6. integrated information theory

Integrated information theory of consciousness

- Starts from phenomenology, identifies five axioms (1. existence, 2. composition, 3. information, 4. integration, 5. exclusion)
- Tries to seek for potential physical mechanisms that can support the axioms

Tononi 2004, 2008, Oizumi et al 2014 PLoS Comp Bio

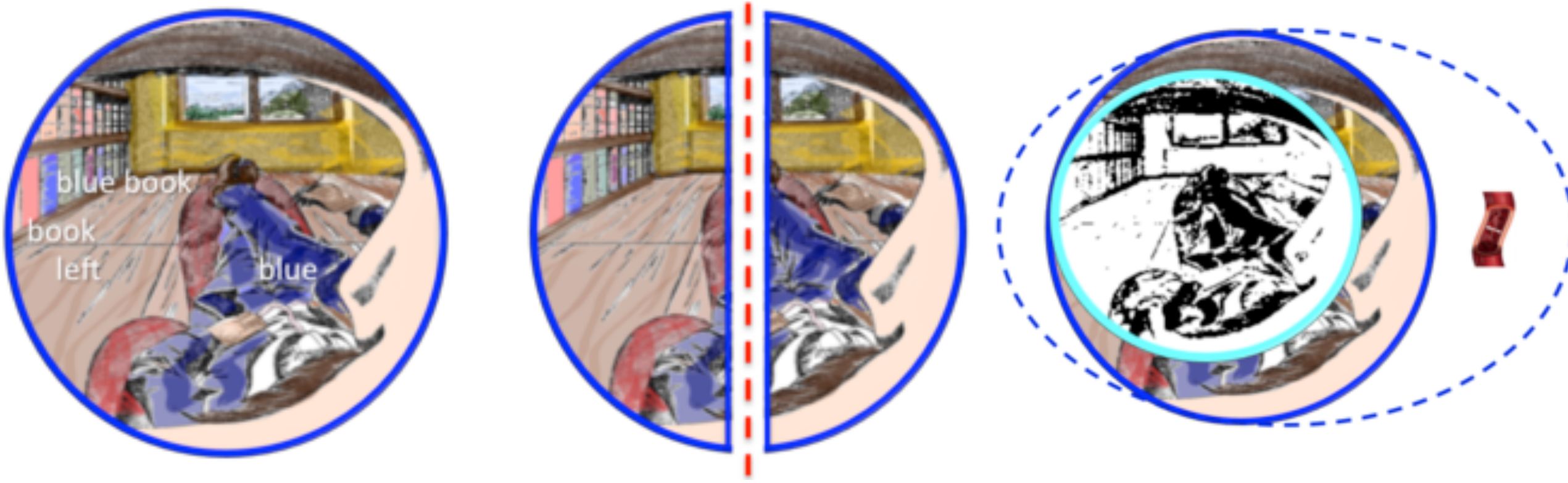
Consciousness is highly **informative**

- Each experience is highly unique and excludes all other possible experiences

Consciousness is **composed** of various aspects

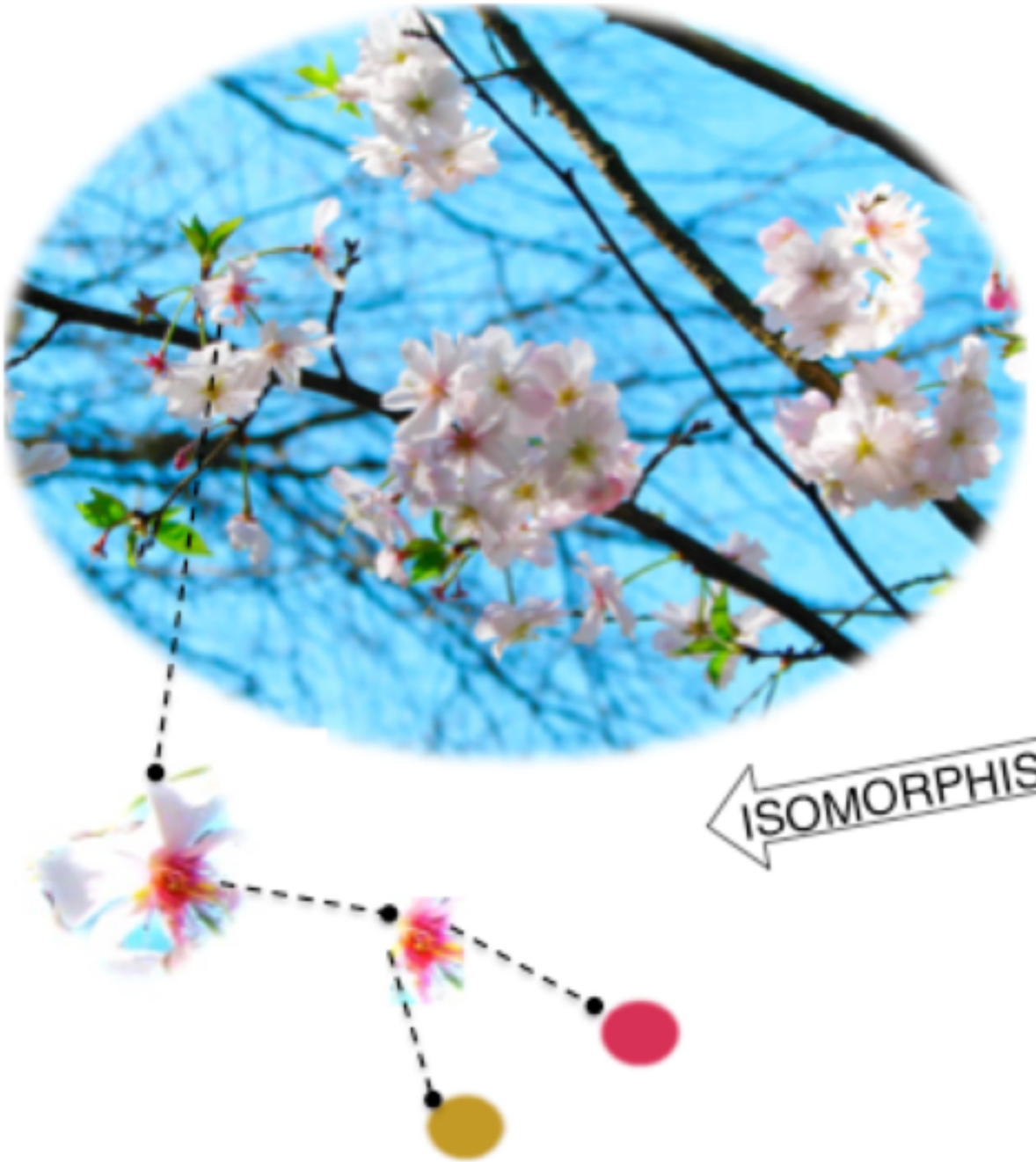
Consciousness is **integrated**

Consciousness is **excluded** outside of the certain boundary



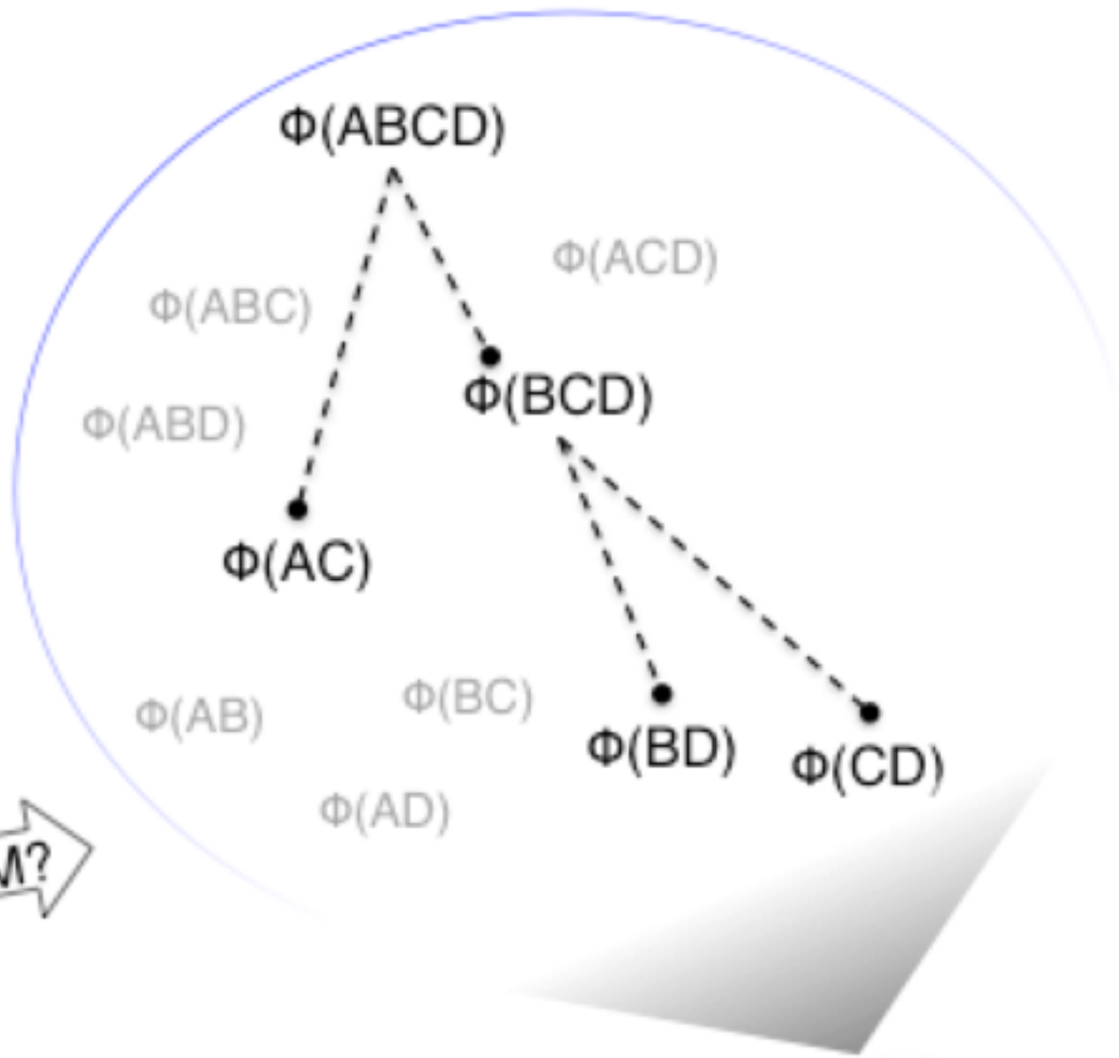
- Based on the 5 phenomenological axioms, IIT derives **5 mathematical postulates.**
- Using the postulates, IIT predicts a particular **level and contents of consciousness** of a system, based on **how the network is connected and which state the system is in.**

Conscious experience:
Intrinsic, integrated, hierarchical structure

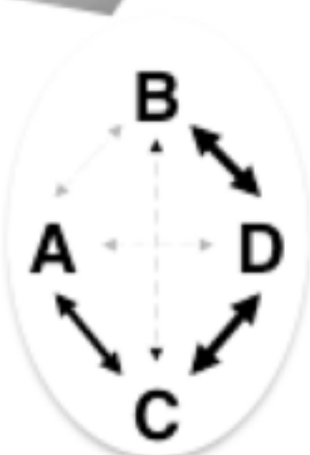


ISOMORPHISM?

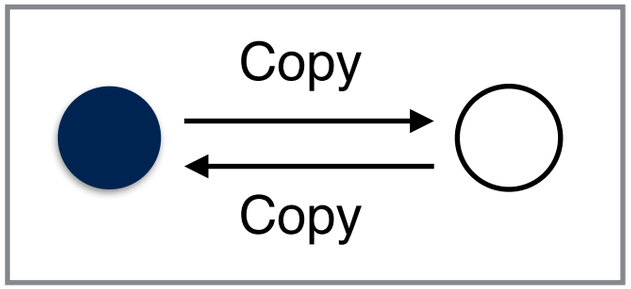
Integrated information:
Intrinsic, integrated, hierarchical structure

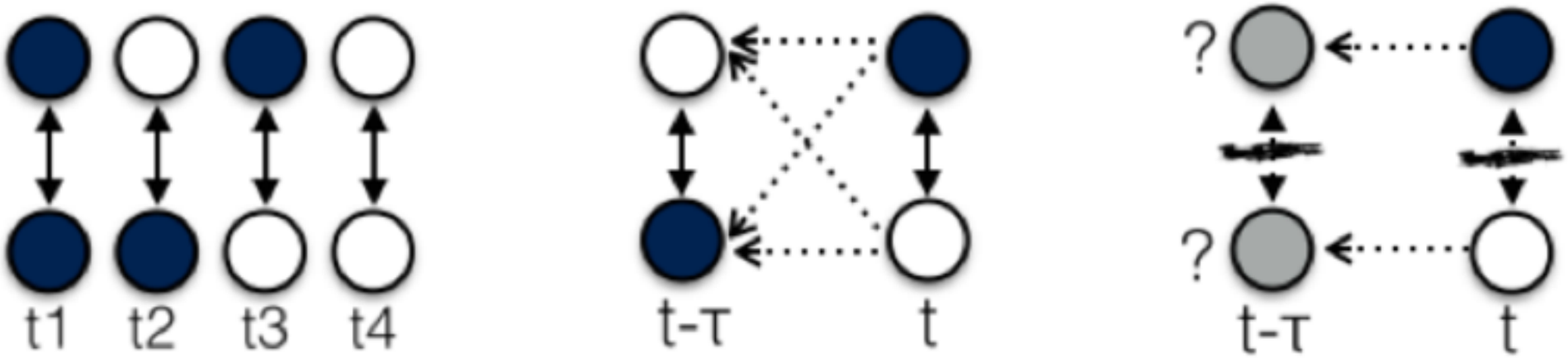


Physical substrate:
interacting elements



Integrated information theory in a nut shell



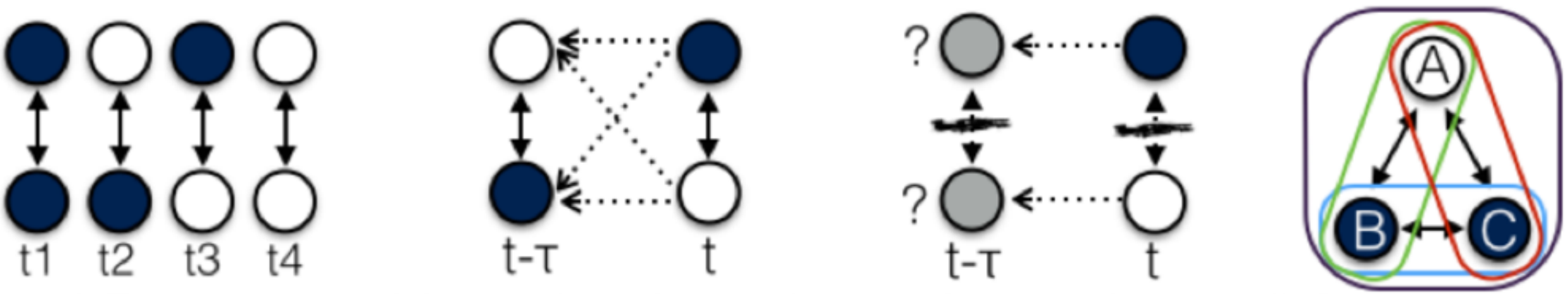


a) **Entropy** b) **Intrinsic information** c) **Integration**

$H=log_2(4)=2$ $I=H-H^*=2$ $\Phi=I-I^*=2$

Tsuchiya 2017 Philosophy Compass

Integrated information theory in a nut shell



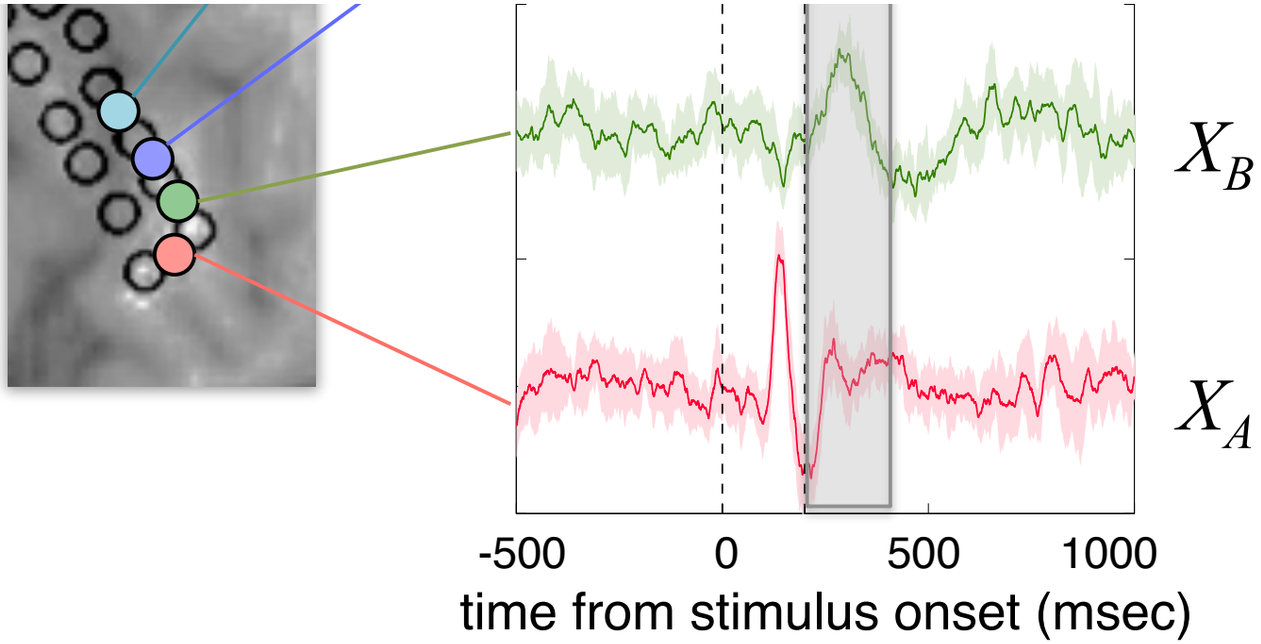
a) **Entropy** b) **Intrinsic information** c) **Integration** d) **Composition**

