

Perspective

An integrative, multiscale view on neural theories of consciousness

Johan F. Storm,^{1,*} P. Christiaan Klink,^{2,3,4} Jaan Aru,⁵ Walter Senn,⁶ Rainer Goebel,⁷ Andrea Pigorini,⁸ Pietro Avanzini,⁹ Wim Vanduffel,^{10,11,12,13} Pieter R. Roelfsema,^{2,4,14,15} Marcello Massimini,^{16,17,18} Matthew E. Larkum,^{19,20} and Cyriel M.A. Pennartz^{21,22}

¹The Brain Signaling Group, Division of Physiology, IMB, Faculty of Medicine, University of Oslo, Domus Medica, Sognsvannsveien 9, Blindern, 0317 Oslo, Norway

²Department of Vision and Cognition, Netherlands Institute for Neuroscience, Royal Netherlands Academy of Arts and Sciences, 1105 BA Amsterdam, the Netherlands

³Experimental Psychology, Helmholtz Institute, Utrecht University, 3584 CS Utrecht, the Netherlands

⁴Laboratory of Visual Brain Therapy, Sorbonne Université, Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, Institut de la Vision, Paris 75012, France

⁵Institute of Computer Science, University of Tartu, Tartu, Estonia

⁶Department of Physiology, University of Bern, Bern, Switzerland

⁷Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Oxfordlaan 55, 6229 EV Maastricht, The Netherlands

⁸Department of Biomedical, Surgical and Dental Sciences, Università degli Studi di Milano, Milan 20122, Italy

⁹Istituto di Neuroscienze, Consiglio Nazionale delle Ricerche, 43125 Parma, Italy

¹⁰Department of Neurosciences, Laboratory of Neuro and Psychophysiology, KU Leuven Medical School, 3000 Leuven, Belgium

¹¹Leuven Brain Institute, KU Leuven, 3000 Leuven, Belgium

¹²Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA 02129, USA

¹³Department of Radiology, Harvard Medical School, Boston, MA 02144, USA

¹⁴Department of Integrative Neurophysiology, VU University, De Boelelaan 1085, 1081 HV Amsterdam, the Netherlands

¹⁵Department of Neurosurgery, Academisch Medisch Centrum, Postbus 22660, 1100 DD Amsterdam, the Netherlands

¹⁶Department of Biomedical and Clinical Sciences “L. Sacco”, Università degli Studi di Milano, Milan 20157, Italy

¹⁷Istituto di Ricovero e Cura a Carattere Scientifico, Fondazione Don Carlo Gnocchi, Milan 20122, Italy

¹⁸Azieli Program in Brain, Mind and Consciousness, Canadian Institute for Advanced Research (CIFAR), Toronto, ON M5G 1M1, Canada

¹⁹Institute of Biology, Humboldt University Berlin, Berlin, Germany

²⁰Neurocure Center for Excellence, Charité Universitätsmedizin Berlin, Berlin, Germany

²¹Swammerdam Institute for Life Sciences, Center for Neuroscience, Faculty of Science, University of Amsterdam, Sciencepark 904, Amsterdam 1098 XH, the Netherlands

²²Research Priority Program Brain and Cognition, University of Amsterdam, Amsterdam, the Netherlands

*Correspondence: j.f.storm@medisin.uio.no

<https://doi.org/10.1016/j.neuron.2024.02.004>

SUMMARY

How is conscious experience related to material brain processes? A variety of theories aiming to answer this age-old question have emerged from the recent surge in consciousness research, and some are now hotly debated. Although most researchers have so far focused on the development and validation of their preferred theory in relative isolation, this article, written by a group of scientists representing different theories, takes an alternative approach. Noting that various theories often try to explain different aspects or mechanistic levels of consciousness, we argue that the theories do not necessarily contradict each other. Instead, several of them may converge on fundamental neuronal mechanisms and be partly compatible and complementary, so that multiple theories can simultaneously contribute to our understanding. Here, we consider unifying, integration-oriented approaches that have so far been largely neglected, seeking to combine valuable elements from various theories.

INTRODUCTION

Comparing theories

The mind-brain problem, i.e., how our conscious experience is related to material brain processes, has been debated by philosophers for centuries and remains one of the deepest unsolved problems in science.^{1–3} The last decades have seen a surge of theoretical and empirical consciousness research. A variety of

neuroscientific theories of consciousness have been proposed and are hotly debated (e.g., Seth and Bayne⁴). Efforts to test competing theories against each other through “adversarial collaboration” have been undertaken, e.g., by consortia sponsored by the Templeton World Charity Foundation (e.g., Cogitate Consortium et al.⁵).

