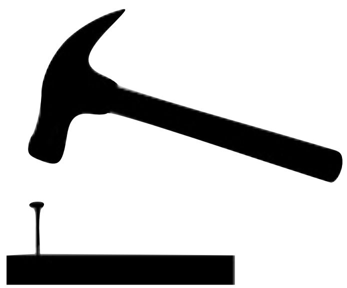
Kognitionspsychologie II: Session 10
Consciousness

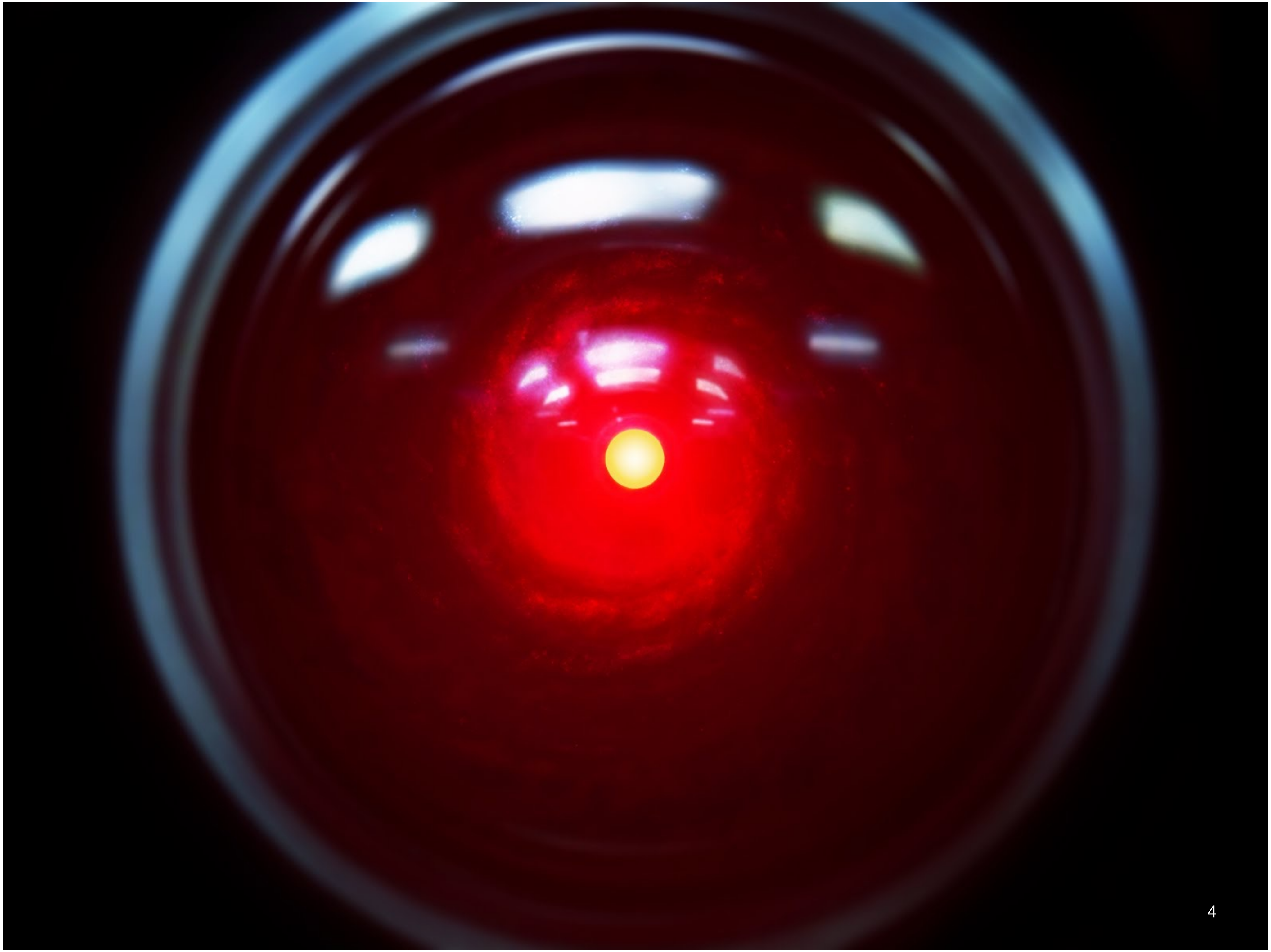
Rui Mata, FS 2022

Learning Objectives

- What is consciousness? Should we think of it as an epiphenomenon or a central (adaptive) characteristic of human cognition?
- Discuss the **adaptive significance** of consciousness by asking “can machines be conscious?”
- Learn about **comparative approaches** to consciousness and ask “are animals conscious?”
- Learn about **developmental patterns** of consciousness (as in self-awareness, autobiographical self)
- Learn about the **neural basis** of consciousness

Consciousness

Ontogeny	Mechanism
Phylogeny	 Adaptive Significance



Can machines have consciousness?

The chinese room argument

John Searle (1980) proposed the Chinese room argument to suggest that a digital computer executing a program cannot be shown to have “consciousness" regardless of how intelligently or human-like the program may make the computer behave (i.e., passes the Turing test - machine shows intelligent behaviour equivalent to, or indistinguishable from, that of a human). More information: https://plato.stanford.edu/entries/chinese-room/

Searle, J. R. (1980). Minds, brains, and programs. ***Behavioral and Brain Sciences***, ***3***(3), 417–424. http://doi.org/10.1017/S0140525X00005756

Can machines have consciousness?

