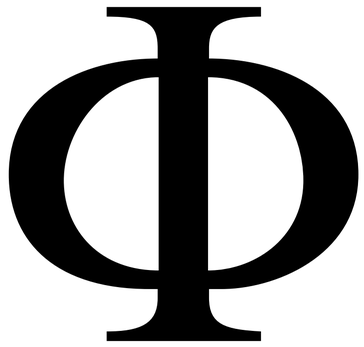
Integrated Information Theory and Wigner's Friend



Andrea Di Biagio
Wigner's Friends Workshop

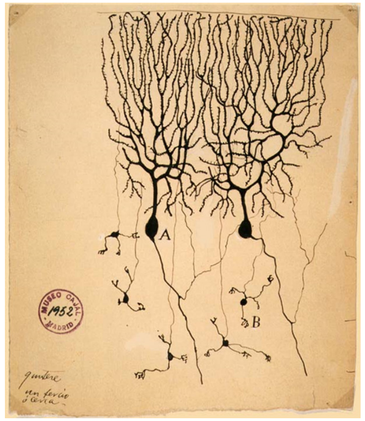
How do you know you have a friend?

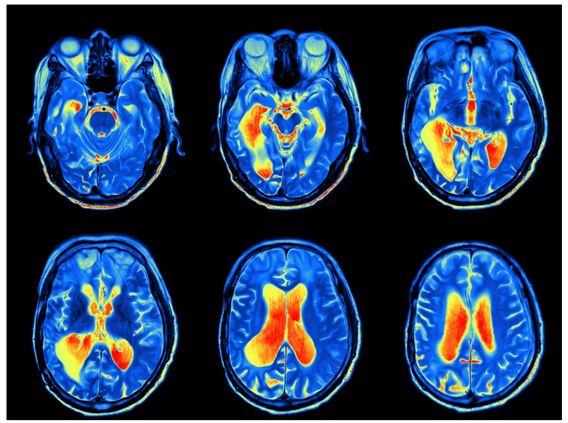
> FRIENDLINESS. If, by open-ended communication, an *independent* party displays cognitive ability at least on par with my own, then they have *thoughts*, and any thought they communicate is *as real as* any communicable thought of my own.

What if did use consciousness as a criterion?

Integrated Information Theory

"*Easy*" vs "*Hard*"





form and function

how come experience?

Integrated Information Theory

Main idea of IIT

Treat consciousness as fundamental and find laws to describe its behaviour

Start from the phenomenology of experience (empirical data)

Then ask what it would take for physical systems to account for its properties.

- Tononi, G., Boly, M., Massimini, M., Koch, C., 2016. Integrated information theory: from consciousness to its physical substrate. Nat Rev Neurosci 17, 450–461. DOI 10.1038/nrn.2016.44
- Oizumi, M., Albantakis, L., Tononi, G., 2014. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. PLOS Computational Biology 10, e1003588. DOI 10.1371/journal.pcbi.1003588

