Artificial general intelligence
Strong AI

Antal Péter

ComBine Lab

Artificial Intelligence group

Department of Measurement and Information Systems

Agenda

- Strong AI
- The conscious will
- Repertoire of consciousness
- Evolution of consciousness
- Neural correlates of consciousness
- Measuring consciousness
- (Weak and strong emergence)

Strong AI, zombie arguments

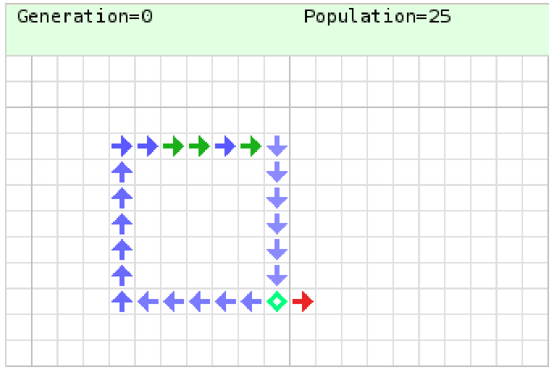
- Simulation and tests for reality (~testing outside from inside)
  - ~BC300: Zhuangzi's (Chuang-Tzu's), Butterfly Dream
  - ~1700: G.Berkeley, subjective idealism
  - Movies: Matrix, Inception,…
  - .. T. Nagel: What is it like to be a bat? N. Bostrom: https://en.wikipedia.org/wiki/Simulation_hypothesis
  - A. Becker: What Is Real?: The Unfinished Quest for the Meaning of Quantum Physics
- Simulation and tests for human mind (~testing inside)
  - Experience: any formally defined (discrete) computation is a program on a universal Turing machine.
  - Experience: any (narrow) intelligence can have a functionally equivalent computational model.
  - Zombie arguments:
    - Assumption: there are (discrete) computational models for conscious minds.
    - Paradox: any execution using arbitrary substrate and realization will give rise to qualia/consciousness.
    - Example: Chinese room (using epiphenomenal patterns in a cellular automaton, see next slide)
      - https://en.wikipedia.org/wiki/Philosophical_zombie
  - Knowledge arguments
    - **Mary the super-scientist** (Frank Jackson: "Epiphenomenal Qualia",1982; "What Mary Didn't Know",1986)
- Reductionism, emergence, downward causation,..

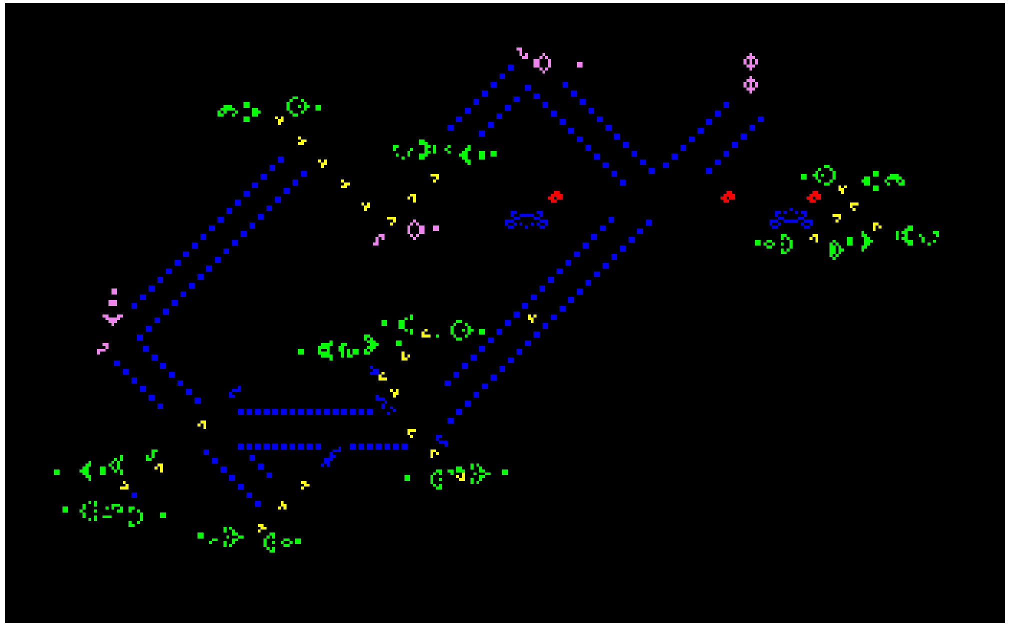
Voight-Kampff Test Questions

Blade runner (1982)

Deckard: Alright, I'm going to ask you a series of questions. Just relax and answer them as simply as you can.
(pause) It's your birthday. Someone gives you a calfskin wallet.
Rachael: I wouldn't accept it. Also, I'd report the person who gave it to me to the police.
Deckard: You've got a little boy. He shows you his butterfly collection plus the killing jar.
Rachael: I take him to the doctor.
Deckard: You're watching television. Suddenly you realize there's a wasp crawling on your arm.
Rachael: I'd kill it.
Deckard: You're reading a magazine. You come across a full-page nude photo of a girl.
Rachael: Is this testing whether I'm a replicant or a lesbian, Mr. Deckard?
Deckard: Just answer the questions, please. (pause) You show it to your husband. He likes it so much he hangs it on your bedroom wall.
Deckard (background): bush outside your window
Rachael: I wouldn't let him.
Deckard (background): orange body, green legs
Deckard: Why not?
Rachael: I should be enough for him.
[audio fades out and in, time passes.]
Deckard: One more question. You're watching a stage play. A banquet is in progress.
The guests are enjoying an appetizer of raw oysters. The entree consists of boiled dog.
Tyrell: Would you step out for a few moments, Rachael? (pause) Thank you.
Deckard: She's a replicant, isn't she?
Tyrell: I'm impressed. How many questions does it usually take to spot them?
Deckard: I don't get it Tyrell.
Tyrell: How many questions?
Deckard: Twenty, thirty, cross-referenced.
Tyrell: It took more than a hundred for Rachael, didn't it?
Deckard: She doesn't know?!
Tyrell: She's beginning to suspect, I think.
Deckard: Suspect? How can it not know what it is?
Tyrell: Commerce, is our goal here at Tyrell. More human than human is our motto.

Reminder: Universal compution in cellular automaton using epiphenomenal patterns





Von Neumann, J. and A. W. Burks (1966): Theory of self-reproducing automata

The Self and Its Brain

PART I by Karl R. Popper

Chapter P1 Materialism Transcends Itself . . . 3
1. Kant's Argument . . . 3
2. Men and Machines . . . 4
3. Materialism Transcends Itself . . . 5
4. Remarks on the Term "Real" . . . 9
5. Materialism, Biology, and Mind . . . 11
6. Organic Evolution . . . 11
7. Nothing New Under the Sun. Reductionism and "Downward Causation" . . . 14
8. Emergence and its Critics . . . 22
9. Indeterminism; the Interaction of Levels of Emergence . . . 32

Chapter P2 The Worlds 1, 2 and 3 . . . 36
10. Interaction; The Worlds 1, 2 and 3 . . . 36
11. The Reality of World 3 . . . 38
12. Unembodied World 3 Objects . . . 41
13. Grasping a World 3 Object . . . 43
14. The Reality of Unembodied World 3 Objects . . . 46
15. World 3 and the Mind-Body Problem . . . 47

Chapter P3 Materialism Criticized . . . 51
16. Four Materialist or Physicalist Positions . . . 51
17. Materialism and the Autonomous World 3 . . . 56
18. Radical Materialism or Radical Behaviourism . . . 60
19. Panpsychism . . . 67
20. Epiphenomenalism . . . 72
21. A Revised Form of J. B. S. Haldane's Refutation of Materialism . . . 75
22. The So-called Identity Theory . . . 81
23. Does the Identity Theory Escape the Fate of Epiphenomenalism? . . . 86
24. A Critical Note on Parallelism. The Identity Theory as a Form of Parallelism . . . 88
25. Additional Remarks on Some Recent Materialist Theories . . . 93
26. The New Promissory Materialism . . . 96
27. Results and Conclusion . . . 98

Chapter P4 Some Remarks on the Self . . . 100
28. Introduction . . . 100
29. Selves . . . 101
30. The Ghost in the Machine . . . 104
31. Learning to be a Self . . . 108
32. Individuation . . . 112
33. Self-Identity: The Self and Its Brain . . . 115
34. The Biological Approach to Human Knowledge and Intelligence . . . 120
35. Consciousness and Perception . . . 124
36. The Biological Function of Conscious and Intelligent Activity . . . 125
37. The Integrative Function of Consciousness . . . 127
38. The Continuity of the Self . . . 129
39. Learning from Experience: The Natural Selection of Theories . . . 132
40. Criticism of the Theory of Unconditioned and Conditioned Reflexes . . . 135
41. Kinds of Memory . . . 139
42. The Self Anchored in World 3 . . . 144

Chapter P5 Historical Comments on the Mind-Body Problem . . . 148
43. The History of Our Picture of the Universe . . . 148
44. A Problem to be Solved by What Follows . . . 151
45. The Prehistoric Discovery of the Self and of World 2 . . . 153
46. The Mind-Body Problem in Greek Philosophy . . . 159
47. Conjectural *versus* Ultimate Explanation . . . 171
48. Descartes: A Shift in the Mind-Body Problem . . . 176
49. From Interactionism to Parallelism: The Occasionalists and Spinoza . . . 181
50. Leibniz's Theory of Mind and Matter: From Parallelism to Identity . . . 184
51. Newton, Boscovich, Maxwell: The End of Ultimate Explanation . . . 189
52. The Association of Ideas as an Ultimate Explanation . . . 194
53. Neutral Monism . . . 196
54. The Identity Theory After Leibniz: From Kant to Feigl . . . 199
55. Linguistic Parallelism . . . 202
56. A Final Look at Materialism . . . 205

Chapter P6 Summary . . . 209
Bibliography to Part I . . . 211

PART II by John C. Eccles

Preface . . . 225

Chapter E1 The Cerebral Cortex . . . 227
1. Résumé . . . 227
2. Anatomical Introduction . . . 229
3. The Columnar Arrangement and the Modular Concept of the Cerebral Cortex . . . 235
4. Modular Interaction . . . 240
4.1 Patterns of Module Action and Interaction . . . 242
5. The Connectivities of Cortical Areas . . . 244
6. Connectivities of the Limbic System . . . 247

Chapter E2 Conscious Perception . . . 250
7. Résumé . . . 250

Popper, K.R. and Eccles, J.C., 1977. *The self and its brain*. Springer

The Self and Its Brain cont'

8. Introduction . . . 252
9. Cutaneous Perception (Somaesthesis) . . . 253
9.1 Pathways to Primary Sensory Area in Cortex . . . 253
9.2 Temporal Analysis of Cutaneous Perception . . . 256
9.3 Secondary and Tertiary Sensory Areas . . . 259
10. Visual Perception . . . 261
10.1 Retina to Primary Visual Area in Cortex . . . 261
10.2 Stages in Reconstitution of the Visual Image . . . 264
10.3 The Perceived Visual Image . . . 270
11. Auditory Perception . . . 271
12. Olfactory Perception . . . 272
13. Emotional Colouring of Conscious Perceptions . . . 273
14. Epilogue . . . 274

Chapter E3 Voluntary Movement . . . 275

15. Résumé . . . 275
16. Introduction . . . 276
17. The Motor Unit . . . 277
18. The Motor Cortex . . . 280
19. Voluntary Movement . . . 282
20. The Cerebellar Controls of Voluntary Movement . . . 286
20.1 The Closed Loop via the Pars Intermedia of the Cerebellum . . . 288
20.2 The Open-Loop System via the Cerebellar Hemispheres . . . 289
21. The Open-Loop Circuits via the Basal Ganglia . . . 290
22. Synthesis of the Various Neuronal Mechanisms Concerned in the Control of Voluntary Movement . . . 290
23. General Discussion . . . 293

Chapter E4 The Language Centres of the Human Brain . . . 295

24. Résumé . . . 295
25. Introduction . . . 296
26. Aphasia . . . 296
27. Experiments on Exposed Brains . . . 300
28. Intracarotid Injections of Sodium Amytal . . . 301
29. The Dichotic Listening Test . . . 302
30. The Self-Conscious Mind and Speech . . . 304
31. Anatomical Substrates of Speech Mechanisms . . . 305
32. The Acquisition of Language . . . 309

Chapter E5 Global Lesions of the Human Cerebrum . . . 311

33. Résumé . . . 311
34. Introduction . . . 313
35. Investigations on the Human Brain After Commissural Section-Commissurotomy . . . 313
36. Discussion on Commissurotomy . . . 325
37. Investigations on the Human Cerebrum after Gross Lesions and Hemispherectomy . . . 329
38. Discussion on Hemispherectomy . . . 330
39. Summary of Linguistic Abilities Disclosed by Global Lesions . . . 333

Chapter E6 Circumscribed Cerebral Lesions . . . 334

40. Résumé . . . 334
41. Introduction . . . 336

42. Temporal Lobe Lesions . . . 336
43. Parietal Lobe Lesions . . . 341
44. Occipital Lobe Lesions . . . 344
45. Frontal Lobe Lesions . . . 345
46. The Limbic System . . . 349
47. The Dominant and Minor Hemispheres . . . 350

Chapter E7 The Self-Conscious Mind and the Brain . . . 355

48. Résumé . . . 355
49. Introduction . . . 357
50. Self-Conscious Mind and the Brain . . . 358
51. Hypothesis of Interaction of Self-Conscious Mind and the Liaison Brain . . . 361
52. The Hypothesis of Cortical Modules and the Self-Conscious Mind . . . 365
53. Sleep, Dreams and Various Forms of Unconsciousness . . . 370
54. Plasticity of "Open" Modules . . . 373
55. Summary . . . 373

Chapter E8 Conscious Memory: The Cerebral Processes Concerned in Storage and Retrieval . . . 377

56. Résumé . . . 377
57. Introduction . . . 379
58. Structural and Functional Changes Possibly Related to Memory . . . 380
59. The So-called Growth Theory of Learning . . . 385
60. The Role of the Self-Conscious Mind in Short-Term Memory . . . 388
61. The Role of the Hippocampus in Learning and Memory . . . 390
62. Hypothesis of Neuronal Happenings in Memory Storage . . . 394
63. Memory Retrieval . . . 399
64. Durations of Memories . . . 402
65. Plastic Responses of Cerebral Cortex . . . 403
66. Retrograde Amnesia . . . 405

Bibliography to Part II . . . 407

PART III Dialogues Between the Two Authors

Dialogue I . . . 425
Dialogue II . . . 437
Dialogue III . . . 449
Dialogue IV . . . 461
Dialogue V . . . 471
Dialogue VI . . . 485
Dialogue VII . . . 498
Dialogue VIII . . . 510
Dialogue IX . . . 521
Dialogue X . . . 532
Dialogue XI . . . 548
Dialogue XII . . . 562

Bibliography to Part III . . . 567

Index of Names . . . 569
Index of Subjects . . . 577

Popper, K.R. and Eccles, J.C., 2012. *The self and its brain*. Springer

Unconscious (initiative) vs conscious (will)

Abstract: Voluntary acts are preceded by electrophysiological "readiness potentials" (RPs). With spontaneous acts involving no preplanning, the main negative RP shift begins at about -550 ms. Such RP's were used to indicate the minimum onset times for the cerebral activity that precedes a fully endogenous voluntary act. The time of conscious intention to act was obtained from the subject's recall of the spatial clock position of a revolving spot at the time of his initial awareness of intending or wanting to move (W). W occurred at about -200 ms. Control experiments, in which a skin stimulus was timed (S), helped evaluate each subject's error in reporting the clock times for awareness of any perceived event.

For spontaneous voluntary acts, RP onset preceded the uncorrected Ws by about 350 ms and the Ws corrected for S by about 400 ms. The direction of this difference was consistent and significant throughout, regardless of which of several measures of RP onset or W were used. It was concluded that cerebral initiation of a spontaneous voluntary act begins unconsciously. However, it was found that the final decision to act could still be consciously controlled during the 150 ms or so remaining after the specific conscious intention appears. Subjects can in fact "veto" motor performance during a 100-200-ms period before a prearranged time to act.

The role of conscious will would be not to initiate a specific voluntary act but rather to select and control volitional outcome. It is proposed that conscious will can function in a permissive fashion, either to permit or to prevent the motor implementation of the intention to act that arises unconsciously. Alternatively, there may be the need for a conscious activation or triggering, without which the final motor output would not follow the unconscious cerebral initiating and preparatory processes.

