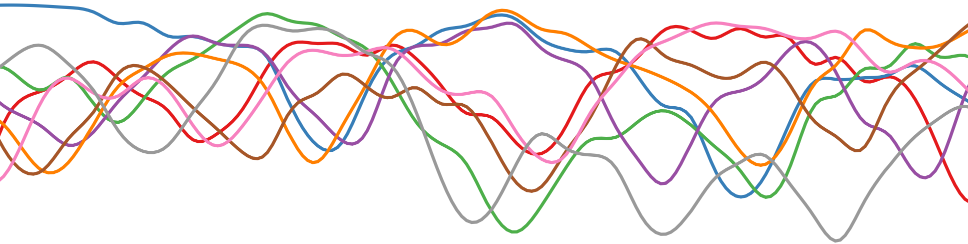
Multi-target Information Decomposition

And applications to Integrated Information Theory



Pedro A.M. Mediano

plogp@pm.me

University of Cambridge

Imperial College London

ACKNOWLEDGEMENTS

Fernando Rosas

Daniel Bor

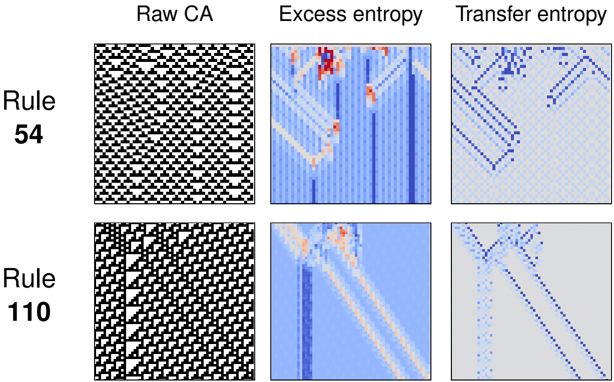
Adam Barrett

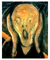
Thanks to:

- Andrea Luppi
- Anil Seth
- Robin Carhart-Harris

INFORMATION DYNAMICS

- **Goal**: Study how information is *stored, transferred, and modified* in a multivariate complex system.





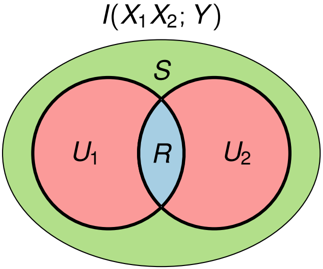
Storage and transfer are kinda sorted – *what about modification*?

PARTIAL INFORMATION DECOMPOSITION

Enter **Partial Information Decomposition** (PID).

In PID, mutual information is split into:

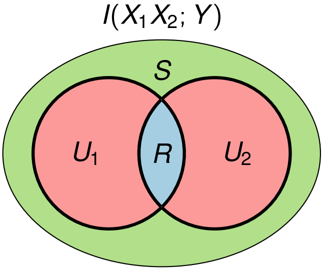
- Redundancy
- Unique information
- **Synergy**

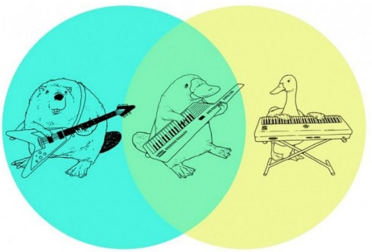


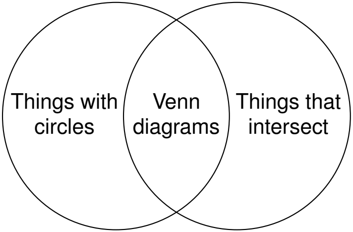
Synergy can help us understand information modification.

(Williams & Beer, 2010)