$H=log_2(4)=2$ $I=H-H^*=2$ $\Phi=I-I^*=2$ ΦAB, ΦAC, ΦBC, ΦABC

Tsuchiya 2017 Philosophy Compass

Continuous variables with Gaussian assumption

https://figshare.com/articles/phi_toolbox_zip/3203326



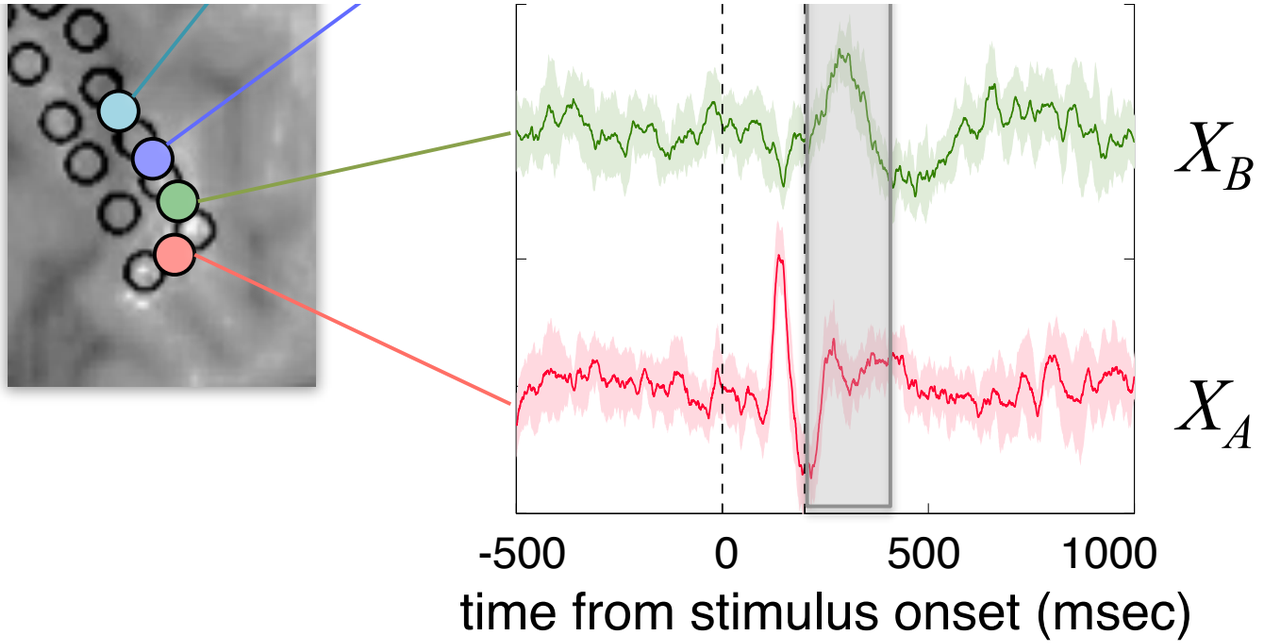
$$H(X) = \frac{1}{2}\log|\Sigma(X)| + \frac{1}{2}N\log(2\pi e),$$

Measures variability of responses

Oizumi et al 2016 PLoS Comp

Continuous variables with Gaussian assumption

https://figshare.com/articles/phi_toolbox_zip/3203326



$$H(X) = \frac{1}{2}\log|\Sigma(X)| + \frac{1}{2}N\log(2\pi e),$$

$$I(X^{t-\tau}; X^t) = H(X^{t-\tau}) - H(X^{t-\tau}|X^t)$$

$$\boldsymbol{\Phi^* = I - I^*}$$

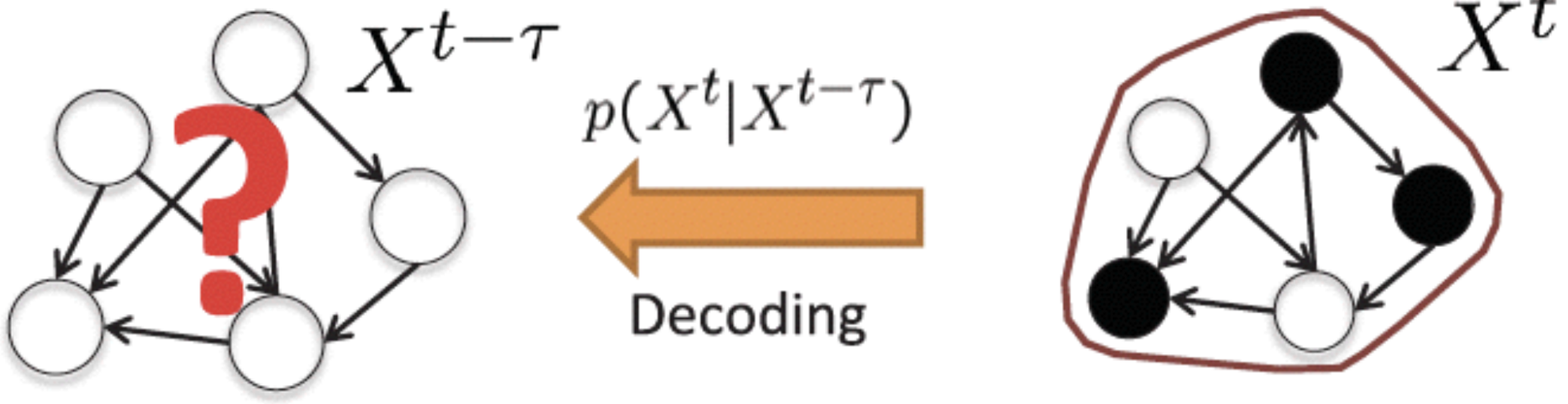
Oizumi et al 2016 PLoS Comp

Intuitive explanation 1

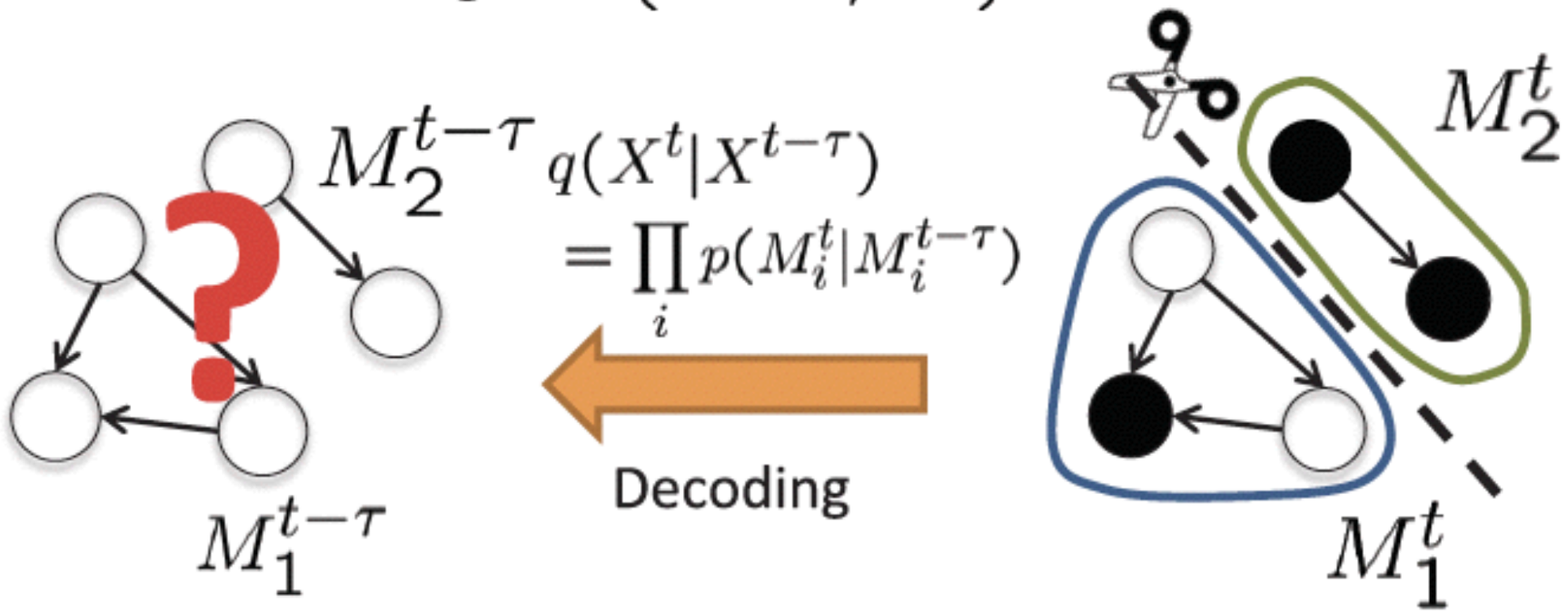
Integrated information
=
loss of predictability of past states based on current states, when system is cut

(Oizumi et al 2016 PLoS Comp)

Matched decoding $I(X^{t-\tau}; X^t)$



Mismatched decoding $I^*(X^{t-\tau}; X^t)$



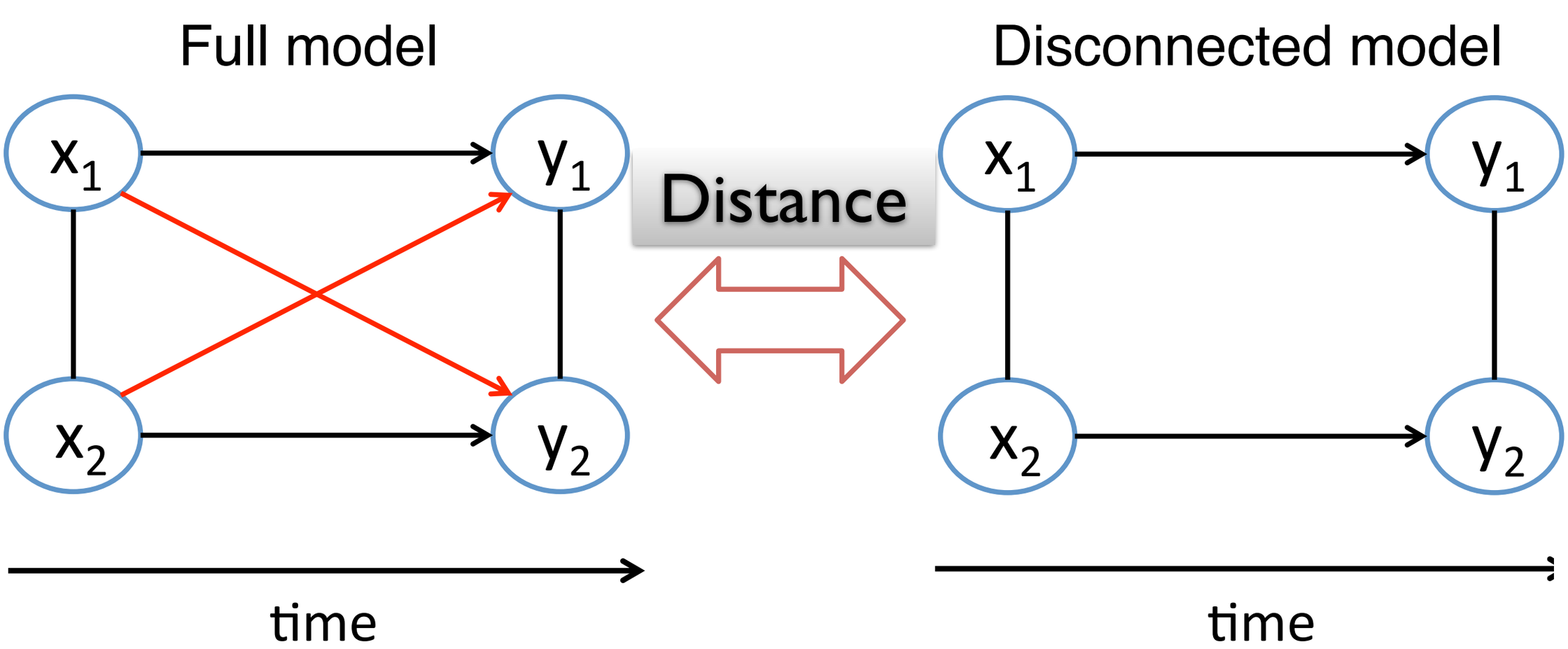
Integrated information $\phi^* = I(X^{t-\tau}; X^t) - I^*(X^{t-\tau}; X^t)$

(Oizumi et al 2016 PLoS Comp)