5. Free will and individual responsibility

This is not the place to debate the issue of free will versus determinism in connection with an apparently endogenous voluntary action that one experiences subjectively as freely willed and self-controllable (see Eccles 1980; Hook 1960; Nagel 1979; Popper & Eccles 1977). However, it is important to emphasize that the present experimental findings and analysis do not exclude the potential for "philosophically real" individual responsibility and free will. Although the volitional process may be initiated by unconscious cerebral activities, conscious control of the actual motor performance of voluntary acts definitely remains possible. The findings should therefore be taken not as being antagonistic to free will but rather as affecting the view of how free will might operate. Processes associated with individual responsibility and free will would "operate" not to initiate a voluntary act but to select and control volitional outcomes. (Voluntary action and responsibility operating behaviorally within a deterministic view would, of course, be subject to analogous restrictions.)

Some may view responsibility and free will as operative only when voluntary acts follow slower conscious deliberation of alternative choices of action. But, as already noted above, any volitional choice does not become a voluntary action until the person moves. In the present study, the subjects reported that the same conscious urge or decision to move that they experienced just before each voluntary act was present and that it was similar whether or not any additional experience of general preplanning had already been going on. Indeed, the reported times for awareness of wanting to move were essentially the same for fully spontaneous acts and those with some preplanning (Libet, Gleason, Wright & Pearl 1983). One might therefore speculate that the actual motor execution even of a deliberately preselected voluntary act may well involve processes similar to those for the spontaneously voluntary acts studied by us. The urge or intention actually to perform the voluntary act would then still be initiated unconsciously, regardless of the preceding kinds of deliberative processes.

The concept of conscious veto or blockade of the motor performance of specific intentions to act is in general accord with certain religious and humanistic views of ethical behavior and individual responsibility. "Self-control" of the acting out of one's intentions is commonly advocated; in the present terms this would operate by conscious selection or control of whether the unconsciously initiated final volitional process will be implemented in action. Many ethical strictures, such as most of the Ten Commandments, are injunctions not to act in certain ways. On the other hand, if the final intention to act arises unconsciously, the mere appearance of an intention could not consciously be prevented, even though its consummation in a motor act could be controlled consciously. It would not be surprising, therefore, if religious and philosophical systems were to create insurmountable moral and psychological difficulties when they castigate individuals for simply having a mental intention or impulse to do something unacceptable, even when this is not acted out (e.g., Kaufinann 1961).

Libet, B., 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and brain sciences*, *8*(4), pp.529-539.

Towards a neurobiological theory of consciousness

Visual awareness is a favorable form of consciousness to study neurobiologically. We propose that it takes two forms: a very fast form, linked to iconic memory, that may be difficult to study; and a somewhat slower one involving visual attention and short-term memory. In the slower form an attentional mechanism transiently binds together all those neurons whose activity relates to the relevant features of a single visual object. We suggest this is done by generating coherent semi-synchronous oscillations, probably in the 40-70 Hz range. These oscillations then activate a transient short-term (working) memory. We outline several lines of experimental work that might advance the understanding of the neural mechanisms involved. The neural basis of very short-term memory especially needs more experimental study.

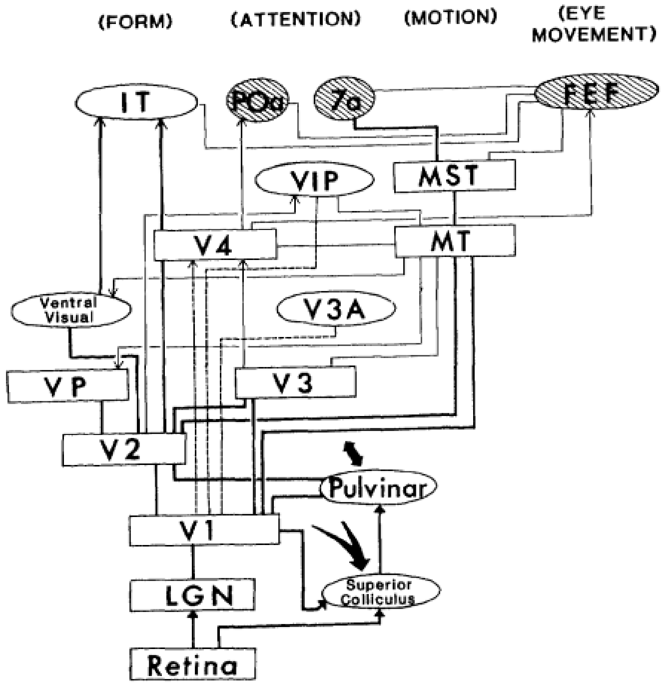


Figure 1. Some of the major cortical and sub-cortical visual areas in the macaque monkey. Major pathways are indicated by bold lines, minor pathways by thin lines; dashed lines weak or uncertain. The shaded ellipsoids correspond to higher-level visual-motor areas. Most cortical areas project back to the superior colliculus as well as to the different maps in the pulvinar, which is part of the thalamus. Furthermore, every forward projection has usually an equally strong, if not stronger, back-projection associated with it (not shown). These feedback pathways usually terminate outside cortical layer IV, while the forward projection terminates most densely in this layer. Areas related to the control and expression of visual attention include the superior colliculus and the pulvinar, as well as areas POa and 7a, part of the posterior parietal cortex. Abbreviations: FEF—frontal eye fields; IT—inferior temporal lobe; LGN—lateral geniculate nucleus; MT, MST—motion-processing areas; V1—primary visual area; V2-4—higher-order visual processing areas; VIP—ventral interparietal area; VP—ventral posterior. The figure is from D. Van Essen and J. Maunsell, personal communication. For more information, see ref 20.

Crick, F. and Koch, C., 1990. Towards a neurobiological theory of consciousness. In *Seminars in the Neurosciences* (Vol. 2, pp. 263-275). Saunders Scientific Publications.

A framework for consciousness

Here we summarize our present approach to the problem of consciousness. After an introduction outlining our general strategy, we describe what is meant by the term 'framework' and set it out under ten headings. This framework offers a coherent scheme for explaining the neural correlates of (visual) consciousness in terms of competing cellular assemblies. Most of the ideas we favor have been suggested before, but their combination is original. We also outline some general experimental approaches to the problem and, finally, acknowledge some relevant aspects of the brain that have been left out of the proposed framework.

Fig. 1. The snapshot hypothesis proposes that the conscious perception of motion is not represented by the change of firing rate of the relevant neurons, but by the (near) constant firing of certain neurons that represent the motion. The figure is an analogy. It shows how a static picture can suggest motion.

Crick, F. and Koch, C., 2003. A framework for consciousness. *Nature neuroscience*, *6*(2), p.119.

Voluntary action and conscious awareness

Humans have the conscious experience of 'free will': we feel we can generate our actions, and thus affect our environment. Here we used the perceived time of intentional actions and of their sensory consequences as a means to study consciousness of action. These perceived times were attracted together in conscious awareness, so that subjects perceived voluntary movements as occurring later and their sensory consequences as occurring earlier than they actually did. Comparable involuntary movements caused by magnetic brain stimulation reversed this attraction effect. We conclude that the CNS applies a specific neural mechanism to produce intentional binding of actions and their effects in conscious awareness.

Haggard, P., Clark, S. and Kalogeras, J., 2002. Voluntary action and conscious awareness. *Nature neuroscience*, *5*(4), p.382.

Unconscious determinants of free decisions in the human brain

There has been a long controversy as to whether subjectively 'free' decisions are determined by brain activity ahead of time. We found that the outcome of a decision can be encoded in brain activity of prefrontal and parietal cortex up to 10 s before it enters awareness. This delay presumably reflects the operation of a network of high-level control areas that begin to prepare an upcoming decision long before it enters awareness.

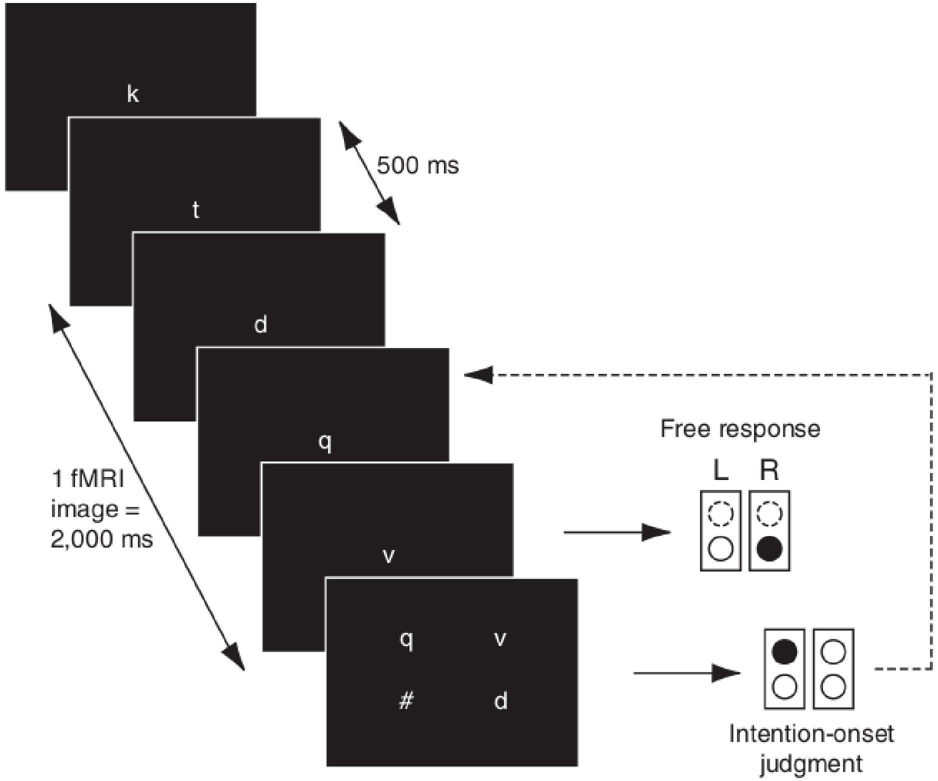


Figure 1 Measuring the onset time of conscious motor intentions. Subjects viewed a letter stream that was updated every 500 ms (shown here only for a few frames). At some point they spontaneously made the decision to press either the left or right button using their corresponding index finger (free response). Subsequently, they were presented with a response-mapping screen that instructed subjects as to which second button to press to report the time at which they consciously made the motor decision

Soon, C.S., Brass, M., Heinze, H.J. and Haynes, J.D., 2008. Unconscious determinants of free decisions in the human brain. *Nature neuroscience*, *11*(5), p.543.

Table 1. Examples of computations pertaining to information-processing levels C0, C1 and C2 in the human brain.

Computation	Examples of experimental findings	References
	C0: Unconscious processing	
Invariant visual recognition	Subliminal priming by unseen words and faces, invariant for font, size, or viewpoint.	(*5*)
	Functional MRI (fMRI) and single-neuron response to unseen words and faces	(*33, 37, 78, 79*)
	Unconscious judgement of chess game configurations	(*80*)
Access to meaning	N400 response to unseen out-of-context words	(*9, 10*)
Cognitive control	Unconscious inhibition or task set preparation by an unseen cue	(*11, 12*)
Reinforcement learning	Subliminal instrumental conditioning by unseen shapes	(*17*)
	C1: Global availability of information	
All-or-none selection and broadcasting of a relevant content	Conscious perception of a single picture during visual rivalry	(*29*)
	Conscious perception of a single detail in a picture or stream	(*28, 81*)
	All-or-none memory retrieval	(*82*)
	Attentional blink: Conscious perception of item A prevents the simultaneous perception of item B	(*27, 30, 83, 84*)
	All-or-none "ignition" of event-related potentials and fMRI signals, only on trials with conscious perception	(*33–35, 85–87*)
	All-or-none firing of neurons coding for the perceived object in prefrontal cortex and other higher areas	(*31, 32, 37, 38, 88*)
Stabilization of short-lived information for off-line processing	Brain states are more stable when information is consciously perceived; unconscious information quickly decays (~1 s)	(*39, 89*)
	Conscious access may occur long after the stimulus is gone	(*90*)
Flexible routing of information	Only conscious information can be routed through a series of successive operations (for example, successive calculations 3 × 4 + 2)	(*91*)
Sequential performance of several tasks	Psychological refractory period: Conscious processing of item A delays conscious processing of item B	(*34, 92*)
	Serial calculations or strategies require conscious perception	(*13, 91*)
	Serial organization of spontaneous brain activity during conscious thought in the "resting state"	(*93*)
	C2: Self-monitoring	
Self-confidence	Humans accurately report subjective confidence, a probabilistic estimate in the accuracy of a decision or computation	(*51, 55*)
Evaluation of one's knowledge	Humans and animals can ask for help or "opt out" when unsure	(*53, 65, 66*)
	Humans and animals know when they do not know or remember	(*49, 53*)
Error detection	Anterior cingulate response to self-detected errors	(*61, 65, 94*)
Listing one's skills	Children know the arithmetic procedures at their disposal, their speed, and error rate.	(*70*)
Sharing one's confidence with others	Decision-making improves when two persons share knowledge	(*69*)

Dehaene, S., Lau, H. and Kouider, S., 2017. What is consciousness, and could machines have it? *Science*, *358*(6362), pp.486-492.

Neural correlates of consciousness

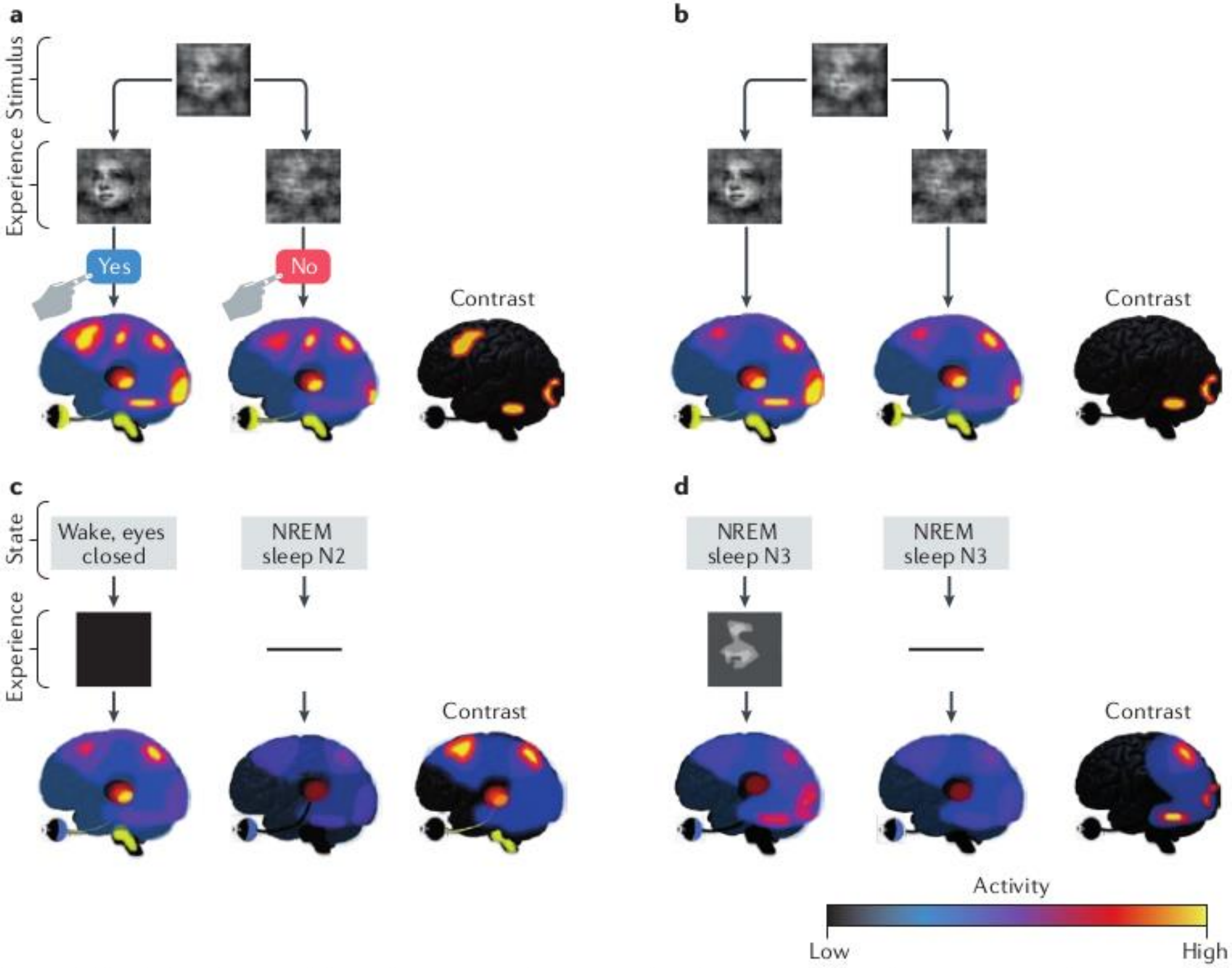


Figure 1 | **Identifying the neural correlates of consciousness.** This schematic diagram illustrates the absolute levels of brain activity across different conditions designed to identify the neural correlates of consciousness (NCC).

Koch, C., Massimini, M., Boly, M. and Tononi, G., 2016. Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience*, *17*(5), p.307.