INTERLUDE: VENN DIAGRAMS

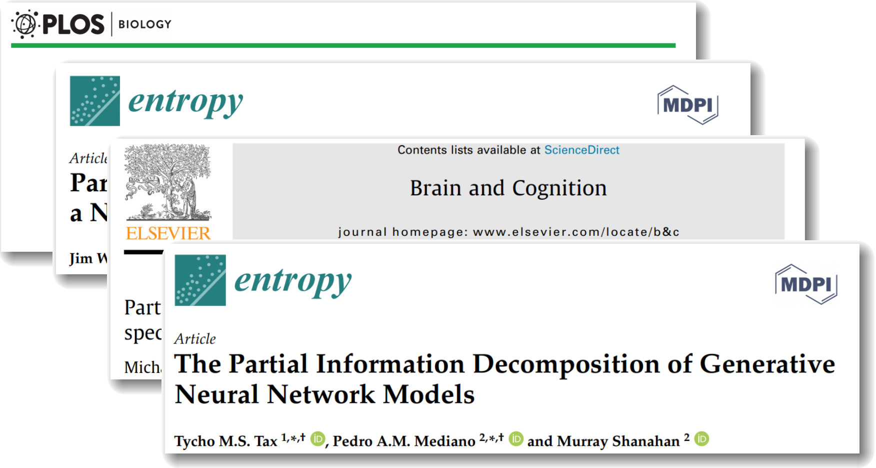






PARTIAL INFORMATION DECOMPOSITION

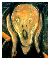
SUCCESS STORIES



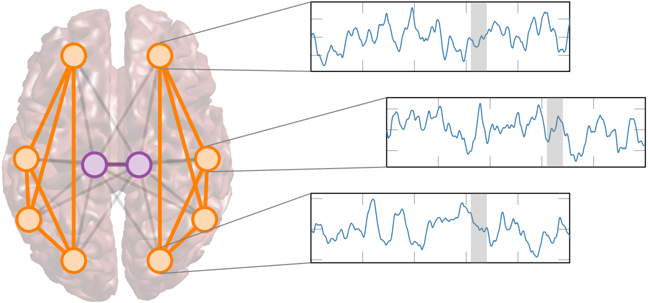
WHEN PID IS NOT ENOUGH

MULTIVARIATE COMPLEX SYSTEMS

- We need to designate one variable as *target* and others as *sources*.



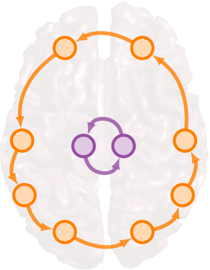
In multivariate systems, there is no natural source/target division.



WHEN PID IS NOT ENOUGH

CONSCIOUSNESS NEUROSCIENCE

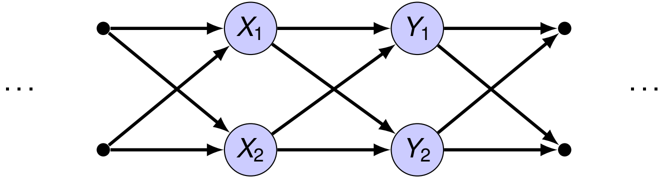
- **Consciousness** is an endogenous process, especially during REM sleep and dreams.

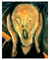


Goal: describe information flows from the system's intrinsic perspective.

INFORMATION AND DYNAMICAL SYSTEMS

We care about multivariate systems evolving *jointly* over time.





Problem: PID cannot deal with multiple targets!

INFORMATION DECOMPOSITION

Can we extend PID to multivariate time series?

Yes! With *Integrated Infomation Decomposition*, ΦID.

$$I(\boldsymbol{X};\boldsymbol{Y}) = \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in\mathcal{A}} I_\partial^{\boldsymbol{\alpha}\rightarrow\boldsymbol{\beta}}$$

Beyond integrated information: A taxonomy of information dynamics phenomena

Pedro A.M. Mediano,[1] Fernando Rosas,[2,3,4] Robin L. Carhart-Harris,[2] Anil K. Seth,[5] and Adam B. Barrett[5,6]

[1] *Department of Computing, Imperial College London, London SW7 2RH*[*]
[2] *Center for Psychedelic Research, Department of Medicine, Imperial College London, London SW7 2DD*
[3] *Data Science Institute, Imperial College London, London SW7 2AZ*
[4] *Center for Complexity Science and Department of Mathematics, Imperial College London, London SW7 2AZ*[*]
[5] *Sackler Center for Consciousness Science, Department of Informatics, University of Sussex, Brighton BN1 9RH*
[6] *The Data Intensive Science Centre, Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK*

(Dated: September 6, 2019)

Take-home message

- ✓ New extension of PID to multi-target setting, applicable to dynamical systems.
- ✓ Solves known problems within Integrated Information Theory (IIT).
- ? New opportunities and challenges for PID research.

THIS TALK

PART I

Multi-target PID and integrated information decomposition (ΦID)

PART II

ΦID and Integrated Information Theory (IIT-2.0)

PART III

A formal theory of causal emergence based on ΦID

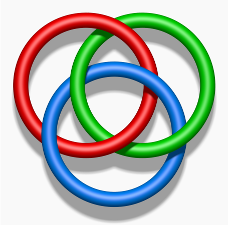
Main track
O14

INFORMATION DECOMPOSITION

Back to basics

- Two neurons can either be correlated or not, and this is measured by *mutual information*.

- However, with 3 or more neurons there are **qualitatively different modes of interaction.**



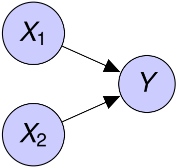
Synergy

Redundancy

PARTIAL INFORMATION DECOMPOSITION

Two predictors X_1, X_2 and one target Y.

