

Теория сознания Тонони

Александр Панов

ИСА РАН

3 марта 2016 г.

Жулио Тонони

- Жулио Тонони — нейробиолог, специалист по исследованию сна и сознания.
- Профессор психиатрии и заслуженный профессор науки о сознании Университета Висконсина, факультет психиатрии, руководитель центра сна и сознания.
- Scopus: 285 статей, 19407 цитирований, h-индекс — 72.



Основные публикации:

- "Sleep function and synaptic homeostasis". In: *Sleep Medicine Reviews* 10.1 (2006), pp. 49–62.
- "Local sleep and learning". In: *Nature* 430.6995 (2004), pp. 78–81.
- "Consciousness and complexity". In: *Science* 282.5395 (1998), pp. 1846–1851.
- "A measure for brain complexity: Relating functional segregation and integration in the nervous system". In: *Proceedings of the National Academy of Sciences of the United States of America* 91.11 (1994), pp. 5033–5037.

Теория интегрированной информации (IIT)

Oizumi, Masafumi, Larissa Albantakis, and Giulio Tononi. "From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0". In: *PLoS Computational Biology* 10.5 (2014), e1003588.



- Разрабатывается с 2004 года
- Цель: создать теоретический подход, обладающий объясняющей и предсказывающей силой
- Основные вопросы:
 - почему сознание генерируется кортико-таламической системой, а не мозжечком?
 - почему сознание исчезает во время сна, хотя мозг активен?
 - почему теряется сознание во время конвульсий, хотя нейронная активность высокая и синхронная?
 - почему нет прямого вклада нейронной активности в сознание, хотя доказано, что она влияет на содержание переживания (experience)?
- Сайт с демонстрациями

Сложные вопросы

Только теория может ответить на сложные вопросы:

- Обладает ли новорожденный сознанием? В какой степени?
- Обладает ли сознанием пациент с остаточной нейронной активностью в одной из областей коры?
- Обладают ли сознанием интеллектуальные машины, способные отлично распознавать изображения и отвечать на сложные вопросы?

Нужно двигаться от феноменологии, от основных принципов, к тому, как они реализуются физическими системами, а не от нейронных механизмов к условиям, при которых на их основе возникает сознание.

Аксиомы

Самоочевидные (по Декарту) истины о сознании:

- **существование**: «Я переживаю, значит я существую»,
- **структурированность**: переживания характеризуются несколькими аспектами в разных сочетаниях (голубой квадрат справа),
- **информационность**: одно переживание может быть отличено от другого переживания,
- **интегрированность**: переживание строго не разделимо на независимые части (красный треугольник),
- **единственность**: одно переживание исключает наличие других в то же самое время, т.е. обладает определенными пространственно-временными разрешением.

Постулаты

Недоказуемые предположения о физическом субстрате сознания.

Постулаты, относящиеся к свойствам физических систем
(механизмов):

- **существование:** системы существуют в определенном состоянии, система состоит из подсистем,
- **структурированность:** элементарные подсистемы могут быть объединены в системы более высокого порядка.

Постулаты

Постулаты, характеризующие элементы системы или подсистемы:

- **информационность**: подсистема вносит вклад в сознательное переживание (генерирует информацию), если она ограничивает множество состояний системы, ее причинно-следственный репертуар (определяет «имеющие значение отличия»),
- **интегративность**: подсистема вносит вклад в сознание, если она определяет причинно-следственный репертуар, неразделимый на независимые компоненты,
- **единственность**: подсистема вносит вклад в сознание, если она определяет *не более чем один* причинно-следственный репертуар, имеющий максимальное значение интеграции ϕ^{Max} и называющийся максимально несократимым причинно-следственным репертуаром (MICE) (в этом случае подсистема образует концепт).

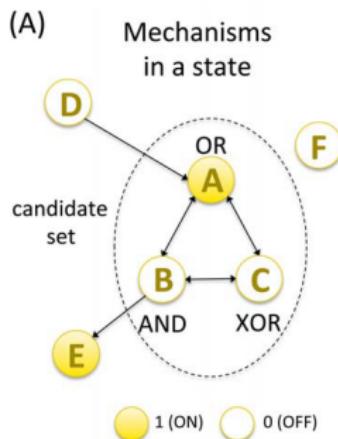
Постулаты

Постулаты, характеризующие систему в целом:

- **информационность**: множество элементов системы обладает сознанием, если соответствующие им подсистемы определяют «имеющие значение отличия», т.е. образуют концептуальную структуру («созвездие» точек в пространстве связанных прошлых и будущих состояний системы),
- **интегративность**: множество элементов системы обладает сознанием, если соответствующие им подсистемы определяют несократимую концептуальную структуру (интеграция Φ оценивается путем разделения элементов на подмножества однонаправленными разрезами),
- **единственность**: среди всех пересекающихся множеств элементов, только одно множество может обладать сознанием: то, которое определяет максимально несократимую концептуальную структуру (MICS, quale) (локальный максимум интегрированной информации Φ^{Max} называется комплексом).

Подсистема: матрица переходов

Подсистемой (механизмами) могут быть что угодно, что обладает каузальной ролью в некоторой системе: нейрон в мозге или логический вентиль в компьютере.