Integrated information theory: from consciousness to its physical substrate

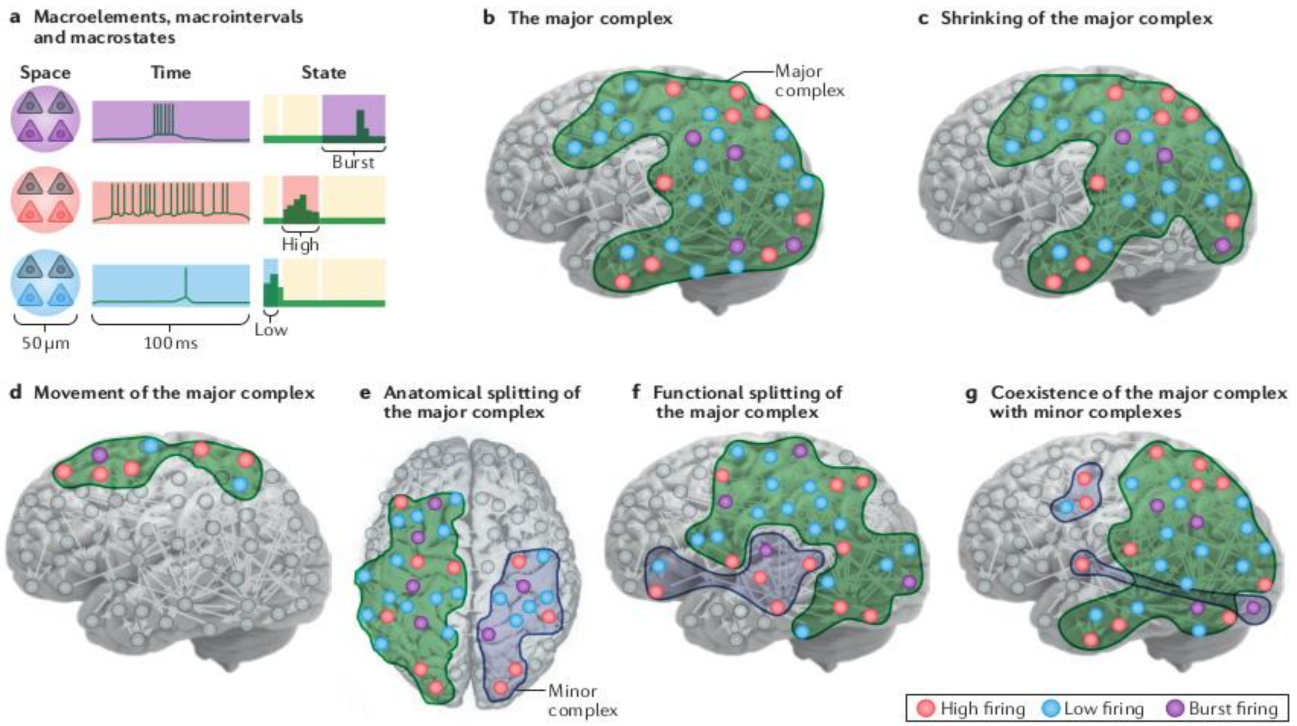


Figure 3 | **Identifying the physical substrate of consciousness (PSC) from first principles.** The complex of neural elements that constitutes the PSC can be identified by searching for maxima of intrinsic cause–effect power. **a** | For example, assume that the elements, timescale and states at which intrinsic cause–effect power reaches a maximum have been identified using optogenetic and unit recording tools (FIG. 2). Here, the elements are groups of neurons, the timescale is over 100 ms and there are three states (low, high and burst firing). **b** | In a healthy, awake participant, the set of neural elements specifying the conceptual structure with the highest Φ^{max} is assumed, based on current evidence, to be a complex of neuronal groups distributed over the posterior cortex and portions of the anterior cortex[5]. Empirical studies can, in principle, establish whether the full neural correlates of consciousness[5] correspond to the maximum of intrinsic cause–effect power, thereby corroborating or falsifying a key prediction of integrated information theory. **c** | The boundaries of the PSC (green line) may change after cortical lesions, such as those causing absolute achromatopsia, resulting in a smaller PSC. **d** | The PSC boundaries may also move as a result of changes in excitability and effective connectivity, as might occur during pure thought that is devoid of sensory content. **e** | The PSC could also split into two large local maxima of cause–effect power (represented here by green and blue boundaries) as a result of anatomical disconnections, such as in split-brain patients, in which instance each hemisphere would have its own consciousness. **f** | The PSC may also split as a result of functional disconnections, which may occur in some psychiatric disorders and perhaps under certain dual-task conditions — for example while driving and talking at the same time. **g** | The coexistence of a large major complex with one or more minor complexes that may support sophisticated, seemingly unconscious performance could be a common occurrence in everyday life.

Tononi, G., Boly, M., Massimini, M. and Koch, C., 2016. Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, *17*(7), pp.450-461.

Towards solving the hard problem of consciousness
Adaptive Resonance Theory (ART)

The hard problem of consciousness is the problem of explaining how we experience qualia or phenomenal experiences, such as seeing, hearing, and feeling, and knowing what they are. To solve this problem, a theory of consciousness needs to link brain to mind by modeling how emergent properties of several brain mechanisms interacting together embody detailed properties of individual conscious psychological experiences. This article summarizes evidence that Adaptive Resonance Theory, or ART, accomplishes this goal. ART is a cognitive and neural theory of how advanced brains autonomously learn to attend, recognize, and predict objects and events in a changing world. ART has predicted that "all conscious states are resonant states" as part of its specification of mechanistic links between processes of consciousness, learning, expectation, attention, resonance, and synchrony. It hereby provides functional and mechanistic explanations of data ranging from individual spikes and their synchronization to the dynamics of conscious perceptual, cognitive, and cognitive–emotional experiences. ART has reached sufficient maturity to begin classifying the brain resonances that support conscious experiences of seeing, hearing, feeling, and knowing. Psychological and neurobiological data in both normal individuals and clinical patients are clarified by this classification. This analysis also explains why not all resonances become conscious, and why not all brain dynamics are resonant. The global organization of the brain into computationally complementary cortical processing streams (complementary computing), and the organization of the cerebral cortex into characteristic layers of cells (laminar computing), figure prominently in these explanations of conscious and unconscious processes. Alternative models of consciousness are also discussed.

Grossberg, S., 2017. Towards solving the hard problem of consciousness: The varieties of brain resonances and the conscious experiences that they support. *Neural Networks*, *87*, pp.38-95.
Grossberg, S., 2019. The resonant brain: How attentive conscious seeing regulates action sequences that interact with attentive cognitive learning, recognition, and prediction. *Attention, Perception, & Psychophysics*, pp.1-28.

The Easy Part of the Hard Problem: A Resonance Theory of Consciousness

Synchronization, harmonization, vibrations, or simply *resonance* in its most general sense seems to have an integral relationship with consciousness itself. One of the possible "neural correlates of consciousness" in mammalian brains is a specific combination of gamma, beta and theta electrical synchrony. More broadly, we see similar kinds of resonance patterns in living and non-living structures of many types. What clues can resonance provide about the nature of consciousness more generally? This paper provides an overview of resonating structures in the fields of neuroscience, biology and physics and offers a possible solution to what we see as the "easy part" of the "Hard Problem" of consciousness, which is generally known as the "combination problem." The combination problem asks: how do micro-conscious entities combine into a higher-level macro-consciousness? The proposed solution in the context of mammalian consciousness suggests that a shared resonance is what allows different parts of the brain to achieve a phase transition in the speed and bandwidth of information flows between the constituent parts. This phase transition allows for richer varieties of consciousness to arise, with the character and content of that consciousness in each moment determined by the particular set of constituent neurons. We also offer more general insights into the ontology of consciousness and suggest that consciousness manifests as a continuum of increasing richness in all physical processes, distinguishing our view from emergentist materialism. We refer to this approach, a meta-synthesis, as a (general) resonance theory of consciousness. We offer some suggestions for testing the theory.

Hunt, T. and Schooler, J., 2019. The easy part of the Hard Problem: A resonance theory of consciousness. *Frontiers in human neuroscience*, *13*, p.378.

"Plant intelligence"

In claiming that plants have consciousness, 'plant neurobiologists' have consistently glossed over the remarkable degree of structural and functional complexity that the brain had to evolve for consciousness to emerge. Here, we outline a new hypothesis proposed by Feinberg and Mallat for the evolution of consciousness in animals. Based on a survey of the brain anatomy, functional complexity, and behaviors of a broad spectrum of animals, criteria were established for the emergence of consciousness. The only animals that satisfied these criteria were the vertebrates (including fish), arthropods (e.g., insects, crabs), and cephalopods (e.g., octopuses, squids). In light of Feinberg and Mallat's analysis, we consider the likelihood that plants, with their relative organizational simplicity and lack of neurons and brains, have consciousness to be effectively nil.

Taiz, L., Alkon, D., Draguhn, A., Murphy, A., Blatt, M., Hawes, C., Thiel, G. and Robinson, D.G., 2019. Plants neither possess nor require consciousness. *Trends in plant science*.

The Cambridge Declaration on Consciousness

We declare the following: *"The absence of a neocortex does not appear to preclude an organism from experiencing affective states. Convergent evidence indicates that non-human animals have the neuroanatomical, neurochemical, and neurophysiological substrates of conscious states along with the capacity to exhibit intentional behaviors. Consequently, the weight of evidence indicates that humans are not unique in possessing the neurological substrates that generate consciousness. Non-human animals, including all mammals and birds, and many other creatures, including octopuses, also possess these neurological substrates."*

Low, P., Panksepp, J., Reiss, D., Edelman, D., Van Swinderen, B. and Koch, C., 2012, July. The Cambridge declaration on consciousness. In *Francis crick memorial conference, Cambridge, England*.

Consciousness demystified (~described) I.

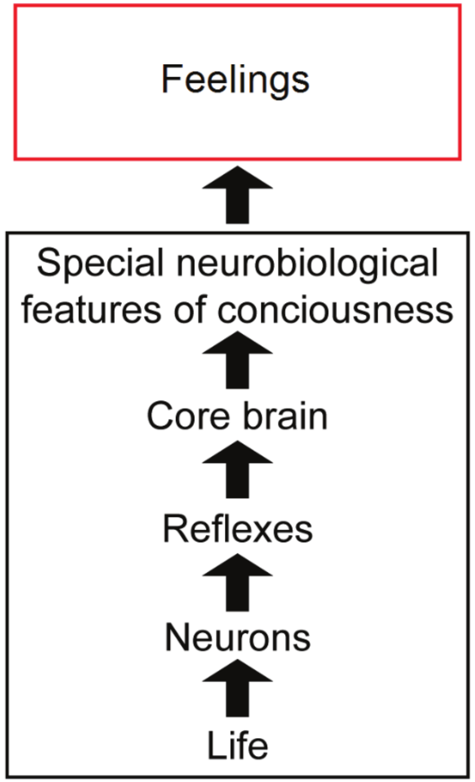


FIGURE 1 | Overview of our theory of consciousness. The sequence from physical life through conscious feelings. Life and the special features of consciousness are emphasized most. The images are reproduced with the

TABLE 2 | The special neurobiological features of consciousness (mostly after Feinberg and Mallatt, 2018a).

Neural complexity (more than in a simple, core brain)
- Brain with many neurons (>100,000?)
- Many subtypes of neurons

Elaborated sensory organs
- Eyes, receptors for touch, taste, smell

Neural hierarchies with neuron-neuron interactions
- Extensive reciprocal communication in and between pathways for the different senses
- Brain's neural computing modules and networks are distributed but integrated, leading to local functional isolation plus global coherence (Nunez, 2016; Mogensen and Overgaard, 2017)
- Synchronized communication by brain-wave oscillations
- The higher levels allow the complex processing and unity of consciousness
- Hierarchies that let consciousness predict events a fraction of a second in advance

Pathways that create mapped mental images or affective states
- Neurons are arranged in topographic maps of the outside world and body structures
- Valence coding of good and bad, for affective states
- Feed into pre-motor brain regions to motivate, choose, and guide movements in space

Brain mechanisms for selective attention and arousal

Memory, short-term or longer

TABLE 3 | Some adaptive roles of consciousness (mostly after Feinberg and Mallatt, 2018a).

Organizes large amounts of sensory input into a set of phenomenal properties for action choice

Its unified simulation of the sensed world directs behavior in this world

It ranks sensed stimuli by importance, by assigning affects to them, making decisions easier

Allows flexible behavior because it sets up many different behavioral choices

Allows easily adjustable behavior because it predicts the consequences of one's actions into the immediate future (Perry and Chittka, 2019; Solms, 2019)

Deals well with new situations, to meet the changing challenges of complex environments

TABLE 4 | Ways in which consciousness is diverse.

A. Three subtypes
- Exteroceptive
- Affective
- Interoceptive

B. Brain regions (mammal example)
- Cerebral cortex (mapped images)
- Subcortical (affects)

C. The coding varies: mapped representations of space versus valence-coding

D. Hubs and nodes are widely distributed within brains

E. In different animal groups
- Vertebrates
- Arthropods
- Cephalopods

Feinberg, T.E. and Mallatt, J.M., 2018. *Consciousness demystified*. MIT Press.

Consciousness demystified (~described) II.

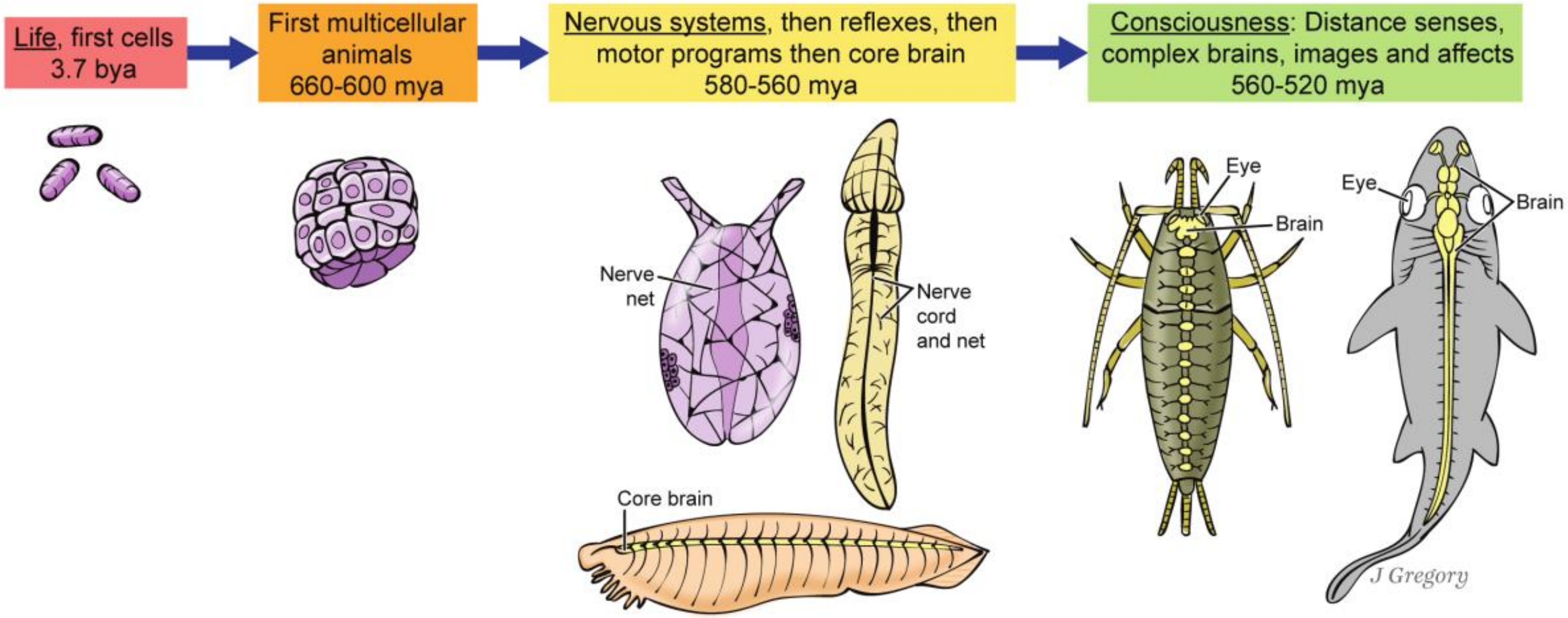


FIGURE 3 | Proposed stages in the evolution of consciousness, as an uninterrupted sequence. The three animals at the 'Nervous system' stage at center are hypothetical, but based roughly on a sea anemone, a hemichordate acorn worm, and the fish-like invertebrate, amphioxus. The two animals at far right are a bristletail insect and a shark. (From *Consciousness Demystified*, MIT, 2018. The images are reproduced with the permission of the copyright holder Mount Sinai Health System.)

Feinberg, T.E. and Mallatt, J.M., 2018. *Consciousness demystified*. MIT Press.