- Joint predictability: $I(X_1 X_2; Y)$
- Marginal predictability: $I(X_1; Y),\ I(X_2; Y)$



However, sometimes:

$$\underbrace{I(X_1 X_2; Y)}_{\text{“the whole“}} > \underbrace{I(X_1; Y) + I(X_2; Y)}_{\text{“the parts“}}$$

The *Partial Information Decomposition* (PID) postulate:

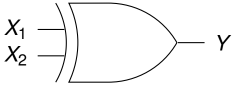
$$I(X_1 X_2; Y) = \underbrace{I_\partial^{\{1\}\{2\}}}_{\text{redundancy}} + \underbrace{I_\partial^{\{1\}} + I_\partial^{\{2\}}}_{\text{unique info}} + \underbrace{I_\partial^{\{12\}}}_{\text{synergy}}$$

(Williams & Beer, 2010)

EXAMPLE: XOR LOGIC GATE

Perfect example of synergy: XOR.

X_1	X_2	Y
0	0	0
0	1	1
1	0	1
1	1	0



- Knowing one input tells you nothing:

$$I(X_1; Y) = I(X_2; Y) = 0$$

- Knowing both inputs tells you everything:

$$I(X_1 X_2; Y) = 1$$

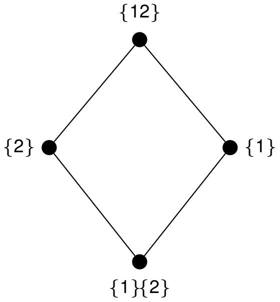
PARTIAL INFORMATION DECOMPOSITION

- Start with *intersection information*, $I_\cap^{\boldsymbol{\alpha}}$.
- Define a set of nodes $\mathcal{A}$ with a partial ordering $\preceq$, such that

$$I_\cap^{\boldsymbol{\alpha}} \leq I_\cap^{\boldsymbol{\beta}} \quad \text{iff} \quad \boldsymbol{\alpha} \preceq \boldsymbol{\beta}$$

- Each $I_\cap^{\boldsymbol{\alpha}}$ can be decomposed into *partial information atoms*:

$$I_\cap^{\boldsymbol{\alpha}} = \sum_{\boldsymbol{\beta} \preceq \boldsymbol{\alpha}} I_\partial^{\boldsymbol{\beta}}$$



- This yields an *underdetermined* system of equations:

$$I(X_1; Y) = I_\partial^{\{1\}\{2\}} + I_\partial^{\{1\}}$$
$$I(X_2; Y) = I_\partial^{\{1\}\{2\}} + I_\partial^{\{2\}}$$
$$I(X_1 X_2; Y) = I_\partial^{\{1\}\{2\}} + I_\partial^{\{1\}} + I_\partial^{\{2\}} + I_\partial^{\{12\}}$$

INTEGRATED INFORMATION DECOMPOSITION

Key ingredients for an information decomposition:

	Lattice	Intersection function
PID	$\mathcal{A}$	$I_\cap^{\boldsymbol{\alpha}}$
ΦID	$\mathcal{A}^2$	$I_\cap^{\boldsymbol{\alpha}\to\boldsymbol{\beta}}$

Product lattice

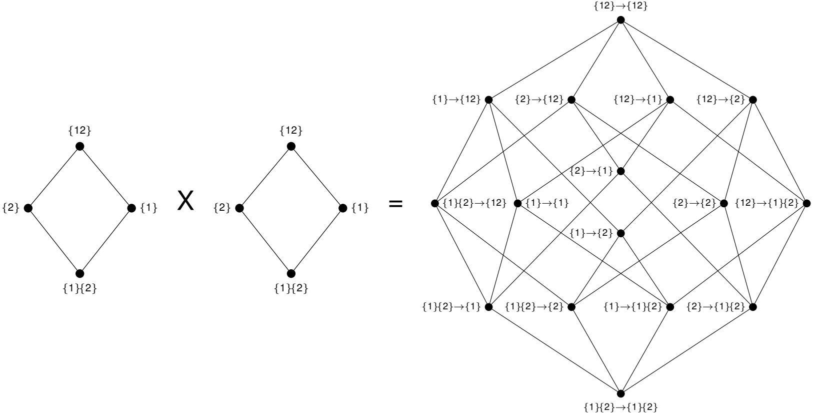
Double-redundancy function

INTEGRATED INFORMATION DECOMPOSITION

PRODUCT LATTICE

Nodes of the **product lattice** are denoted as $\boldsymbol{\alpha} \to \boldsymbol{\beta}$ for $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathcal{A}$, and

$$\boldsymbol{\alpha} \to \boldsymbol{\beta} \preceq \boldsymbol{\alpha}' \to \boldsymbol{\beta}' \quad \text{iff} \quad \boldsymbol{\alpha} \preceq \boldsymbol{\alpha}' \text{ and } \boldsymbol{\beta} \preceq \boldsymbol{\beta}'.$$