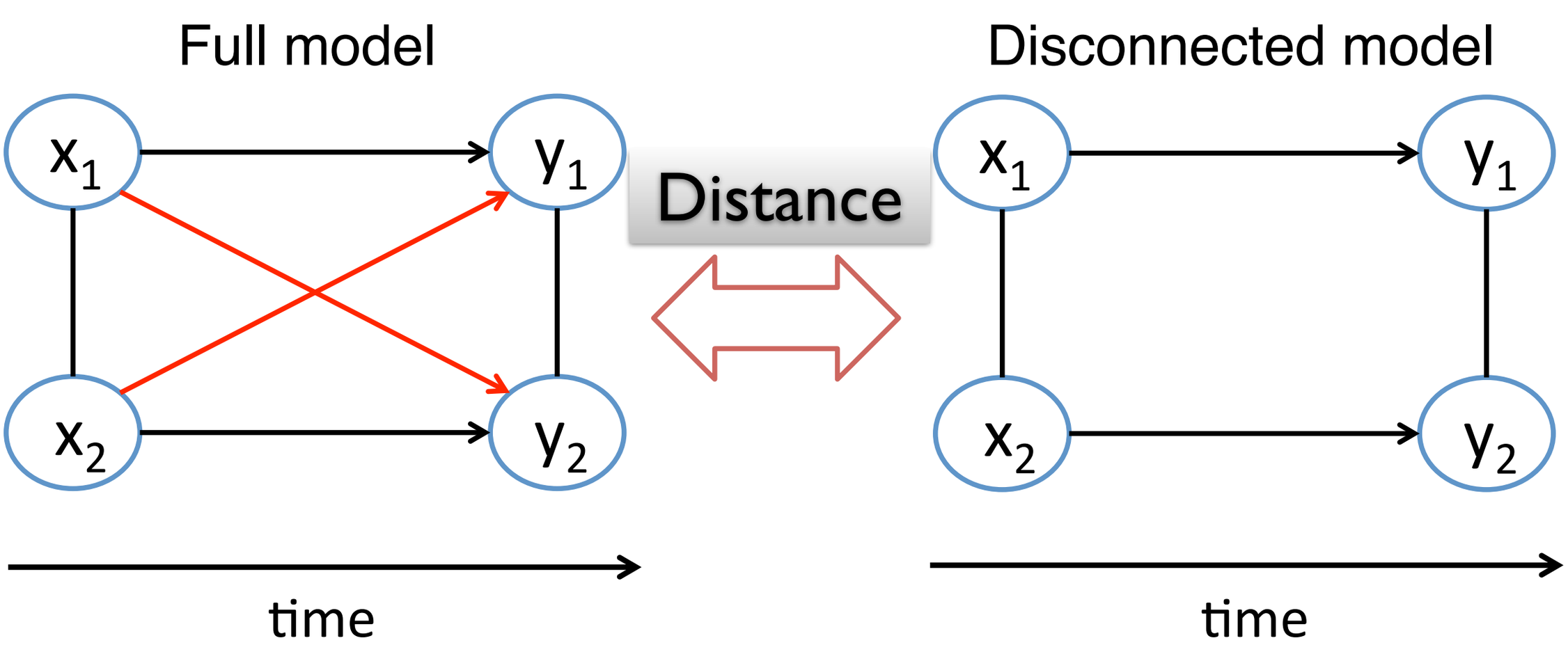
Intuitive explanation 2

Integrated information
=
a *distance* between the actual and disconnected model

Oizumi, Tsuchiya, Amari 2016 PNAS



Oizumi, Tsuchiya, Amari 2016 PNAS

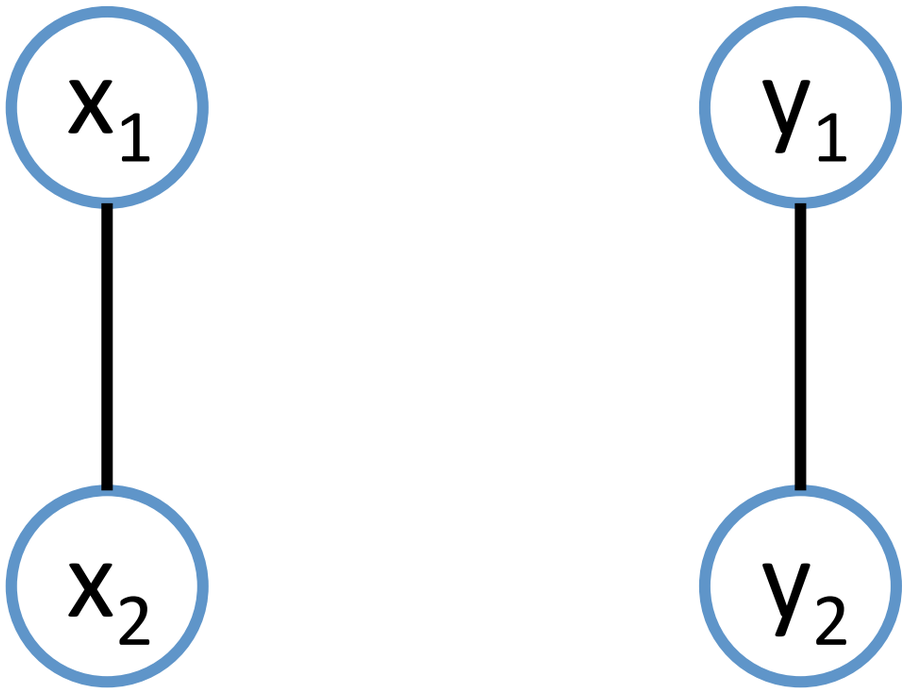


Integrated information

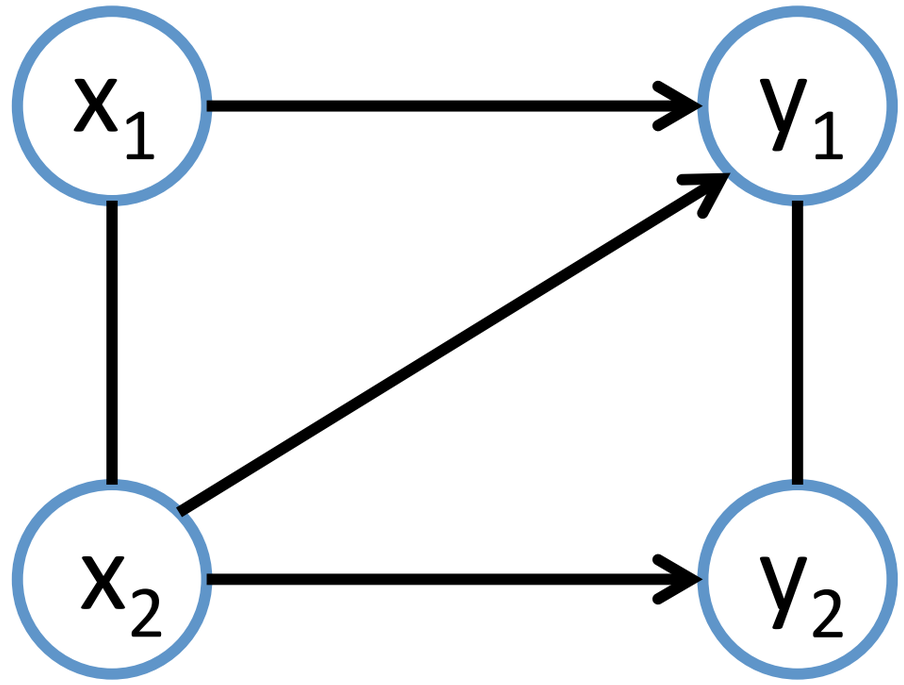
$$\Phi_G = \min_{q(X,Y)} D\left(p(X,Y) || q(X,Y)\right)$$