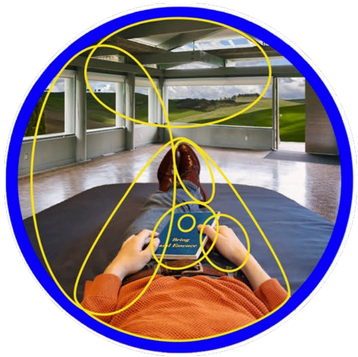
Integrated Information Theory

Methodology of IIT

phenomenal axioms of experience

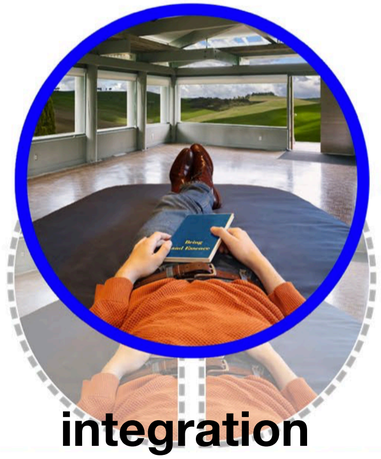
Tononi et al 2206.02069

existence

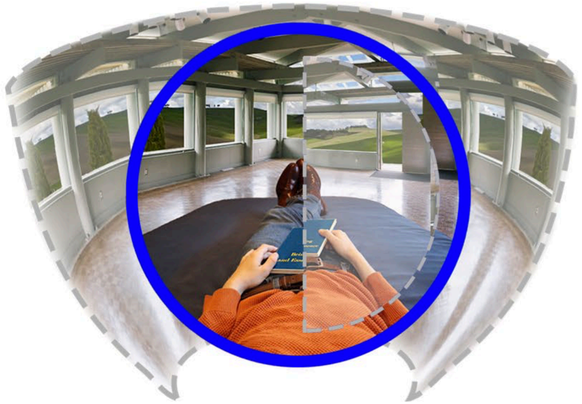


composition

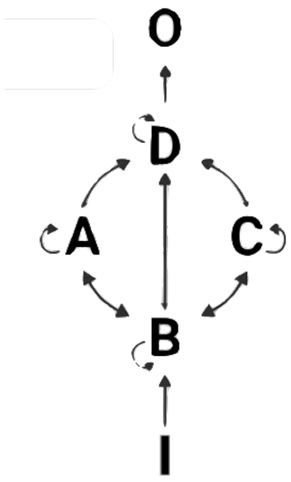
information

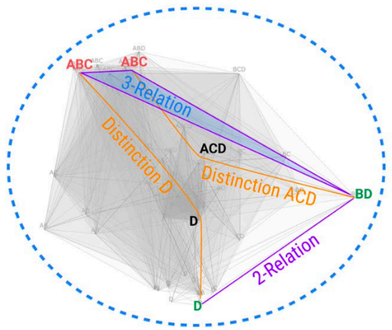


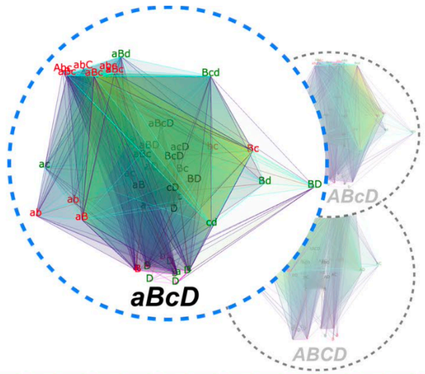
integration

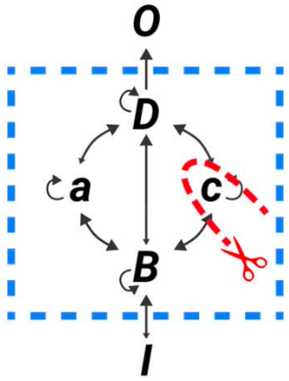


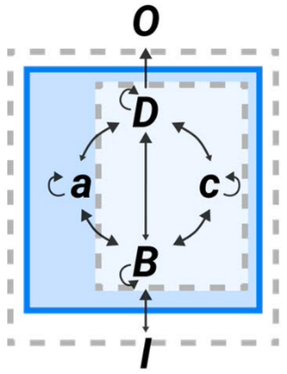
exclusion











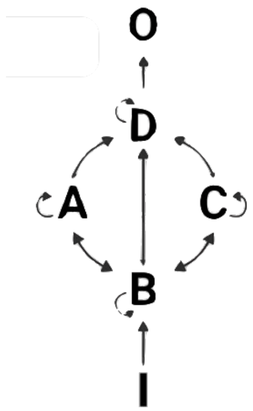
physical postulates about substratum

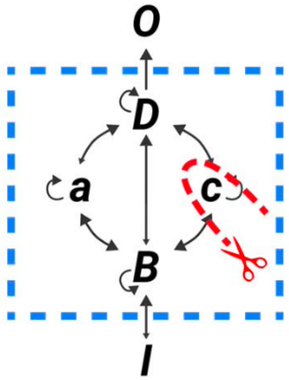
Integrated Information Theory

What is Φ?

Measures of the causal power a system has on itself,
how many states it has,
how irreducible its dynamics.

