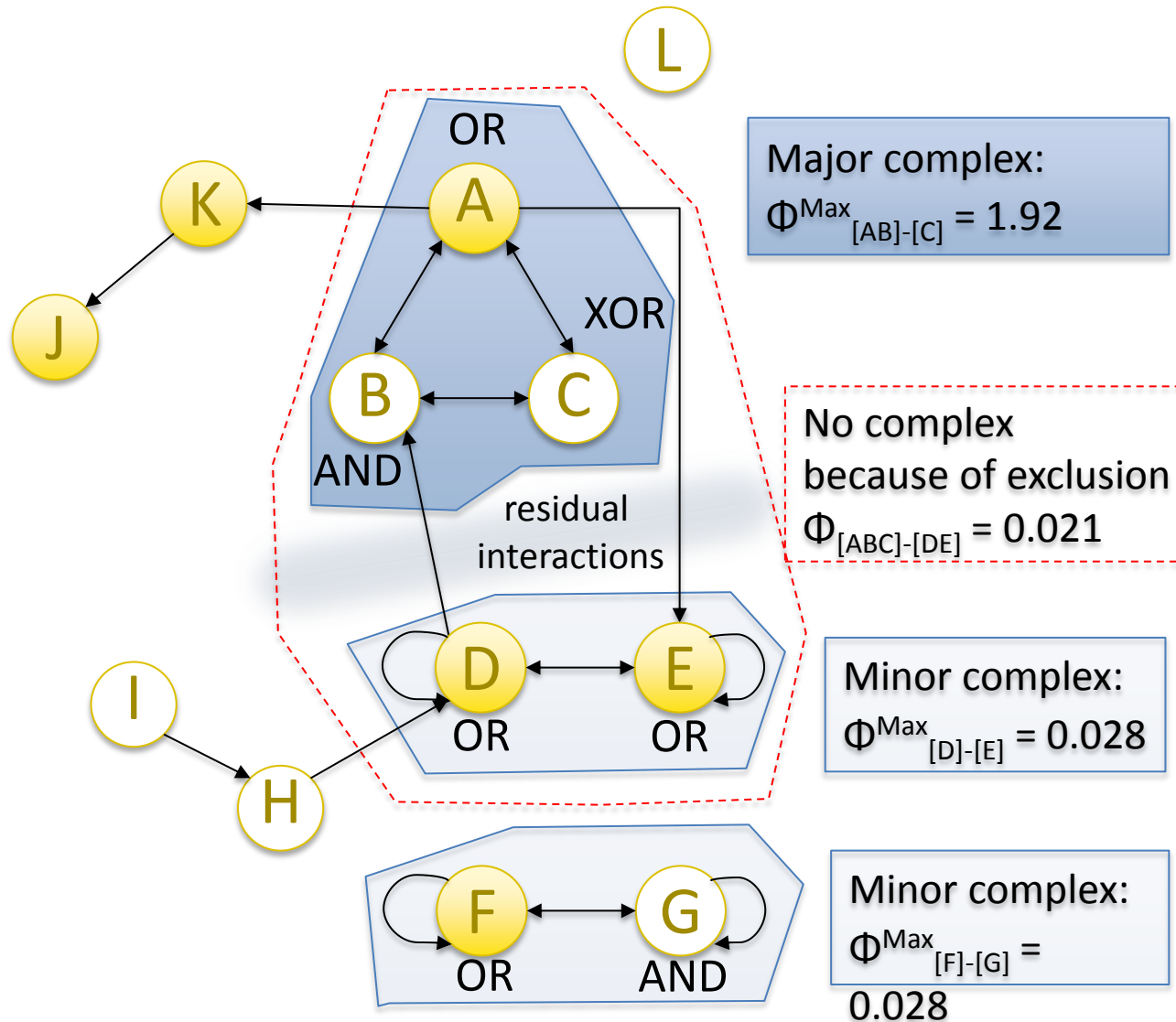
Axiom

Postulate