INTEGRATED INFORMATION DECOMPOSITION

DOUBLE-REDUNDANCY

- **Compatibility axiom**: in the following cases a double-redundancy can be reduced to a PID redundancy or the mutual information:

$$I_\cap^{\boldsymbol{\alpha}\to\boldsymbol{\beta}} = \begin{cases} \texttt{Red}(\boldsymbol{X}^{\alpha_1}, \dots, \boldsymbol{X}^{\alpha_J}; \boldsymbol{Y}^{\beta_1}) & \text{if} \quad K = 1, \\ \texttt{Red}(\boldsymbol{Y}^{\beta_1}, \dots, \boldsymbol{Y}^{\beta_K}; \boldsymbol{X}^{\alpha_1}) & \text{if} \quad J = 1, \\ I(\boldsymbol{X}^{\alpha_1}; \boldsymbol{Y}^{\beta_1}) & \text{if} \quad J = K = 1. \end{cases}$$

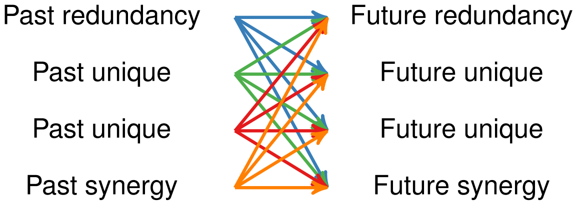
- **15-for-free lemma**: ΦID axioms provide unique values for the 16 atoms of the product lattice after one defines (i) a PID redundancy function $\texttt{Red}(\cdot)$, and (ii) an expression for $I_\partial^{\{1\}\{2\}\to\{1\}\{2\}}$.

Integrated information decomposition

- In PID there are 4 terms: redundancy, unique (2x), and synergy:

$$\underbrace{I_\partial^{\{1\}\{2\}}}_{\text{redundancy}}, \underbrace{I_\partial^{\{1\}}, I_\partial^{\{2\}}}_{\text{unique info}}, \underbrace{I_\partial^{\{12\}}}_{\text{synergy}}$$

- In ΦID, we have all combinations of past and future PID:



- In total, $4 \times 4 = 16$ terms.

INTEGRATED INFORMATION DECOMPOSITION

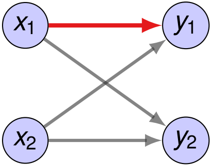
Examples:

- $I_{\partial}^{\{1\}\{2\}\to\{1\}\{2\}}$: redundant stored information
- $I_{\partial}^{\{1\}\to\{2\}}$: unique transferred information
- $I_{\partial}^{\{1\}\to\{1\}\{2\}}$: “duplicated” information
- ...

Integrated information decomposition

Examples

Active information storage:



$$\text{AIS} = I(X_1; Y_1)$$

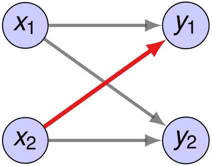
$$\overset{\text{PID}}{=} I_\partial^{\{1\}\{2\}}(\boldsymbol{X}; Y_1) + I_\partial^{\{1\}}(\boldsymbol{X}; Y_1)$$

$$\overset{\Phi\text{ID}}{=} I_\partial^{\{1\}\{2\}\to\{1\}\{2\}} + I_\partial^{\{1\}\{2\}\to\{1\}} + I_\partial^{\{1\}\to\{1\}\{2\}} + I_\partial^{\{1\}\to\{1\}}$$

INTEGRATED INFORMATION DECOMPOSITION

EXAMPLES

Transfer entropy:



$$\text{TE} = I(X_2; Y_1 | X_1)$$

$$\overset{\text{PID}}{=} I_\partial^{\{12\}}(\boldsymbol{X}; Y_1) + I_\partial^{\{2\}}(\boldsymbol{X}; Y_1)$$

$$\overset{\Phi\text{ID}}{=} I_\partial^{\{12\}\to\{1\}\{2\}} + I_\partial^{\{12\}\to\{1\}} + I_\partial^{\{2\}\to\{1\}\{2\}} + I_\partial^{\{2\}\to\{1\}}$$

INTEGRATED INFORMATION DECOMPOSITION

MEASURES


What does a double-redundancy function look like?


Start from a PID redundancy, then make a multi-target extension.