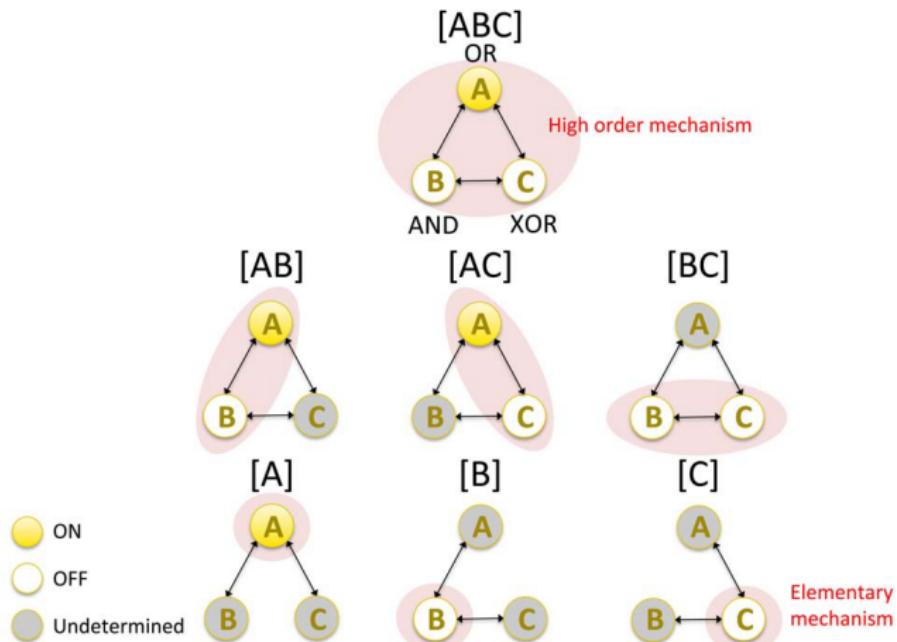
(B) Transition Probability Matrix (TPM)

		t ₁								
		ABC	000	100	010	110	001	101	011	111
t ₀	ABC	000	1	0	0	0	0	0	0	0
		100	0	0	0	0	1	0	0	0
010		0	0	0	0	0	0	1	0	0
110		0	1	0	0	0	0	0	0	0
001		0	1	0	0	0	0	0	0	0
101		0	0	0	0	0	0	0	0	1
011		0	0	0	0	0	1	0	0	0
111		0	0	0	1	0	0	0	0	0

current state $s_0(ABC) = 100$

Подсистема: структурированность

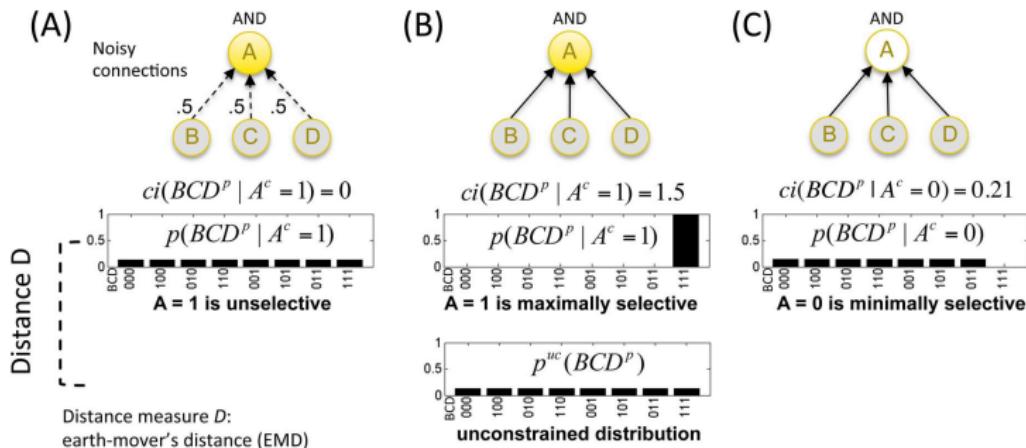
Подсистемы формируют системы второго порядка, которые, в свою очередь, составляют системы третьего порядка.



Подсистема: информативность

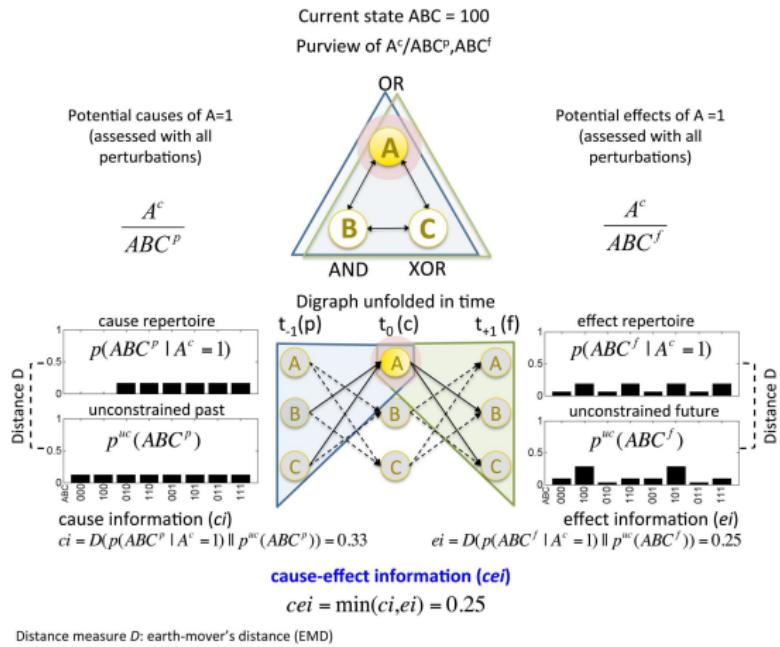
Информация по Тонони предназначена для определения отличий, имеющих («внутреннее») значение в системе, что отличает ее от «внешней» информации по Шеннону.

