Title: Being vs. Happening: information from the intrinsic perspective of the system itself  
Date: Apr 12, 2018 03:30 PM  
URL: http://pirsa.org/18040120  

Abstract: When applied to a physical system, the two main, established notions of information, Shannon Information and Algorithmic Information, explicitly neglect the mechanistic structure of the system under evaluation. Shannon information treats the system as a channel and quantifies correlations between the system’s inputs and outputs, or between its past and future states. Algorithmic information quantifies the length of the shortest program capable of reproducing the system’s outputs or dynamics. The goal in both cases is to predict the system’s behavior from the perspective of an extrinsic investigator. From the intrinsic perspective of the system itself, however, information must be physically instantiated to be causally relevant. For every bit, there must be some mechanism that is in one of two (or several) possible states, and which state it is in must matter to other mechanisms. In other words, the state must be a difference that makes a difference and implementation matters. By examining the informational and causal properties of artificial organisms (animats) controlled by small, adaptive neural networks (Markov Brains), I will discuss necessary requirements for intrinsic information, autonomy, and agency.
Being vs. Happening:
Information from the intrinsic perspective of the system itself

04/12/18 – Larissa Albantakis
Motivation: Consciousness
To be conscious is to have an experience

Integrated information theory (IIT) - Giulio Tononi
Consciousness is what goes away when one falls into dreamless sleep

Seurat: Le dormeur
Consciousness is not just self-reflection

Rodin: Le penseur
Consciousness ≠ Responsiveness

Imagine playing tennis in the ‘vegetative state’

Intelligence vs. Consciousness

CONSCIOUSNESS

INTELLIGENCE

0
Consciousness is not just of the environment

Magritte: Le dormeur teméraire
Perception: “Information processing”? 

Extrinsic Information  
(deviations from independence) 

Intrinsic Information  
(differences that make a difference) 

Experience is intrinsic

Extrinsic Information
(deviations from independence)

Intrinsic Information
(differences that make a difference)

Neural correlates of consciousness (NCC)

The minimal neuronal mechanisms jointly sufficient for any one specific conscious percept.

Crick and Koch, Nat Neurosci 2003
Level and content of consciousness

Level of consciousness:
- irrespective of content
- Wake vs dreamless sleep, anesthesia, or coma (brain damaged patients)

Content of consciousness:
- experience is structured, it has many aspects
- Subjects report perceiving stimulus or not (Masking, binocular rivalry, inattentional blindness)
The neural substrate of consciousness?

Cerebral Cortex: 16 billion neurons, 100 trillion synapses

Basal Ganglia: 0.4 billion neurons

Cerebellum: 70 billion neurons

Spinal cord: 1 billion neurons

Macro, Micro
From BCC to NCC, and back ... to the hard problem

Kouider and Dehaene, 2007

500 ms fixation
500 ms mask
50 ms prime
33 ms mask
700 ms target

I don’t see it
I see it

Kouider and Dehaene

invisible

visible

P3b: 576 ms

Dehaene and Changeux, 2011

Visibility rating
P3 size

Chalmers 1995

?
One cannot start from matter and “squeeze” consciousness out of it.
But one can start from consciousness itself and ask what physical system could account for its properties.
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 *explanatory*, *predictive* and *inferential* power
- Leads to *measures* of both quantity and quality of consciousness
- Proposes a *formal framework* to characterize the intrinsic cause-effect structure of discrete dynamical systems

Oizumi*, Albantakis* and Tononi, 2014; Tononi, Scholarpedia 2015; Tononi et al. 2016 in Nat Rev Neurosci
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 *explanatory*, *predictive* and *inferential* power
- Leads to *measures* of both quantity and quality of consciousness

- Proposes a *formal framework* to characterize the intrinsic cause-effect structure of discrete dynamical systems

Oizumi*, Albantakis* and Tononi, 2014; Tononi, Scholarpedia 2015; Tononi et al. 2016 in Nat Rev Neurosci
Integrated Information Theory (IIT)

Phenomenal existence
Experience

Physical existence

Cause-effect power

- ON
- MANIPULATE
- OFF
- OBSERVE
Essential properties of phenomenal existence
(immediately and indubitably true of every experience)

- Intrinsicality
- Composition
- Information
- Integration
- Exclusion
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 **explanatory**, **predictive** and **inferential** power
- Leads to **measures** of both quantity and quality of consciousness
- Proposes a **formal framework** to characterize the intrinsic cause-effect structure of discrete dynamical systems

Oizumi*, Albantakis* and Tononi, 2014; Tononi, Scholarpedia 2015; Tononi et al. 2016 in Nat Rev Neurosci
A system of mechanisms in a state

Extrinsic Information
(deviations from independence)

Intrinsic Information
(differences that make a difference)

Information: Extrinsic perspective and intrinsic perspective

**Extrinsic:** signal transmission across a channel (I/O correlations) from the extrinsic perspective of an **observer**

**Intrinsic:** differences that make a difference (causes/effects) from the intrinsic perspective of a system

**Environment**

**Behavior**

Correlational measures, such as mutual information between I and O

set of concepts: Cause-effect structure (C)