Project debater: Smart but not conscious

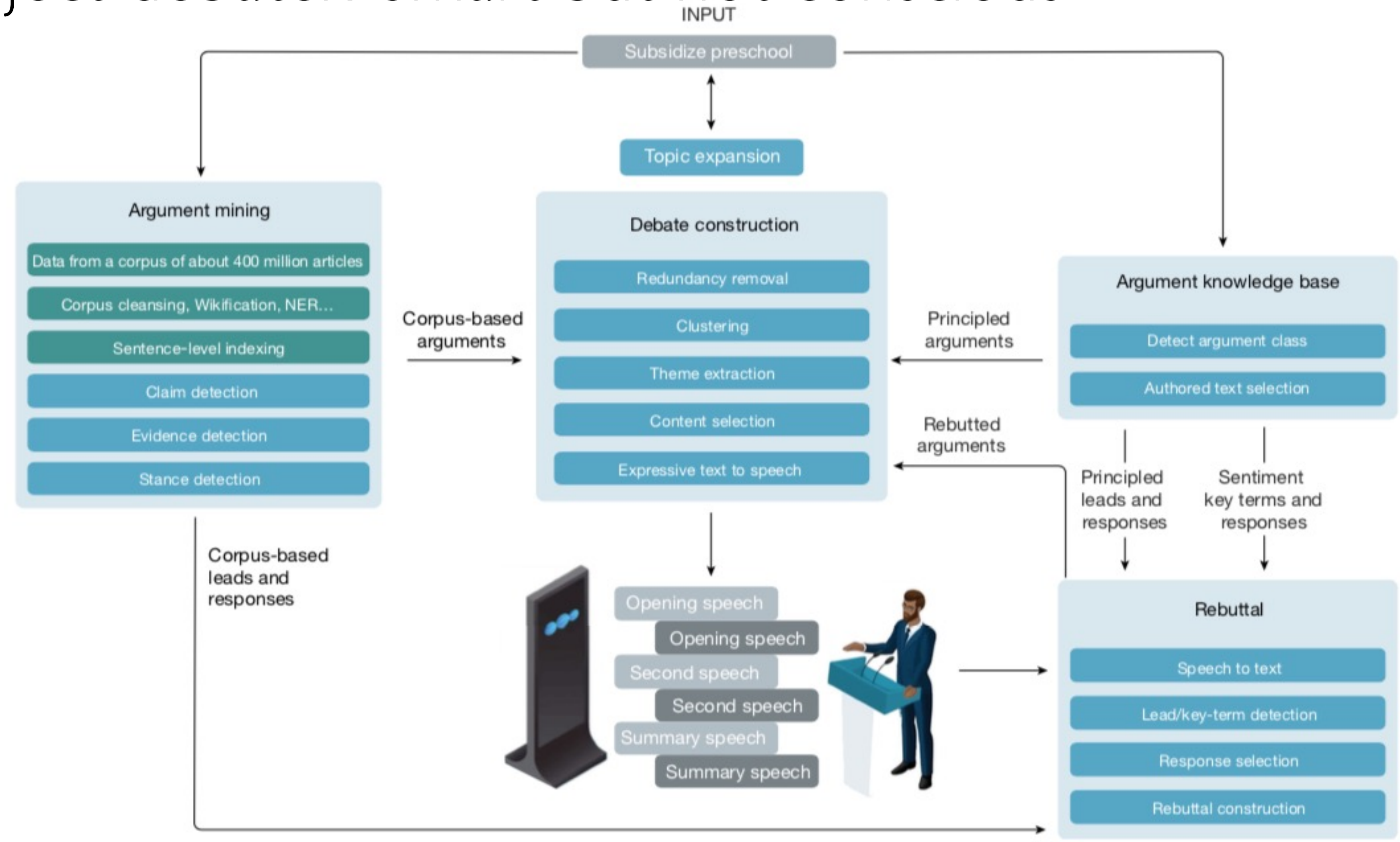


Fig. 2 | System architecture. Description of Project Debater components. Offline analysis is shaded in green; online analysis is shaded in blue. NER stands for named entity recognition.

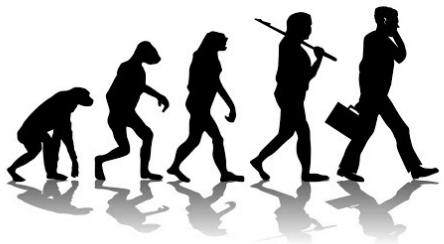
Slonim, N., Bilu, Y., Alzate, C., Bar-Haim, R., Bogin, B., Bonin, F., et al. (2021). An autonomous debating system. Nature, 591(7850), 379–384. http://doi.org/10.1038/s41586-021-03215-w

Can machines have consciousness?

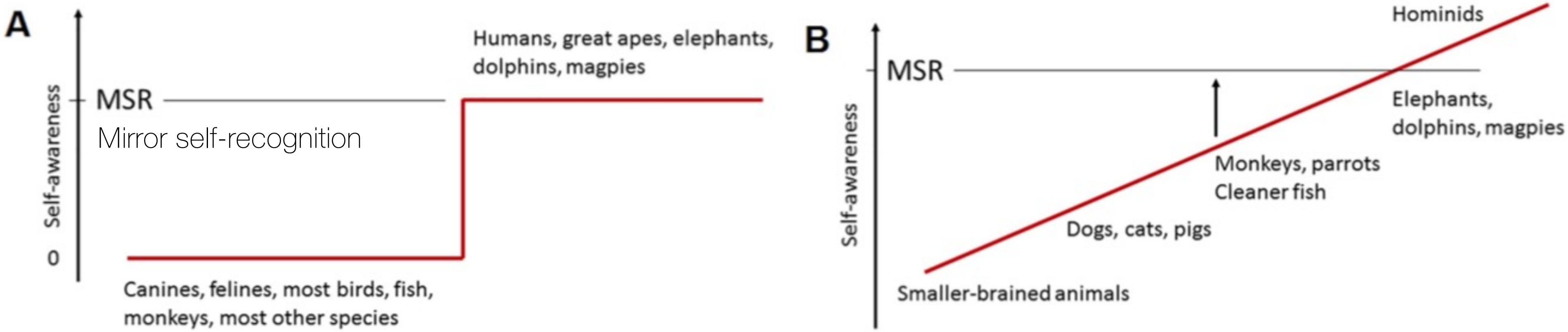
Type	Description	Example(s)
C0: Unconscious processing	Information processing can be realized by (mindless) automatons	face or speech recognition, priming, debating(!)
C1: Global availability	Selection of information for global broadcasting, making it robust, and available for computation and report	reportable aspects of sensory experience
C2: Self-monitoring	Self-monitoring of computations, leading to a subjective sense of certainty or error.	confidence, error-monitoring, knowledge of strategy efficacy

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? ***Science***, 358(6362), 486–492. http://doi.org/10.1126/science.aan8871

Consciousness

Ontogeny	Mechanism
 Phylogeny	Adaptive Significance

Do animals have consciousness?