Ограничено элементом A распределение прошлых (будущих) состояний системы BCD (перспективы) называется причинным (следственным) репертуаром элемента A .



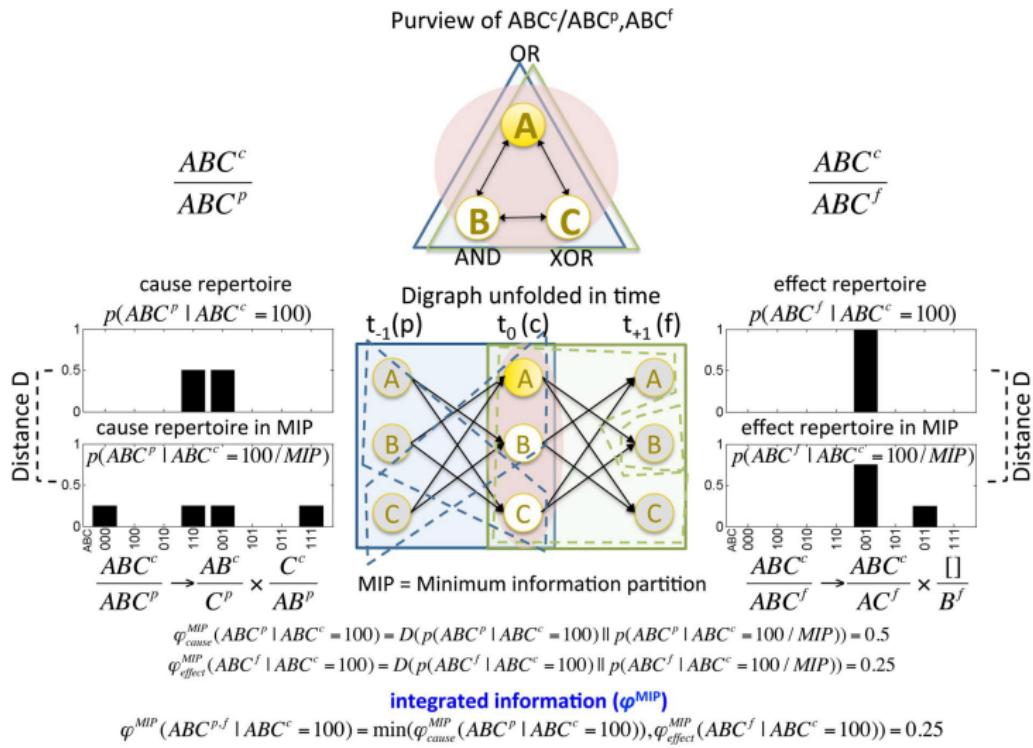
Подсистема: информативность

ci, ce - причинная и следственная информация элемента $A = 1$ относительно перспективы ABC .



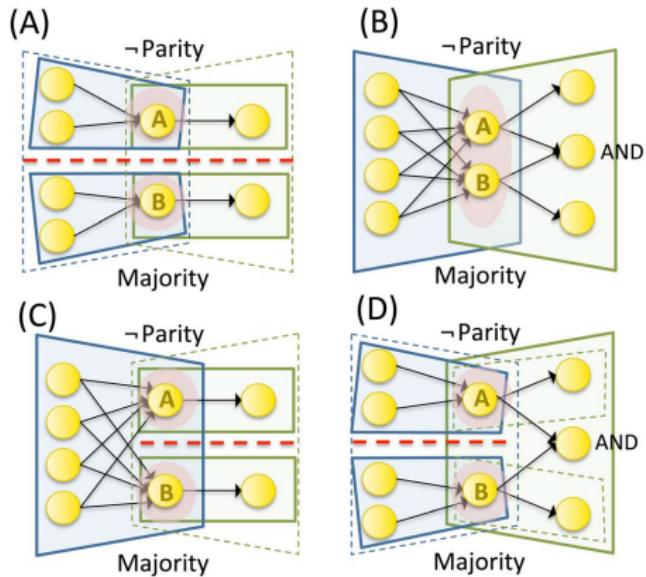
Подсистема: интегративность

МИР - минимально информативное разделение.



Подсистема: интегративность

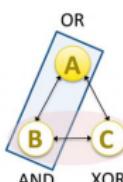
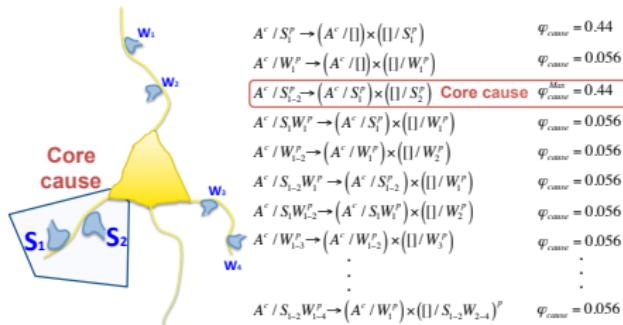
Подсистемы, не генерирующие информации, не существуют с внутренней точки зрения всей системы.