Historically, most researchers have emphasized the development and validation of their preferred theoretical framework in



Table 1. Terms and concepts

Term	Definition or explanation	References	Comments/[synonyms or similar concepts]
Conscious / consciousness (general and philosophical terms)	“‘Conscious’ and ‘consciousness’ are umbrella terms that cover a wide variety of mental phenomena. ... used with a diversity of meanings, ... ‘conscious’ is heterogeneous in its range, being applied both to whole organisms... and to particular mental states and processes.”	Stanford Encyclopedia of Philosophy	
Consciousness (C) (as used in this article)	Despite the lack of any agreed upon definition of the term “consciousness,” we here use the term in a sense that is roughly synonymous with the term “experience,” including both PC and AC (see below) and thus including “subjective experience,” “inner experience.”	Chalmers, ^{1,15} Lamme, ⁸ Pennartz, ¹² Koch et al., ¹⁶ and Sanders et al. ¹⁷	The concept of consciousness that we use here includes both awake, alert states with their experiences, including perception of the external world, but also “inner experiences”: thoughts, feelings, dreams, imagery, hallucinations, etc. that may sometimes occur in unresponsive states (e.g., REM sleep, etc.).
Phenomenal consciousness (PC)	PC refers to those properties of experience that correspond to “what it is like” for a subject to have those experiences (qualia ¹⁸). “These features are apparent to the subject from the inside, so tracking them arguably depends on one’s having the relevant experience.”	Stanford Encyclopedia of Philosophy; Block ¹⁹	That an item is part of PC does not imply that it is part of AC.
Access consciousness (AC)	AC refers to those properties of experience that are accessible for use in reasoning, report, and the control of action (i.e., in voluntary (“rational”) control of behavior).	Stanford Encyclopedia of Philosophy; Block ¹⁹	That an item is part of AC does not imply that it is part of PC.
State of consciousness	Overall conscious and unconscious states (related to brain states), e.g., wakefulness, dreaming, psychedelic states, dreamless sleep, minimally conscious or “vegetative” states, coma, general anesthesia.	Stanford Encyclopedia of Philosophy	[“Generic Consciousness: Stanford Encyclopedia of Philosophy”]
Working memory (WM)	A low-capacity memory system for temporary maintenance and manipulation of information in view of a task to be performed. It usually stores information for up to ~10–15 s for use in reasoning, decision-making, and guiding of behaviors.	Wang ²⁰ and Hempel and Oppenheim ²¹	[<i>Short-term memory</i> (STM) is often used as a synonym of WM, but some researchers hold that STM is only short-term storage of information, whereas WM allows manipulation of the stored information in service of completing a task.]
Sensory memory (SM)	A large-capacity store, briefly storing (~0.1–1 s) sensory data, separate for each sensory modality, providing a snapshot of the current, overall sensory input; automatic, outside cognitive control, weakly dependent on attention.	Clark, ⁹ Pennartz, ¹³ and Carlson ²²	Includes <i>iconic memory</i> (for vision), <i>echoic memory</i> (for hearing), <i>haptic memory</i> (for touch stimuli).
Explanandum	The phenomenon to be explained by the theory; a sentence “describing the phenomenon to be explained” by the theory.	Stanford Encyclopedia of Philosophy; Bastos et al. ²³	

(Continued on next page)

Table 1. Continued

Term	Definition or explanation	References	Comments/[synonyms or similar concepts]
Scientific theory	An explanation of an aspect of the natural world and universe that can be (or a fortiori that has been) repeatedly tested and corroborated in accordance with the scientific method, using accepted protocols of observation, measurement, and evaluation of results.	Wikipedia (https://en.wikipedia.org/w/index.php?title=Scientific_theory&oldid=1183411663)	There is no universally accepted definition of “scientific theory.” Similar definitions are, e.g., given by Encyclopedia Britannica (https://www.britannica.com/science/scientific-theory), Stanford Encyclopedia of Philosophy (https://plato.stanford.edu/entries/structure-scientific-theories/), National Academy of Sciences USA, ²⁴ and American Association for the Advancement of Science (https://www.aaas.org/sites/default/files/0219board_statement.pdf).

isolation. This article, however, takes a different approach. It arises from researchers within the Human Brain Project (HBP; 2013–2023) working on topics related to consciousness and representing different theories. Based on recent empirical progress, we show that many aspects of the various theories of consciousness do not necessarily contradict each other, as sometimes claimed; instead, theories often try to explain different aspects of consciousness and tend to converge on fundamental neuronal mechanisms and processes. Here, we argue that several theories are at least partly compatible and complementary and we consider approaches toward convergence that have so far been largely neglected.