- For example, take Barrett's (2015) *Minimum Mutual Information* redundancy function. In PID:

$$I_{\partial}^{\{1\}\{2\}}(\boldsymbol{X}; Y) = \min_i I(X_i; Y)$$

- And in ΦID:

$$I_{\partial}^{\{1\}\{2\}\to\{1\}\{2\}}(\boldsymbol{X}; \boldsymbol{Y}) = \min_{i,j} I(X_i; Y_j)$$

INTEGRATED INFORMATION DECOMPOSITION

MEASURES

An operational information decomposition via synergistic disclosure

Fernando Rosas,[1,2,3,*] Pedro Mediano,[4,*] Borzoo Rassouli,[5] and Adam Barrett[6]

- Define *synergistic channels* $\mathcal{C}(\boldsymbol{X})$ as channels that disclose information about the whole, but not about the parts.
- Define synergy as the information extractable from a synergistic channel. In PID:

$$I_\partial^{\{12\}}(\boldsymbol{X};Y) = \sup_{\substack{p_{V|\boldsymbol{X}} \in \mathcal{C}(\boldsymbol{X}): \\ V - \boldsymbol{X} - Y}} I(V;Y) \,.$$

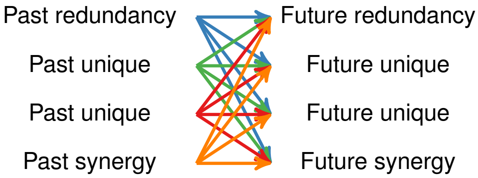
- And in ΦID, from two synergistic channels:

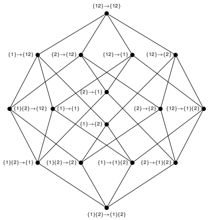
$$I_\partial^{\{12\}\to\{12\}}(\boldsymbol{X};\boldsymbol{Y}) = \sup_{\substack{p_{V|\boldsymbol{X}} \in \mathcal{C}(\boldsymbol{X}), \\ p_{U|\boldsymbol{Y}} \in \mathcal{C}(\boldsymbol{Y}): \\ V - \boldsymbol{X} - \boldsymbol{Y} - U}} I(V;U) \,.$$

INTERIM SUMMARY

INTEGRATED INFORMATION DECOMPOSITION

- ✓ ΦID: extension of PID to dynamical systems, applicable to multivariate time series analysis.
- ✓ Straightforward generalisation of available software and PID measures.

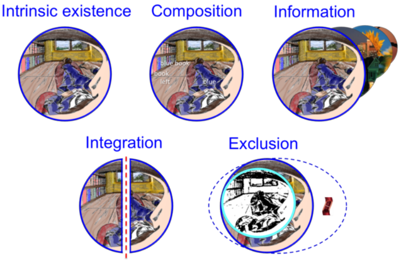




Integrated Information Theory

INTEGRATED INFORMATION THEORY

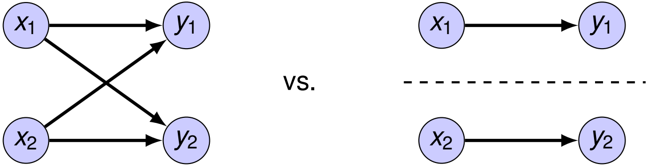
- Candidate theory of consciousness, very influential in the consciousness neuroscience community.
- Aims to develop a quantitative measure of consciousness, Φ.



(Tononi, 2015)

INTEGRATED INFORMATION THEORY

- Recent versions of IIT focus on *deviations from independence* or *irreducibility*.



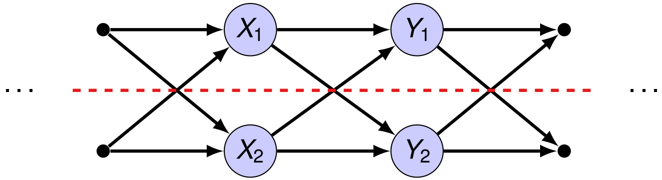
BUT there are many ways in which a system can deviate from independence.

(Oizumi, Tsuchiya & Amari, 2016)

MEASURING INTEGRATED INFORMATION

- Let's take one of the early IIT measures:

$$\Phi^{\mathrm{WMS}} = \underbrace{I(X_1 X_2; Y_1 Y_2)}_{\text{predict future with whole}} - \underbrace{I(X_1; Y_1) - I(X_2; Y_2)}_{\text{predict future with parts}}$$



- $\Phi^{\mathrm{WMS}} > 0$ for systems that are "more than the sum of its parts."
- $\Phi^{\mathrm{WMS}} \leq 0$ for systems that are independent or strongly correlated.

(Barrett & Seth 2011)
(Balduzzi & Tononi 2008)

INTEGRATED INFORMATION DECOMPOSITION

- Given the formula for integrated information

$$\Phi^{\mathrm{WMS}} = I(X_1 X_2; Y_1 Y_2) - I(X_1; Y_1) - I(X_2; Y_2) ,$$

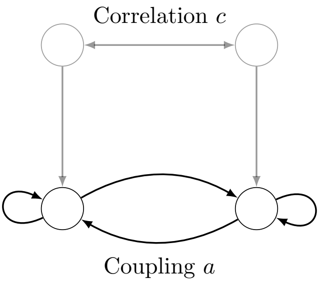
We can decompose it using ΦID:

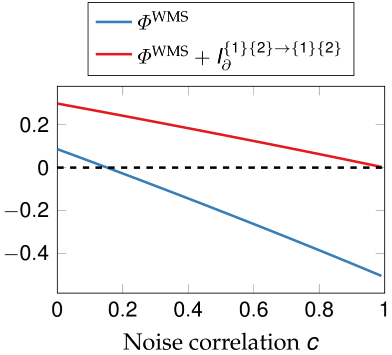
$$\begin{aligned} \Phi^{\mathrm{WMS}} = & \left. -I_\partial^{\{1\}\{2\}\to\{1\}\{2\}} \right\} \text{Red} \\ & \left. \begin{aligned} & + \mathrm{Syn}(X_1, X_2; Y_1 Y_2) + I_\partial^{\{1\}\{2\}\to\{12\}} \\ & + I_\partial^{\{1\}\to\{12\}} + I_\partial^{\{2\}\to\{12\}} \end{aligned} \right\} \text{Syn} \\ & \left. + I_\partial^{\{1\}\to\{2\}} + I_\partial^{\{2\}\to\{1\}} \right\} \text{Un} \end{aligned}$$

INTEGRATED INFORMATION DECOMPOSITION

Now we can fix Φ^{WMS} to be non-negative:

$$\Phi^{\text{C}} = \Phi^{\text{WMS}} + I_\partial^{\{1\}\{2\}\to\{1\}\{2\}}$$

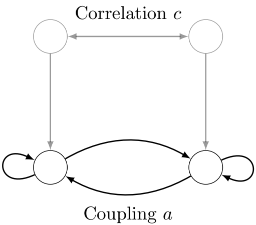


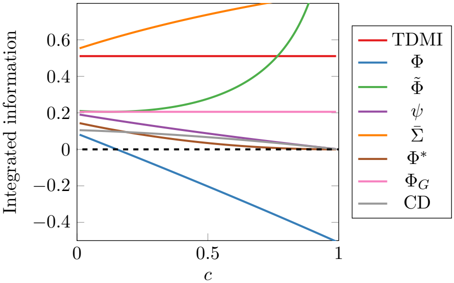


INTEGRATED INFORMATION DECOMPOSITION

Many measures proposed, still no consensus.

Measure	Description	Reference
Φ	Predictive information lost after splitting the system	Balduzzi, 2008
$\tilde{\Phi}$	Uncertainty gained after splitting the system	Barrett, 2011
ψ	Synergistic information between parts of the system	Griffith, 2014
Φ^*	Decoding accuracy lost after splitting the system	Oizumi, 2015
Φ_G	Distance to system with disconnected parts	Oizumi, 2015

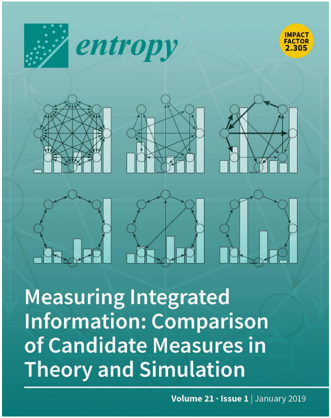




INTEGRATED INFORMATION DECOMPOSITION

- Different measures capture different effects.

They are different not only in practice, but *even in principle*.

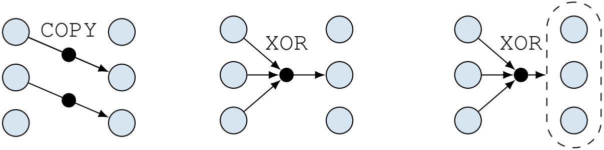


ΦID atoms	**Measures**			
	Φ	CD	ψ	Φ_G
$I_\partial^{\{1\}\{2\}\to\{1\}\{2\}}$	-	0	0	0
$I_\partial^{\{1\}\{2\}\to\{i\}}$	0	0	0	0
$I_\partial^{\{1\}\{2\}\to\{12\}}$	+	0	0	0
$I_\partial^{\{i\}\to\{1\}\{2\}}$	0	+	0	+
$I_\partial^{\{i\}\to\{i\}}$	0	0	0	0
$I_\partial^{\{i\}\to\{j\}}$	+	+	0	+
$I_\partial^{\{i\}\to\{12\}}$	+	0	0	0
$I_\partial^{\{12\}\to\{1\}\{2\}}$	+	+	+	+
$I_\partial^{\{12\}\to\{i\}}$	+	+	+	+
$I_\partial^{\{12\}\to\{12\}}$	+	0	+	0

INTEGRATED INFORMATION DECOMPOSITION

Φ captures **fundamentally different** phenomena.