$$\Phi_G = \frac{1}{2} \log \frac{|\Sigma(E)'|}{|\Sigma(E)|}.$$

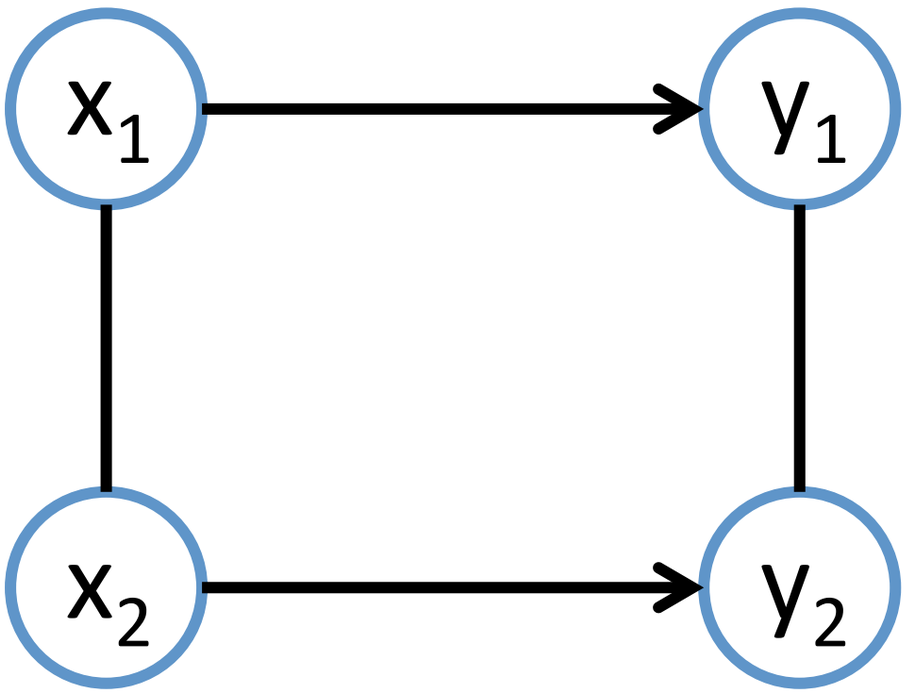
Mutual information



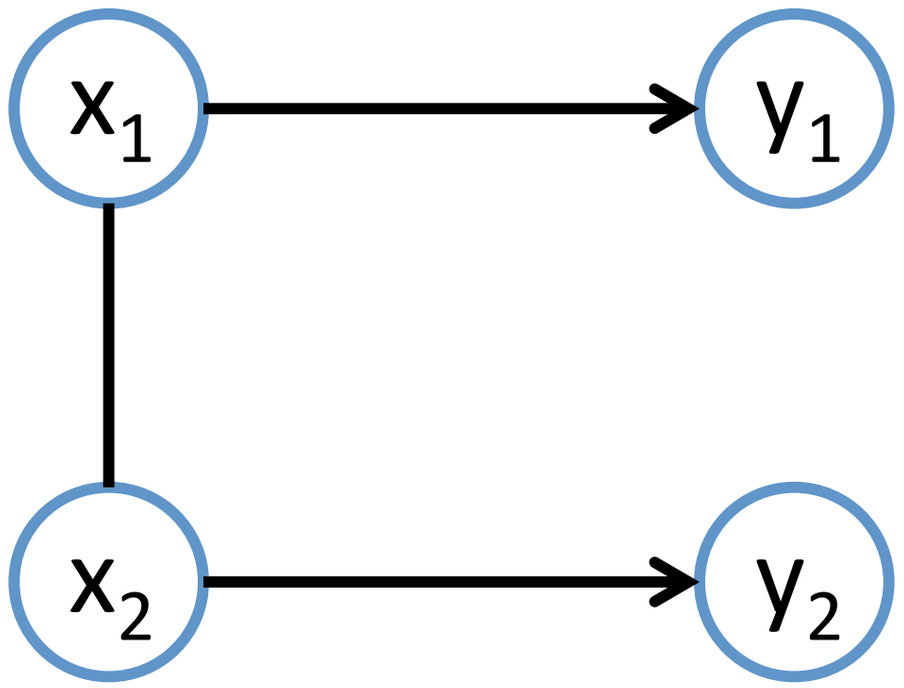
Transfer entropy



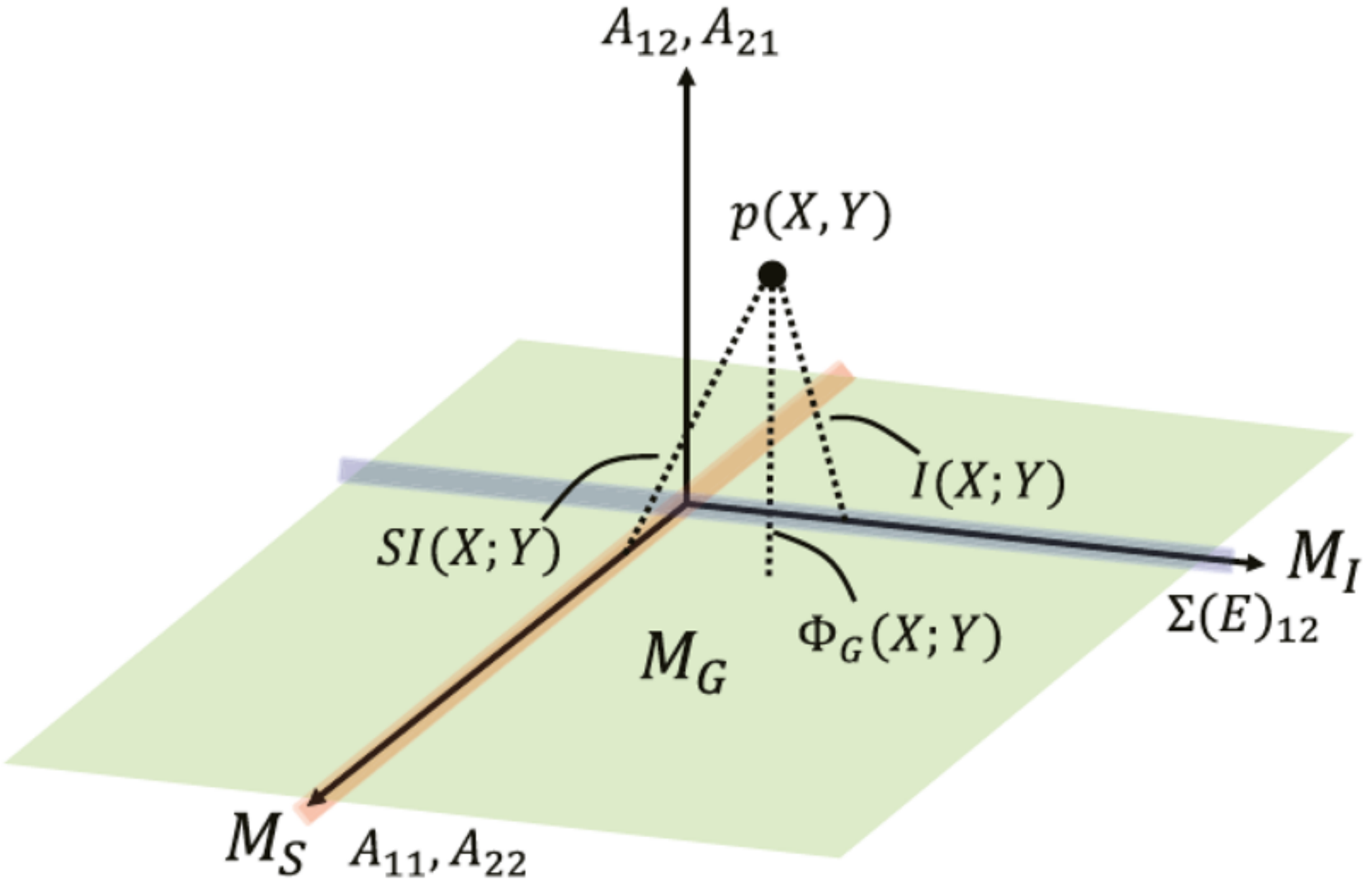
Integrated information



Stochastic interaction

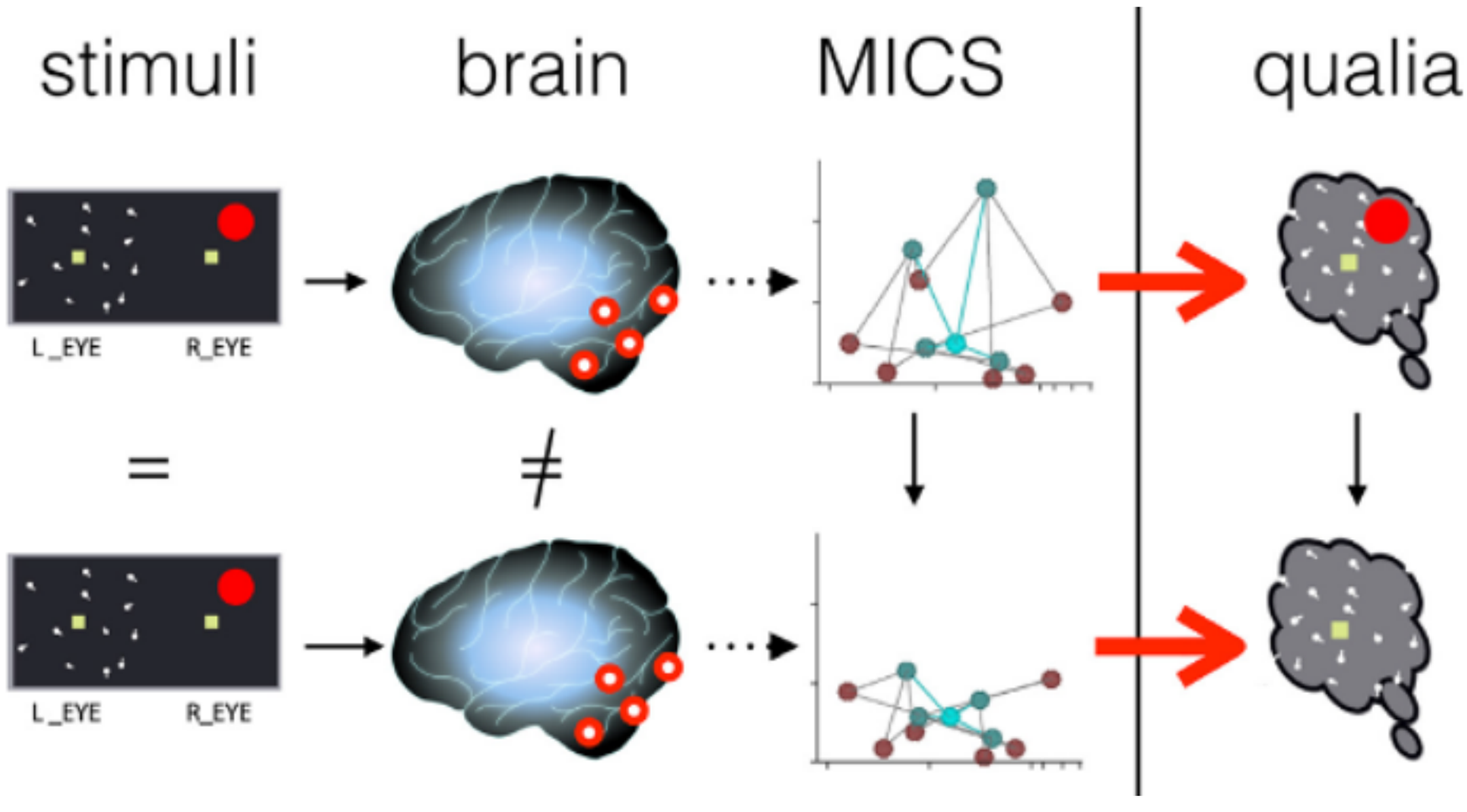


A unified framework for causal interactions based on information geometry



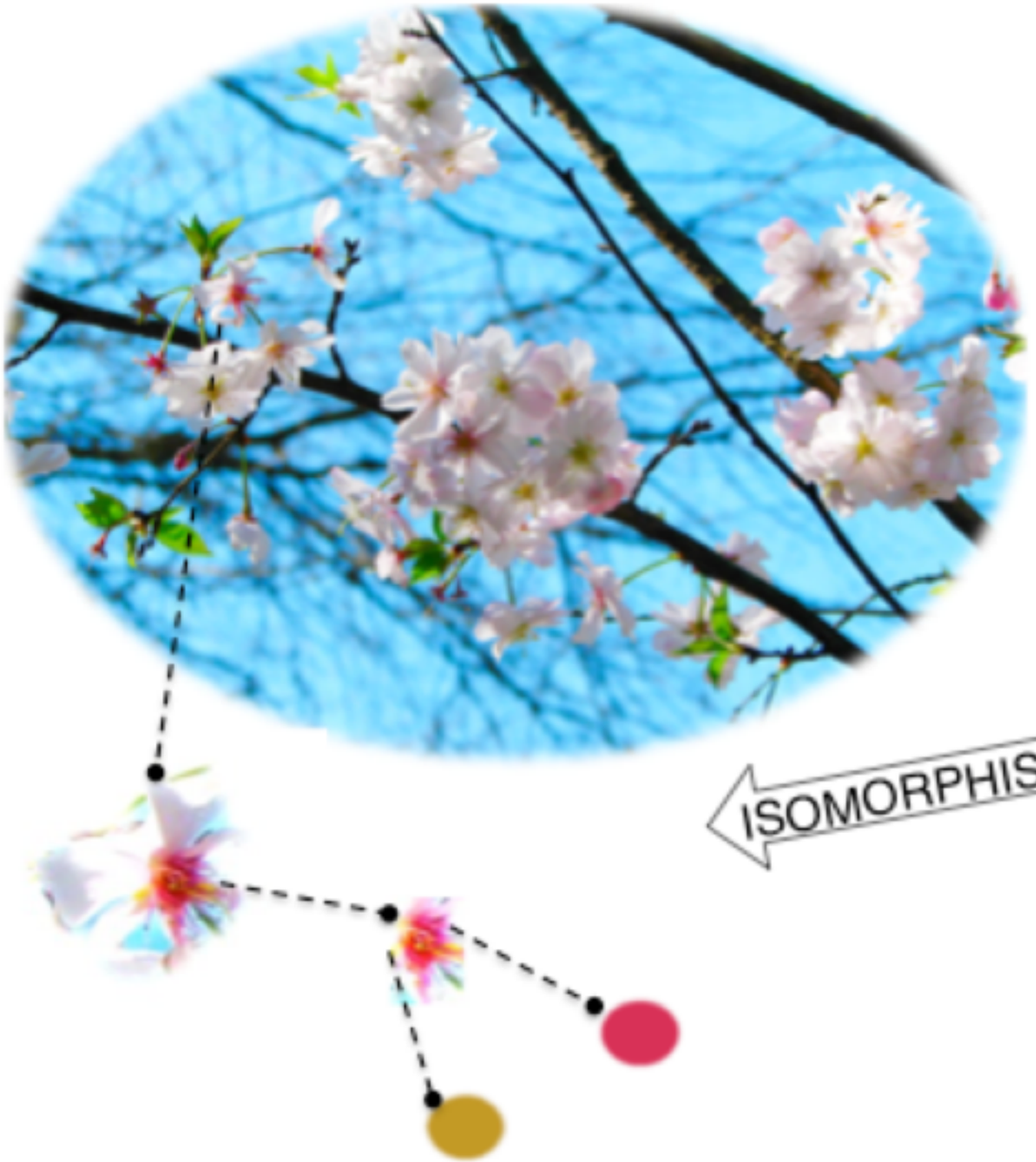
Oizumi, Tsuchiya, Amari 2016 PNAS

- Empirical tests of (a subset of predictions from) IIT on contents of consciousness
  - Patterns of integrated information should be *isomorphic* to contents of consciousness



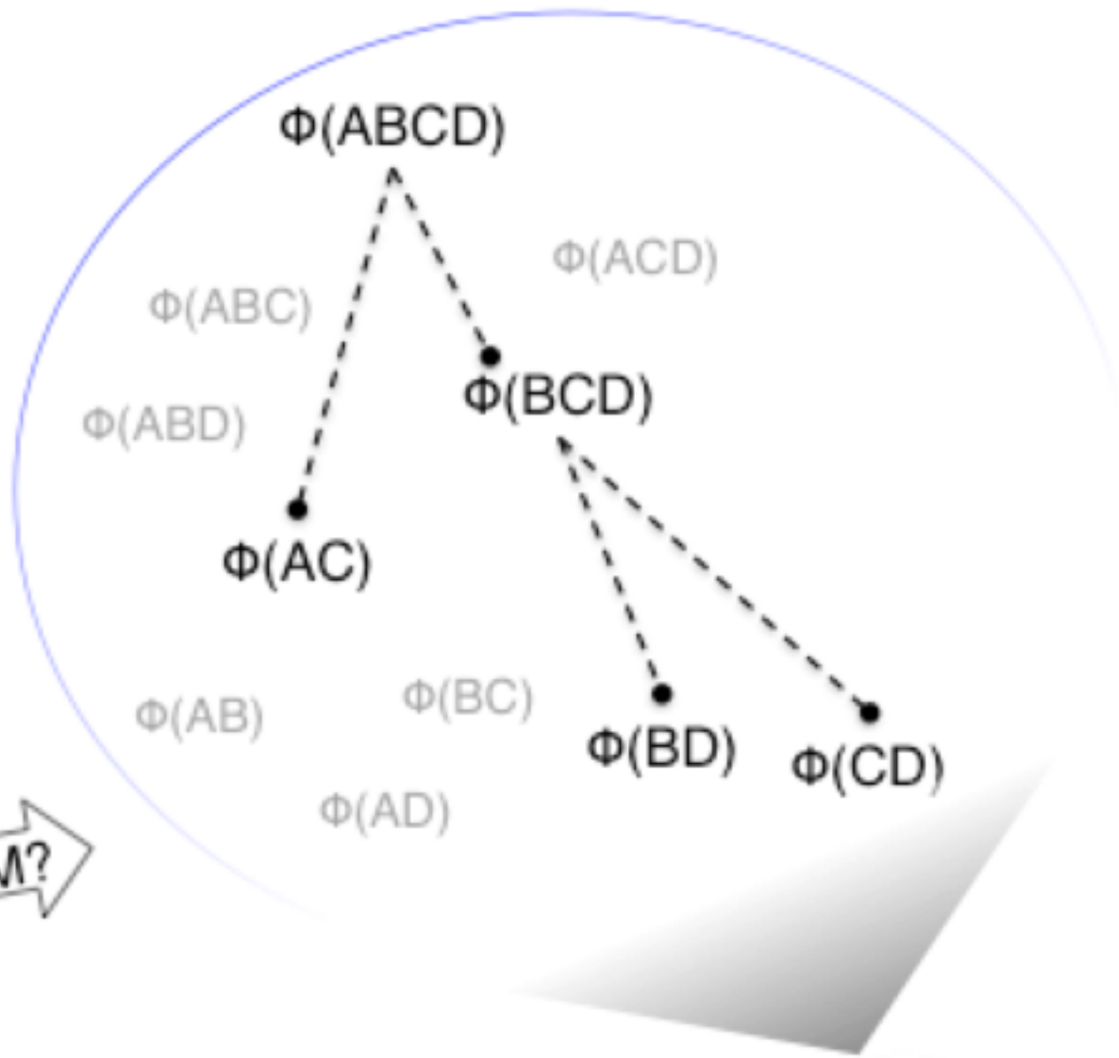
Tsuchiya, Taguchi, Saigo 2016 Neuroscience Research

Conscious experience:
Intrinsic, integrated, hierarchical structure

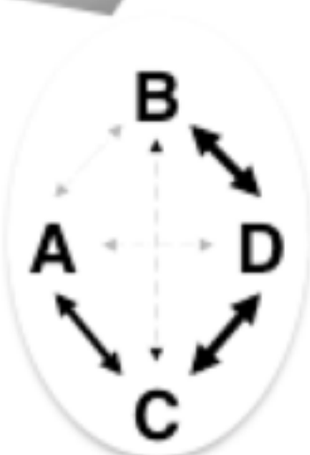


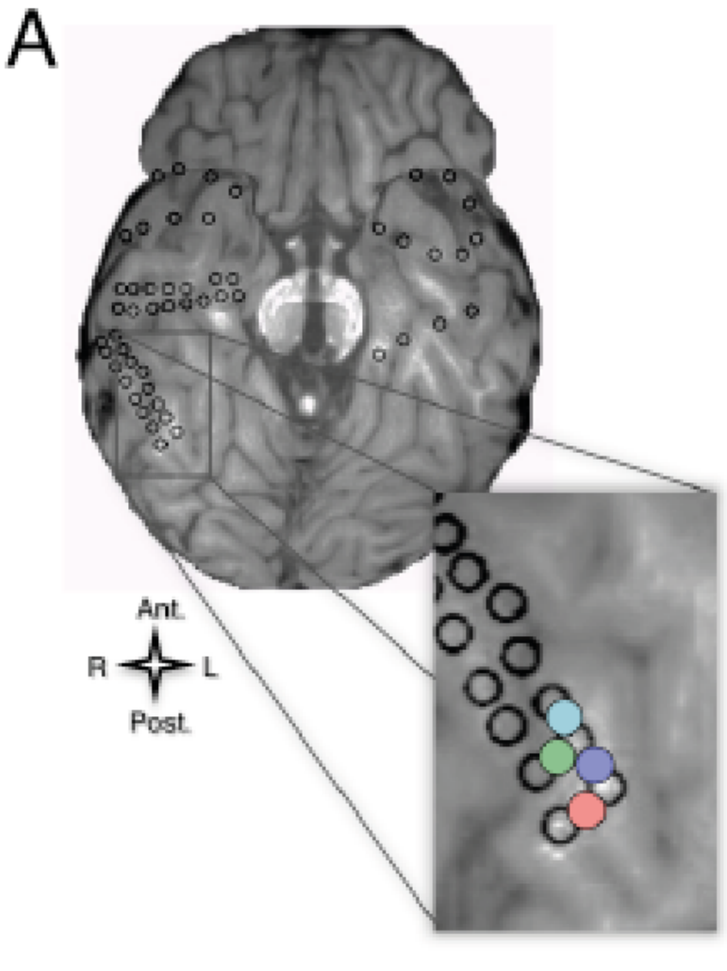
ISOMORPHISM?

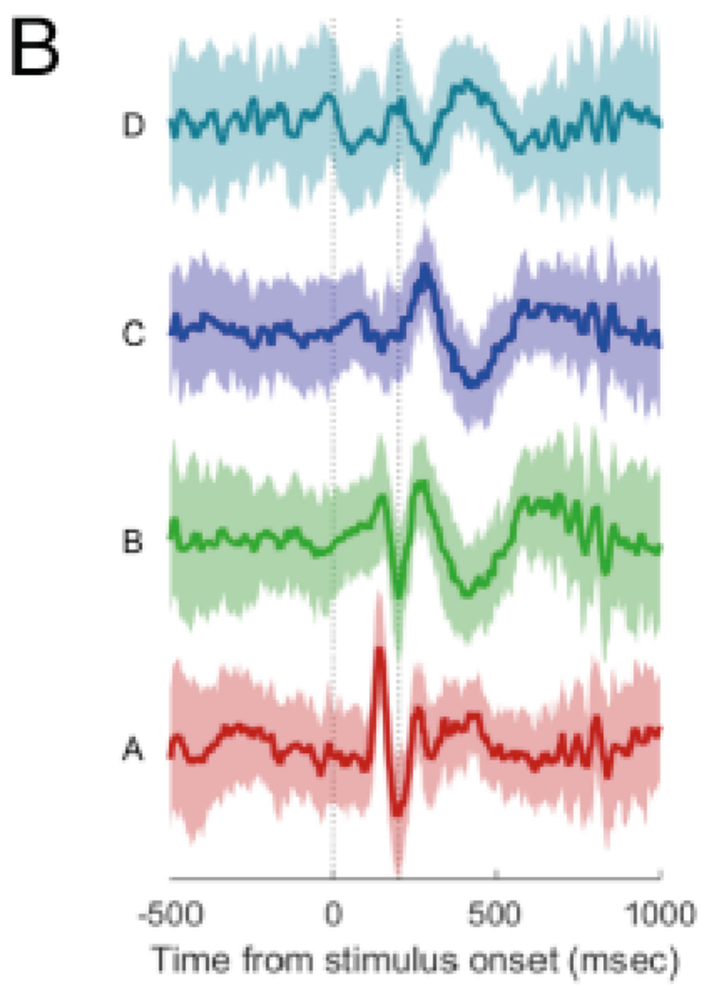
Integrated information:
Intrinsic, integrated, hierarchical structure