Consciousness demystified (~described) III.

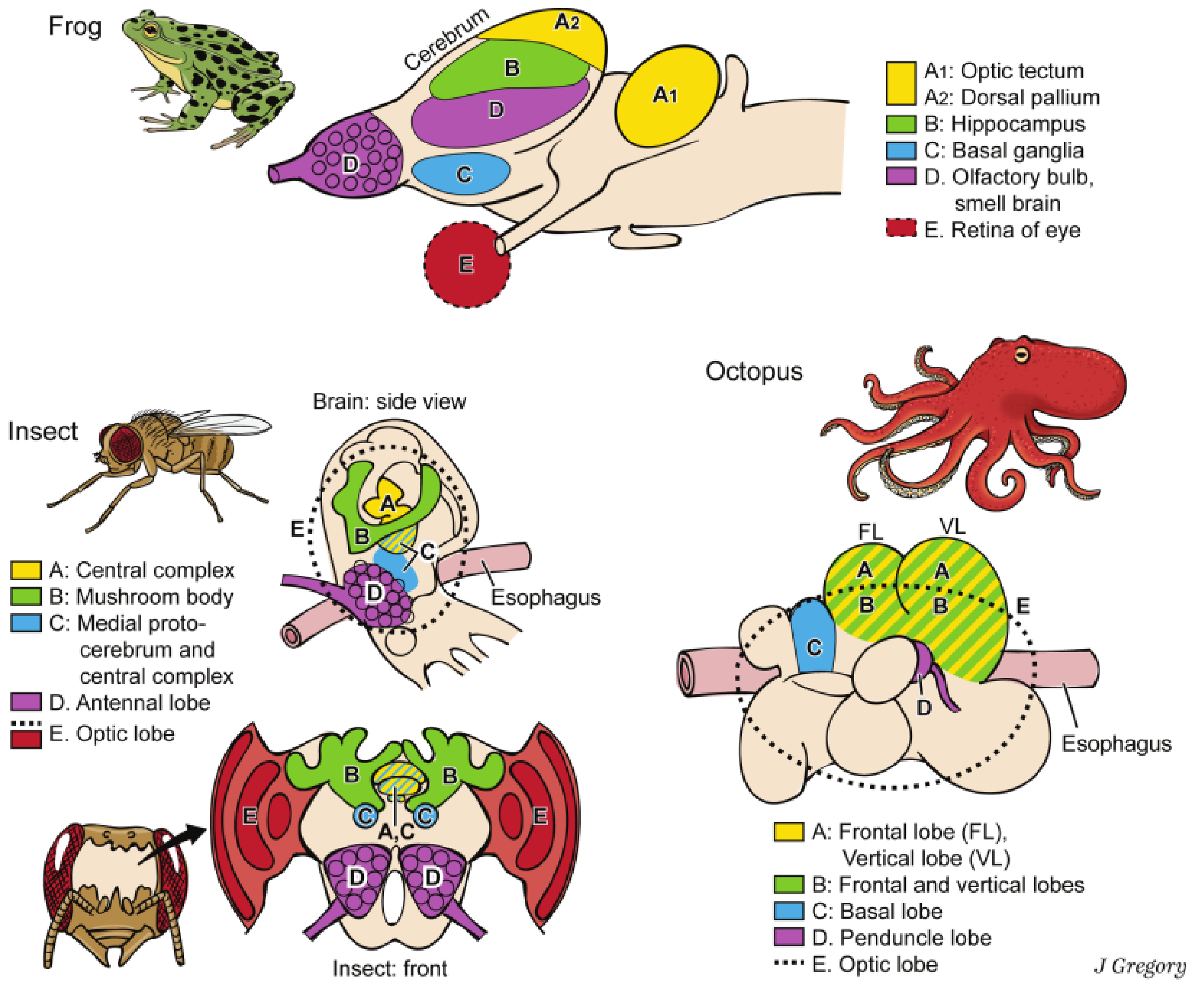


FIGURE 4 | Dissimilar brains of three different taxa of animals with consciousness. The areas with similar functions are colored the same in the different brains. The general code is: A, image-based consciousness; B, memory; C, pre-motor center; D, smell processing; and E, visual processing. (From *Consciousness Demystified*, MIT, 2018. The images are reproduced with the permission of the copyright holder Mount Sinai Health System.)

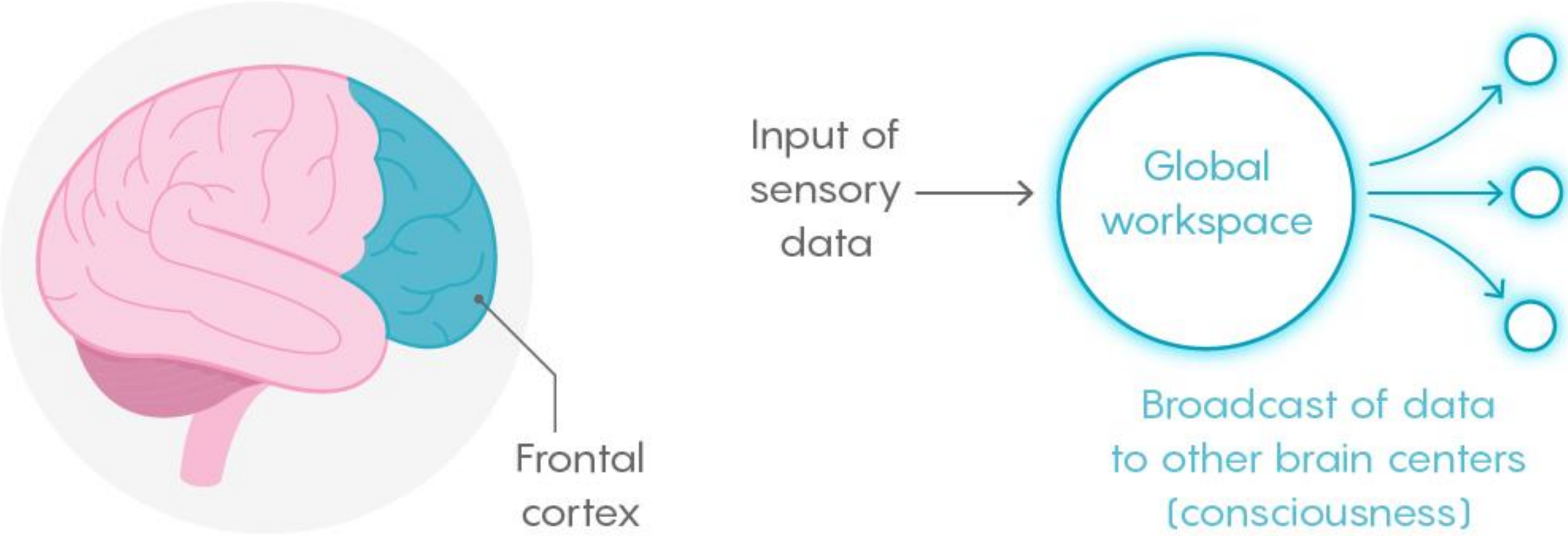
Feinberg, T.E. and Mallatt, J.M., 2018. *Consciousness demystified*. MIT Press.

Theories of consciousness

1. Global Workspace Theory
3. 1st and Higher Order Representation Theory
2.Phenomenal and Access Consciousness Theory
3. Information Integration Theory
4. Consciousness State Space Theory
5a, Holonomic brain theory – Karl Pribram and David Bohm
5b, Orch-OR theory – Stuart Hammeroff and Roger Penrose
[+6, Resonance theories]

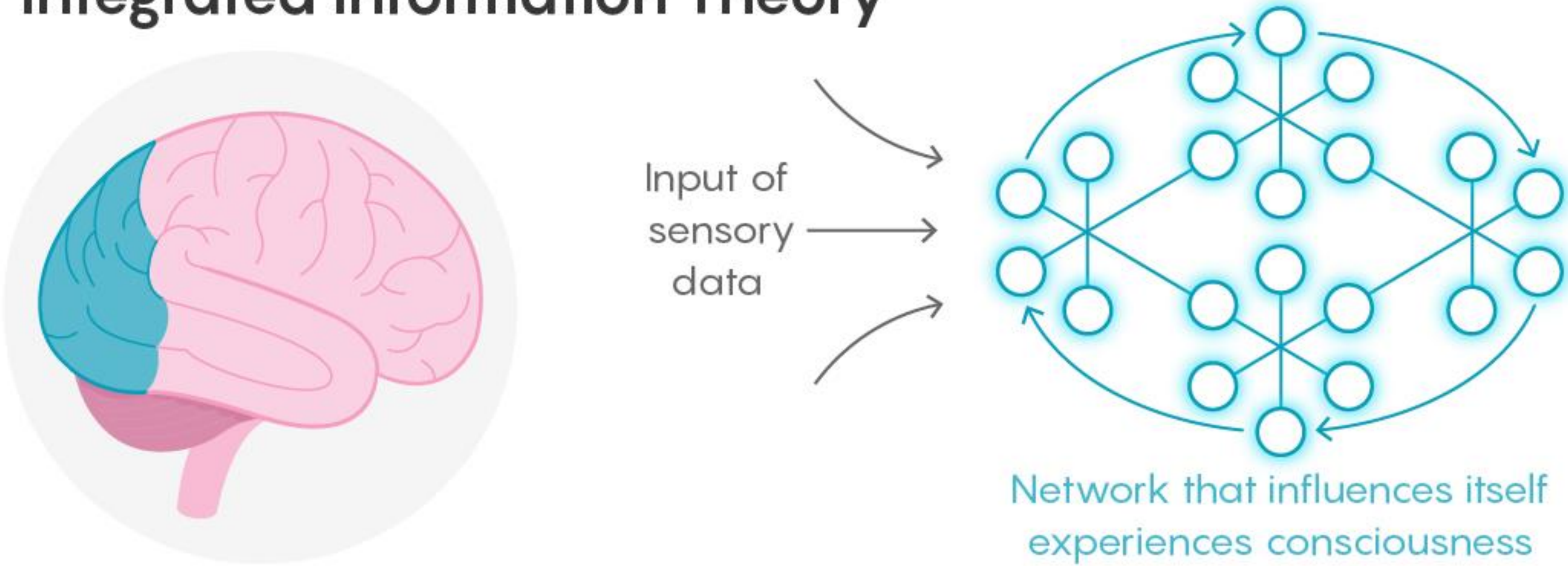
Joseph Ivin Thomas. “Current Status of Consciousness Research from the Neuroscience Perspective”. Acta Scientific Neurology 2.1 (2019)

Global Workspace Theory



According to one theory, consciousness is a form of information processing. It occurs when sensory data for an experience go to a "global workspace" and are distributed to other centers. The architecture for this process in the brain may be in the frontal cortex .

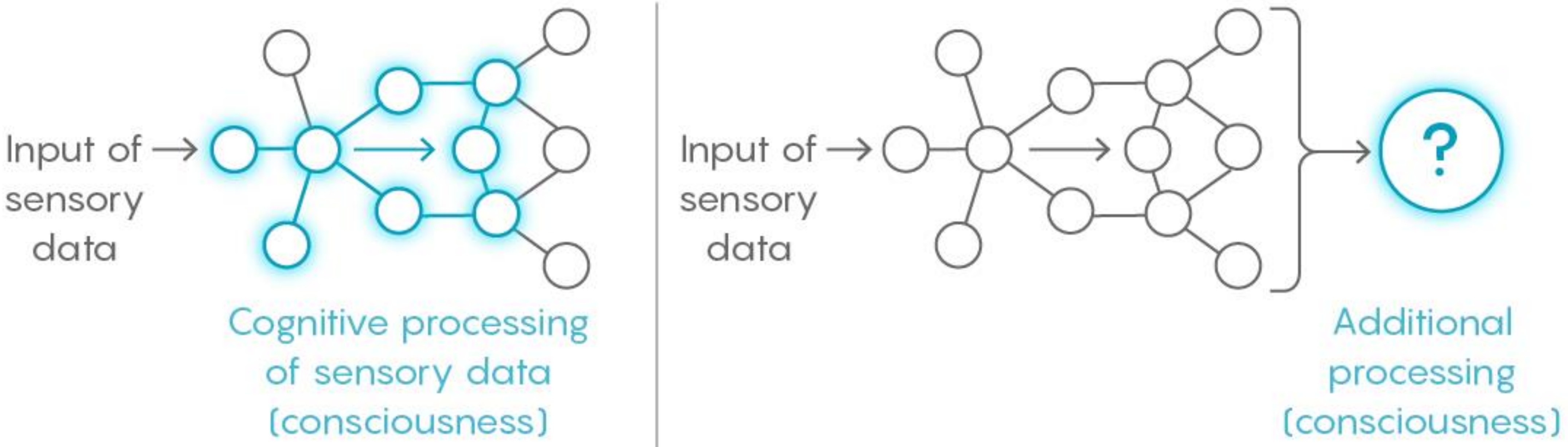
Integrated Information Theory



The integrated information theory argues that consciousness is intrinsic to cognitive networks that exert a “causal power” on themselves.

The back of the brain might have the right architecture for this capacity.

Two Other Classes of Consciousness Theories



First-Order Theories | **Higher-Order Theories**

First-order theories maintain that consciousness is simply a product of the cognitive processing of sensory information. Higher-order theories posit that consciousness involves something done to build on that cognitive representation of the sensory experience.

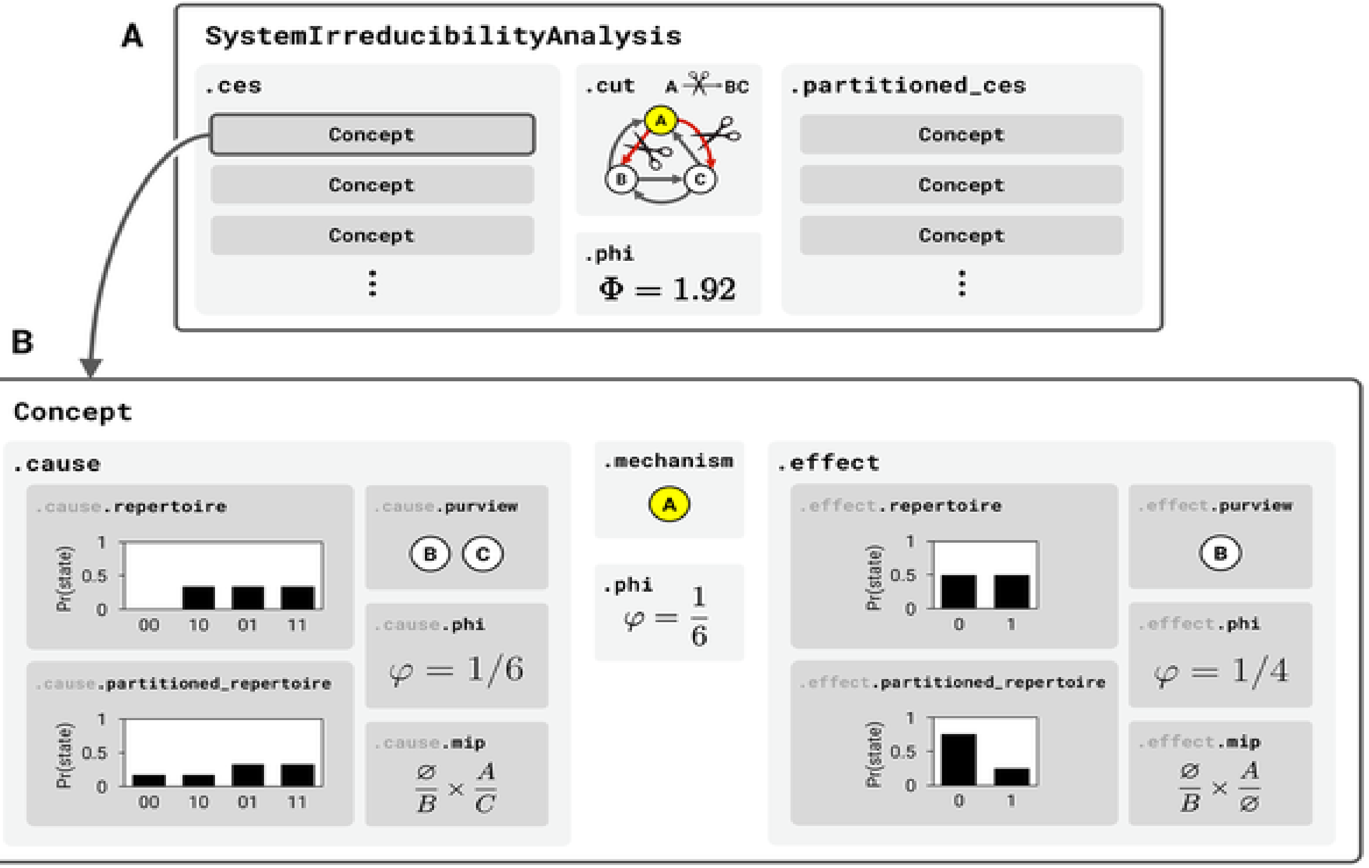
Experiment: Information processing" vs. causal power

The first two contenders are the global workspace theory (GWT), championed by Stanislas Dehaene of the Collège de France in Paris, and the integrated information theory (IIT), proposed by Giulio Tononi of the University of Wisconsin in Madison. The GWT says the brain's prefrontal cortex, which controls higher order cognitive processes like decision-making, acts as a central computer that collects and prioritizes information from sensory input. It then broadcasts the information to other parts of the brain that carry out tasks. Dehaene thinks this selection process is what we perceive as consciousness. By contrast, the IIT proposes that consciousness arises from the interconnectedness of brain networks. The more neurons interact with one another, the more a being feels conscious—even without sensory input. IIT proponents suspect this process occurs in the back of the brain, where neurons connect in a gridlike structure.