- These systems have the same Φ, but are qualitatively different:

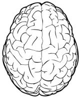
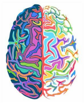


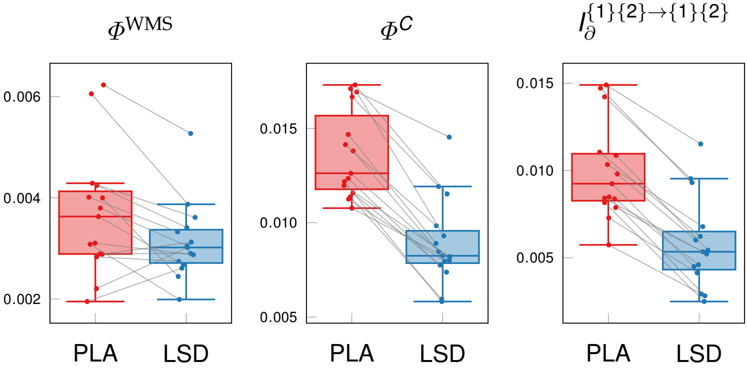
$\Phi^{\text{WMS}} = 1$ $\Phi^{\text{WMS}} = 1$ $\Phi^{\text{WMS}} = 1$

$I_\partial^{\{1\}\to\{2\}} = 1$ $I_\partial^{\{123\}\to\{2\}} = 1$ $I_\partial^{\{123\}\to\{123\}} = 1$

WORK IN PROGRESS: PSYCHEDELIC STATE

- Source-localised MEG data from subjects under the influence of psychedelic drugs.

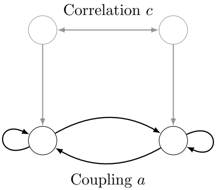
PLA  vs.  LSD

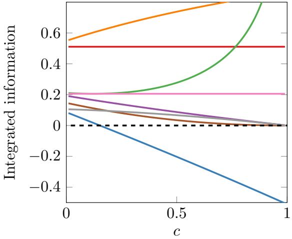


CONCLUSIONS

INTEGRATED INFORMATION THEORY

- ✓ ΦID shows why Φ^{WMS} can be negative, analogous to a "dynamical co-information."
- ✓ Φ measures a mix of different effects that can be disentangled with ΦID.





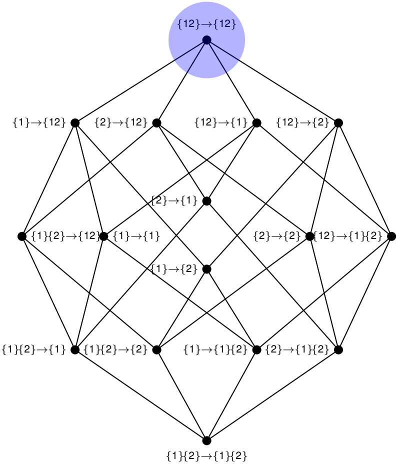
Causal emergence

CAUSAL EMERGENCE

Let's go back to the ΦID lattice...

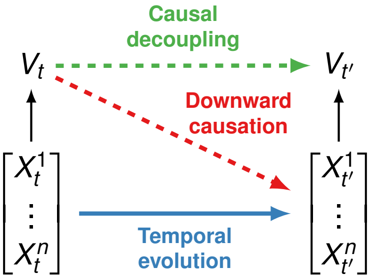
... and focus on the top node.

$I_\partial^{\{12\}\to\{12\}}$ represents **causal emergence**.



CAUSAL EMERGENCE

OUTLINE



Main track

O14

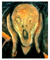
1. Provide a formal definition of causal emergence.
2. Distinguish between two different *kinds* of emergence.
3. Propose a practical criterion and show it in action.

DEFINING EMERGENCE

Definition (causal emergence):

A supervenient feature $V_t = F(X_t)$ exhibits causal emergence if $\texttt{Un}(V_t; \boldsymbol{X}_{t'} | \boldsymbol{X}_t) > 0$.

- Informally: A feature that says something about the future that individual micro elements don't.



BUT we need to know V_t in advance.

Solution: use more PID!

Result:

A system has causally emergent features if and only if $\texttt{Syn}(\boldsymbol{X}_t; \boldsymbol{X}_{t'}) > 0$.

A TAXONOMY OF EMERGENCE

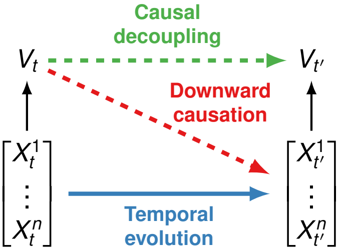
- We can further divide this into two *kinds* of emergence:

Downward causation:

$$\texttt{Un}(V_t; X^i_{t'} | \boldsymbol{X}_t) > 0$$

Causal decoupling:

$$\texttt{Un}(V_t; V_{t'} | \boldsymbol{X}_t, \boldsymbol{X}_{t'}) > 0$$



A TAXONOMY OF EMERGENCE

We can decompose causal emergence using ΦID!