The mirror mark test has encouraged a binary view of self-awareness according to which a few species possess this capacity and others do not (A). The gradualist view (B), in contrast, assigns the highest level of self-awareness to hominids, who spontaneously explore and play with their reflection and care about their appearance, and assigns intermediate or lower levels to other species, but no zero level because all animals need a self-concept. Reactions to mirrors range from permanent confusion about one's reflection to only brief or no confusion between one's reflection and a stranger. Some species, such as macaques, seem to possess this intermediate level and can therefore, with the aid of training, be "lifted" (arrow) to a level of mirror understanding closer to self-awareness.

de Waal, F. B. M. (2019). Fish, mirrors, and a gradualist perspective on self-awareness. *Plos Biology*, *17*(2), e3000112–8. http://doi.org/10.1371/journal.pbio.3000112

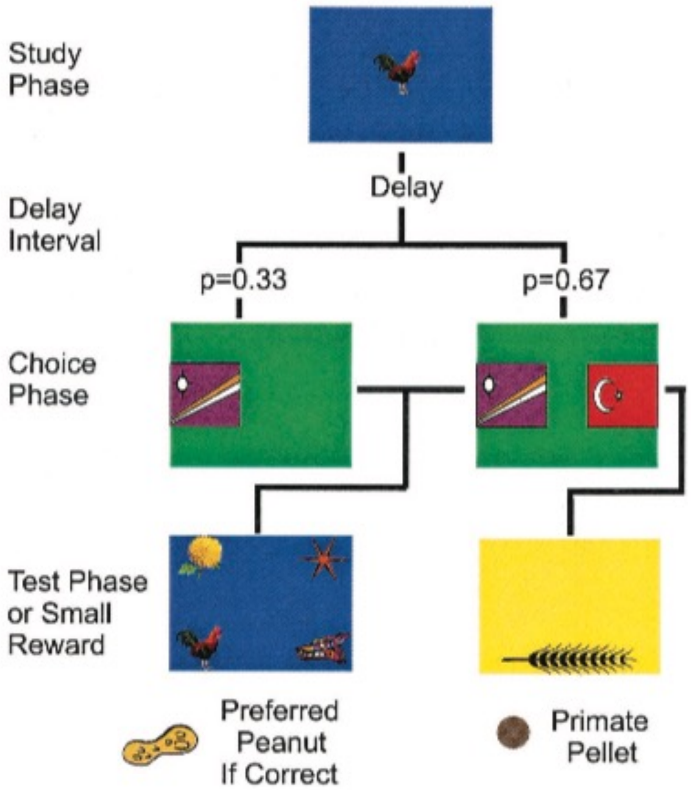
Do animals have consciousness?

This paper reviews evidence that increases the probability that many animals experience at least simple levels of consciousness. First, the search for neural correlates of consciousness has not found any consciousness- producing structure or process that is limited to human brains. Second, appropriate responses to novel challenges for which the animal has not been prepared by genetic programming or previous experience provide suggestive evidence of animal consciousness because such versatility is most effectively organized by conscious thinking.

Global availability (C1): Animal intelligence/creativity as evidence of global broadcasting necessary for making new associations?

Griffin, D. R., & Speck, G. B. (2004). New evidence of animal consciousness. *Animal Cognition*, *7*(1), 5–18. http://doi.org/10.1007/s10071-003-0203-x

Do animals have consciousness?



Method for assessing whether monkeys know when they remember. Each colored panel represents what monkeys saw on a touch-sensitive monitor at a given stage in a trial. At the start of each trial, monkeys studied a randomly selected image. A delay period followed over which monkeys often forgot the studied image. In two-thirds of trials, animals chose between taking a memory test (Right, left-hand stimulus) and declining the test (Right, right-hand stimulus). In one-third of trials, monkeys were forced to take the test (Left). Better accuracy on chosen than on forced tests indicated that monkeys know when they remember and reject tests when they have forgotten, if given the option.

Self-monitoring (C2): Non-human primates can report the presence or absence of memory: "this study documents in monkeys one important objective functional feature of human conscious cognition: the ability to make adaptive decisions about future behavior contingent on the current availability of knowledge."

Hampton, R. R. (2001). Rhesus monkeys know when they remember. ***PNAS***, 98(9), 5359–5362. http://doi.org/10.1073/pnas.071600998

Consciousness

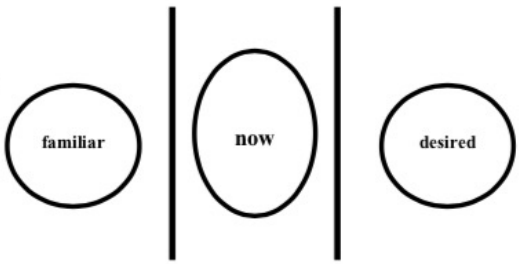
Ontogeny 	Mechanism
Phylogeny	Adaptive Significance

The ontogeny of consciousness

(a)

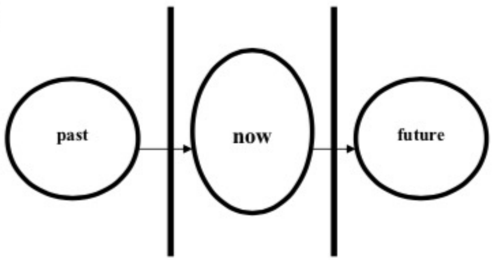
Minimal consciousness: Restricted to present intero- and exteroreceptor stimulation (Now)

(b)



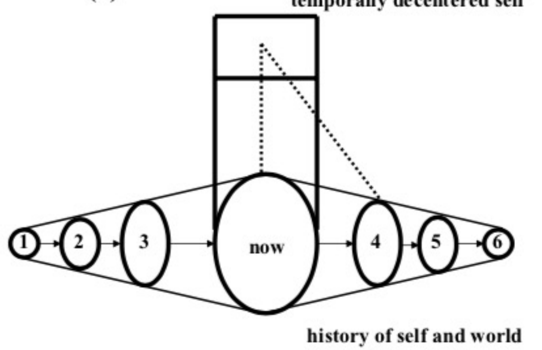
Recursive consciousness: Past and future events can be considered but toddlers cannot simultaneously represent the present when representing past or future

(c)