Подсистема: единственность

Подсистема, определяющая максимально несократимый причинно-следственный репертуар (MICE) образует концепт.

«Не следует множить причины без необходимости» (каузальный вариант бритвы Оккама): только один набор синапсов может вызвать активность нейрона.

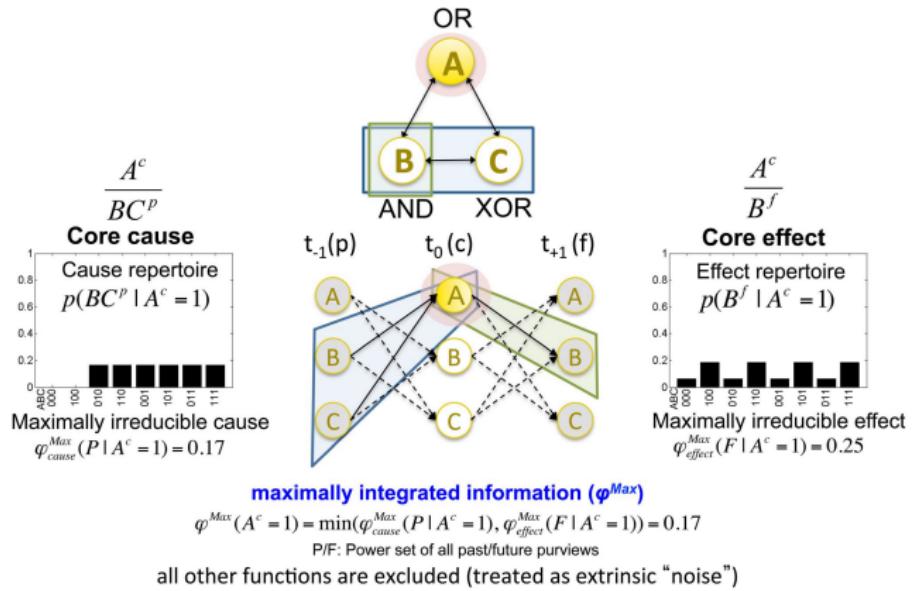


Purviews of mechanism BC (Past)

$BC^c / A^p \rightarrow (C^c / []) \times (B^c / A^p)$	$\varphi_{cause} = 0$
$BC^c / B^p \rightarrow (C^c / []) \times (B^c / B^p)$	$\varphi_{cause} = 0$
$BC^c / C^p \rightarrow (C^c / []) \times (B^c / C^p)$	$\varphi_{cause} = 0$
$BC^c / AB^p \rightarrow (C^c / AB^p) \times (B^c / [])$	$\varphi_{cause}^{Max} = 0.33$
$BC^c / BC^p \rightarrow (C^c / []) \times (B^c / BC^p)$	$\varphi_{cause} = 0$
$BC^c / AC^p \rightarrow (C^c / []) \times (B^c / AC^p)$	$\varphi_{cause} = 0$
$BC^c / ABC^p \rightarrow (C^c / []) \times (B^c / AB^p)$	$\varphi_{cause} = 0.17$

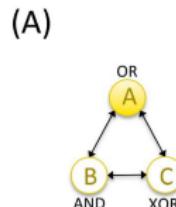
Концепт

Максимально несократимый причинно-следственный репертуар (MICE) определяет, что такое концепт, а значение ϕ^{Max} определяет меру интеграции.



Система: информативность

Множество всех концептов рассматриваемого множества образует его концептуальную структуру. Ось концептуального пространства соответствует одному из возможных сочетаний прошлого и будущего состояний системы.

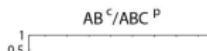
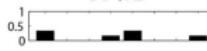
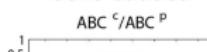


A set of elements
in their current state s_0

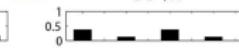
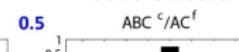
(C)

$$\varphi^{\text{Max}}(P, F | X = s_0)$$

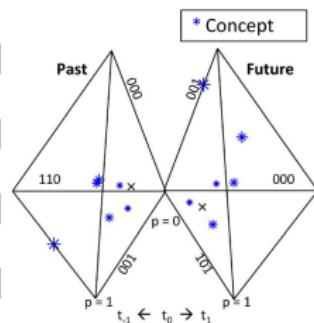
Core causes



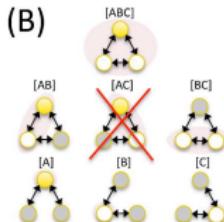
Core effects



(D)