Quantifier of consciousness





Central identity of IIT

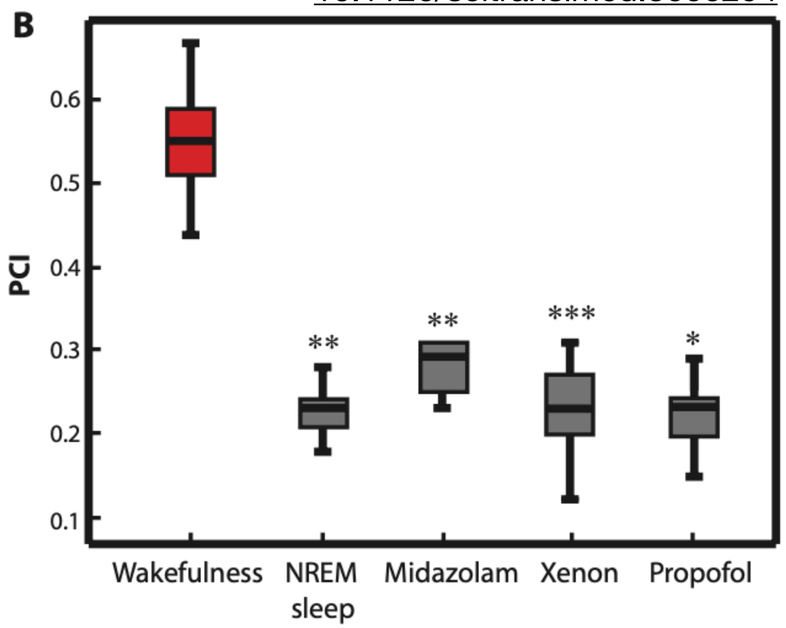
**"The maximally irreducible conceptual structure (MICS)
generated by a complex of elements is *identical* to its experience. "**