$$\underbrace{\texttt{Syn}(\boldsymbol{X}_t;\boldsymbol{X}_{t'})}_{\text{total emergence capacity}} = \underbrace{\mathcal{G}(\boldsymbol{X}_t)}_{\text{causal decoupling}} + \underbrace{\mathcal{D}(\boldsymbol{X}_t)}_{\text{downward causation}}$$

- Example; for $n = 2$ variables:

$$\texttt{Syn}(\boldsymbol{X}_t;\boldsymbol{X}_{t'}) = \underbrace{I_\partial^{\{12\}\to\{12\}}}_{\text{synergy to synergy}} + \underbrace{I_\partial^{\{12\}\to\{1\}\{2\}} + I_\partial^{\{12\}\to\{1\}} + I_\partial^{\{12\}\to\{2\}}}_{\text{synergy to the rest}}$$

PRACTICAL TOOLS

A feature V_t is causally emergent **if** $\Psi > 0$.

$$\Psi^{(k)}_{t,t'}(V) := I(V_t; V_{t'}) - \sum_{j=1}^{n} I(X^j_t; V_{t'})$$

Pros:

- ✓ Uses only standard mutual information.
- ✓ Uses only pairwise marginals (no curse of dimensionality).
- ✓ No false positives.

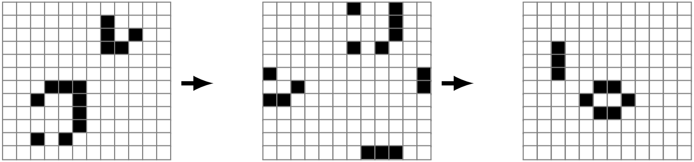
Cons:

- ✗ Needs a candidate feature V_t.
- ✗ Double-counts redundancy (which reduces sensitivity).
- ✗ Inconclusive if $\Psi \leq 0$.

CAUSAL EMERGENCE

RESULTS

Canonical example of emergence: the **Game of Life.**



- **Micro variable**: cell states, $\boldsymbol{X}_t \in \{0, 1\}^n$.
- **Macro variable**: particle type, $V_t \in \{\texttt{blinker}, \texttt{glider}, \ldots\}$.

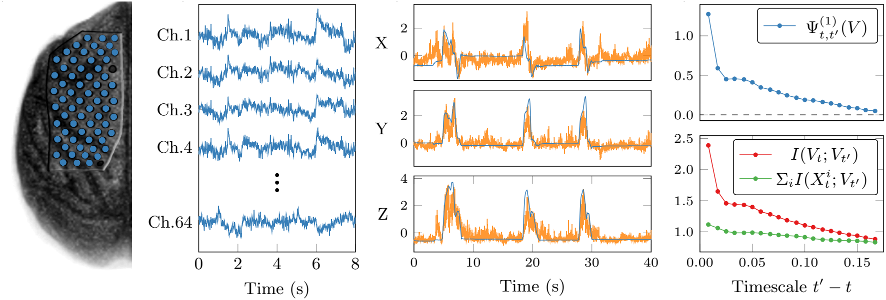
$\longrightarrow$ $\Psi_{t,t'}(V) = 0.58$ bit

CAUSAL EMERGENCE

RESULTS

Example of emergence: **neural activity** during motor control.

- **Micro variable**: ECoG channels, $\boldsymbol{X}_t \in \mathbb{R}^{64}$.
- **Macro variable**: decoded hand motion, $\boldsymbol{V}_t \in \mathbb{R}^3$.



CONCLUSIONS

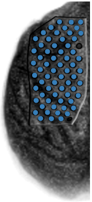
CAUSAL EMERGENCE

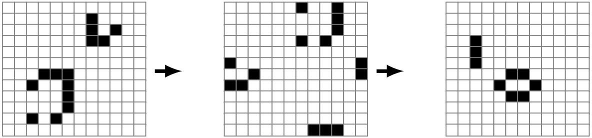
✓ Rigorous, quantitative theory of causal emergence with practical tools.

Main track

O14

? Initial steps towards testing emergence in neural data.





Wrap-up

WRAP-UP

- ✓ ΦID: extension of PID to dynamical systems, applicable to multivariate time series analysis.
- ✓ ΦID allows us to dissect and compare Φ measures, and to make new ones.
- ✓ We proposed a quantitative, rigorous theory of causal emergence.
- ? Many questions open: extensions, algorithms, applications, and more.

plogp@pm.me

Thank you for listening!

Back-up slides

PID NOTATION

- We need to define a *coarse-grained* PID:
  - $\texttt{Un}(X; Y|\boldsymbol{Z})$: unique information that X has about Y that no individual variable Z^i has.
  - $\texttt{Syn}(\boldsymbol{X}; Y)$: information about Y that no individual X^i has (but $\boldsymbol{X}$ as a whole does).

COARSE-GRAINED PID