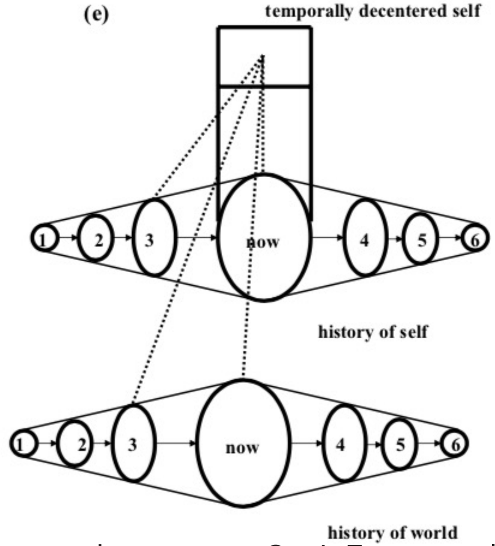
Self-consciousness: Children (2-year olds) can consider descriptions of past or future-oriented events in relation to a present experience; but no consideration of events occurring at different times

(d)



Reflective consciousness 1: 3-year-olds can consider two events occurring at two different times, including an event occurring in the present (i.e., now, EventA is occurring, but Yesterday, EventB occurred). History of self and of the world are confounded.

(e)



Reflective consciousness 2: 4-5 year-olds have a temporally decentered perspective: Children can coordinate two series, the history of the self and the history of the world.

Zelazo, P.D. Gao, H. & Todd, R. (2007). The development of consciousness. In Zelazo, Moscovitch & Thompson (eds.), Cambridge Handbook of Consciousness. Cambridge University Press. pp. 405-432.

The ontogeny of consciousness

Description	Age	Example(s)
	newborns	attend significantly more toward someone else's finger touching their cheek (single touch) than toward their own hand touching their cheek (double touch).
Implicit self-awareness	8 weeks	self-defensive reactions to impending collisions with objects moving toward them
	4 months	infants reach for objects within their reach, but inhibit such action when object is too large or out of reach
Social referencing	7 months	focus on other individuals' gaze toward novel things; use emotional cues from others in their decisions to either approach or avoid novel things
Self-awareness	18 months	mirror test, infants who pass the mirror mark test show social emotions such as embarrassment or pride

Some explanations emphasise cognitive abilities (e.g., multiple representations, memory; cf. Zelazo et al.) others put more emphasis on social aspects (cf. Rochat) - the two are not incompatible!

Rochat, P. (2018). The Ontogeny of Human Self-Consciousness. *Current Directions in Psychological Science*, *27*(5), 345–350. http://doi.org/10.1177/0963721418760236

Consciousness

Ontogeny	Mechanism 
Phylogeny	Adaptive Significance

The mechanism(s) of consciousness

There have been many contributions to understanding the neural basis of consciousness but three key syndromes were likely central:

Amnesia	The analysis of amnesia patients (most prominently HM) led to the conclusion that memory deficits resulting from temporal lobe damage was limited to declarative memory, that is, memory that could be consciously experienced. Such lesion studies showed it was possible to understand the contribution of different brain areas to different aspects of consciously accessible memory.
Blindsight	Damage to the primary visual cortex produces an apparent blindness in the visual field opposite to the lesion. Yet, when requested to do so, blindsight patients can make guesses about the identity or presence of visual stimuli presented to the "blind" field at accuracy levels that are well above chance. They are consciously blind but can "see" sufficiently to control behavior.
Split brain	Split-brain surgery involves surgical section of the corpus callosum and other lesser cerebral commissures in an effort to help relieve intractable epilepsy. The lack of communication between the hemispheres can lead to interesting patterns of behaviour and cognition revealing hemispheric specialisation with each hemisphere not only has separate behavioral control capacities but possibly, separate mental systems—two conscious beings.

LeDoux, J. E., Michel, M., & Lau, H. (2020). A little history goes a long way toward understanding why we study consciousness the way we do today. *Proceedings of the National Academy of Sciences of the United States of America*, *2*(13), 201921623–6984. http://doi.org/10.1073/pnas.1921623117

Blindsight

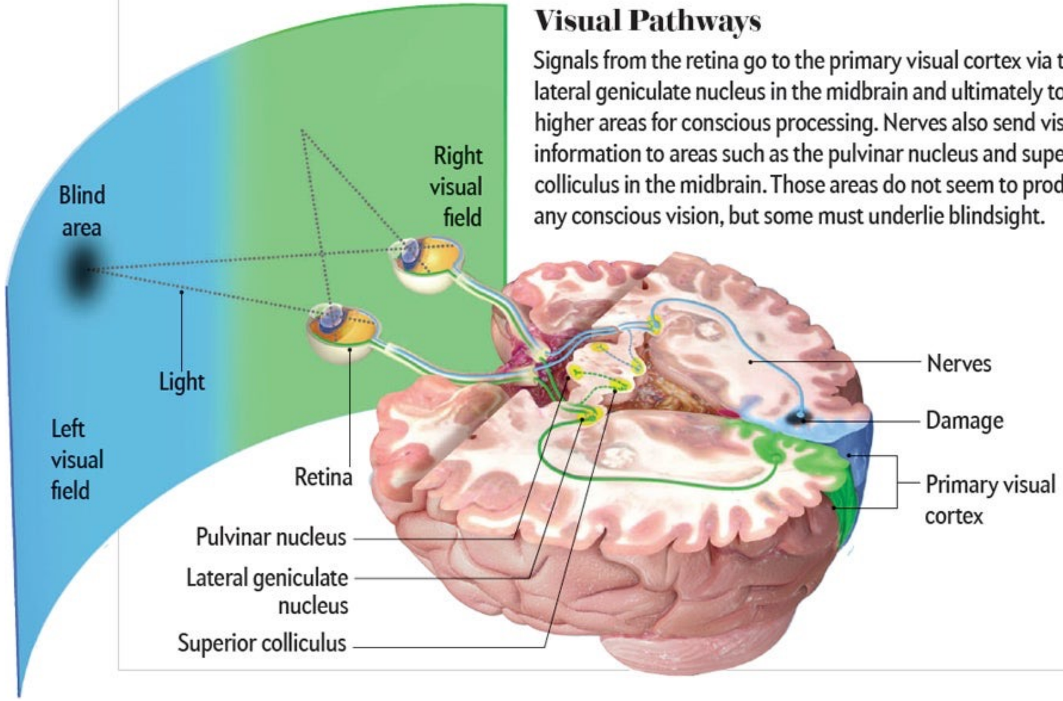
BASICS

What Is Blindsight?