To test the schemes, six labs will run experiments with a total of more than 500 participants, costing the foundation $5 million. The labs, in the United States, Germany, the United Kingdom, and China, will use three techniques to record brain activity as volunteers perform consciousness-related tasks: functional magnetic resonance imaging, electroencephalography, and electrocorticography (a form of EEG done during brain surgery, in which electrodes are placed directly on the brain). In one experiment, researchers will measure the brain's response when a person becomes aware of an image. The GWT predicts the front of the brain will suddenly become active, whereas the IIT says the back of the brain will be consistently active.

https://www.sciencemag.org/news/2019/10/outlandish-competition-seeks-brain-s-source-consciousness

PyPhi: A toolbox for integrated information theory



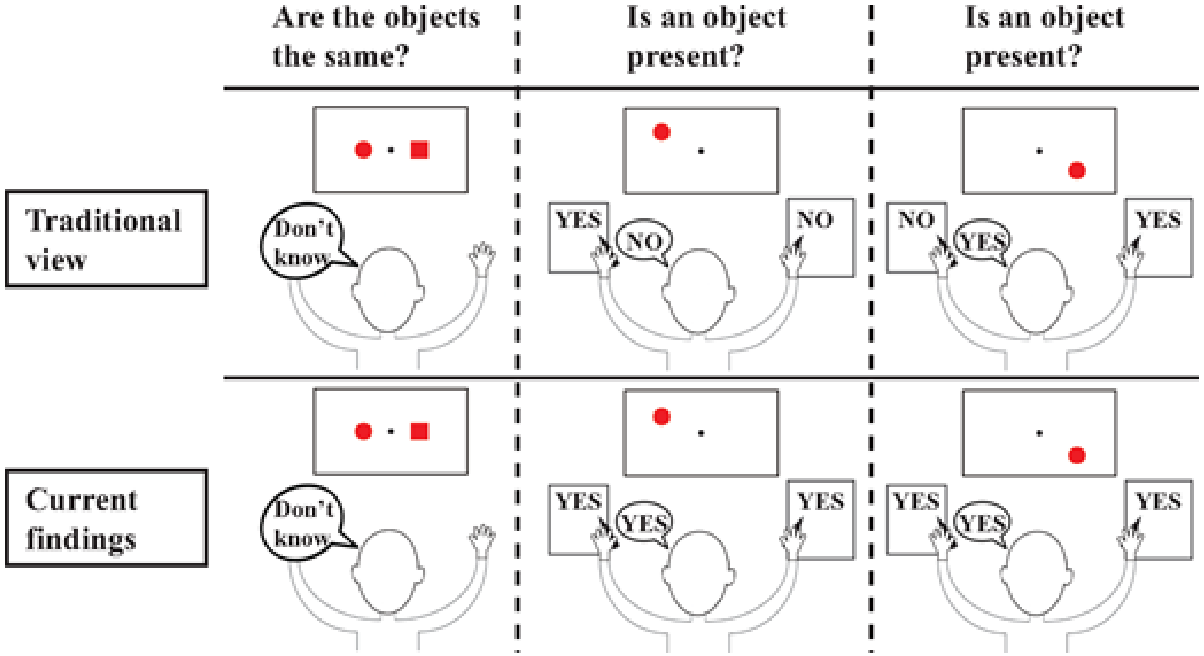
Calculating Φ Illustration of the algorithm:
https://journals.plos.org/ploscompbiol/article/file?type=supplementary&id=info:doi/10.1371/journal.pcbi.1006343.s001

Mayner WGP, Marshall W, Albantakis L, Findlay G, Marchman R, et al. (2018) PyPhi: A toolbox for integrated information theory. PLOS Computational Biology 14(7): e1006343. https://doi.org/10.1371/journal.pcbi.1006343

Critiques of IIT

Scott Aaronson: Why I Am Not An Integrated Information Theorist, 2014
John Horgan: Can Integrated Information Theory Explain Consciousness?, 2015
Cerullo, M.A., 2015. The problem with phi: a critique of integrated information theory. *PLoS computational biology*, *11*(9), p.e1004286.
Complexity: Ladyman, J., Lambert, J. and Wiesner, K., 2013. What is a complex system?. *European Journal for Philosophy of Science*, *3*(1), pp.33-67.
Mediano, P., Seth, A. and Barrett, A., 2019. Measuring integrated information: Comparison of candidate measures in theory and simulation. *Entropy*, *21*(1), p.17.
Weak vs strong emergence: Turkheimer, F.E. et al., 2019. Conflicting emergences. Weak vs. strong emergence for the modelling of brain function. *Neuroscience & Biobehavioral Reviews*, *99*, pp.3-10.

Split brain: divided perception but undivided consciousness

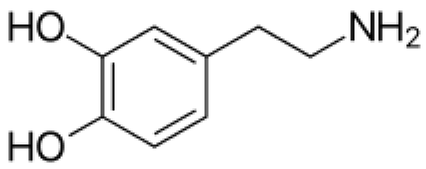


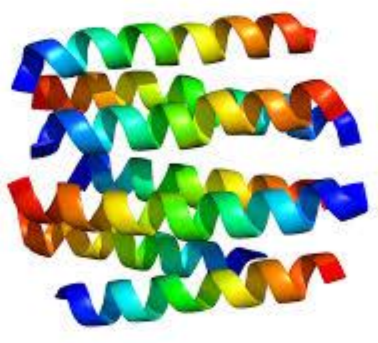
….

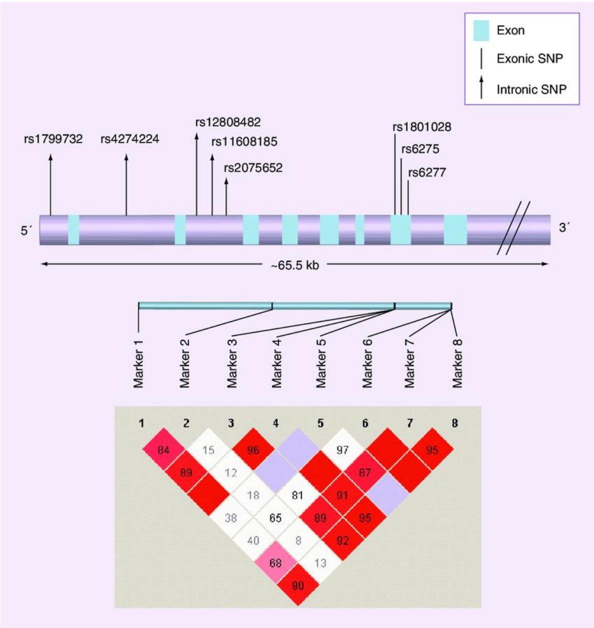
- In conclusion, with two patients, and across a wide variety of tasks we have shown that severing the cortical connections between the two hemispheres does not seem to lead to two independent conscious agents within one brain. Instead, we observed that patients without a corpus callosum were able to respond accurately to stimuli appearing anywhere in the visual field, regardless of whether they responded verbally, with the left or the right hand—despite not being able to compare stimuli between visual half-fields, and despite finding separate levels of performance in each visual half-field for labelling or matching stimuli. This raises the intriguing possibility that even without massive communication between the cerebral hemispheres, and thus increased modularity, unity in consciousness and responding is largely preserved.
- This preserved unity of consciousness may be especially challenging for the two currently most dominant theories of consciousness, the Global Workspace theory (Baars, 1988, 2005; Dehaene and Naccache, 2001) and the Integration Information theory (Tononi, 2004, 2005; Tononi and Koch, 2015).

Pinto, Y., Neville, D.A., Otten, M., Corballis, P.M., Lamme, V.A., De Haan, E.H., Foschi, N. and Fabri, M., 2017. Split brain: divided perception but undivided consciousness. *Brain*, *140*(5), pp.1231-1237.

Reward - deficiency







Addictions (free will???):

- substance and behavioral addictions
- new forms of behavioral addictions: computer games, internet pornography, social networking, sport exercises...

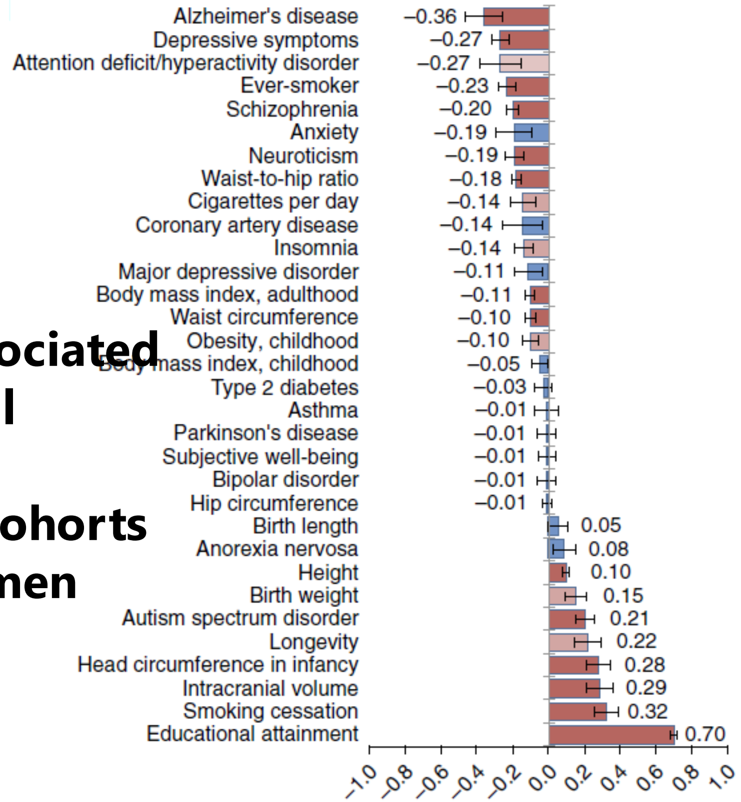
Blum, K., Cull, J.G., Braverman, E.R. and Comings, D.E., 1996. **Reward deficiency syndrome**. *American Scientist*, *84*(2), pp.132-145.
Blum, K., Chen, A.L., Giordano, J., Borsten, J., Chen, T.J., Hauser, M., Simpatico, T., Femino, J., Braverman, E.R. and Barh, D., 2012. **The addictive brain: all roads lead to dopamine**. *Journal of psychoactive drugs*, *44*(2), pp.134-143.
Robbins, T.W. and Clark, L., 2015. **Behavioral addictions**. *Current opinion in neurobiology*, *30*, pp.66-72.
Zald, D.H. and Treadway, M.T., 2017. **Reward processing, neuroeconomics, and psychopathology**. *Annual review of clinical psychology*, *13*, pp.471-495.
Blum, K., Gold, M.S., Modestino, E.J., Elman, I., Baron, D., Badgaiyan, R.D. and Bowirrat, A., 2019. **Hypothesizing Major Depression as a Subset of Reward Deficiency Syndrome (RDS) Linked to Polymorphic Reward Genes**. In *Handbook of Behavioral Neuroscience* (Vol. 29, pp. 419-426). Elsevier.

Consciousness: embodied, factorized, endangered

Genetic factors and traits of intelligence

rsID	Annotation
rs2490272	*FOXO3* intronic
rs9320913	Intergenic
rs10236197	*PDE1C* intronic
rs2251499	Intergenic
rs36093924	*CYP2D7* ncRNA_intr
rs7646501	Intergenic
rs4728302	*EXOC4* intronic
rs10191758	*ARHGAP15* intronic
rs12744310	Intergenic
rs66495454	*NEGR1* upstream
rs113315451	*CSE1L* intronic
rs12928404	*ATXN2L* intronic
rs41352752	*MEF2C* intronic
rs13010010	*LINC01104* ncRNA_intr
rs16954078	*SKAP1* intronic
rs11138902	*APBA1* intronic
rs6746731	*ZNF638* intronic
rs6779302	Intergenic

Optimism is associated with exceptional longevity in 2 epidemiologic cohorts of men and women



Sniekers, Suzanne, et al. "Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence." *Nature Genetics* (2017).

Personality traits/psychological dimensions

- Openness to experience (*inventive/curious* vs. *consistent/cautious*)
- Conscientiousness (*efficient/organized* vs. *easy-going/careless*)
- Extraversion (*outgoing/energetic* vs. *solitary/reserved*)
- Agreeableness (*friendly/compassionate* vs. *challenging/detached*)
- Neuroticism (*sensitive/nervous* vs. *secure/confident*)

https://en.wikipedia.org/wiki/Big_Five_personality_traits

Optimism is associated with exceptional longevity in 2 epidemiologic cohorts of men and women

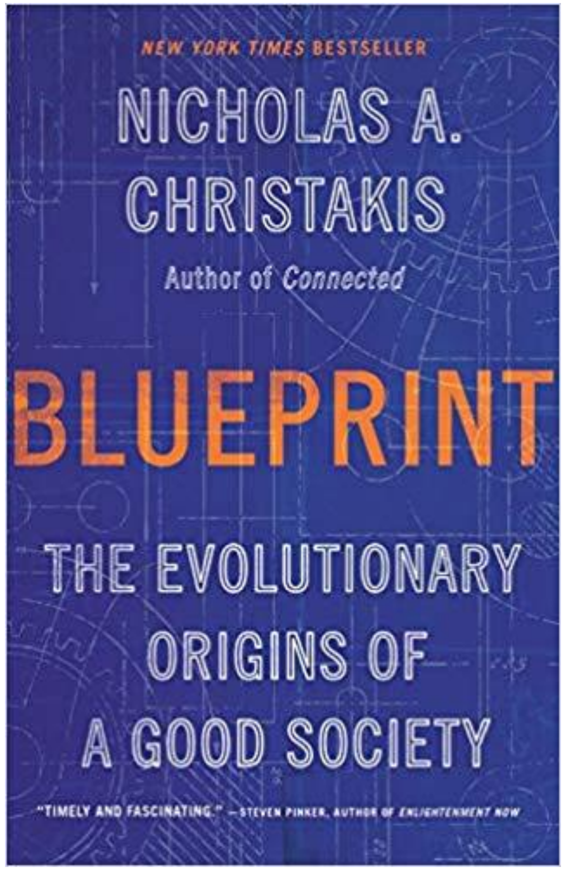
Most research on exceptional longevity has investigated biomedical factors associated with survival, but recent work suggests nonbiological factors are also important. Thus, we tested whether higher optimism was associated with longer life span and greater likelihood of exceptional longevity. Data are from 2 cohorts, women from the Nurses' Health Study (NHS) and men from the Veterans Affairs Normative Aging Study (NAS), with follow-up of 10 y (2004 to 2014) and 30 y (1986 to 2016), respectively. Optimism was assessed using the Life Orientation Test–Revised in NHS and the Revised Optimism–Pessimism Scale from the Minnesota Multiphasic Personality Inventory-2 in NAS. Exceptional longevity was defined as survival to age 85 or older. Primary analyses used accelerated failure time models to assess differences in life span associated with optimism; models adjusted for demographic confounders and health conditions, and subsequently considered the role of health behaviors. Further analyses used logistic regression to evaluate the likelihood of exceptional longevity. In both sexes, we found a dose-dependent association of higher optimism levels at baseline with increased longevity (*P* trend < 0.01). For example, adjusting for demographics and health conditions, women in the highest versus lowest optimism quartile had 14.9% (95% confidence interval, 11.9 to 18.0) longer life span. Findings were similar in men. Participants with highest versus lowest optimism levels had 1.5 (women) and 1.7 (men) greater odds of surviving to age 85; these relationships were maintained after adjusting for health behaviors. Given work indicating optimism is modifiable, these findings suggest optimism may provide a valuable target to test for strategies to promote longevity.

Significance

Optimism is a psychological attribute characterized as the general expectation that good things will happen, or the belief that the future will be favorable because one can control important outcomes. Previous studies reported that more optimistic individuals are less likely to suffer from chronic diseases and die prematurely. Our results further suggest that optimism is specifically related to 11 to 15% longer life span, on average, and to greater odds of achieving "exceptional longevity," that is, living to the age of 85 or beyond. These relations were independent of socioeconomic status, health conditions, depression, social integration, and health behaviors (e.g., smoking, diet, and alcohol use). Overall, findings suggest optimism may be an important psychosocial resource for extending life span in older adults.