Conceptual structure
plotted as a **constellation**
of concepts
in concept space



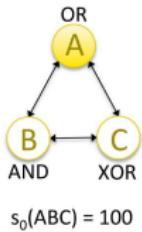
Power set of mechanisms

concepts

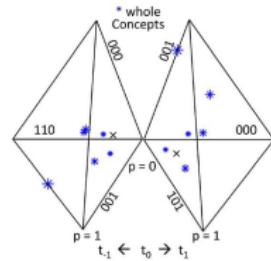
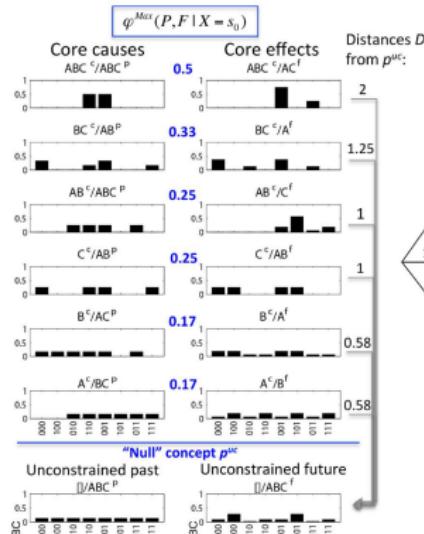
Semiotic schemas

Система: информативность

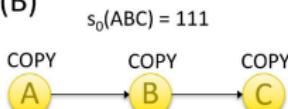
(A)



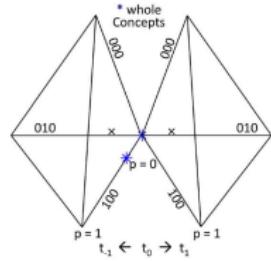
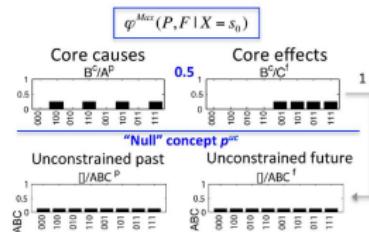
$$CI(C|s_0) = D(C\|p^{uc}) = 2.11$$



(B)

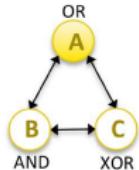


$$CI(C|s_0) = D(C\|p^{uc}) = 0.5$$

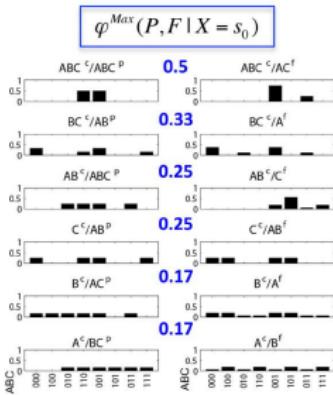


Система: интегративность

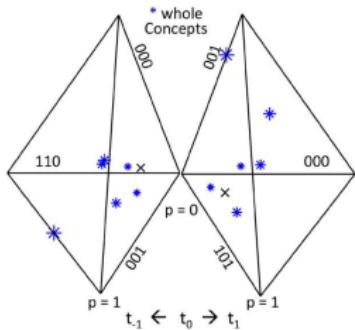
(A)



$$s_0(ABC) = 100$$

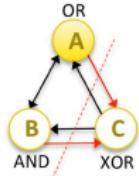


Distances D:
(Earth mover's
distance)



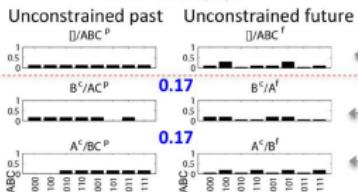
(B)

Partitioned
constellation (MIP)

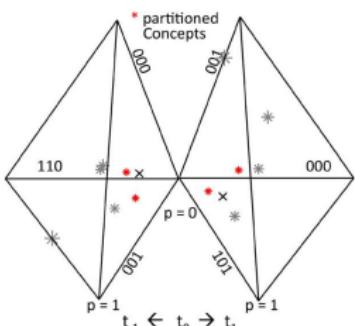


$$\Phi^{\text{MIP}}(C \mid s_0) = D(C \parallel C_{\rightarrow}^{\text{MIP}})$$

$$\rightarrow \Phi^{\text{MIP}} = 1.92$$



Unconstrained past Unconstrained future



MIP = Minimum information partition (unidirectional)

Панов (ИСА РАН)

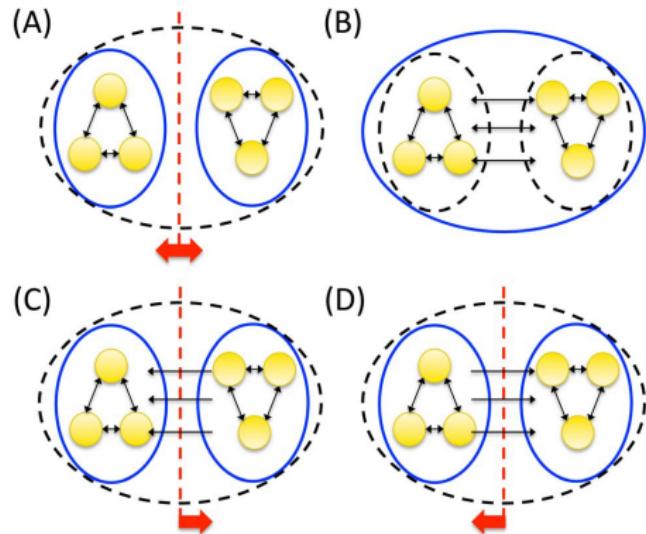
Semiotic schemas

3 марта 2016 г.