- **cause repertoires**
  - $p(ABC|AB^c=10) = 0.33$
  - $p(AB|C^c=1) = 0.25$
  - $p(AC|B^c=0) = 0.25$
  - $p(BC|A^c=1) = 0.17$

- **effect repertoires**
  - $p(C|AB^c=10) = 0.33$
  - $p(AB|C^c=1) = 0.25$
  - $p(A|B^c=0) = 0.25$
  - $p(B|A^c=1) = 0.17$

within the system ABC state-dependent
**INTRINSICALITY**

From the intrinsic perspective of a system, only “differences that make a difference” within the system matter. Therefore, the system’s mechanisms must constrain the past and future of the system.
Formalism to assess intrinsic cause-effect structure

IIT principles:
- **Composition**
- **Information**
  (Differences that make a difference)
- **Integration**
  (Irreducible to parts)
- **Exclusion**
  (Only maxima of integrated information count)
Formalism to assess intrinsic cause-effect structure

IIT principles:
- Composition
- Information
  (Differences that make a difference)
- Integration
  (Irreducible to parts)
- Exclusion
  (Only maxima of integrated information count)

\[ D(p(AC^{t-1}|B^t = 1), p(AC^{t-1})) \]

*Interventional probabilities!!! + max entropy!
Formalism to assess intrinsic cause-effect structure

IIT principles:
• Composition
• Information
  (Differences that make a difference)
• Integration
  (Irreducible to parts)
• Exclusion
  (Only maxima of integrated information count)

\[ D(p(ABC^{t-1}|AB^t = 10), p(ABC^{t-1})) \]

*Interventional probabilities!!!
  + max entropy!
Formalism to assess intrinsic cause-effect structure

IIT principles:
- Composition
- Information
  (Differences that make a difference)
- Integration
  (Irreducible to parts)
- Exclusion
  (Only maxima of integrated information count)

\[
D(p(ABC^{t-1}|AB^t = 10), p(ABC^{t-1})) \\
D(p(ABC^{t+1}|AB^t = 10), p(ABC^{t+1}))
\]

*Interventional probabilities!!!*  
*+ max entropy!*
Formalism to assess intrinsic cause-effect structure

IIT principles:
• Composition
• Information
  (Differences that make a difference)
• Integration
  (Irreducible to parts)
• Exclusion
  (Only maxima of integrated information count)

integrated cause information ($\varphi$):

$$D\left(\begin{array}{c}
p(ABC^{t+1}|AB^t = 10), \\
p(BC^{t+1}|A^t = 1) \ast (p(A^{t+1}|B^t = 0)\end{array}\right)$$

Interventional probabilities!!!
+ max entropy!
Formalism to assess intrinsic cause-effect structure

IIT principles:
- Composition
- Information
  (Differences that make a difference)
- Integration
  (Irreducible to parts)
- Exclusion
  (Only maxima of integrated information count)
**Formalism to assess intrinsic cause-effect structure**

**IIT principles:**
- **Composition**
- **Information**
  (Differences that make a difference)
- **Integration**
  (Irreducible to parts)
- **Exclusion**
  (Only maxima of integrated information count)
Formalism to assess intrinsic cause-effect structure

IIT principles:

• Composition
• Information
  (Differences that make a difference)
• Integration
  (Irreducible to parts)
• Exclusion
  (Only maxima of integrated information count)

$\Phi^{\text{Max}}(ABC) = 0.92$

$\Phi(AB) = 0$
$\Phi(BC) = 0.5$
$\Phi(AC) = 0.5$
Formalism to assess intrinsic cause-effect structure

IIT principles:
- Composition
- Information (Differences that make a difference)
- Integration (Irreducible to parts)
- Exclusion (Only maxima of integrated information count)
Formalism to assess intrinsic cause-effect structure

IIT principles:
- Composition
- Information
  (Differences that make a difference)
- Integration
  (Irreducible to parts)
- Exclusion
  (Only maxima of integrated information count)

$\Phi^{Max}(ABC) = 0.92$

Cause-effect structure (C)

<table>
<thead>
<tr>
<th>Repertoire</th>
<th>Repertoire</th>
<th>$\phi^{Max}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>$p(ABC\mid AB^C=10)$</td>
<td>0.33</td>
</tr>
<tr>
<td>C</td>
<td>$p(ABC\mid C^C=1)$</td>
<td>0.25</td>
</tr>
<tr>
<td>B</td>
<td>$p(ABC\mid B^C=0)$</td>
<td>0.25</td>
</tr>
<tr>
<td>A</td>
<td>$p(ABC\mid A^C=1)$</td>
<td>0.17</td>
</tr>
</tbody>
</table>
An experience is a global maximum ($\Phi$) of cause-effect power
(a cause-effect structure)