Physical substrate:
interacting elements

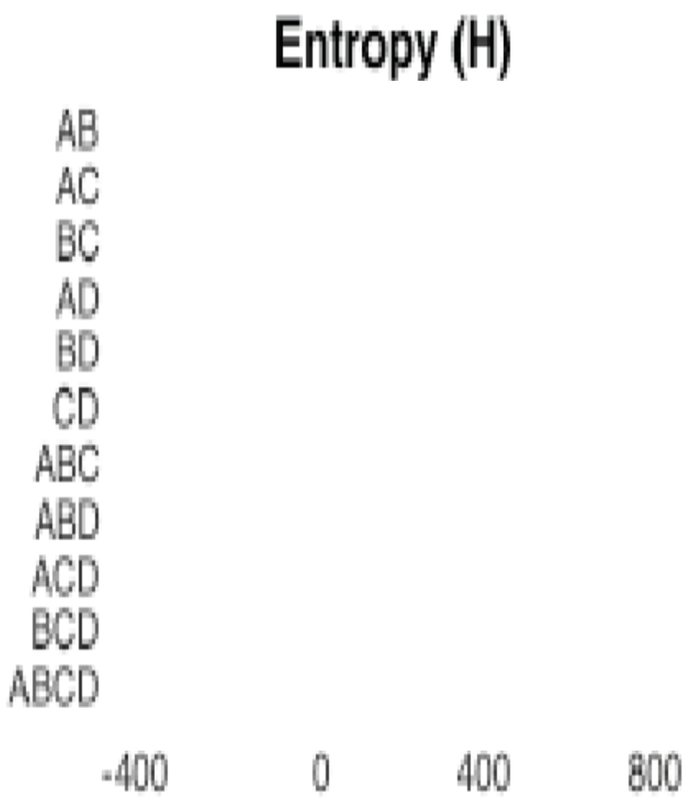




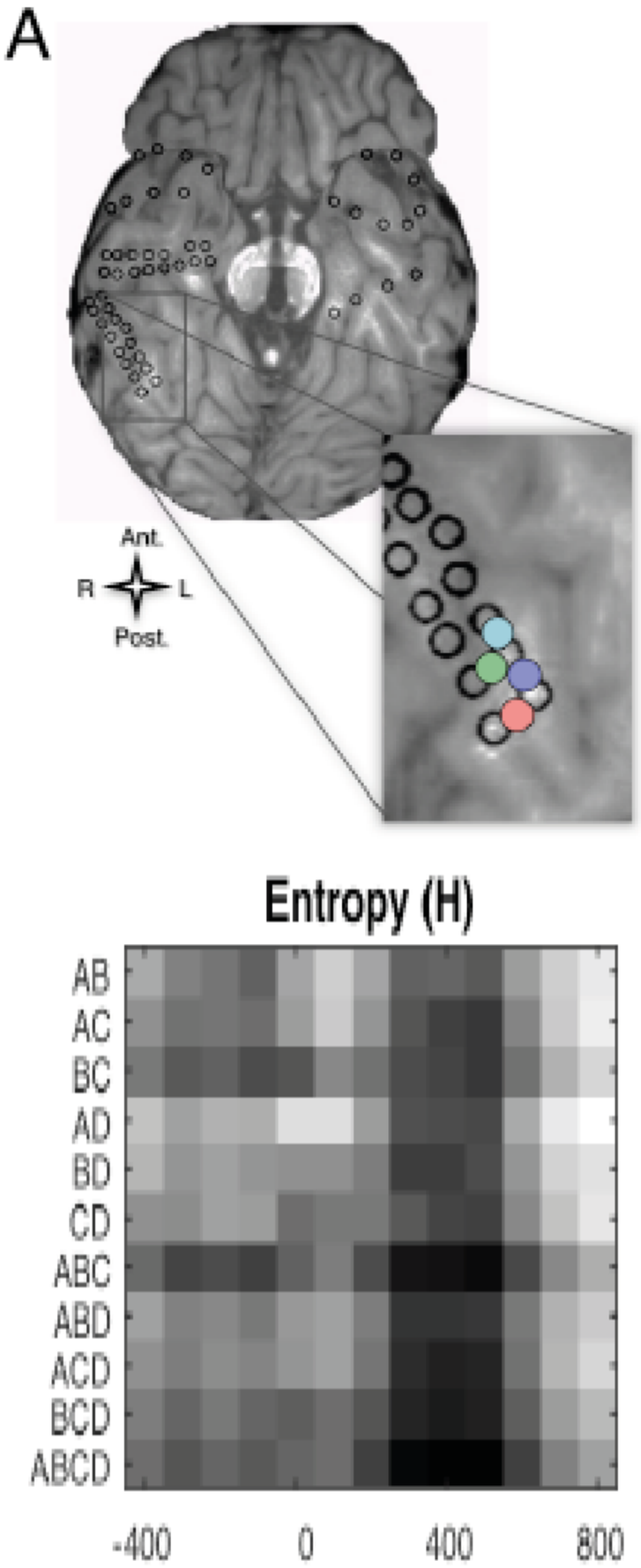


11 possible subsets
for 4 channels:

6 x 2 channels (AB, AC,..)
4 x 3 channels (ABC, …)
1 x 4 channels (ABCD)



Haun et al 2016 bioRxiv

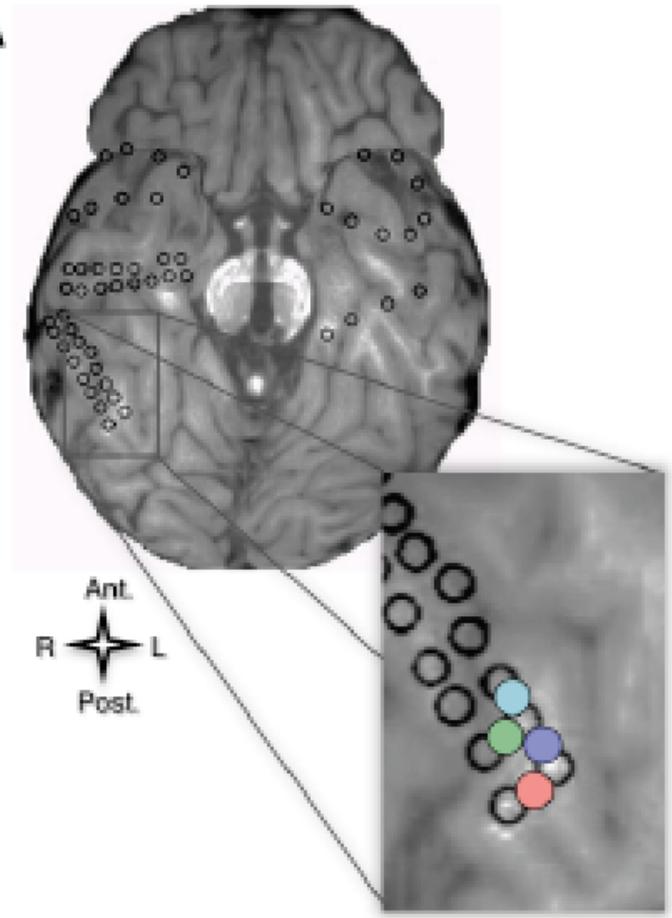


Gaussian assumption

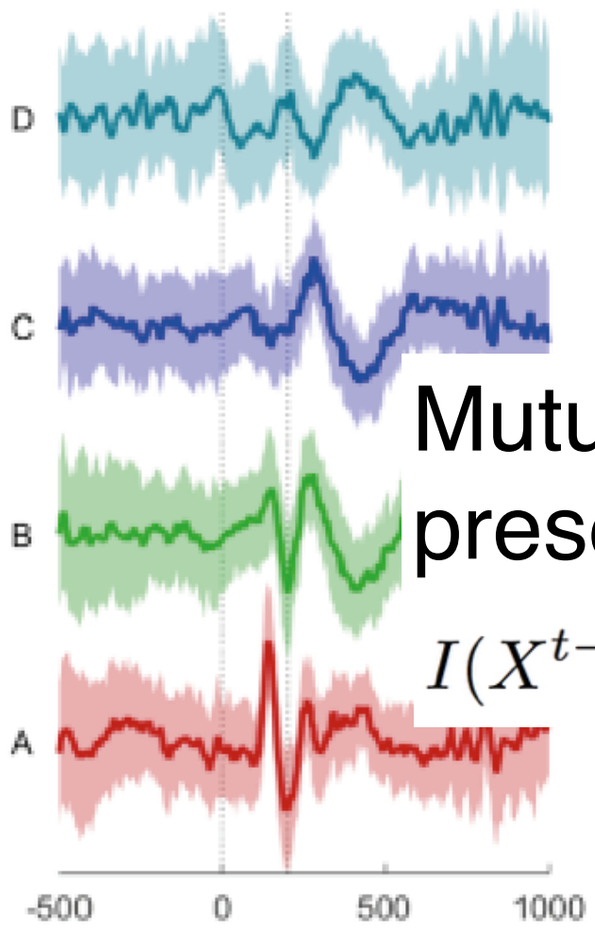
Randomness of the system
- entropy: $H(X^t)$

Haun et al 2016 bioRxiv

A

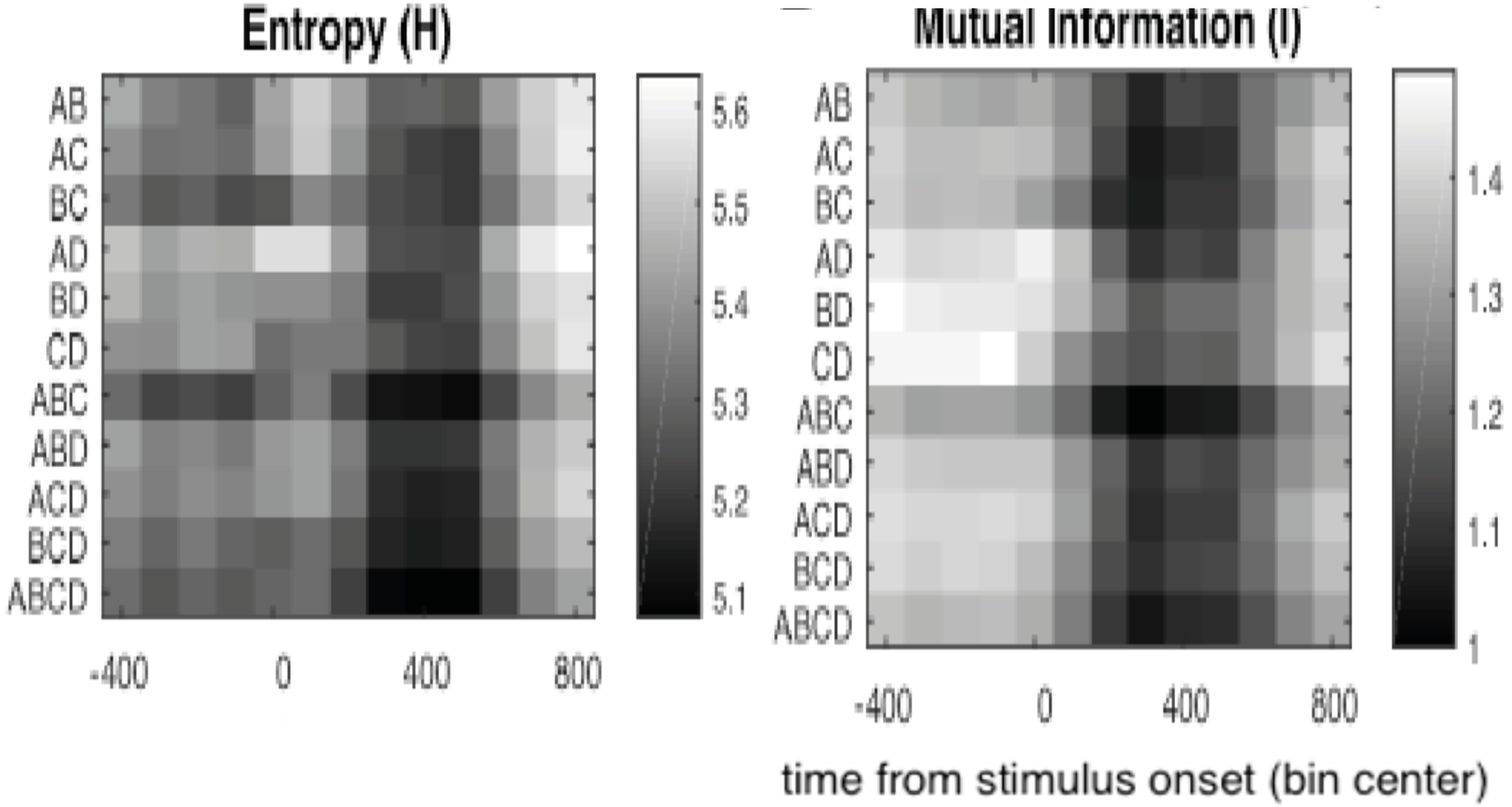


B



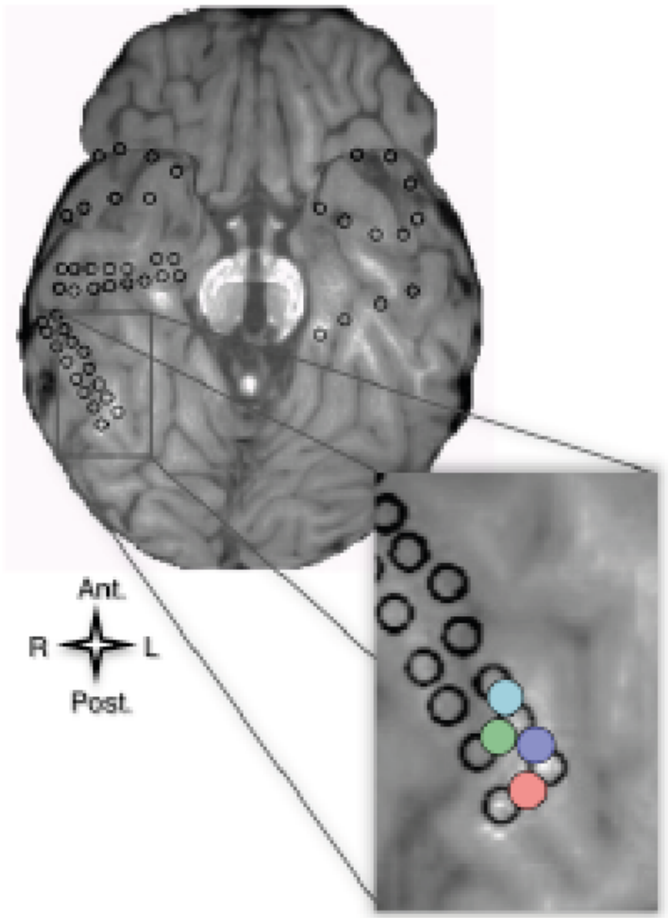
Mutual information between present and past (tau=3 ms)