Conscious vision in humans depends on the integrity of a region of the brain called the primary visual cortex. Damage there causes blindness in corresponding areas of the visual field. “Blindsight” occurs when patients respond in some way to an item displayed in their blind area, where they cannot consciously see it. In a dramatic demonstration of the phenomenon, a patient called TN navigated an obstacle course despite his total blindness.

Visual Pathways

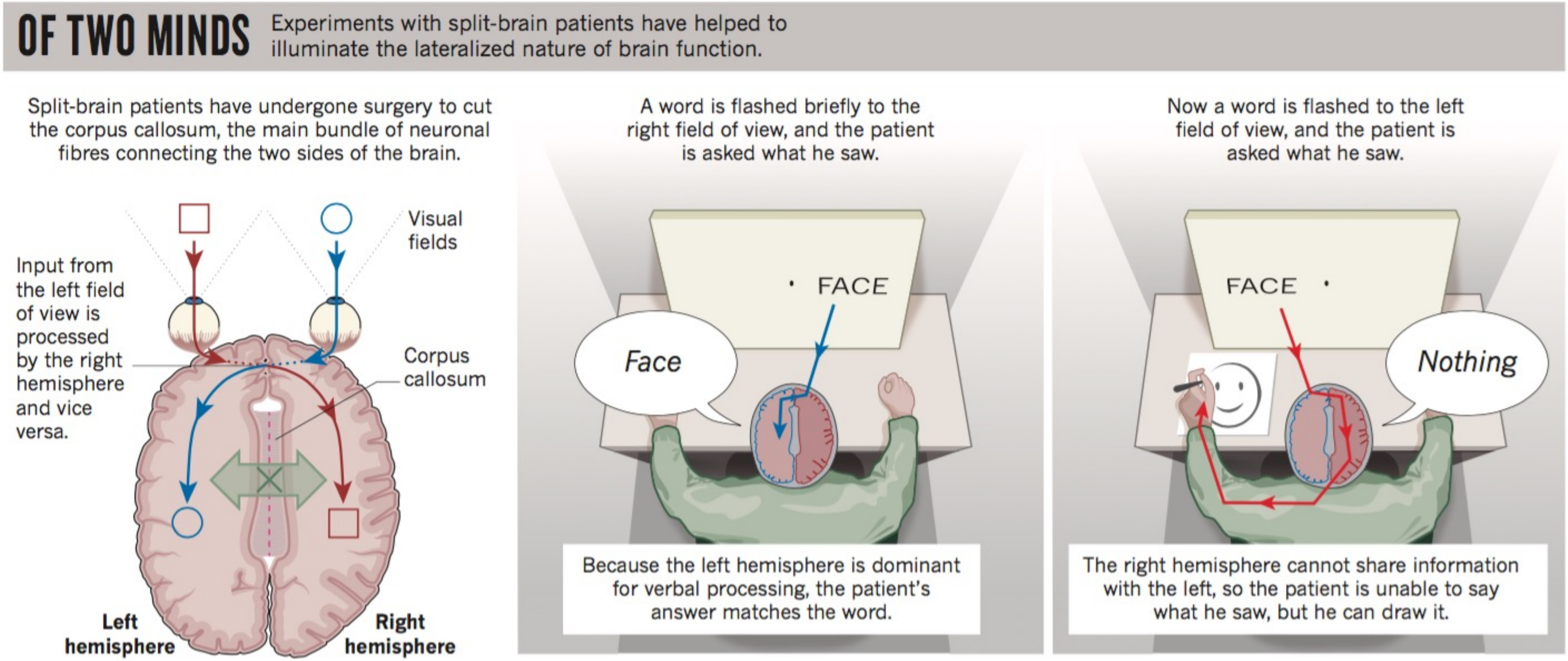
Signals from the retina go to the primary visual cortex via the lateral geniculate nucleus in the midbrain and ultimately to higher areas for conscious processing. Nerves also send visual information to areas such as the pulvinar nucleus and superior colliculus in the midbrain. Those areas do not seem to produce any conscious vision, but some must underlie blindsight.



Navigating Blind
Suspecting patient TN might exhibit blindsight, researchers, including Lawrence Weiskrantz (*shown with TN*), asked him to walk down a cluttered hallway, telling him it was empty. TN avoided all the obstacles, even though he remained unaware of them and of his meandering path.

de Gelder, B. (2010). Uncanny sight in the blind. Scientific American, (May), 60-65.

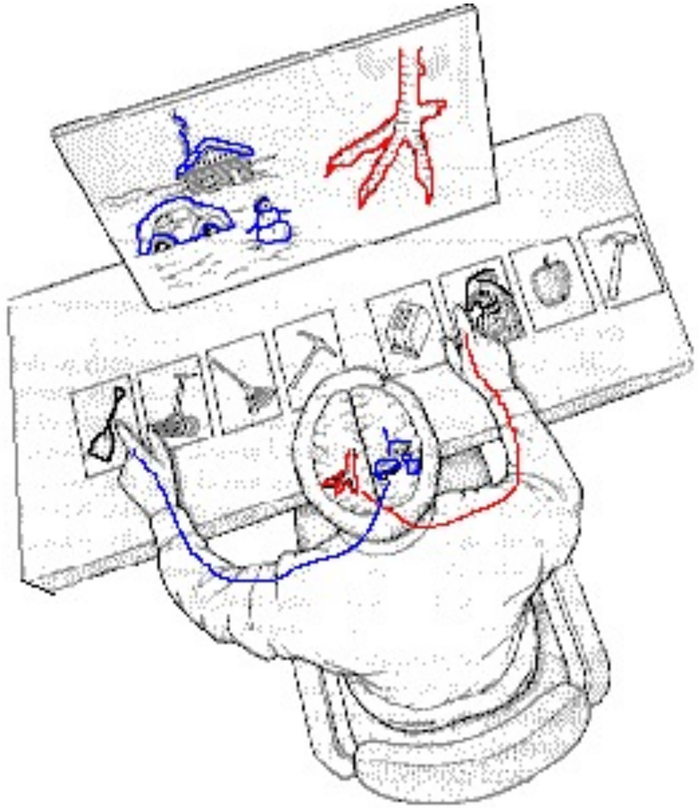
Split brain



Split-brain experiments were helpful in proving an understanding of brain lateralisation as well as a non-unitary view of consciousness: “activities of the brain which underlie conscious experience are distributed over different functional modules”.