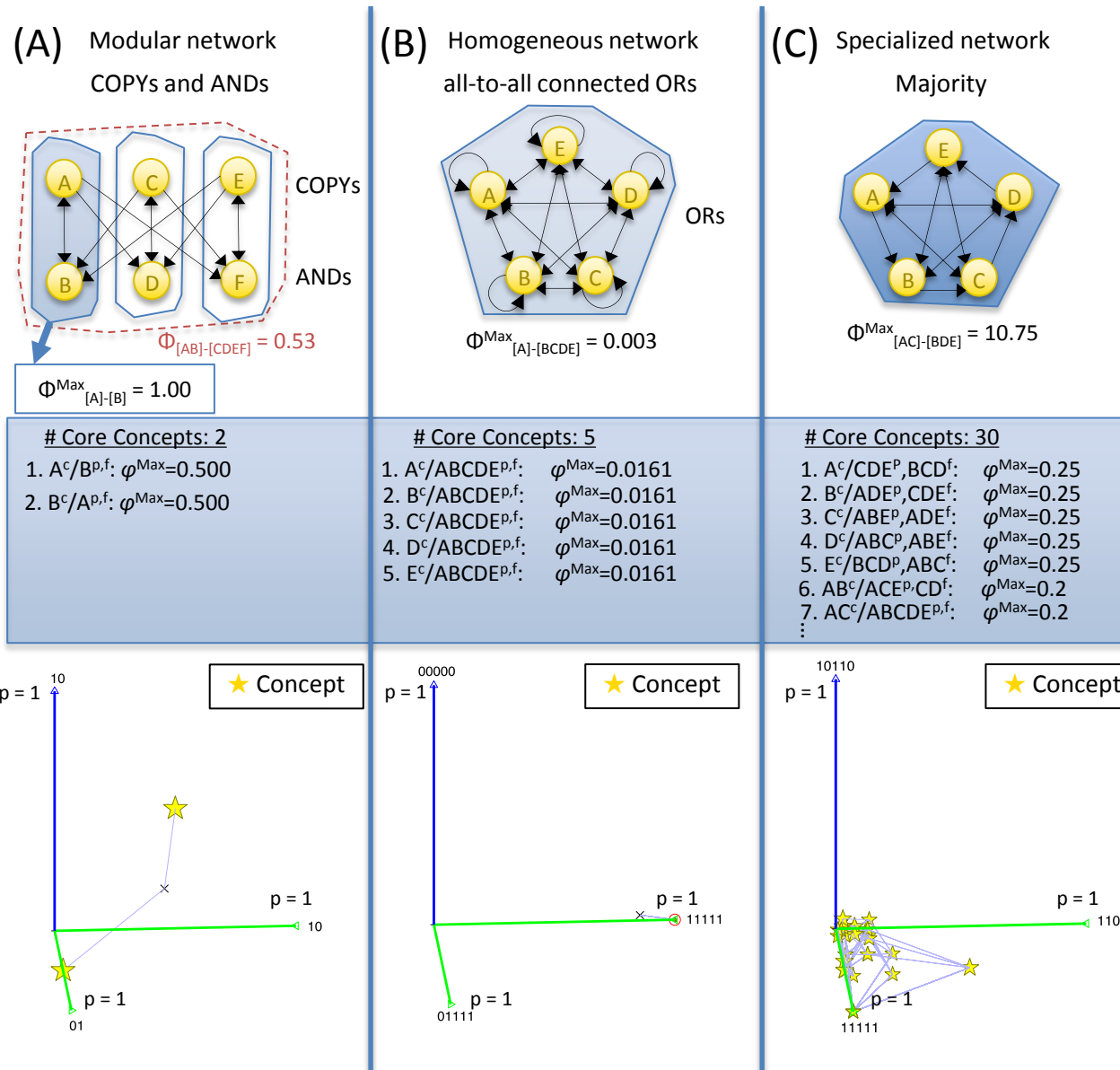
The NCC and maxima of cause-effect power