$$I(X^{t-\tau}; X^t) = H(X^{t-\tau}) - H(X^{t-\tau}|X^t)$$

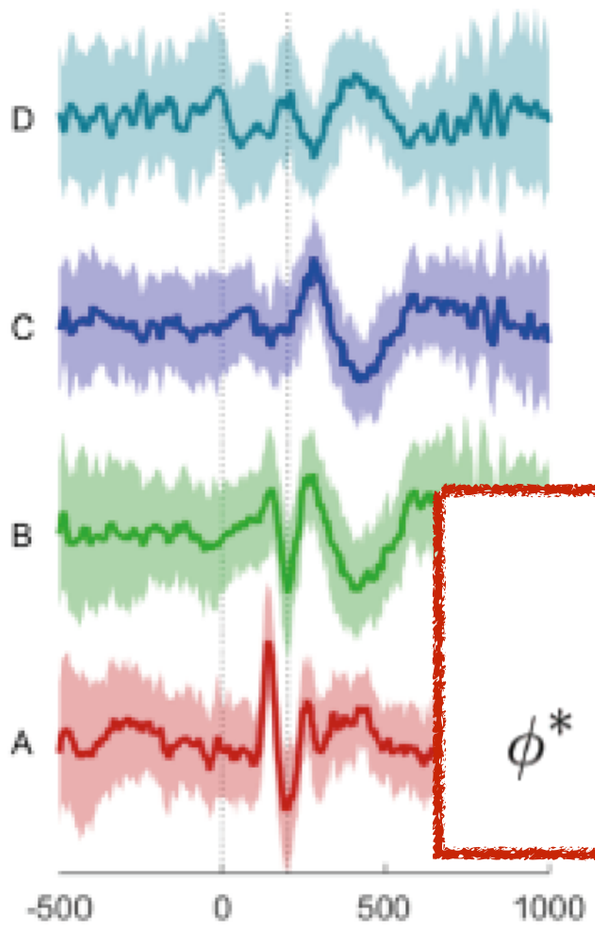


Haun et al 2016 bioRxiv

A

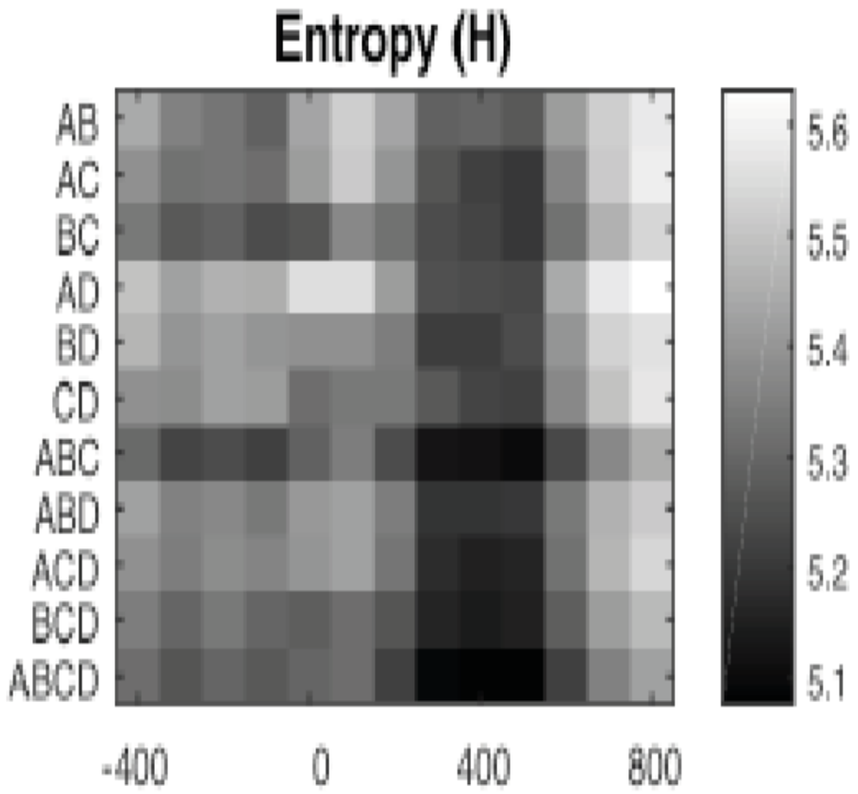


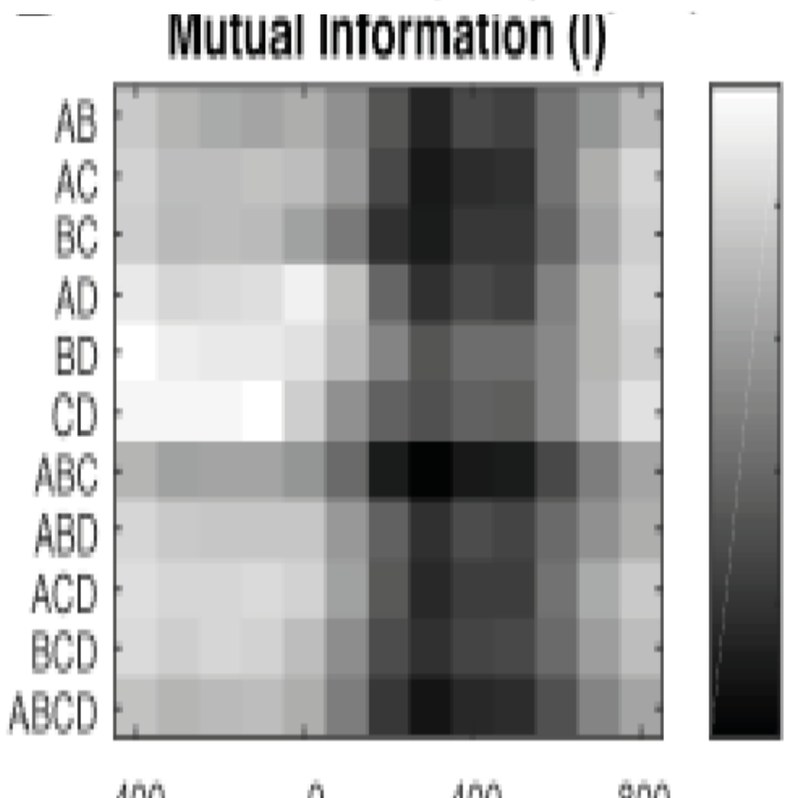
B

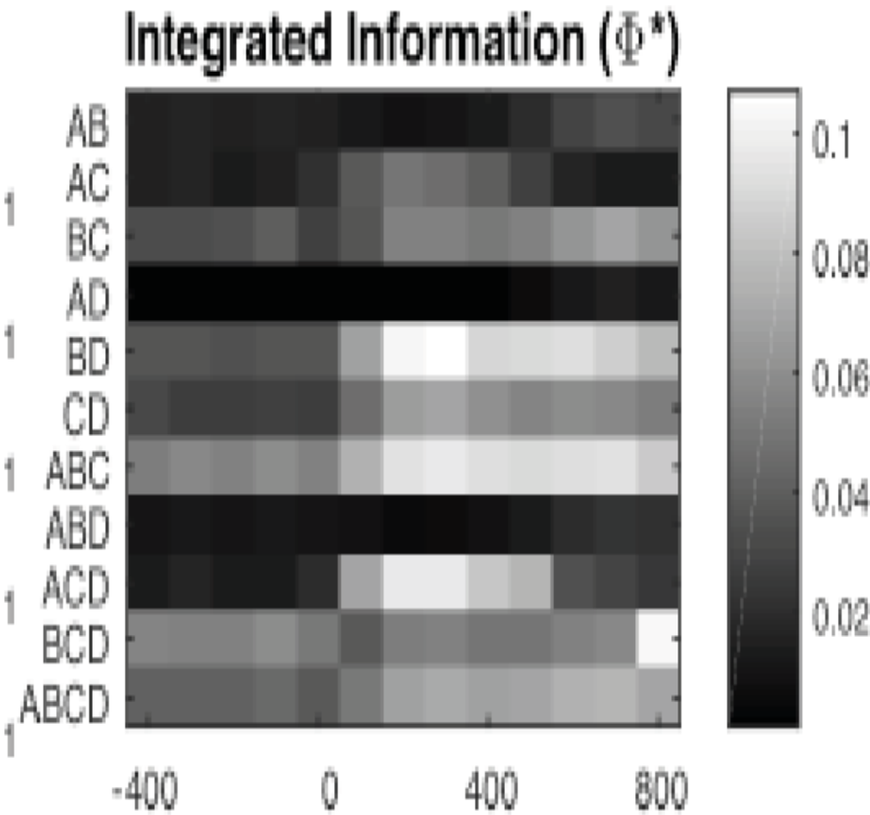


Integrated information

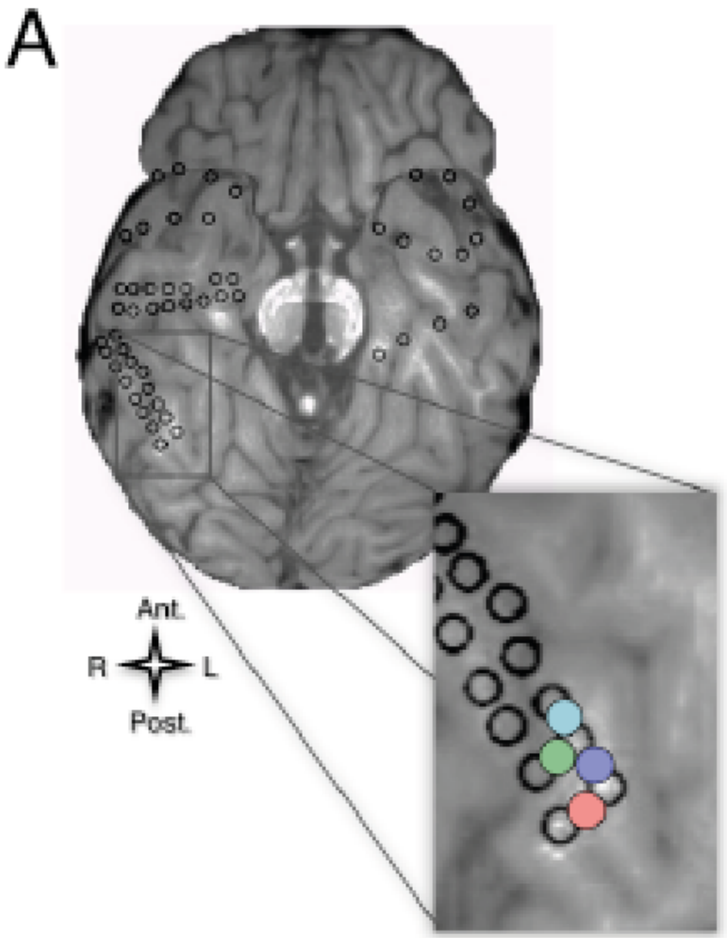
$$\phi^* = I(X^{t-\tau}; X^t) - I^*(X^{t-\tau}; X^t)$$

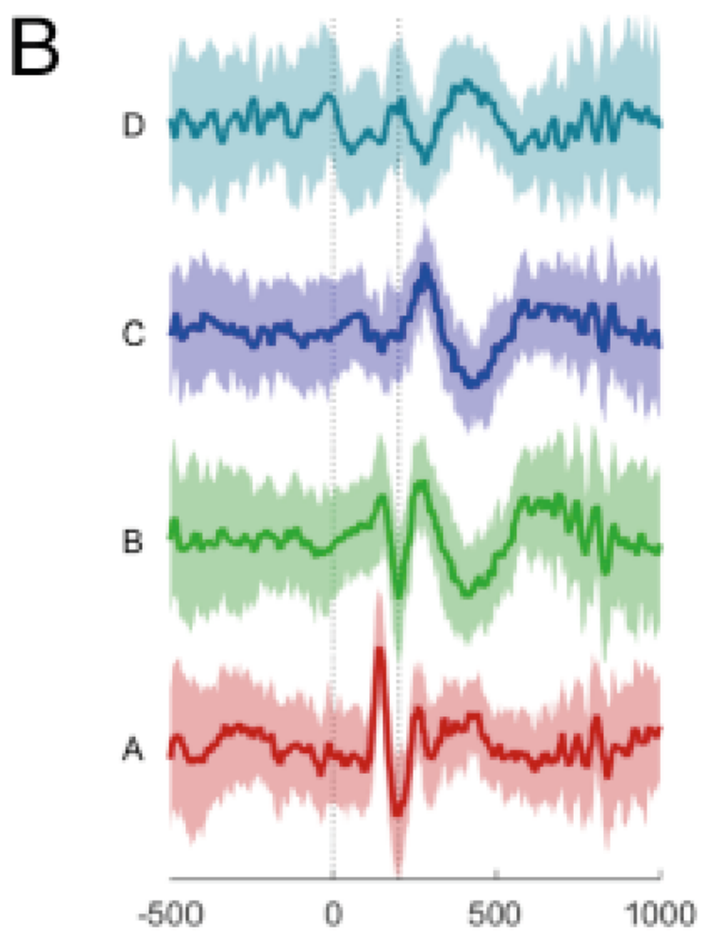


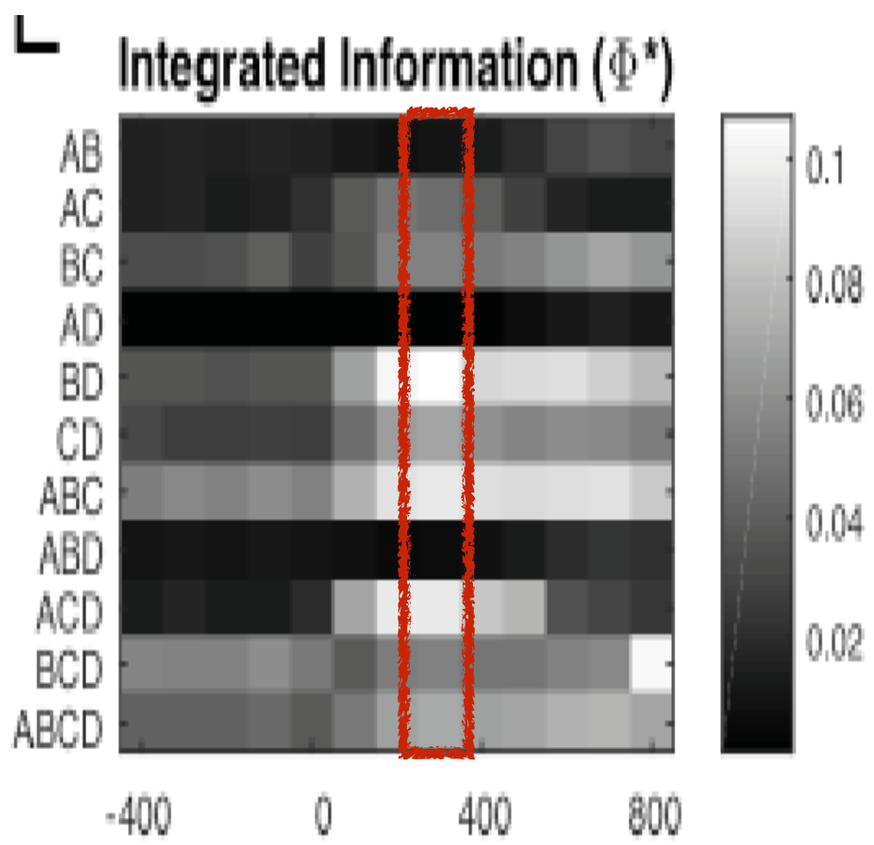


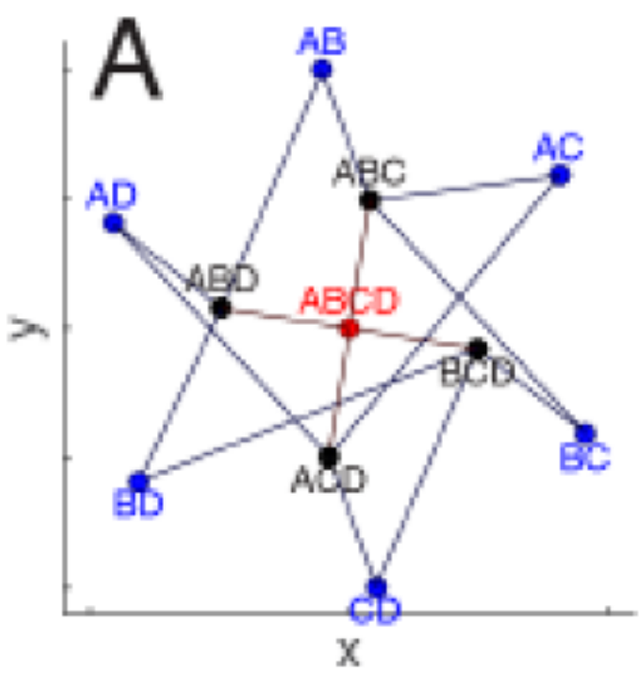


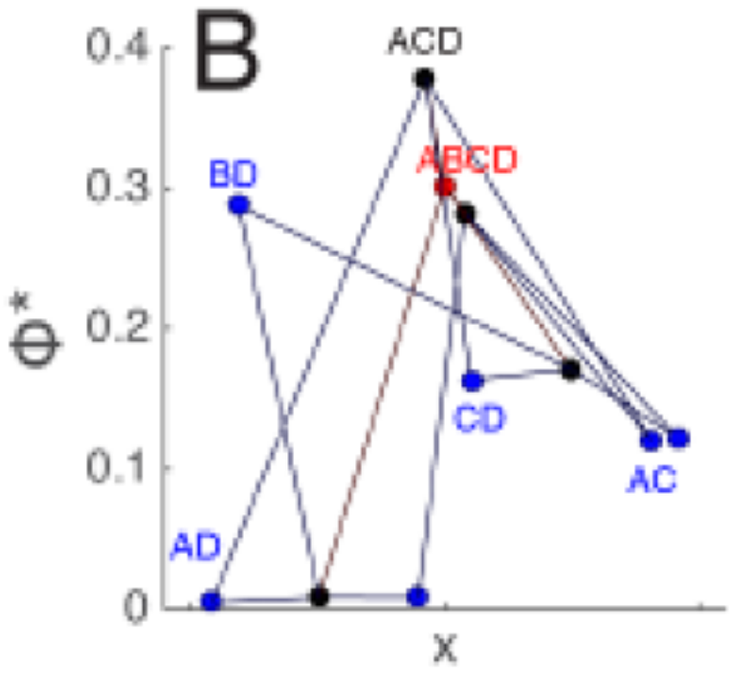
time from stimulus onset (bin center)

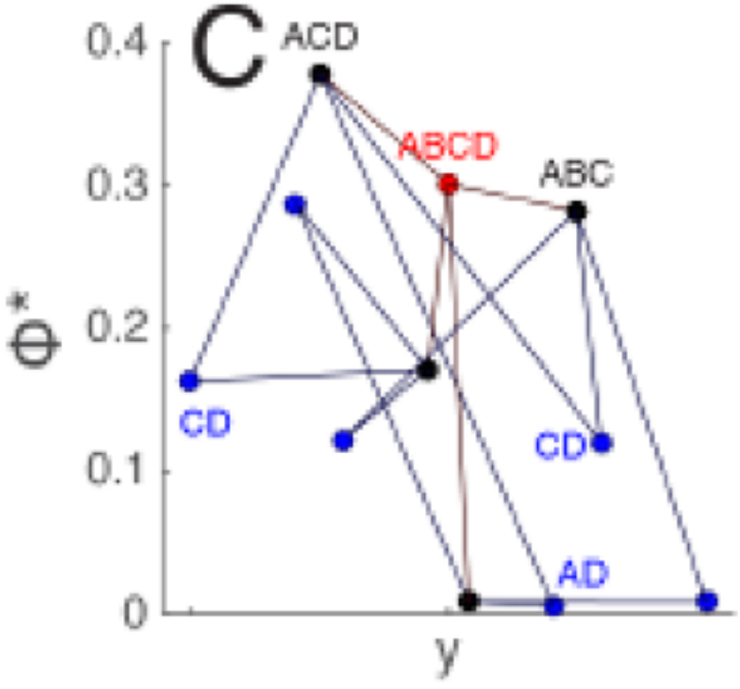

A
Ant.
R
L
Post.