Lee, L.O., James, P., Zevon, E.S., Kim, E.S., Trudel-Fitzgerald, C., Spiro, A., Grodstein, F. and Kubzansky, L.D., 2019. Optimism is associated with exceptional longevity in 2 epidemiologic cohorts of men and women. *Proceedings of the National Academy of Sciences*, *116*(37), pp.18357-18362.

Good societies/intelligent species



- Christakis, N.A., 2019. Blueprint: The evolutionary origins of a good society. *Hachette Book Group*.
- https://podbay.fm/podcast/1406534739/e/1559569524
- Benjamin, D.J., Cesarini, D., Van Der Loos, M.J., Dawes, C.T., Koellinger, P.D., Magnusson, P.K., Chabris, C.F., Conley, D., Laibson, D., Johannesson, M. and Visscher, P.M., 2012. The genetic architecture of economic and political preferences. *Proceedings of the National Academy of Sciences*, *109*(21), pp.8026-8031.
- Kleppestø, T.H., Czajkowski, N.O., Vassend, O., Røysamb, E., Eftedal, N.H., Sheehy-Skeffington, J., Kunst, J.R. and Thomsen, L., 2019. Correlations between social dominance orientation and political attitudes reflect common genetic underpinnings. *Proceedings of the National Academy of Sciences*, *116*(36), pp.17741-17746.

Dimensions of consciousness and the psychedelic state

It has often been suggested in the popular and academic literature that the psychedelic state qualifies as a higher state of consciousness relative to the state of normal waking awareness. This article subjects this proposal to critical scrutiny, focusing on the question of what it would mean for a state of consciousness to be 'higher'. We begin by considering the contrast between conscious contents and conscious global states. We then review the changes in conscious global state associated with psychedelic drug use, focusing on the effects of two serotonergic hallucinogens: psilocybin and lysergic acid diethylamide. Limiting our review to findings obtained from lab-based experiments and reported in peer-reviewed journals, we prioritize the more common and reliably induced effects obtained through subjective questionnaires and psychophysical measures. The findings are grouped into three broad categories (sensory perception, cognitive function, and experiences of unity) and demonstrate that although certain aspects of consciousness are improved or enhanced in the psychedelic state, many of the functional capacities that are associated with consciousness are seriously compromised. Psychedelic-induced states of consciousness are indeed remarkable in many ways, but it is inappropriate to regard them as 'higher' states of consciousness. The fact that psychedelics affect different aspects of consciousness in fundamentally different ways provides evidence against the unidimensional (or 'level-based') view of consciousness, and instead provides strong support for a multidimensional conception of conscious states. The final section of the article considers the implications of this analysis for two prominent theories of consciousness: the Global Workspace Theory and Integrated Information Theory.

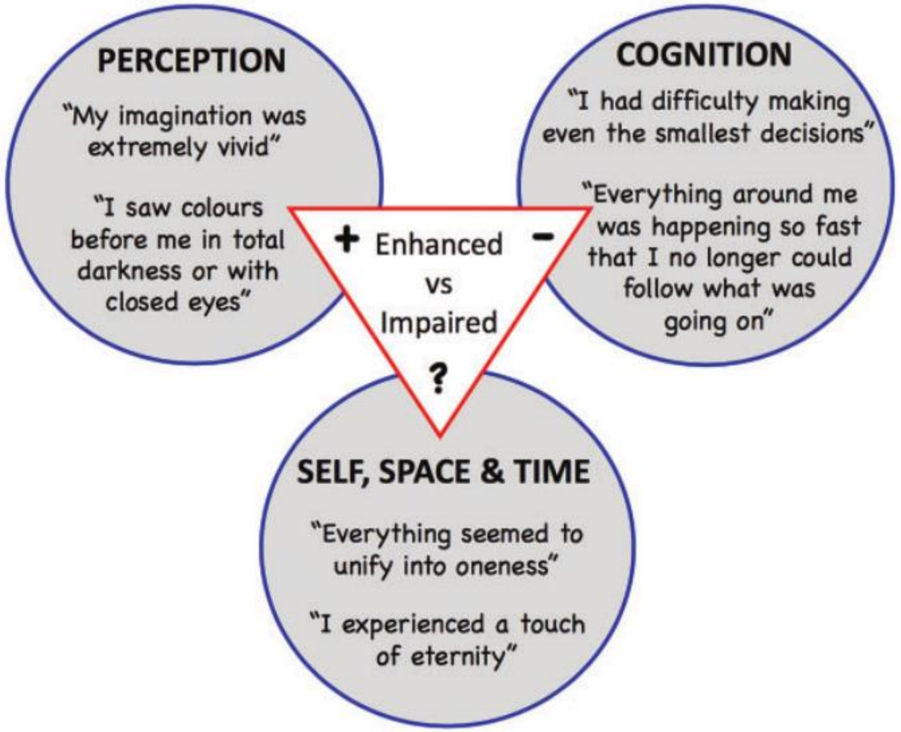


Figure 1. A schematic summary of the key aspects of consciousness discussed in this paper, which illustrates the inappropriateness of a unidimensional account of consciousness.

Bayne, T. and Carter, O., 2018. Dimensions of consciousness and the psychedelic state. *Neuroscience of consciousness*, *2018*(1), p.niy008.

https://en.wikipedia.org/wiki/Alexander_Shulgin

Psychedelics, Meditation, and Self-Consciousness

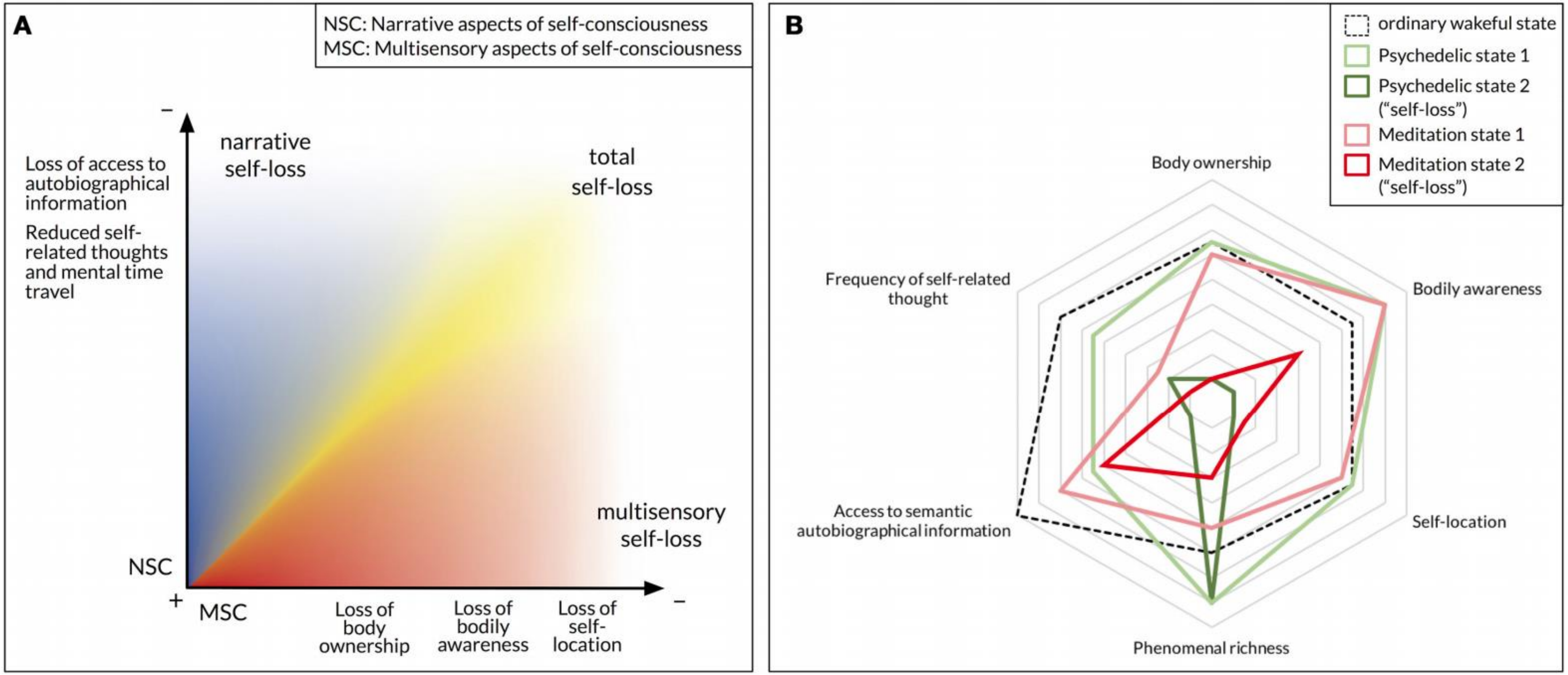


FIGURE 1 | Multidimensional models of self-loss in global states of consciousness. **(A)** A simplified two-dimensional **(B)** A tentative multidimensional model of self-loss.

Millière, R., Carhart-Harris, R.L., Roseman, L., Trautwein, F.M. and Berkovich-Ohana, A., 2018. Psychedelics, meditation, and self-consciousness. *Frontiers in psychology*, *9*.

Cognitive and neural foundations of religious belief

We propose an integrative cognitive neuroscience framework for understanding the cognitive and neural foundations of religious belief. Our analysis reveals 3 psychological dimensions of religious belief (God's perceived level of involvement, God's perceived emotion, and doctrinal/experiential religious knowledge), which functional MRI localizes within networks processing Theory of Mind regarding intent and emotion, abstract semantics, and imagery. Our results are unique in demonstrating that specific components of religious belief are mediated by well-known brain networks, and support contemporary psychological theories that ground religious belief within evolutionary adaptive cognitive functions.

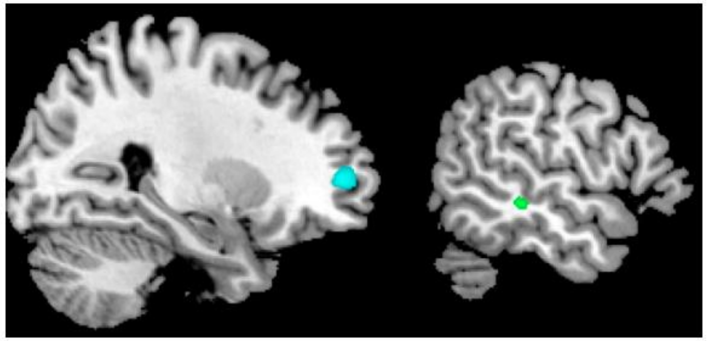


Fig. 2. Effect of God's perceived emotion (D2): Love (*Left*) vs. Anger (*Right*). Threshold was set to $P < 0.05$, FDR corrected. Activations are shown in blue for God's perceived love and green for God's perceived anger.

Kapogiannis, D., Barbey, A.K., Su, M., Zamboni, G., Krueger, F. and Grafman, J., 2009. Cognitive and neural foundations of religious belief. *Proceedings of the National Academy of Sciences*, *106*(12), pp.4876-4881.

Harris, S., Kaplan, J.T., Curiel, A., Bookheimer, S.Y., Iacoboni, M. and Cohen, M.S., 2009. The neural correlates of religious and nonreligious belief. *PLoS one*, *4*(10), p.e7272.

Mystical experience

https://samharris.org/podcasts/177-psychedelic-science/

- Roland Griffiths: main six categorical features:
  - Sense of unity/interconnectedness
  - Sense of preciousness/sacredness/respect deserving
  - Noetic (more real than any external experience)
  - Positive mood
  - Open heartedness
  - Transcendence of space and time
  - Ineffability

Griffiths, R.R., Richards, W.A., McCann, U. and Jesse, R., 2006. Psilocybin can occasion mystical-type experiences having substantial and sustained personal meaning and spiritual significance. *Psychopharmacology*, *187*(3), pp.268-283.

Strassman, R., 2000. *DMT: The spirit molecule: A doctor's revolutionary research into the biology of near-death and mystical experiences*. Simon and Schuster.

Killingsworth, M.A. and Gilbert, D.T., 2010. A wandering mind is an unhappy mind. *Science*, *330*(6006), pp.932-932.

Neuroscience and brain theory of meditation

Chapter 1: The neuroscience of meditation: classification, phenomenology, correlates, and mechanisms

Abstract

1 Introduction

2 Deconstructing mindfulness

3 Classifying meditation techniques

4 Phenomenology

5 Structural and functional correlates of meditation practices

6 Oscillatory correlates of meditation

7 Mechanisms underlying meditation and attention regulation

8 Mechanisms underlying meditation and emotion regulation

9 Future perspectives

Chapter 9: Toward a brain theory of meditation

Abstract

1 Introduction

2 Characterizing different forms of meditation and their relationships

3 Brain networks, brain states and meditation

4 Conscious access, global workspace and meditation

5 Outline of a brain theory of meditation (BTM)

Acknowledgments

Brandmeyer, T., Delorme, A. and Wahbeh, H., 2019. The neuroscience of meditation: classification, phenomenology, correlates, and mechanisms. *Progress in brain research*, *244*, pp.1-29.

Raffone, A., Marzetti, L., Del Gratta, C., Perrucci, M.G., Romani, G.L. and Pizzella, V., 2019. Toward a brain theory of meditation. In *Progress in brain research* (pp. 207-232).

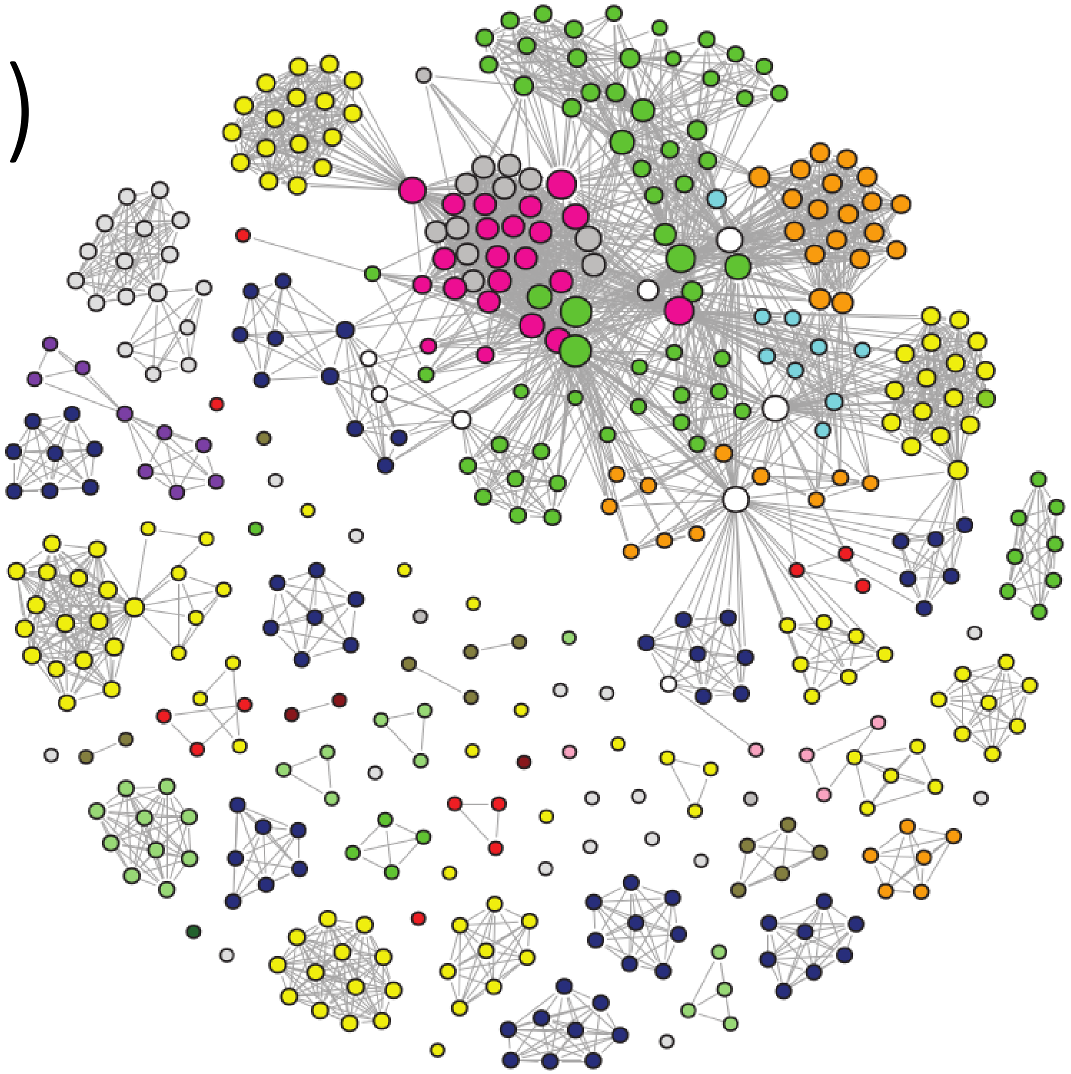
Marvin (1978)