- The NCC is a maximum of cause-effect power (integrated information Φ^{\max}),
eg posterior cortical areas, superficial cortical layers, pyramidal cells, or otherwise
- The NCC can vary (expand, shrink, split, and move)
- The elements of the NCC have a spatial scale that achieves the highest Φ^{\max} ,
eg neurons, local groups of neurons, or otherwise
- The (discrete) time scale of the NCC is that at which its elements achieve highest Φ^{\max} ,
eg sec, hundred msec, or sec
- The activity states that matter to the NCC are differences that make most difference to it,
eg bursting, high mean firing, low mean firing

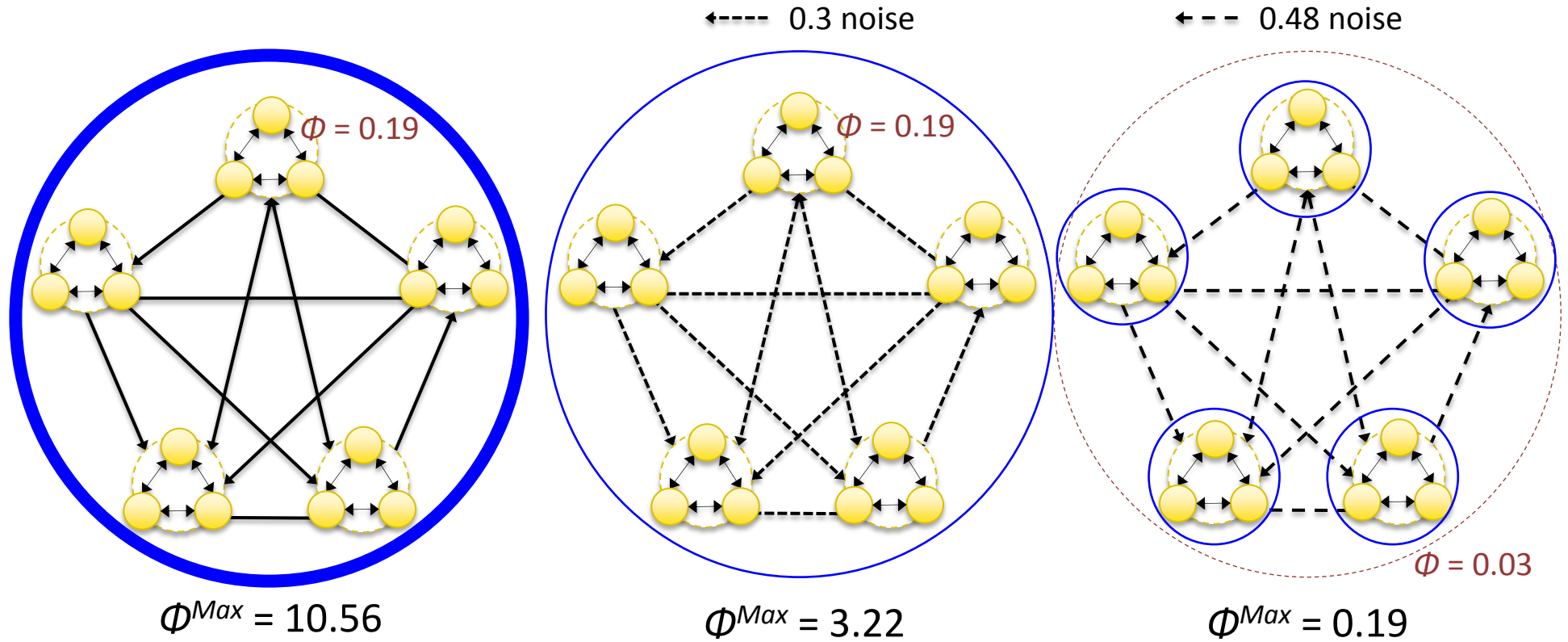
A system can condense into major and minor complexes and their residual interactions