19

Система: интегративность

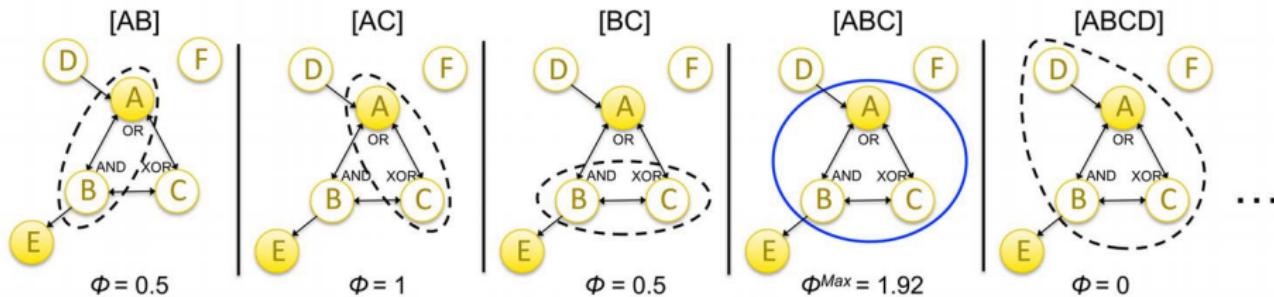
Множество элементов генерируют интегративную концептуальную информацию Φ только в том случае, если у него есть причины и эффекты в остальной части множества.



Система: единственность

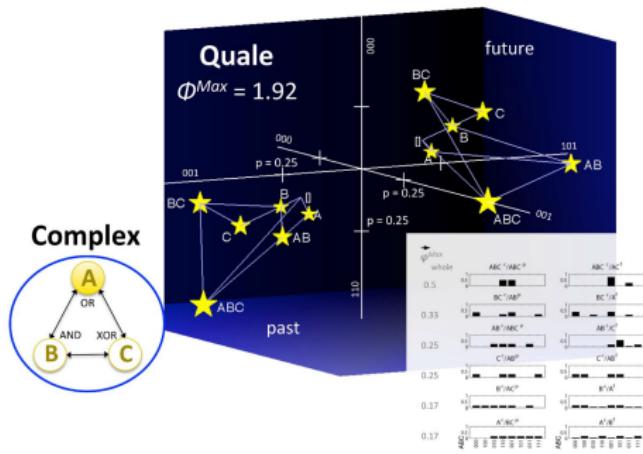
При поиске максимально несократимой концептуальной структуры (MICS), рассматриваются все возможные подсистемы системы

Power set of elements of system ABCDEF



Комплекс

Комплекс определяется как множество элементов в рамках системы, которая генерирует локальный максимум интегративной концептуальной информации Φ^{Max} . Этот максимум рассматривается не только по элементам, но и по пространственно-временному разрешению.



Центральное тождество

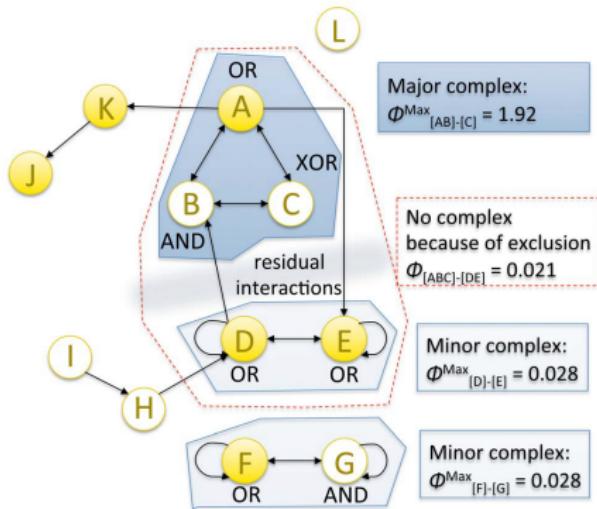
Переживание тождественно максимально несократимой концептуальной структуре (интегративной информационной структуре), определяемой подсистемами комплекса в некотором состоянии.

Количественно уровень сознания соответствует значению Φ^{Max} , качество или содержание переживания соответствует определенному «созвездию» концептов, составляющих quale.

Приложения и выводы

- Доминантный комплекс с наивысшим Φ_{Max} образуется меняющимся со временем множеством нейронов кортикалльной системы и генерирует сознание бодрствования.
- Наряду с этим комплексом есть несколько «минимально осознанных» побочных комплексов и много реактивных неосознанных процессов.
- При лоботомии могут образовываться два доминантных комплекса.

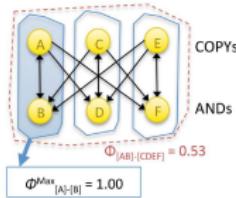
Пример, когда система сводится к одному основному комплексу и нескольким побочным.



Приложения и выводы

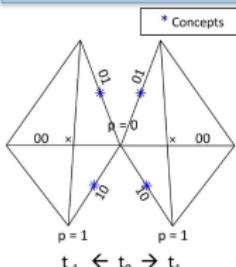
Тип связей определяет уровень сознания: мозжечок (A), кора (C), сон или анестезия (B).