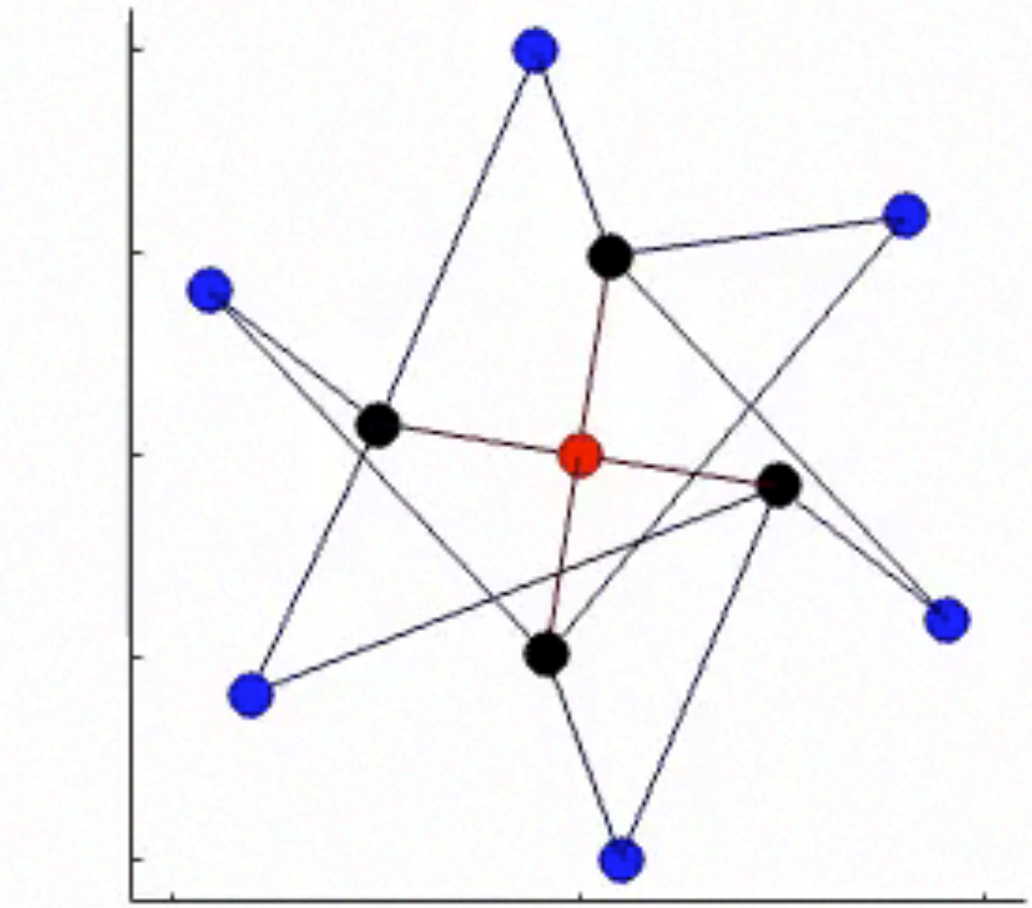

B
D
C
B
A
-500
0
500
1000
Time from stimulus onset (msec)

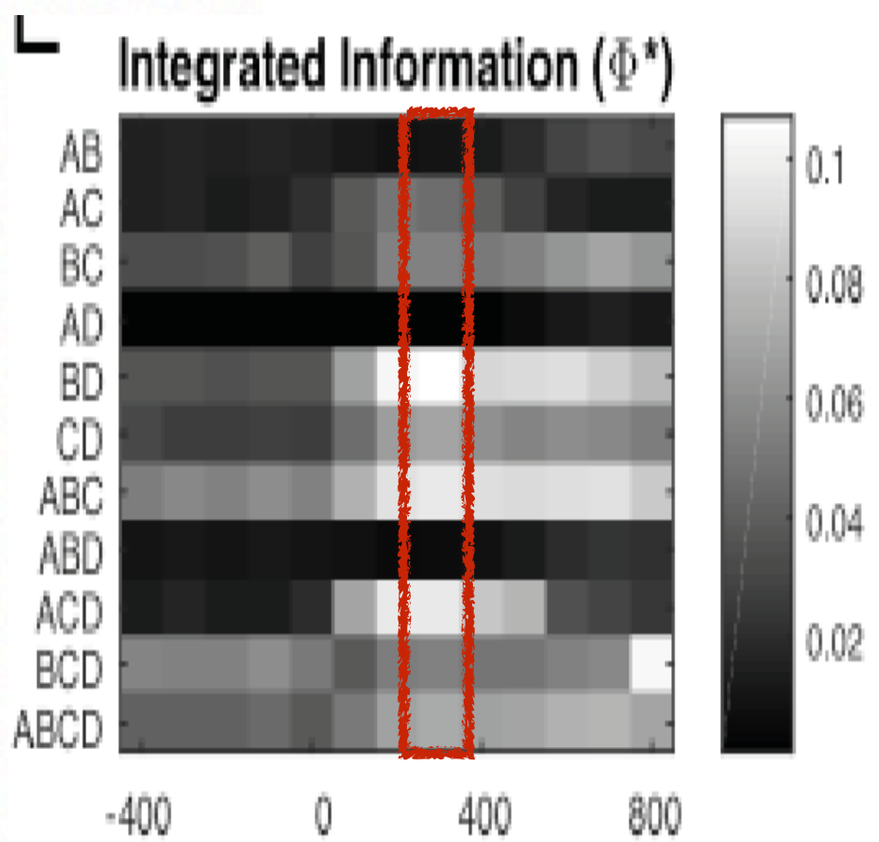

Integrated Information (Φ*)
AB
AC
BC
AD
BD
CD
ABC
ABD
ACD
BCD
ABCD
-400
0
400
800
0.1
0.08
0.06
0.04
0.02

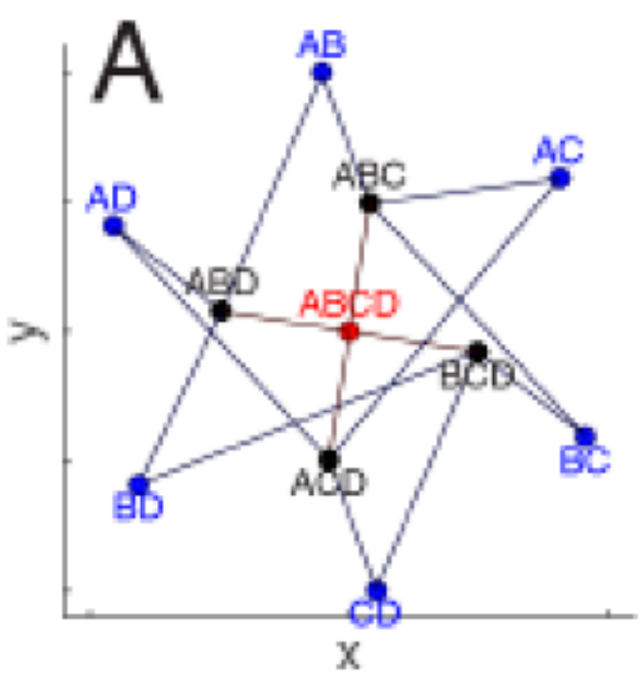

A
AB
AC
AD
ABC
ABD
ABCD
BCD
BC
BD
ACD
CD
y
x

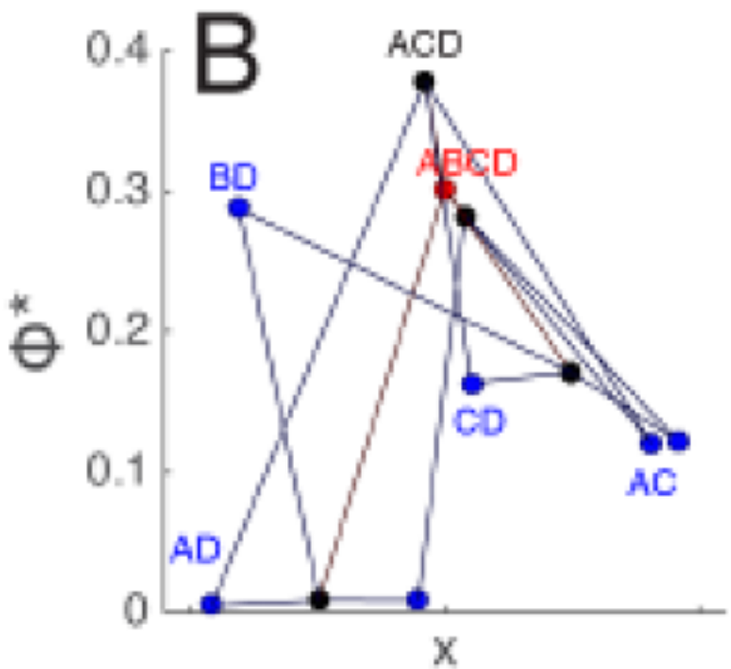

B
0.4
0.3
0.2
0.1
0
ACD
ABCD
BD
CD
AC
AD
x

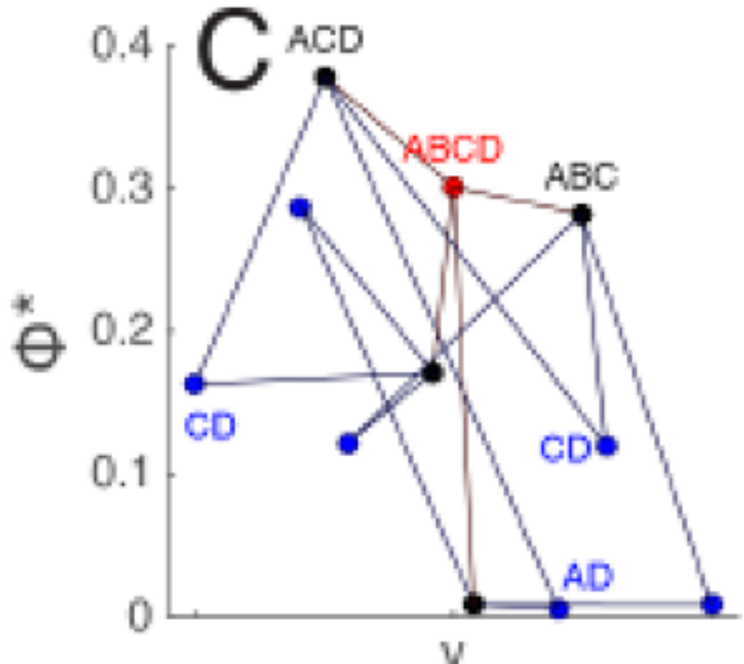

C
0.4
0.3
0.2
0.1
0
ACD
ABCD
ABC
CD
CD
AD
y


CFS trials (Face Vis:4, Contrast:3)

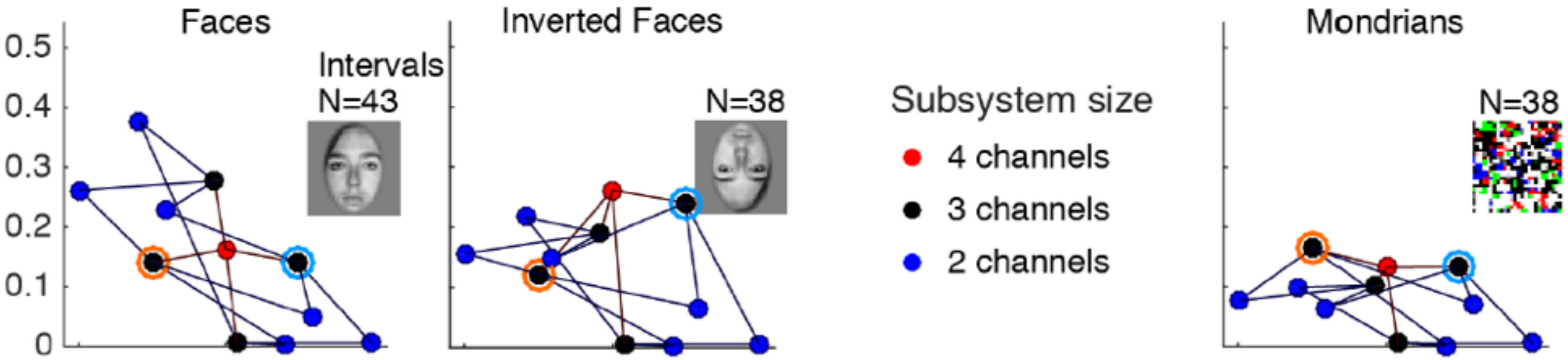

Integrated Information (Φ*)
AB
AC
BC
AD
BD
CD
ABC
ABD
ACD
BCD
ABCD
-400
0
400
800
0.1
0.08
0.06
0.04
0.02