Qualia generated by modular, homogeneous, and specialized networks



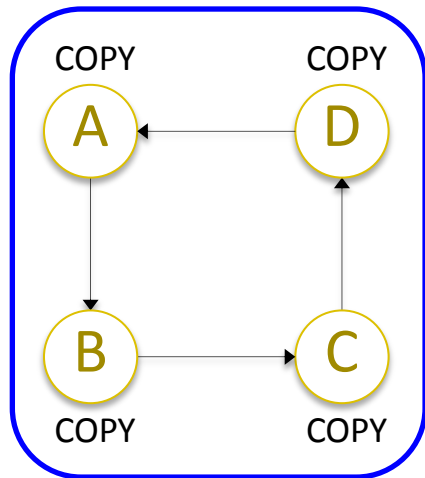
Consciousness can be graded



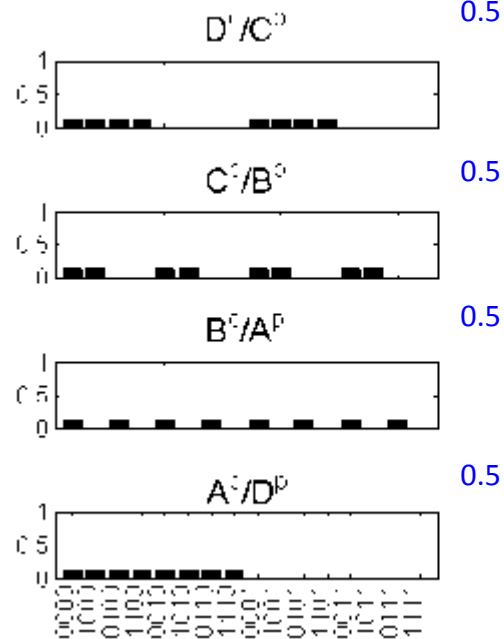
Inactive systems can be conscious

$$j^{Max}(P, F | X = s_0)$$

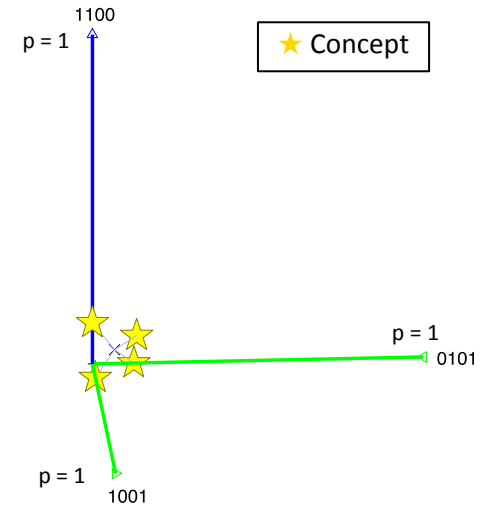
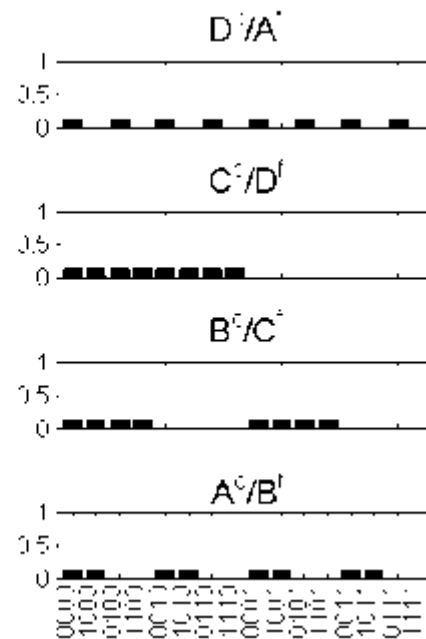
$$j^{Max} = 1$$



Core causes



Core effects



From the extrinsic perspective, conscious and unconscious systems can be functionally equivalent