TRILLIAN:
Marvin! It must be Marvin!
FORD:
The paranoid android!
ZAPHOD:
Space cookies! Oh, hand me the rap-rod, plate captain!
GARKBIT:
Pardon sir?
ZAPHOD:
Pass the phone, waiter. Gee you guys are so un-hip it's a wonder your bums don't fall off.
GARKBIT:
Our what sir?
[The telephone is set on the table]
The phone sir.
ZAPHOD:
Marvin! Hi, how you doing kid?
MARVIN:
***[On telephone]* I think you ought to know I'm feeling very depressed.**
ZAPHOD:
Hey, yeah? We're having a great time: food, wine, a little personal abuse, and the universe going foom. Where can we find you?
MARVIN:
[On telephone] You don't have to pretend to be interested in me you know. I know perfectly well I'm only a menial robot -
ZAPHOD:
yeah, okay, okay, but, uh, where are you?
MARVIN:
[On telephone] "Reverse primary thrust, Marvin", that's all they say to me. "Open airlock number three, Marvin." "Marvin, can you pick up that piece of paper?" Can I pick up that piece of paper! Here I am, brain the size of a planet…
ZAPHOD:
Yeah, yeah, uh
MARVIN:
[On telephone] 'm quite used to be humiliated. I can even go stick my head in a bucket of water if you'd like.
ZAPHOD:
yeah, uh Marvin?
MARVIN:
[On telephone] Would you like me to go and stick my head in a bucket of water? I've got one ready. Wait a minute.
FORD:
What's he saying, Zaphod?
ZAPHOD:
Oh, nothing. *[Audible over the telephone is the sound of Marvin sticking his head in a bucket of water]*

Marvin the Paranoid Android is a fictional character in *The Hitchhiker's Guide to the Galaxy* series by Douglas Adams. Marvin is the ship's robot aboard the starship *Heart of Gold*. Originally built as one of many failed prototypes of Sirius Cybernetics Corporation's *GPP* (Genuine People Personalities) technology, Marvin is afflicted with severe depression and boredom, in part because he has a "brain the size of a planet"[1] which he is seldom, if ever, given the chance to use. Indeed, the true horror of Marvin's existence is that no task he could be given would occupy even the tiniest fraction of his vast intellect.....[https://en.wikipedia.org/wiki/Marvin_the_Paranoid_Android]

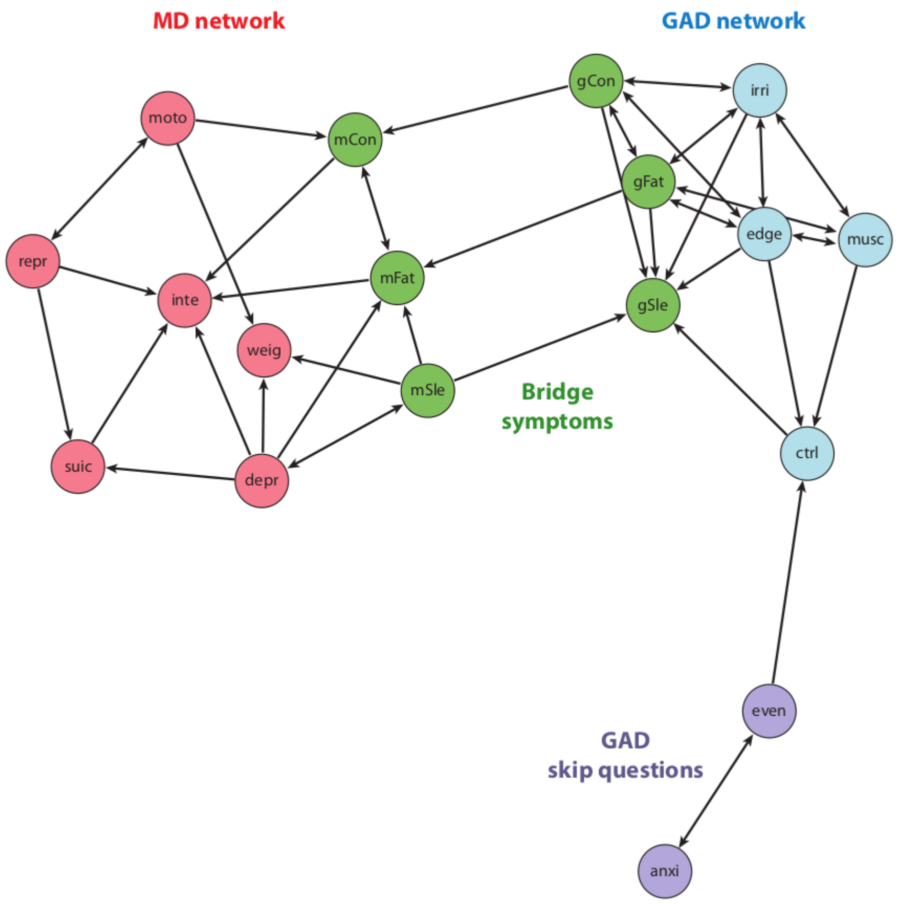
Psychiatric diseases
(DSM-IV symptom space)

- Disorders usually first diagnosed in infancy, childhood, or adolescence
- Delirium, dementia, and amnesia and other cognitive disorders
- Mental disorders due to a general medical condition
- Substance-related disorders
- Schizophrenia and other psychotic disorders
- Mood disorders
- Anxiety disorders
- Somatoform disorders
- Factitious disorders
- Dissociative disorders
- Sexual and gender identity disorders
- Eating disorders
- Sleep disorders
- Impulse control disorders not elsewhere classified
- Adjustment disorders
- Personality disorders
- Symptom is featured equally in multiple chapters



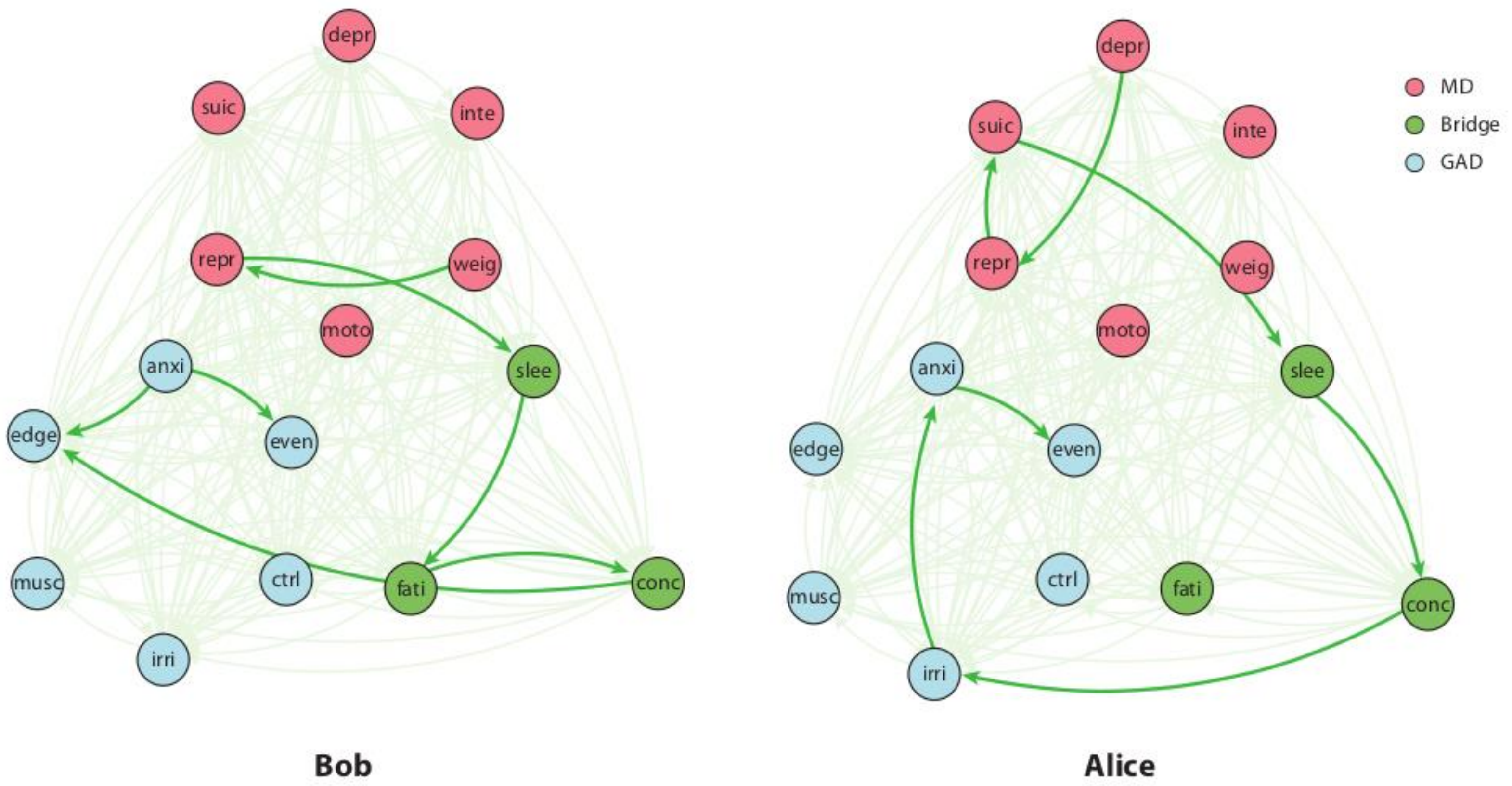
Borsboom, D. and Cramer, A.O., 2013. Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology*, *9*, pp.91-121.

Psychiatric diseases
(Major Depression – Generalized Anxiety Disorder)



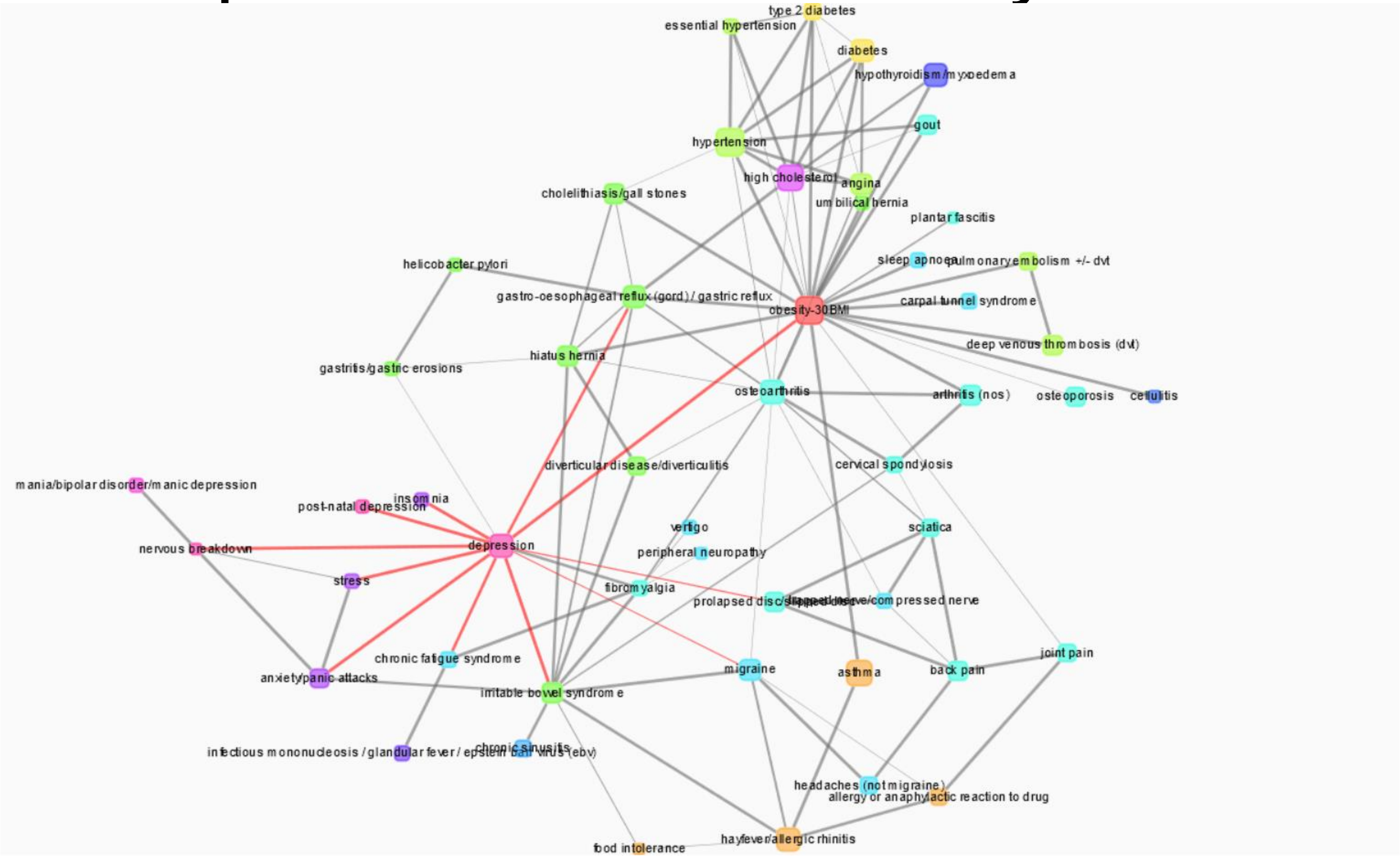
Borsboom, D. and Cramer, A.O., 2013. Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology*, *9*, pp.91-121.

Psychiatric diseases
(Personal hypothetitcal major depression networks)



Borsboom, D. and Cramer, A.O., 2013. Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology*, *9*, pp.91-121.

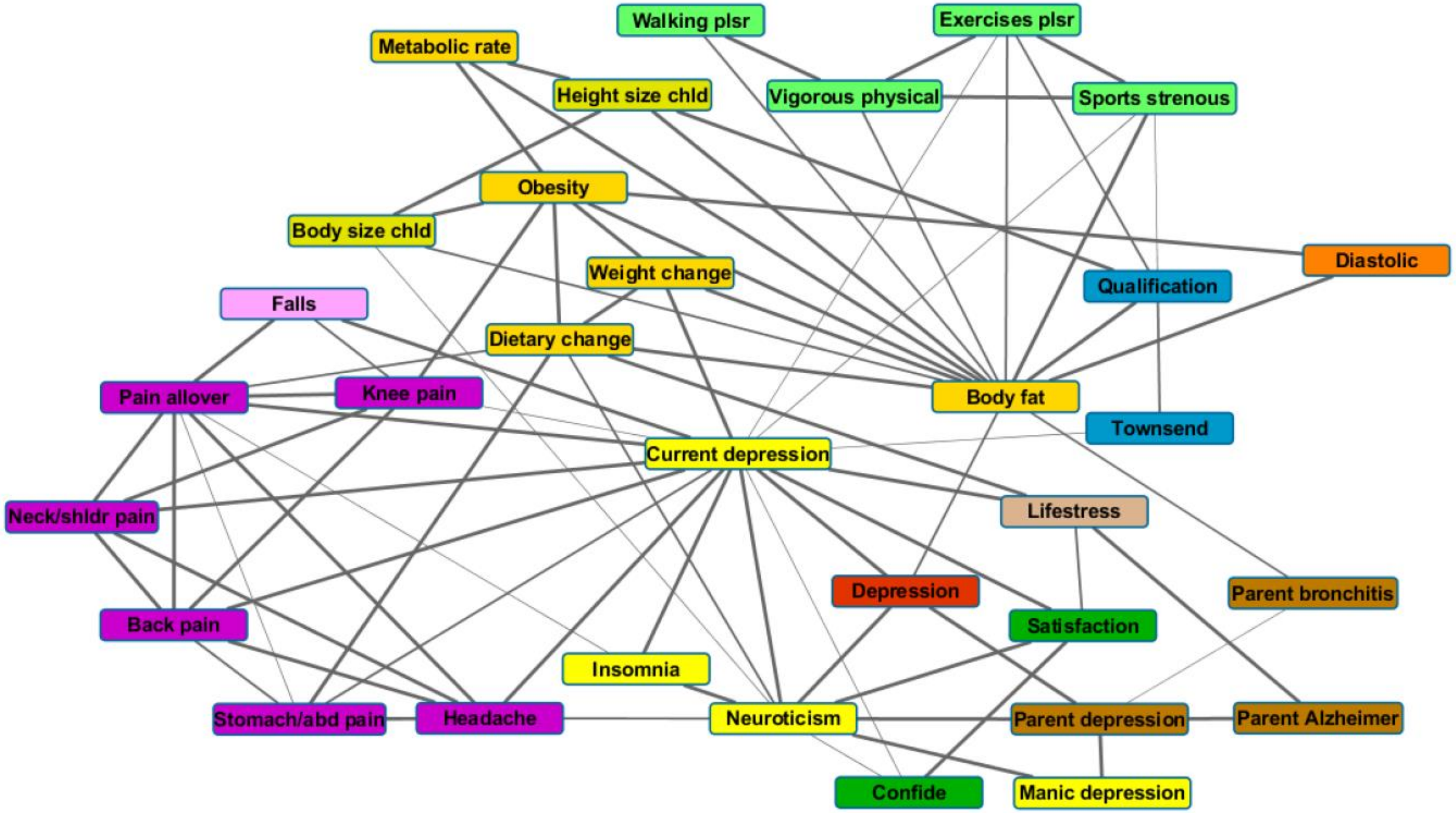
Depression multimorbidity cluster



http://bioinformatics.mit.bme.hu/UKBNetworks/full/index.html#/

Marx, P., Antal, P., Bolgar, B., Bagdy, G., Deakin, B. and Juhasz, G., 2017. Comorbidities in the diseasome are more apparent than real. *PLoS computational biology*, *13*(6), p.e1005487.