Wolman, D. (2012). The split brain: A tale of two halves. ***Nature, 483***(7389), 260–263. http://doi.org/10.1038/483260a

Split brain



From the point of view of the left hemisphere, responses coming from the right hemisphere are generated nonconsciously. Split-brain studies sometimes involved the experimenter asking out loud, “Why did you do/pick that?” The patient responded via his left hemisphere with a verbal answer. The left hemisphere routinely took things in stride, telling a tale that made the responses make sense. For example, when the right hemisphere picked “shovel” in response to a snowy landscape (see figure), the patient (his left hemisphere) would explain his action by saying, “to clean the chicken coop.” This can be seen as evidence of confabulation because the left hemisphere was not privy to the information that instructed him to pick the shovel.

“The narratives weaved by the left hemisphere were viewed as interpretations of situations and were proposed to be an important mechanism used by humans to maintain a sense of mental unity in the face of neural diversity. The narration/interpretation process was later proposed to depend on cognitive functions of prefrontal cortex (…).”

LeDoux, J. E., Michel, M., & Lau, H. (2020). A little history goes a long way toward understanding why we study consciousness the way we do today. *Proceedings of the National Academy of Sciences of the United States of America*, *2*(13), 201921623–6984. http://doi.org/10.1073/pnas.1921623117

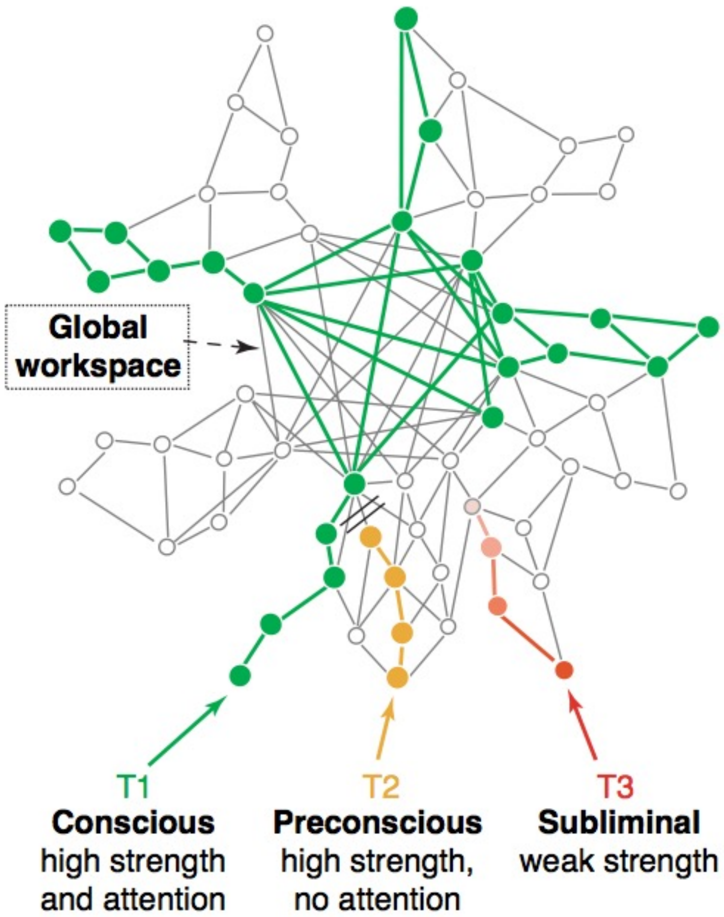
Global Neuronal Workspace Model

The Global Workspace Model was initially proposed by Bernard Baars to account for conscious and unconscious processes and proposes that consciousness corresponds to a "momentarily active, subjectively experienced" event”— the "inner domain in which we can rehearse telephone numbers to ourselves or in which we carry on the narrative of our lives. It is usually thought to include inner speech and visual imagery.” Dehaene et al. adopt Baars view and propose that the flexible dissemination of information is a characteristic property of the conscious state:

”When we say that we are aware of a certain piece of information, what we mean is just this: the information has entered into a specific storage area that makes it available to the rest of the brain”.

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. ***Trends in Cognitive Sciences, 10***(5), 204–211. doi:10.1016/j.tics.2006.03.007