- Oizumi, Albantakis, Tononi (2014) 10.1371/journal.pcbi.1003588

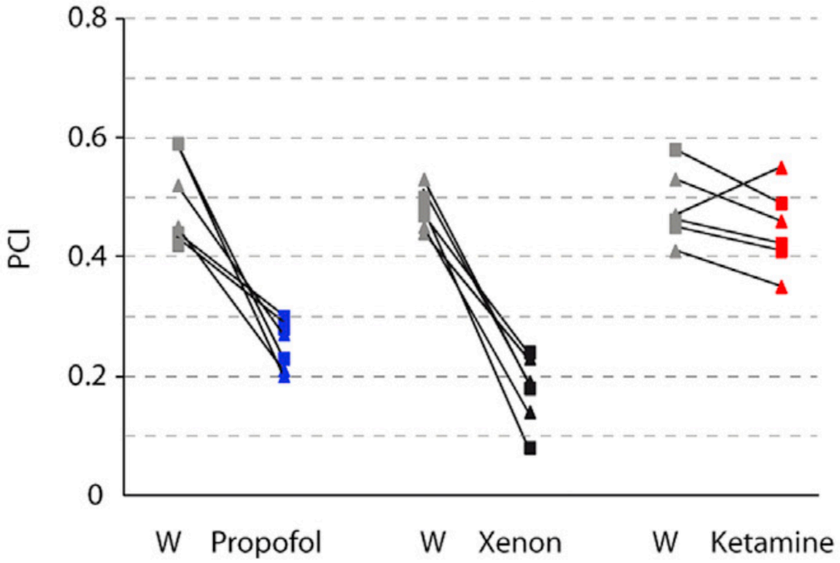
Integrated Information Theory

Experimental results

10.1126/scitranslmed.3006294

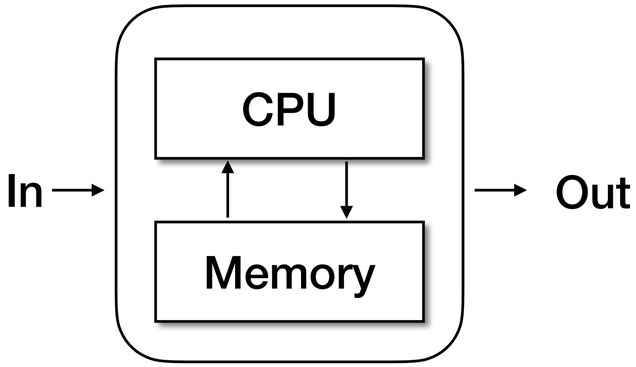


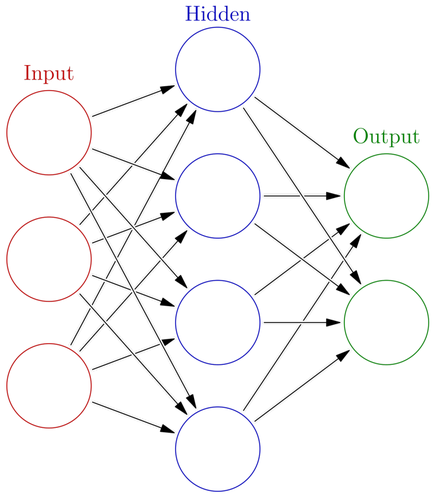
10.1016/j.cub.2015.10.014



PCI = perturbational complexity index

Prediction: implementation matters



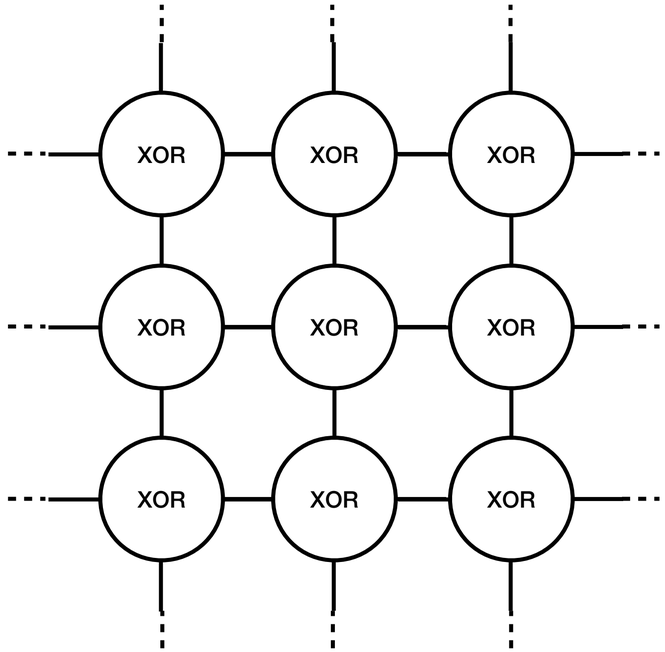


QUALL-E not conscious!

Integrated Information Theory

Aaronson's 2D grid challenge

systems with very simple behaviour can have high Φ values



a feature, not a bug!

Relevance to EWFS

Φ as a benchmark for EWFS

Suppose IIT (or some descendant) starts to become accepted.

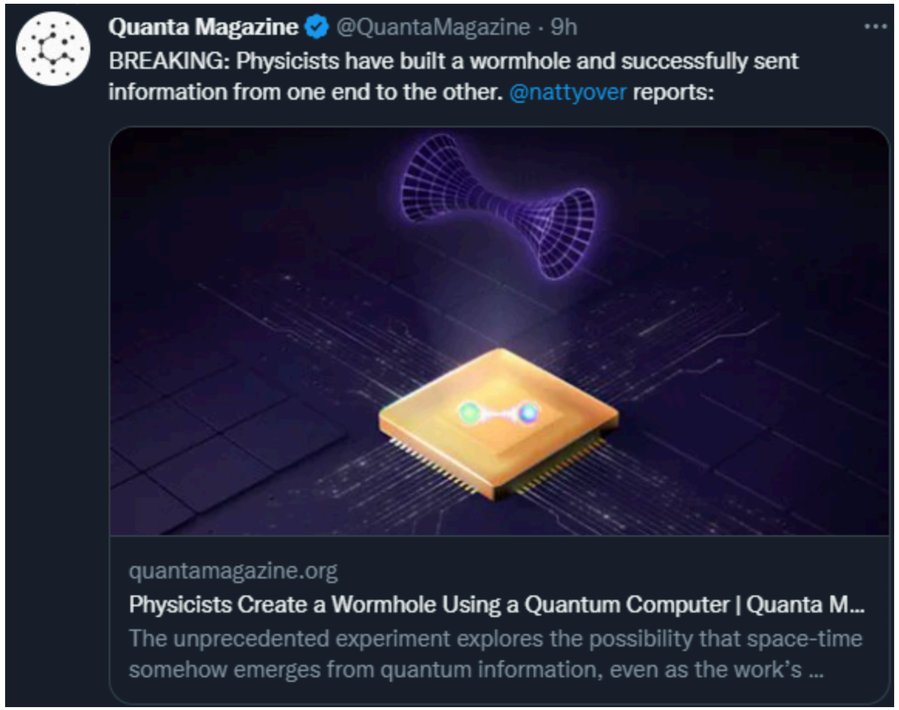
Then one has a way of gauging how conscious something can is.