(A) Modular network
COPYs and ANDs

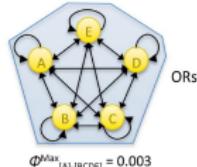


Core Concepts: 2

- 1. A'/B^{nf}: $\varphi^{\text{Max}}=0.500$
- 2. B'/A^{nf}: $\varphi^{\text{Max}}=0.500$

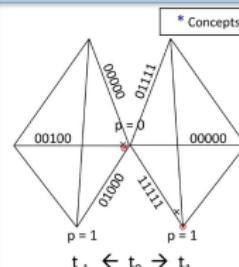


(B) Homogeneous network
all-to-all connected ORs

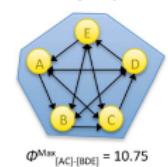


Core Concepts: 5

- 1. A'/ABCDE^{nf}: $\varphi^{\text{Max}}=0.0161$
- 2. B'/ABCDE^{nf}: $\varphi^{\text{Max}}=0.0161$
- 3. C'/ABCDE^{nf}: $\varphi^{\text{Max}}=0.0161$
- 4. D'/ABCDE^{nf}: $\varphi^{\text{Max}}=0.0161$
- 5. E'/ABCDE^{nf}: $\varphi^{\text{Max}}=0.0161$

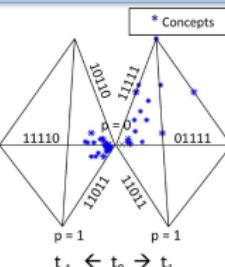


(C) Specialized network
Majority



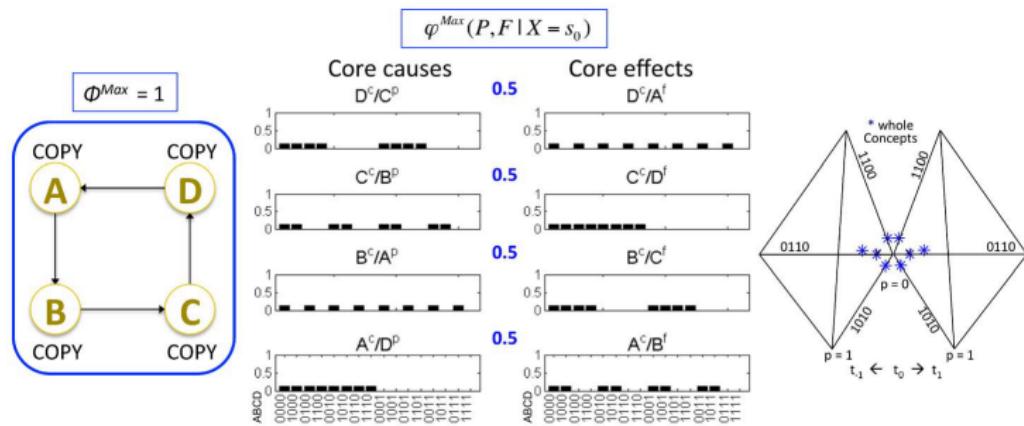
Core Concepts: 30

- 1. A'/CDE^f, BCD: $\varphi^{\text{Max}}=0.25$
- 2. B'/ADE^f, CDE: $\varphi^{\text{Max}}=0.25$
- 3. C'/ABE^f, ADE: $\varphi^{\text{Max}}=0.25$
- 4. D'/ABC^f, ABE: $\varphi^{\text{Max}}=0.25$
- 5. E'/BCD^f, ABC: $\varphi^{\text{Max}}=0.25$
- 6. AB'/ACE^f, CD^f: $\varphi^{\text{Max}}=0.2$
- 7. AC'/ABCDE^f: $\varphi^{\text{Max}}=0.2$
- ⋮



Приложения и выводы

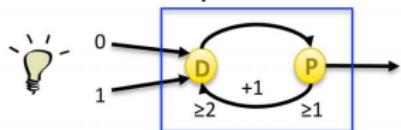
Неактивная система может быть осознанной, т.к. и активные и неактивные элементы вносят вклад в концептуальную структуру (доминантный комплекс только из неактивных нейронов!).



Приложения и выводы

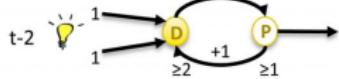
Даже простые системы обладают сознанием: «минимально осознанный» фотодиод.

(A) A minimally conscious photodiode

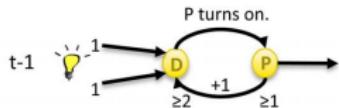


(B) Mode of operation

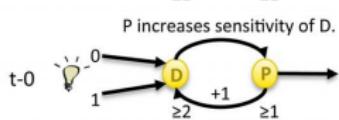
Strong light switches D on.



P turns on.

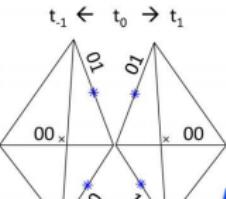


P increases sensitivity of D.