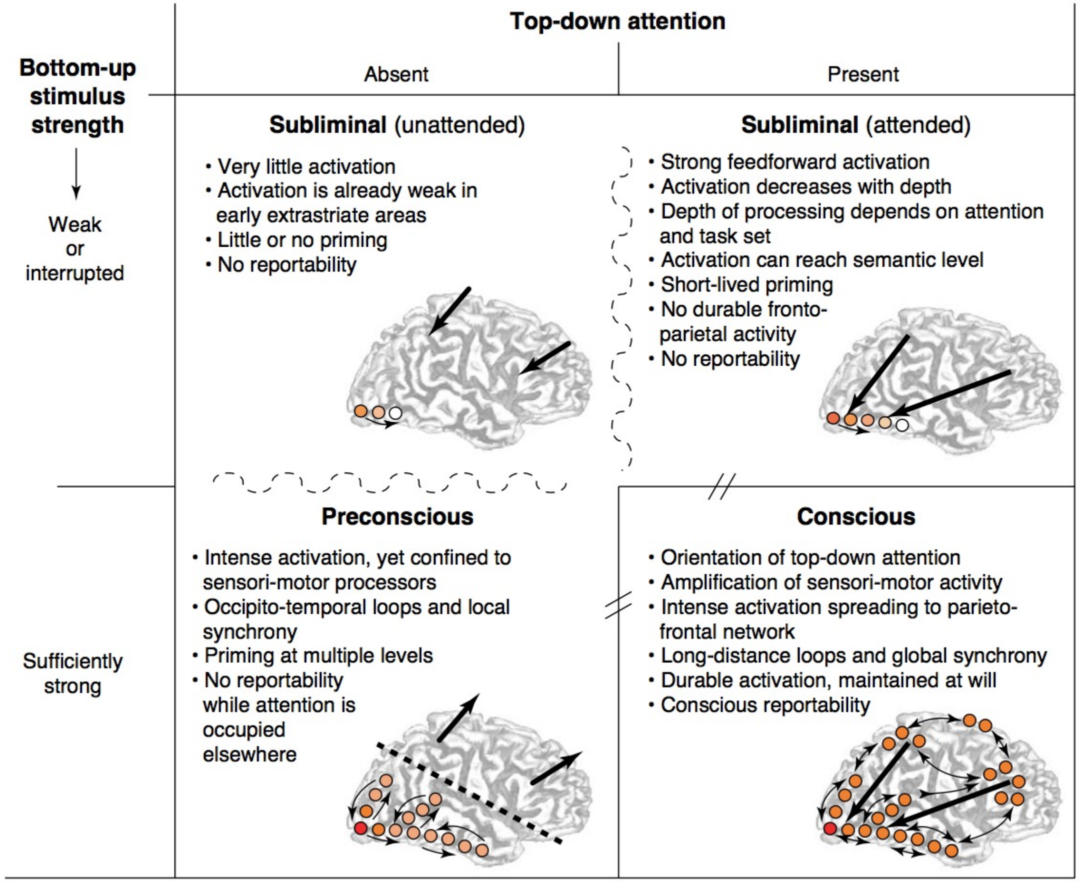
Global Neuronal Workspace Model



A visual target T1 (in green) is consciously accessed when it activates, in a synchronised, reciprocal and long-lasting manner, a set of 'central workspace' neurons particularly dense in **parietal**, **prefrontal** and **cingulate** cortices, and whose long-distance connections enable broadcasting to many distant areas. A stimulus can fail to become conscious for two reasons: (1) it might not have enough bottom-up strength, for example, owing to low-level masking or presentation close to threshold (subliminal stimulus T3, in red); or (2) it might have enough strength to be visible, but still fail to be seen by losing the competition for central access relative to other concurrent stimuli or task sets (preconscious stimulus T2, in orange).

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. ***Trends in Cognitive Sciences,*** ***10***(5), 204–211. doi:10.1016/j.tics.2006.03.007

Global Neuronal Workspace Model



Shades of color illustrate the amount of activation, and small arrows the interactions among them. Large arrows schematically illustrate the orientation of top-down attention to the stimulus, or away from it ('task-unrelated attention'). Dashed curves indicate a continuum of states, and thick lines indicate a sharp transition between states.

Proposed distinction between subliminal, preconscious, and conscious processing. During subliminal processing, activation propagates but remains weak and quickly dissipating (decaying to zero after 1–2 seconds). A continuum of subliminal states can exist, depending on masking strength, top-down attention, and instructions. During preconscious processing, activation can be strong, durable, and can spread to multiple specialised sensori-motor areas. However, when attention is oriented away from the stimulus (large black arrows), activation is blocked from accessing higher parieto-frontal areas and establishing long-distance synchrony. During conscious processing, activation invades a parieto-frontal system, can be maintained ad libidum in working memory, and becomes capable of guiding intentional actions including verbal reports. The transition between preconscious and conscious is sharp, as expected from the dynamics of a self-amplified non-linear system.

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. ***Trends in Cognitive Sciences,*** ***10***(5), 204–211. doi:10.1016/j.tics.2006.03.007

Global Neuronal Workspace Model

What is consciousness good for? (back to adaptive significance!)

"(…) conscious access is the extension of brain activation to higher association cortices interconnected by long-distance connections and forming a reverberating neuronal assembly with distant perceptual areas. Why would this brain state correspond to conscious access? Neurocomputational simulations show that once stimulus-evoked activation has reached highly interconnected associative areas, two important changes occur:

(1) The activation can **reverberate**, thus holding information on-line for a long duration essentially unrelated to the initial stimulus duration;

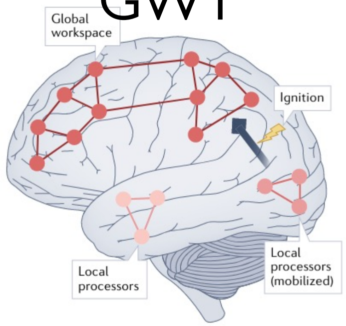
(2) Stimulus information can be rapidly **propagated** to many brain systems.

We argue that both properties are characteristic of conscious information processing which in our view is associated with a distinct internal space, buffered from fast fluctuations in sensory inputs, where information can be shared across a broad variety of processes including evaluation, verbal report, planning and long-term memory."

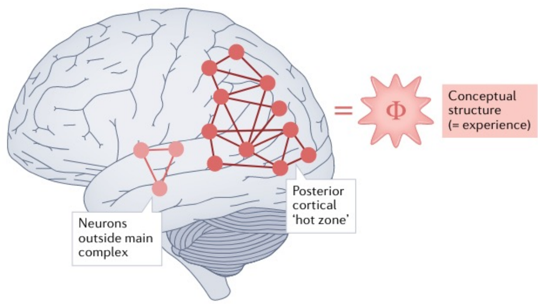
Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. ***Trends in Cognitive Sciences, 10***(5), 204–211. doi:10.1016/j.tics.2006.03.007