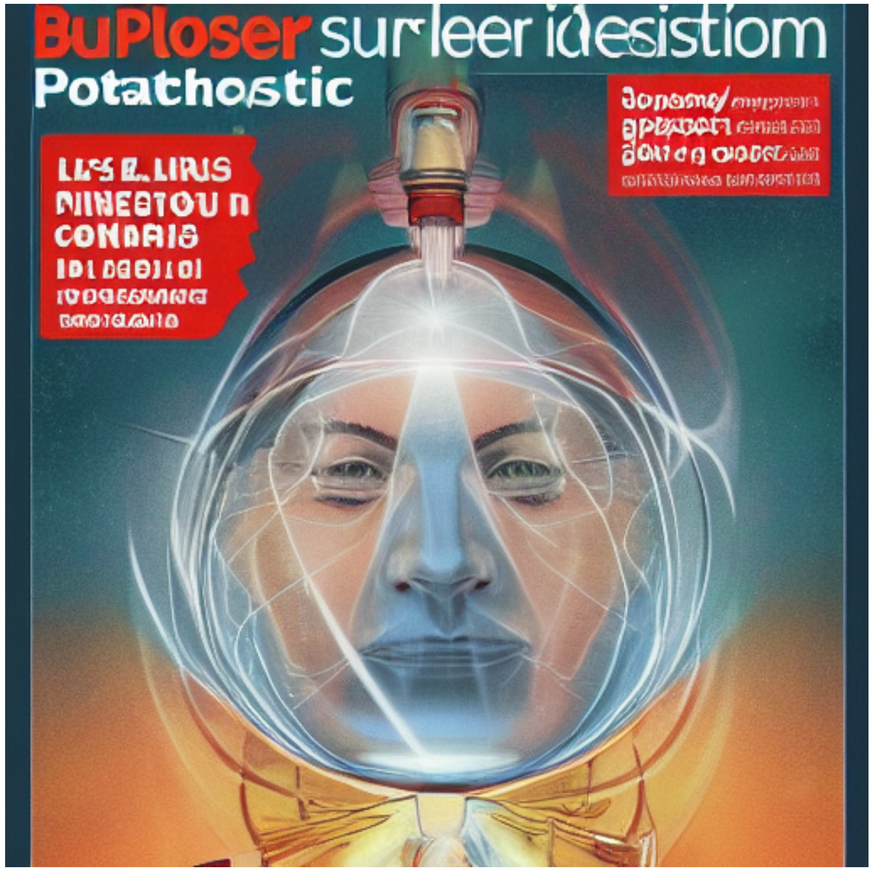
Perhaps high Φ values for a quantum system attainable before HLAI

LF^{Φ} ?

Relevance to EWFS

Note on ethics





Scientists put living being in a quantum superposition!

References

IIT

- Aaronson, S., 2014. Why I Am Not An Integrated Information Theorist (or, The Unconscious Expander). Shtetl-Optimized scottaaronson.blog/?p=1799
- Chalmers, D.J., McQueen, K.J., 2021. Consciousness and the Collapse of the Wave Function arXiv 2105.02314
- Tononi, G., Albantakis, L., Boly, M., Cirelli, C., Koch, C., 2022. Only what exists can cause: An intrinsic view of free will. arXiv 2206.02069
- Tononi, G., Boly, M., Massimini, M., Koch, C., 2016. Integrated information theory: from consciousness to its physical substrate. Nat Rev Neurosci 17, 450–461. DOI 10.1038/nrn.2016.44
- Zanardi, P., Tomka, M., Venuti, L.C., 2018. Towards Quantum Integrated Information Theory. arXiv 1806.01421
- Casali, A.G., Gosseries, O., Rosanova, M., *et. al.*, 2013. A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. Sci. Transl. Med. 5 DOI. 10.1126/scitranslmed.3006294
- Sarasso, S., Boly, M., Napolitani, M., Gosseries, O., *et. a.*, 2015. Consciousness and Complexity during Unresponsiveness Induced by Propofol, Xenon, and Ketamine. Current Biology 25, 3099–3105. DOI 10.1016/j.cub.2015.10.014
- Oizumi, M., Albantakis, L., Tononi, G., 2014. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. PLOS Computational Biology 10, e1003588. DOI 10.1371/journal.pcbi.1003588