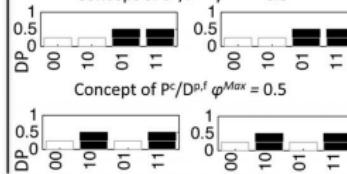
(C) Quale

$$\phi_{Max}(\text{photodiode}) = 1$$

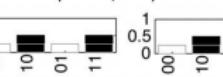


* Concepts

$$\text{Concept of } D^c/p_{p,f} \phi_{Max} = 0.5$$



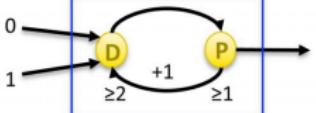
$$\text{Concept of } P^c/D^{p,f} \phi_{Max} = 0.5$$



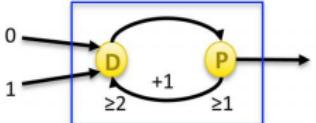
(D)

Same quale, no quality

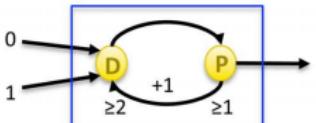
A minimally conscious photodiode



A minimally conscious blue detector



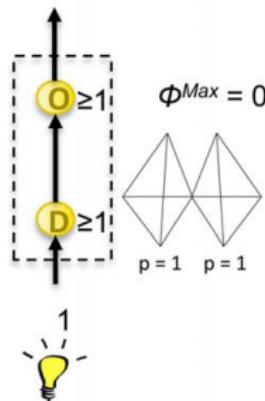
A minimally conscious thermistor



Приложения и выводы

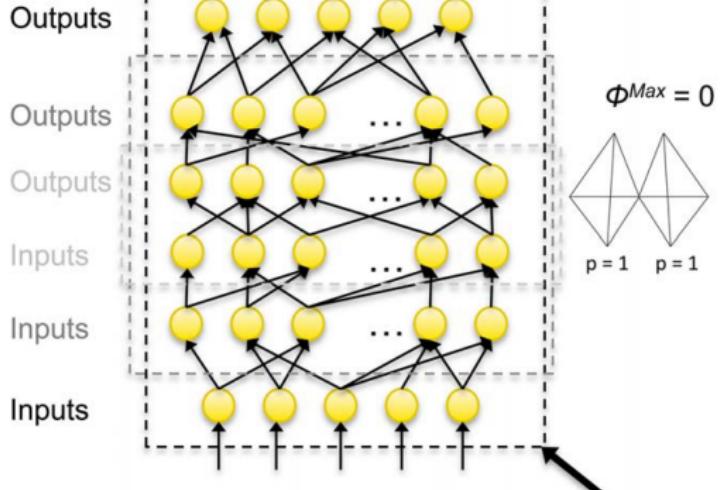
Даже сложные системы могут не обладать сознанием: реактивная «зомби»-сеть. Важна рекурсия, обратная связь, повторный вход, но только их не достаточно.

(A) Unconscious photodiode



(B)

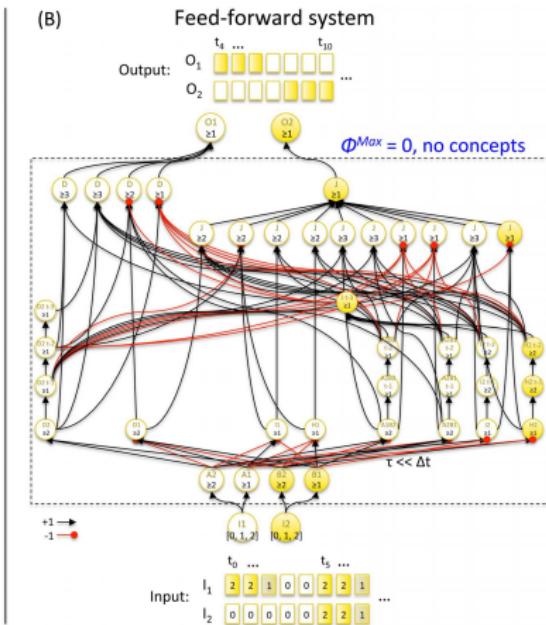
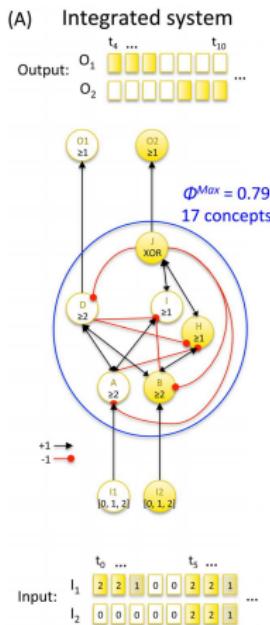
Feed-forward system



Extrinsic definition of a system

Приложения и выводы

Сознательная сложная система и бессознательная «зомби»-система могут быть функционально эквивалентны. Осознанность нельзя определить только по внешним проявлениям.



Ограничения и неполнота теории

Данный подход применить для моделирования реальной физической системы (мозга) невозможно:

- необходимо либо дискретизировать интересующие перемененные или расширять модель до непрерывных значений,
- для биологических систем очень немного наблюдаемых состояний,
- высокая вычислительная сложность уже при более 12 элементов.

Необходимо уточнить следующие моменты:

- не обсуждалась связь MISC и ощущений разной модальности,
- необходимо вычислять пространственно-временное разрешение, а не задавать его заранее оптимальным,
- не обсуждалось как значение комплекса задается через MICS,
- не обсуждается соотнесение внешней каузальной структуры внешней среды и внутренней каузальной структуры MICS (повышение осознанности при адаптации к внешней среде),
- гипотеза о том, что интегрированная информация и каузальность одно и то же, требует уточнения.