Comparing consciousness theories…

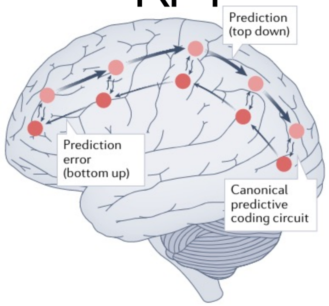
GWT



IIT



RPT

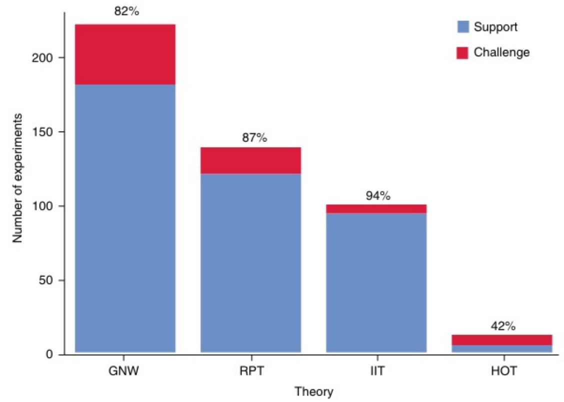


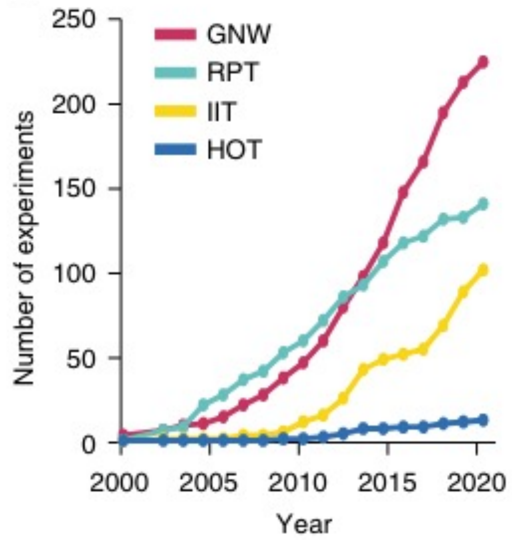
Global workspace theories (GWT). Mental states are conscious when they are broadcast within a global workspace in which fronto-parietal networks play a central hub-like role.

Integrated information theory. Consciousness is identical to the cause–effect structure of a physical system that specifies a maximum of irreducible integrated information; anatomically, IIT is associated with a posterior cortical 'hot zone'.

Re-entry theory and predictive processing. Conscious mental states are associated with top-down signalling (re-entry; thick arrows) that, for predictive processing, convey predictions about causes of sensory signals (t arrows signify bottom-up prediction errors), that continuous minimization of prediction errors implements an approximation to Bayesian inference.

Higher-order thought. Consciousness refers higher-order representations in dorsolateral prefrontal cortex that accompany first-order representations elsewhere in the brain.





Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews Neuroscience*. https://doi.org/10.1038/s41583-022-00587-4;

Yaron, I., Melloni, L., Pitts, M., & Mudrik, L. (2022). The ConTraSt database for analysing and comparing empirical studies of consciousness theories. *Nature Human Behaviour*, *6*(4), 593–604. https://doi.org/10.1038/s41562-021-01284-5

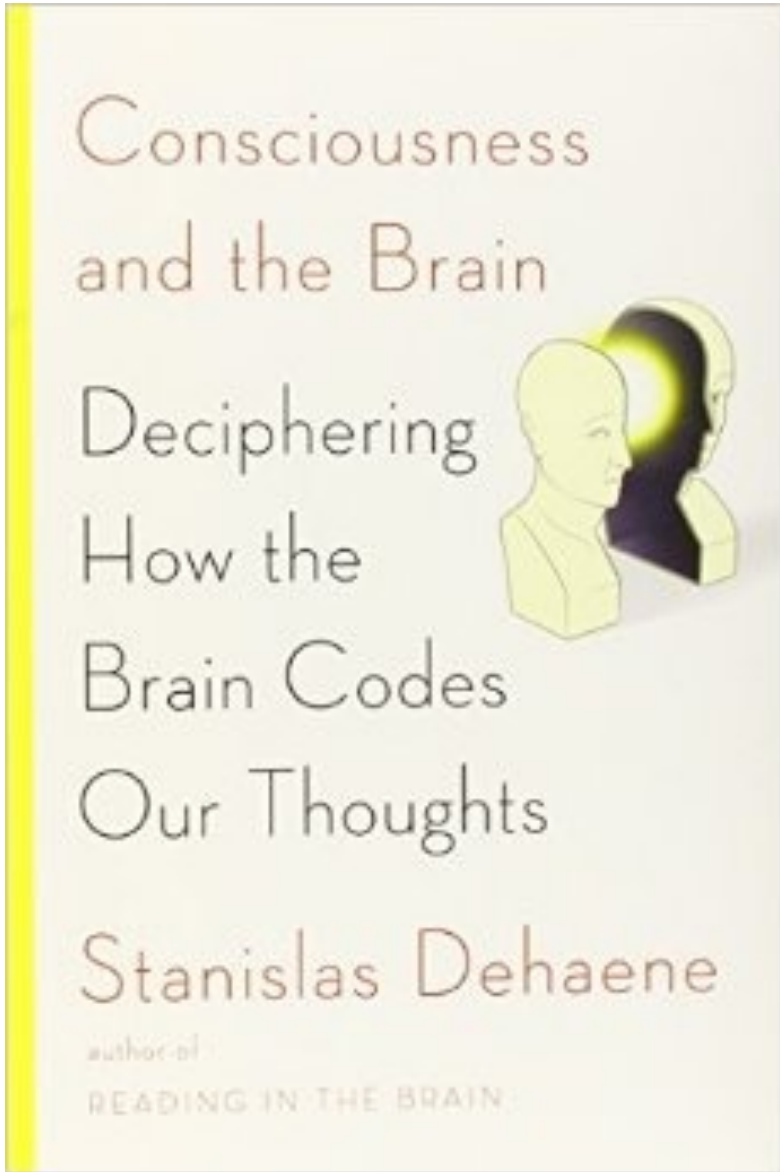
Summary

- **Adaptive Significance**: Different definitions of consciousness allows different answers to the question of whether machines can be conscious. According to recent views, criteria of global availability and self-monitoring are not typically implemented in artificial intelligence. Consciousness seen as an adaptive trait that allows for flexible/versatile cognition.

- **Comparative approaches:** animals are capable of some levels of self-awareness and versatility that can be seen as indicating components of consciousness (global availability, self-monitoring).

- **Development of consciousness:** Progression of implicit to explicit self-awareness and increased complexity in the representation of the self in relation to the world; both socio-emotional and cognitive aspects should be emphasised.

- **Neural basis:** a number of pathologies have made clear that there are important aspects of unconscious processing that are dissociable from conscious access (amnesia, blindsight, split brain); current neural models propose role of prefrontal and parietal areas for maintenance of information necessary for global availability and self-monitoring.


Consciousness and the Brain
Deciphering How the Brain Codes Our Thoughts
Stanislas Dehaene
author of
READING IN THE BRAIN