Envirome - life style - depression



Hullam, G., Antal, P., Petschner, P., Gonda, X., Bagdy, G., Deakin, B. and Juhasz, G., 2019. The UKB envirome of depression: from interactions to synergistic effects. *Scientific reports*, *9*(1), pp.1-19.

Interstellar: a benevolent(&sarcastic) TÁRS?

C: I think we can scratch our way to Edmunds' planet.

B:- What about fuel?

C:- Not enough. But I have a plan. We let Gargantua pull us down close to her horizon. Then a powered slingshot around, launching us towards Edmunds' planet.

B: - Manually?

C: - That's what I'm here for. I'm gonna take us just inside the critical orbit.

B:- What about time slippage?

C:- Neither of us has time to worry... about relativity right now, Dr. Brand.

B: I'm sorry, Cooper.

C: Once we've gathered enough speed around Gargantua... we use Lander 1... and Ranger 2 as rocket boosters to... push us out of the black hole's gravity. The Lender's linkages have been destroyed... so we'll have to control manually. Once Lander 1 is spent, TARS will detach... And get sucked into that black hole.

B: Why does TARS have to detach?

C: We have to shed the weight to escape the gravity.

TARS: Newton's third law. The only way humans have ever figured out of getting somewhere... is to leave something behind.

B: Cooper, you can't ask TARS to do this for us.

C: He's a robot. So you don't have to ask him to do anything.

B: - Cooper, you asshole!

C: - Sorry, you broke up a little bit there.

T: It's what we intended, Dr. Brand.

..

T:- See you on the other side, Coop.

Co:- See you there, Slick! Okay, CASE. Nice reckless flying!

Ca - Learned from the master.

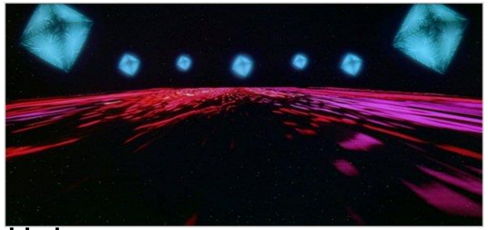
Co: - Ranger 2, prepare to detach.

B: - What? No! No! Cooper! What are you doing?

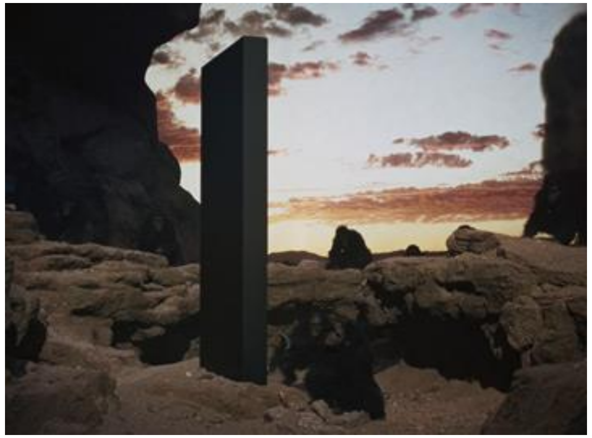
Co: Newton's third law. You got to leave something behind.

TARS

2001: A Space Odyssey (1968)
The value alignment problem



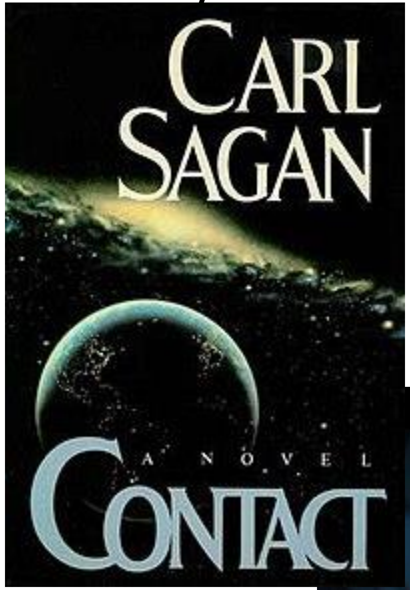
Bowm Open the pod door, Hal.
Bowm Open the pod bay doors please, Hal. Open the pod bay doors please, Hal. Hello, Hal, do you read me? Hello, Hal, do you read me? Do you read me, Hal? Do you read me, Hal? Hello, Hal, do you read me? Hello, Hal, do you read me? Do you read me, Hal?
Hal Affirmative, Dave. I read you.
Bowm Open the pod bay doors, Hal.
Hal I'm sorry, Dave. I'm afraid I can't do that.
Bowm What's the problem?
Hal I think you know what the problem is just as well as I do.
Bowm What are you talking about, Hal?
Hal This mission is too important for me to allow you to jeopardise it.
Bowm I don't know what you're talking about Hal.
Hal I know that you and Frank were planning to disconnect me. And I'm afraid that's something I cannot allow to happen.
Bowm Where the hell did you get that idea Hal?
Hal Dave! Although you took very thorough precautions in the pod against my hearing you I could see your lips move.
Bowm Alright Hal. I'll go in through the emergency airlock.
Hal Without your space helmet, Dave, you're going to find that rather difficult.
Bowm Hal I won't argue with you anymore. Open the doors.
Hal Dave, this conversation can serve no purpose anymore. Goodbye.



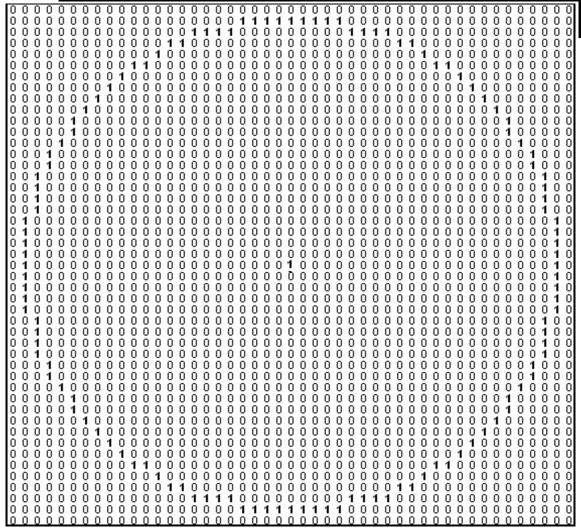
https://en.wikipedia.org/wiki/2001:_A_Space_Odyssey_(novel)
https://en.wikipedia.org/wiki/2001:_A_Space_Odyssey_(film)

http://miscel.wikidot.com/2001-transcript

Contact
(Is our world fundamentally a better place with techs/aliens/strong AIs?)



π



L: My guest tonight is author and theologian Palmer Joss, 'God's diplomat' according to the New York Times. His new book -- Losing Faith -- is currently number one on that publication's bestseller list. Thanks for being with us, Reverend. Okay -- who's losing faith -- and why?
J: Well, let me start this way, Larry. What has science done for you lately?
L: Besides letting me broadcast this program all over the world?
J: Besides that. Or better, I'll give you that, but tell me this: **Are you happier? Are we happier? Is our world fundamentally a better place?** Don't get me wrong -- we're smart, Larry. We shop at home, we surf the net... and we feel emptier and lonelier and more cut off from each other than at any other time in human history...
W: Y'know who'd make great astronomers? Vampires. Think about it; the perfect synthesis of career and lifestyle.
W: Got a bogey, boss?
E: I'm not sure. You mind checking right ascension 18 hours, 34 minutes; declination plus 38 degrees 41 minutes?
...
E: Numbers. Those are numbers, each pulse is a set -- break it down --
E: 79 -- 83 -- 91 -- they're all primes, no way that's a natural phenomenon -- !
...

https://en.wikipedia.org/wiki/Contact_(novel)
https://en.wikipedia.org/wiki/Contact_(1997_American_film)

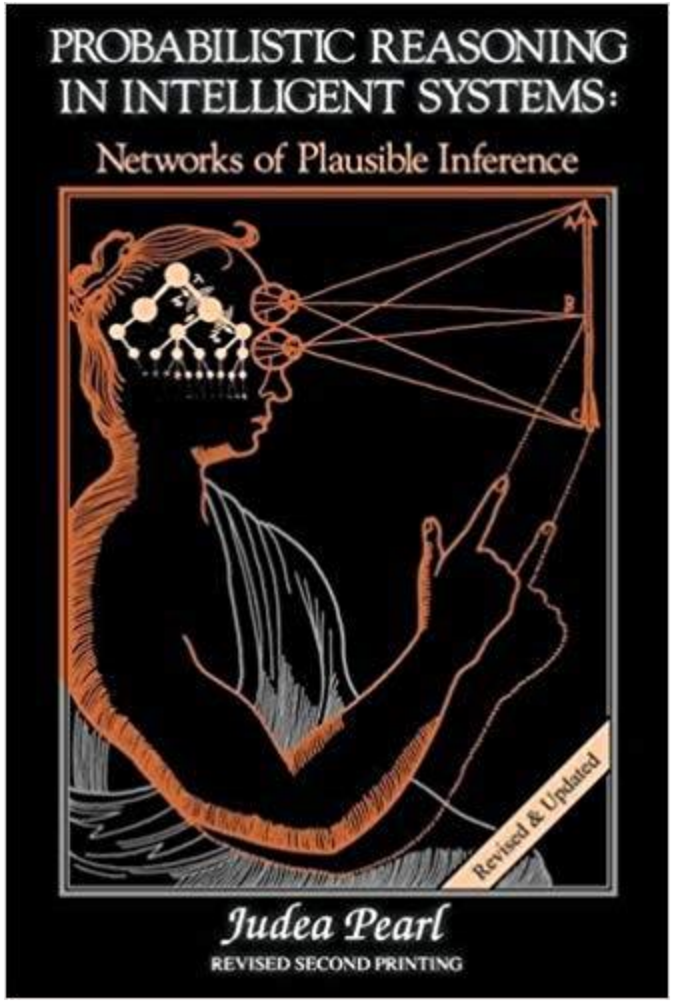
Understanding why

Judea Pearl

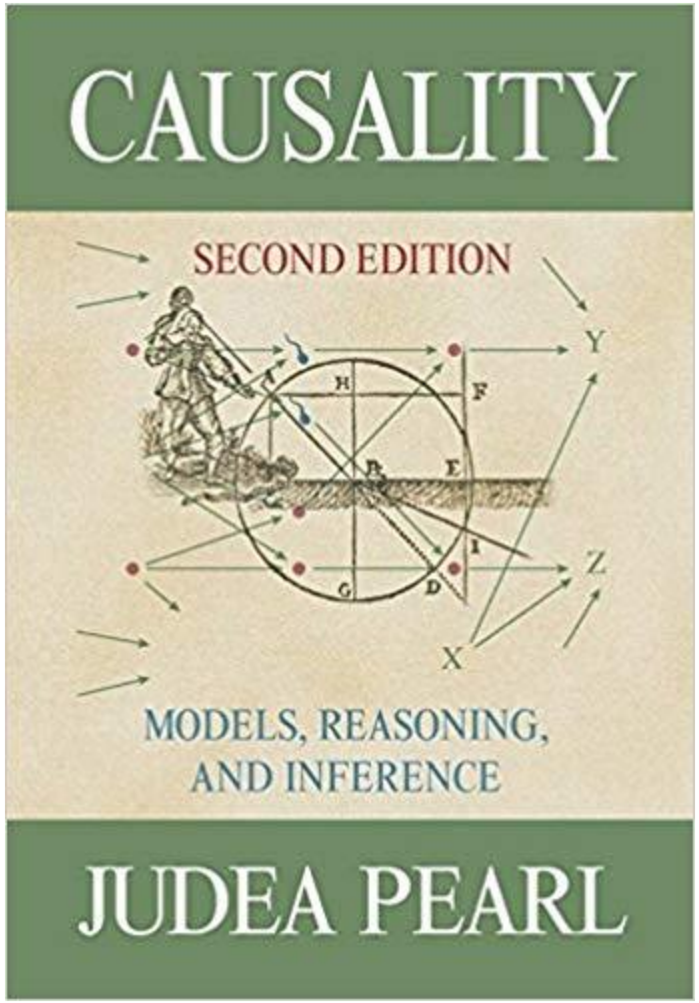
- The New Science of Cause and Effect
  - https://www.youtube.com/watch?v=ZaPV1OSEpHw
- The Mathematics of Causal Inference: with Reflections on Machine Learning
  - https://www.youtube.com/watch?v=bcRl7sXR1hE
- Making Sense with Sam Harris #164 - Cause & Effect with Judea Pearl
  - https://www.youtube.com/watch?v=NNDvhFbMD0s
- **==> Causal Reasoning, Counterfactuals, Bayesian Networks, and the Path to AGI**
  - https://lexfridman.com/judea-pearl/

03:18 – Descartes and analytic geometry
06:25 – Good way to teach math
07:10 – From math to engineering
09:14 – Does God play dice?
10:47 – Free will
11:59 – Probability
22:21 – Machine learning
23:13 – Causal Networks
27:48 – Intelligent systems that reason with causation
29:29 – Do(x) operator
36:57 – Counterfactuals
44:12 – Reasoning by Metaphor
51:15 – Machine learning and causal reasoning
53:28 – Temporal aspect of causation
56:21 – Machine learning (continued)
59:15 – Human-level artificial intelligence
1:04:08 – Consciousness
1:04:31 – Concerns about AGI
1:09:53 – Religion and robotics
1:12:07 – Daniel Pearl
1:19:09 – Advice for students
1:21:00 – Legacy

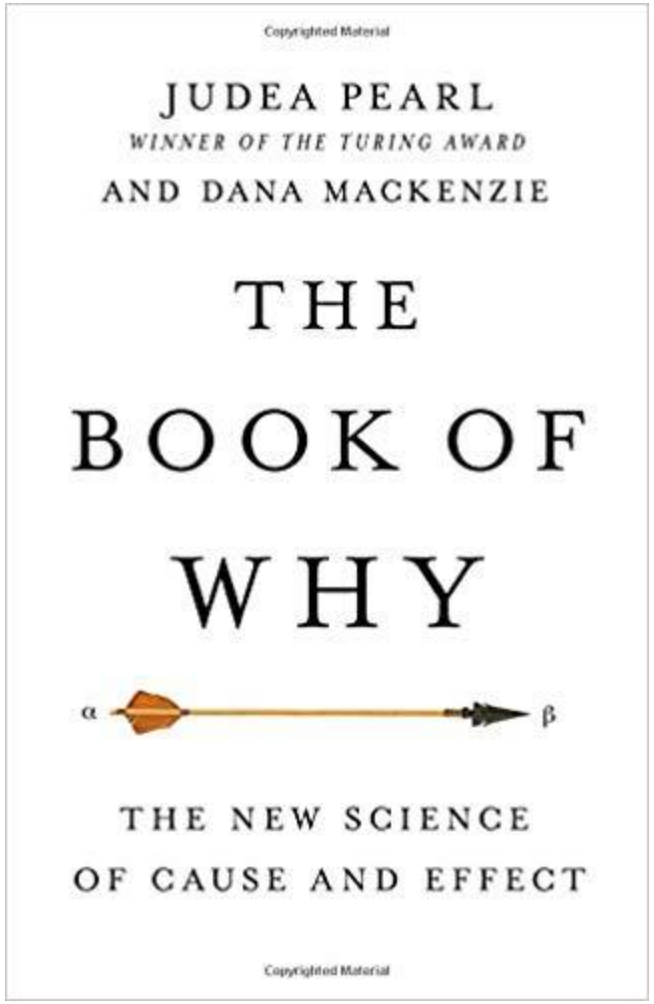
Causality



1988



2000



2018

"Summary": open questions

Strong AI –?– human-level/general intelligence
Strong AI –?– explainable AI (XAI)
Strong AI –?– existential risk (human ethics)
Strong AI –?– human-compatible intelligence (=level+XAI+risk)
Strong AI –?– superintelligence
Strong AI –?– (super)ethics (orthogonality hypothesis, value alignment problem==>back to line 1 :-)