Quantity:
irreducibility $\Phi$
of the cause-effect structure

Quality:
‘form’
of the cause-effect structure

Experience = Cause-effect structure

Physical substrate

Tononi et al
Nature Rev Neurosci 2016
From phenomenology to mechanisms, and back

Intrinsicality

PHENOMENOLOGY

Composition
Information
Window
Integration
Exclusion

Composition
Information
Intrinsicality
Integration
Exclusion

empirical
prediction

inference

cause states
effect states

composition of conceptual structure

Pirsa: 18040120

Page 41/61
**Animats** – adaptive, behaving systems

**Structure**
- Sensors
- Hidden elements
- Motors

**Environment**
- 36 units
- 16 units

**Adaptation**
- Generation #0
  - Fitness: 50%
- Generation #11264
  - Fitness: 78.9%
- Generation #59904
  - Fitness: 97.7%

**Selection & Mutation**
- **Fitness**: % of successfully caught and avoided blocks (out of 128 trials)
- **Point mutations, deletions, and duplications** in the genome after each selection
Integrated information $\Phi$ increases during evolution

Task:
catch: 3, 6
avoid: 4, 5

Animat

S1  S2
A  B
C  D
M1  M2

Generation 2048

Fitness

Phi

Number of Concepts
Integrated information $\Phi$ increases during evolution
Integrated information $\Phi$ increases during evolution

Task:
- catch: [3, 6]
- avoid: [4, 5]

[Diagram showing network and fitness plots]
Integrated information $\Phi$ increases during evolution
Evolved network structures
Simple task can be solved in a modular and integrated manner

Task 1

Environment

Catch 1

Avoid 3

Feed-forward + self-loop network

100% Fitness

$\Phi = 0$

Feed-back network

$\langle \Phi^{max} \rangle = 0.11$

Generation #59904

Generation #59904

Past states

Future states
Robustness and Autonomy

fission yeast cell cycle model

\[ \Phi = 0.431 \]

functionally identical ‘backbone’ network

\[ \Phi = 0.069 \]

maximum of \( \Phi \) in every state in the cycle

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

Task 1 (easy)
- Catch: 1
- Avoid: 3

Task 2 (difficult)
- Catch: 3
- Avoid: 4

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

---

Albantakis et al., 2014
Autonomous Agents with context dependent behavior

Task 2

- catch: 3, 6
- avoid: 4, 5

Main complex:
- $\langle \phi^{Max} \rangle = 1.13$
- $\langle \#\text{concepts} \rangle = 4.4$

Generation #59904
Fitness: 97.7%

Animat is partially autonomous due to memory.

Modular/ Feed-forward system

$\Phi = 0$

Behavior entirely driven by the environment.
Why did consciousness evolve?

In adaptive systems, evolution to an environment with a rich causal structure provides a link between the system’s behavior and its intrinsic cause-effect structures, leading to behaviorally fit systems with high $\Phi$.

Axioms (phenomenal existence)

Postulates (physical existence) $\rightarrow$ IIT formalism

Why did consciousness evolve?

Why did integrated information evolve?
Intelligence vs. Consciousness
Functional equivalence does not imply phenomenal equivalence

$\Phi = 2.31$

Findlay et al., in preparation
Computers simulating our behavior (BCC) are not conscious

 Computers simulating our neural organization (NCC) are not conscious
Question

At what spatiotemporal grain should we study the brain if our goal is to understand conscious experience?
**Macro elements: Coarse-grains of micro elements**

Proof-of-concept that $\Phi$ can peak at a macro level in small networks of logic-gates

---


Macro elements: Black boxes of micro elements

Proof-of-concept that $\Phi$ can peak at a macro level in small networks of logic-gates

$\Phi = 0.453$

- 55 micro elements
- 55 first order concepts
- 0 high order concepts

$\Phi = 0.080$

- 20 macro elements
- 20 first order concepts
- 0 high order concepts

Macro elements: Black boxes of micro elements

Proof-of-concept that $\Phi$ can peak at a macro level in small networks of logic-gates

$\Phi = 0.453$

- 55 micro elements
- 55 first order concepts
- 0 high order concepts

$\Phi = 0.080$

- 20 macro elements
- 20 first order concepts
- 0 high order concepts

$\Phi = 2.33$

- 5 macro elements
- 5 first order concepts
- 25 high order concepts

Increasing spatiotemporal grain size

Remarks on intrinsic information

- Intrinsic information is not about prediction. It’s about (causal) structure. “Being vs. happening”.
- Structural equivalence between phenomenology and cause-effect structure is testable. Not everything that can be decoded is part of phenomenology.
- Redundancy is meaningful.
- Implementation matters.

? 

- Is IIT’s causal analysis compatible with physics or not? QM?
- Is there convergence or real noise in the physical world? (E.g, Do real AND gates exist?)

Integrated Information Theory

- physical information
- causal composition and higher order interactions
- complexity
- identifying causal/informational boundaries
- informational/causal measures of autonomy
- causal exclusion and emergence
- practical approximations of integrated information
- applications

Submission Deadline: **30 November 2018**

**Guest Editors**

**Dr. Larissa Albantakis**
University of Wisconsin–Madison

**Prof. Giulio Tononi**
University of Wisconsin–Madison

**Author Benefits**

- High visibility
- Rapid publication
- Unlimited and free access for readers
- No Copyright Constraints
- Thorough Peer-Review
- No Space Constraints, No Extra Space or Color Charges