A
AB
AC
AD
ABC
ABD
ABCD
BCD
BC
BD
ACD
CD
x
y


B
ACD
BD
ABCD
CD
AC
AD
0.4
0.3
0.2
0.1
0
x

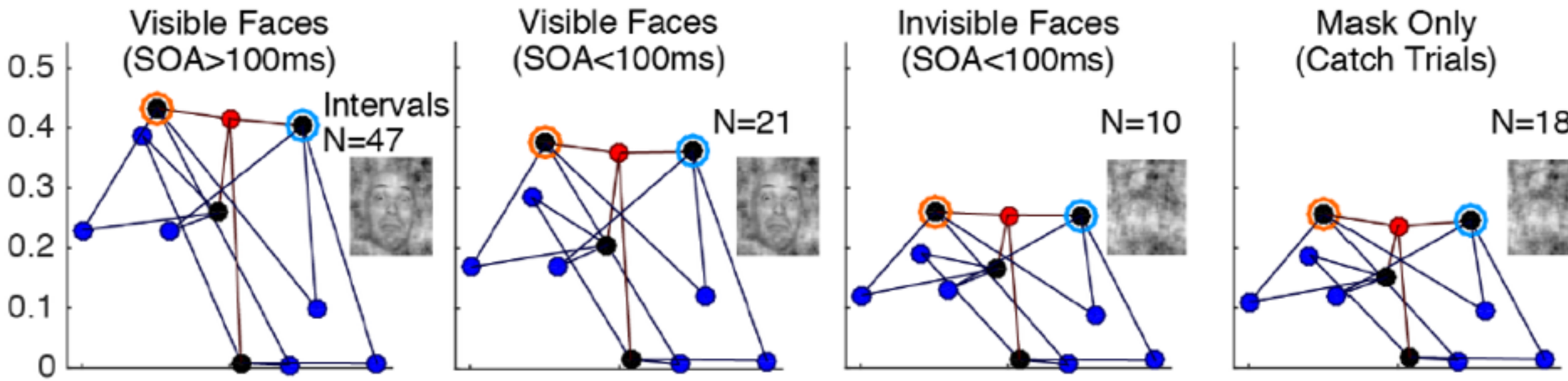

C
ACD
ABCD
ABC
CD
CD
AD
0.4
0.3
0.2
0.1
0
y

Unmasked conditions - 3 of 5 categories shown



Haun et al 2016 bioRxiv

Backward masking - 17 ms face in one of 4 quadrants

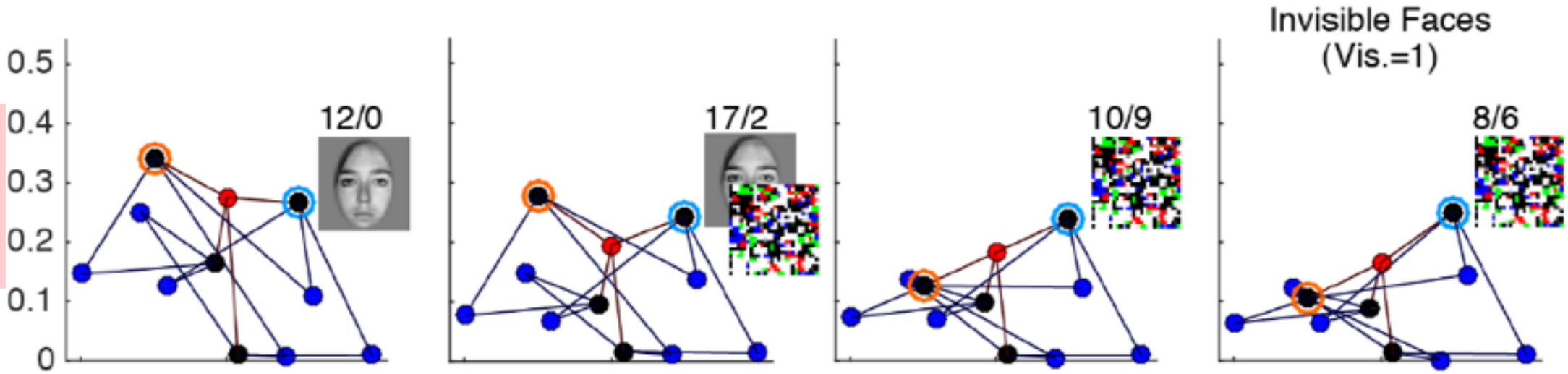


Haun et al 2016 bioRxiv

Visibility=4 Visibility=3 Visibility=2 Visibility=1

Continuous Flash Suppression

Middle contrast face

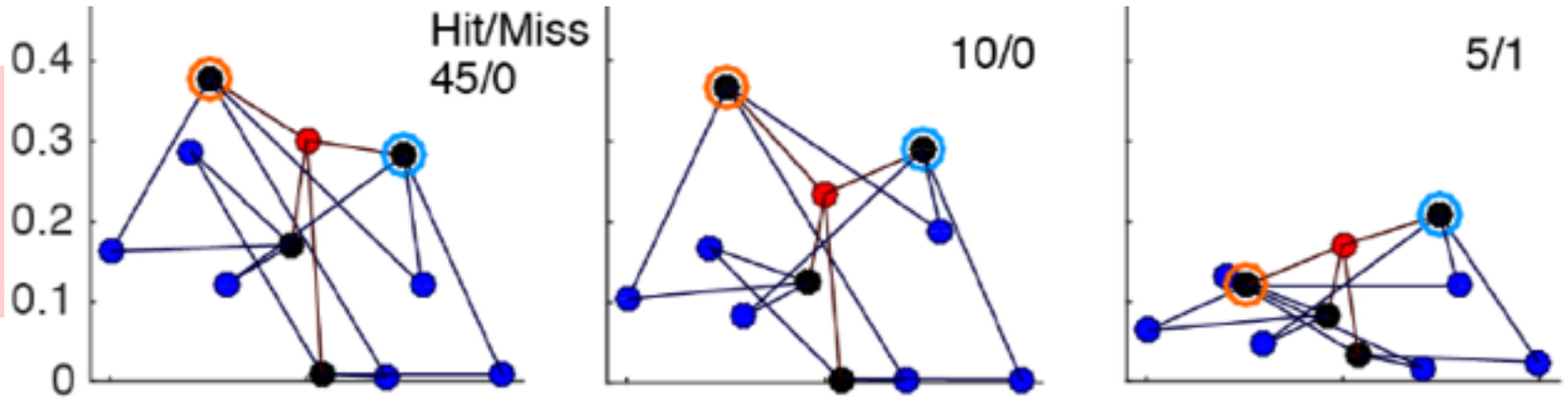


Haun et al 2016 bioRxiv

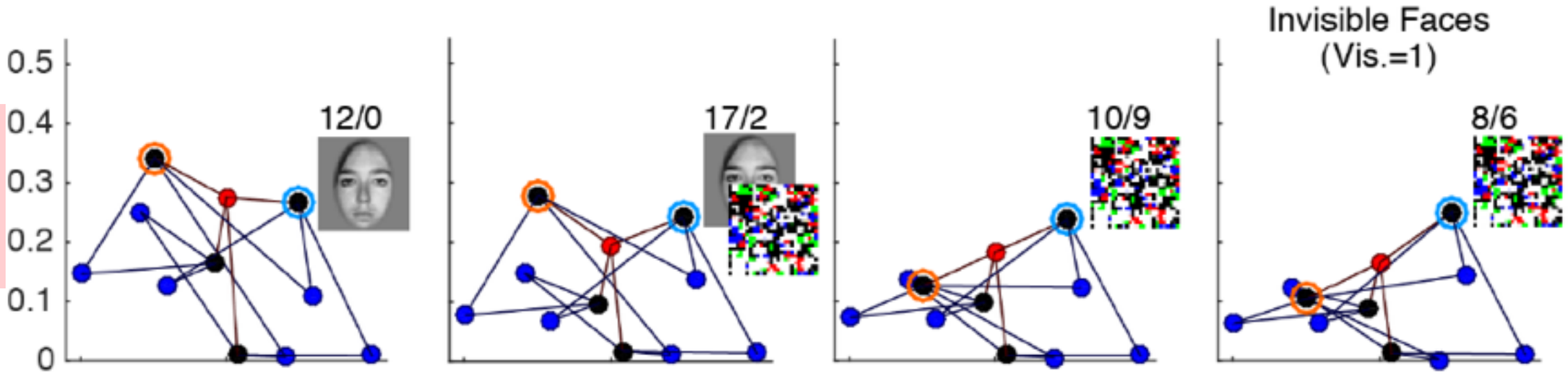
Visibility=4 Visibility=3 Visibility=2 Visibility=1

Continuous Flash Suppression

High contrast face



Middle contrast face

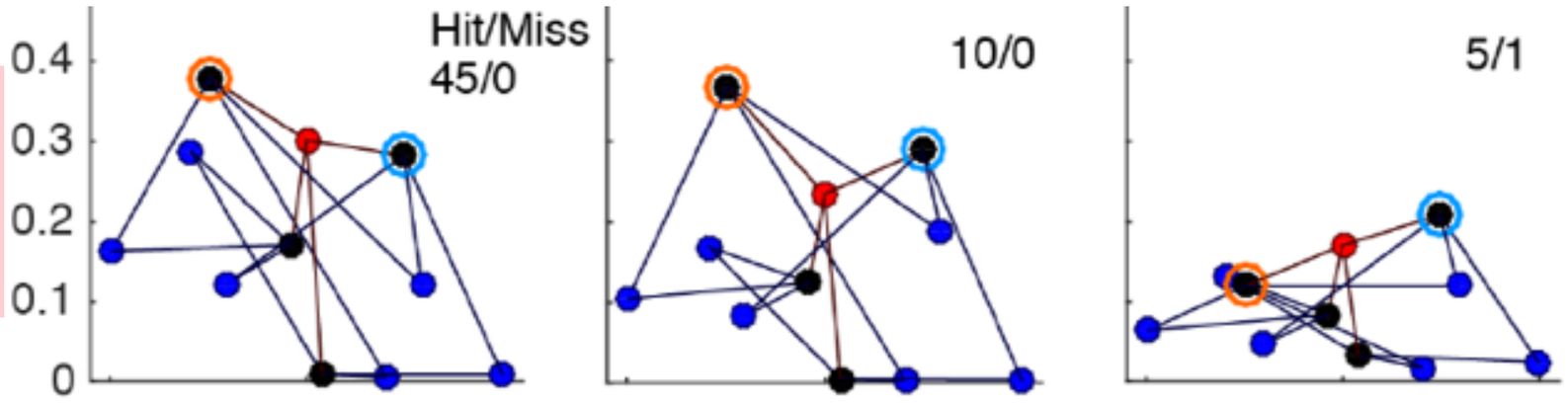


Haun et al 2016 bioRxiv

Visibility=4 Visibility=3 Visibility=2 Visibility=1

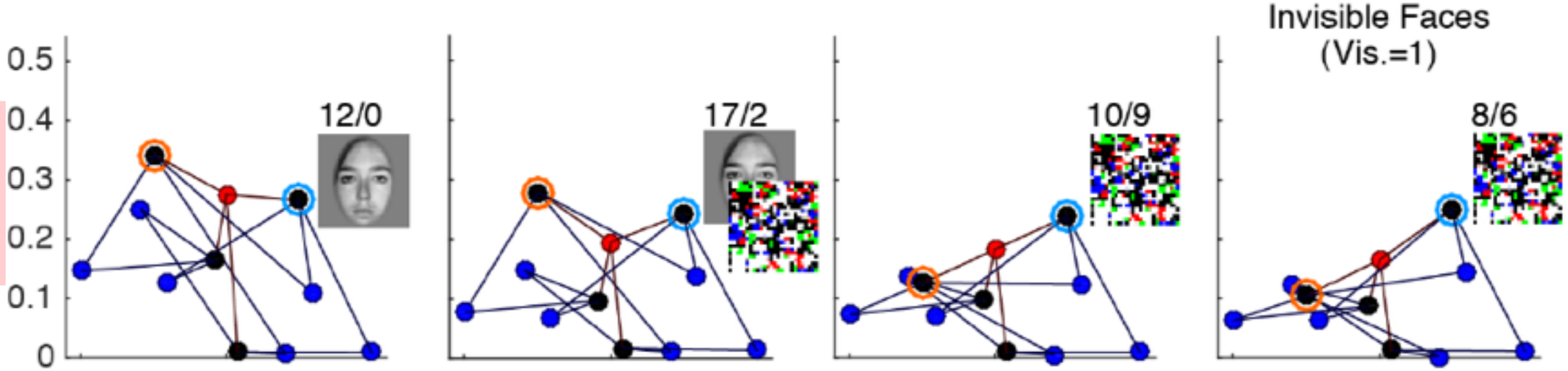
Continuous Flash Suppression

High contrast face

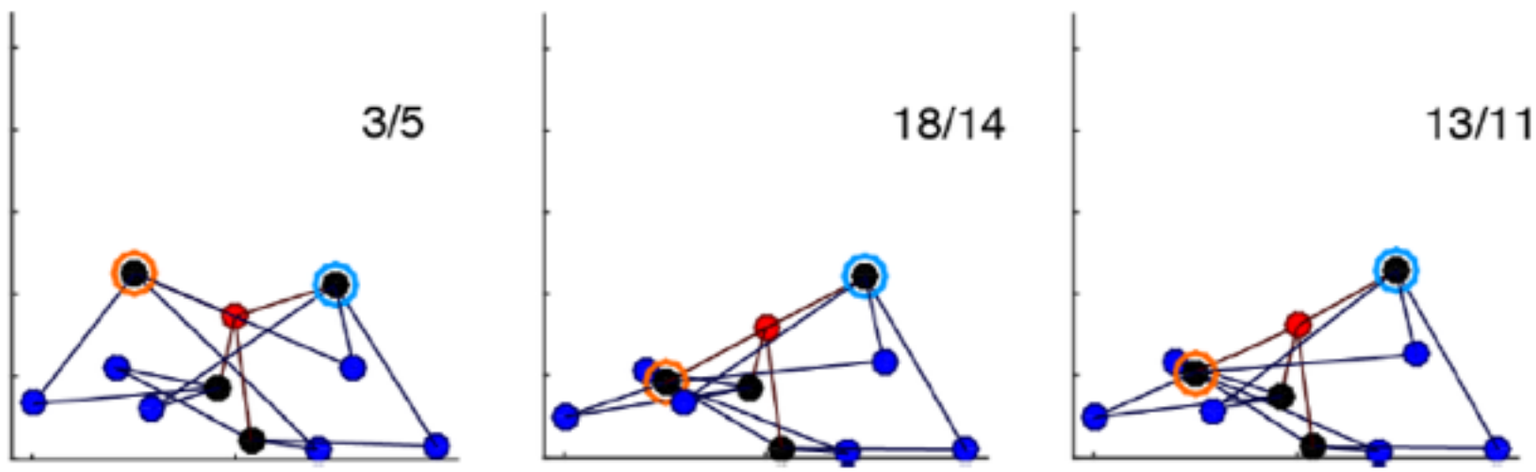


Invisible Faces (Vis.=1)

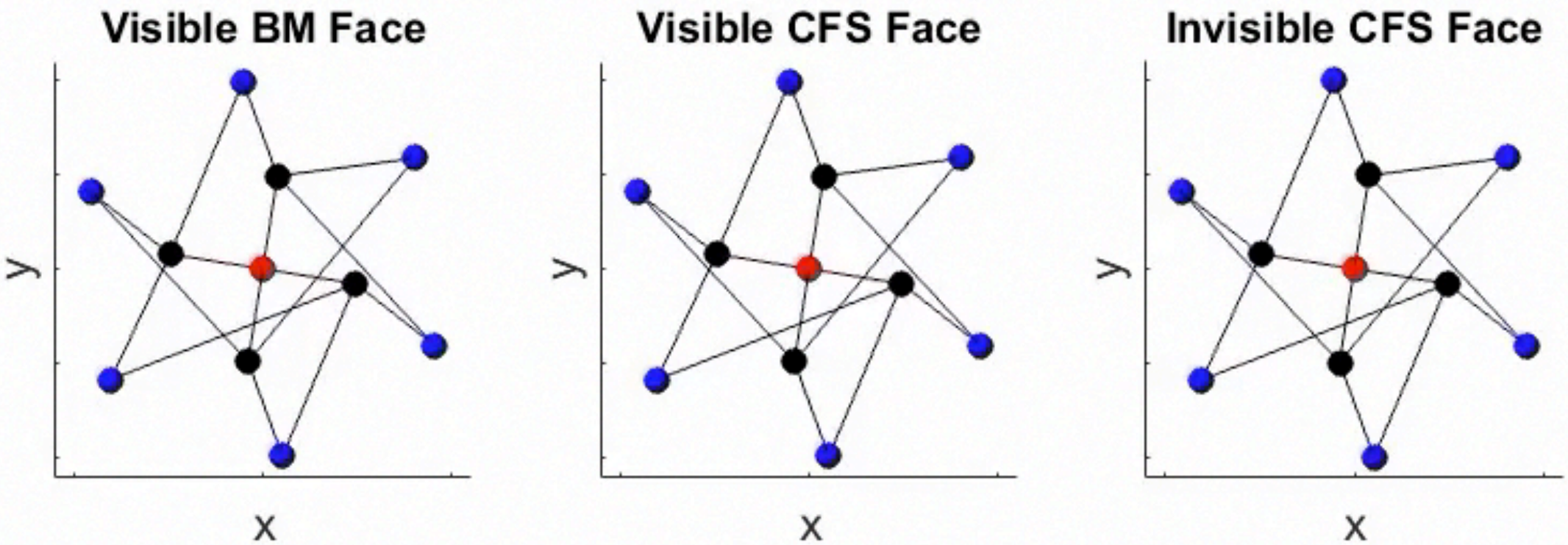
Middle contrast face



Low contrast face



Exemplar isomorphism across different tasks



Haun et al 2016 bioRxiv

Summary of this talk

- Overview of IIT
- Integrated information as:
  - loss of decoding accuracy (Φ^*)
  - distance between full and disconnected models (Φ_G)
- Empirical testing of IIT
  - Phi patterns: hierarchy of causal interactions better correlates with conscious perception than entropy or mutual information.

A path to bats' experience…

- Can be compared between sensory modalities
  - Across individuals
  - Across species
  - Common topological properties of phi* pattern for vision vs audition?
- What is it like to be a bat?
- Dissolution of the Hard Problem?