As it is not feasible to cover all proposed, neuroscience-based theories of consciousness, we focus here on five that we think represent prominent and complementary perspectives: (1) global neuronal workspace theory (GNWT⁶), (2) integrated information theory (IIT⁷), (3) recurrent processing theory (RPT⁸), (4) predictive processing (PP^{9–11}) and neurorepresentationalism (NREP^{12,13}), and (5) dendritic integration theory (DIT¹⁴). These theories cover three of the four broad categories of consciousness theories outlined by Seth and Bayne⁴ and thus allow a broad discussion of how different theories may be reconciled and possibly integrated. Here, we use the term “theories” in a broad sense as there is no agreed-upon definition of this term (Table 1), and other terms (“hypothesis,” “model,” and “framework”) also lack stringent definitions. We focus on the content of these theories and ideas, rather than their status or maturity. Another criterion we used when choosing theories was that each should represent a cumulative body of neuroscientific research brought together under a definable framework of ideas.

By examining how these five theories compare and may be partly reconciled, we aim to start an integration process that could also incorporate valuable elements from other theories in the future. However, some current theories may of course turn out to be entirely or largely erroneous or sterile and will be weeded out by future selection, as has happened repeatedly in

the history of science. Trying to overcome long-standing, seemingly unresolvable philosophical and theoretical debates, we propose a different approach, focusing on empirically based notions of integration, complexity, representation, and recurrent processing at different levels, which appear to be shared features of these theories.

Following Ned Block,¹⁹ we distinguish mechanisms of phenomenal consciousness (PC; immediate, subjective experience, e.g., of seeing a blue sky) from those of access consciousness (AC) and reportability. We discuss how these distinctions are related to proposed empirical measures of consciousness (Boxes 1 and 2) and how they relate to attention, expectations, planned behavior, and different forms of memory.^{8,12,25–28} We also distinguish between levels of explanation from low-level, mechanistic features to high-level, system-wide properties.

Concepts and explananda

When comparing theories, it is essential to first clarify the key concepts, define the terms used, and identify which phenomena (explananda) the theories aim to explain. A point that is immediately highlighted by our comparative approach is that different theories tend to focus on distinct aspects or subtypes of “consciousness,” which may cause conceptual confusion and miscommunication. Opposing theories may promote different experimental approaches that are not directly comparable, or different interpretations of data. For example, GNWT has focused on AC¹⁹ (Table 1), primarily trying to explain how conscious content can be accessed and used by the subject, and on conscious vs. unconscious “information processing” (as opposed to phenomenal experience), typically studying responses to a particular sensory input, such as seeing a face. By contrast, integrated information theory (IIT), recurrent processing theory (RPT), and predictive processing (PP)/NREP have primarily focused on “PC,”¹⁹ (i.e., what “it is like” to experience something¹⁸), irrespective of cognitive functions and whether it can be reported or not.^{50,51} Thus, in contrast to

Box 1. Methods for assessing states and levels of consciousness

Identifying brain-based measures to reliably index the presence and the absence of consciousness in humans is the focus of a substantial body of literature, a fundamental concern for clinicians, and arguably an important first step for understanding the relationship between consciousness and its physical substrate. Practically, brain-based measures are mandated when the behavioral assessment of consciousness is unreliable, such as in subjects who are paralyzed during general anesthesia or are unresponsive because of severe brain injuries.

A first class of markers involves by-passing motor behavior and directly measuring brain responses to sensory stimuli, such as verbal commands or complex sequences of sounds. A prominent example of this class of markers is the P3b elicited by global violations of expectation in auditory sequences.²⁹ According to GNWT, the P3b indexes the broadcasting (following ignition) of sensory information to frontal cortical processors and has been used to detect the presence of consciousness in brain-injured patients. Although the P3b provides a specific marker (i.e., it is present only in subjects who are aware of the stimulus), it is often absent in conscious brain-injured patients and conscious healthy subjects.^{30,31} Such low sensitivity of the P3b is an empirical reminder of the potential dissociation between the neuronal correlates of AC (indexing connectedness to the environment and/or engagement of high-level cognitive functions) and those of PC.

A second class of markers focuses directly on brain activity, without necessarily relying on sensory processing and cognitive functions. A recent empirical convergence has emerged on measures of integration and differentiation holding the promise of a more accurate detection of consciousness.³² Notably, this convergence involved many independent researchers endorsing disparate theoretical frameworks and employing different techniques across various conditions such as sleep, dreaming, anesthesia, epilepsy, hallucinatory states, and coma. In general, the prospect of reliably detecting consciousness independently of connectedness to the external environment, behavioral performance, cognitive, and memory abilities has clinical implications and fosters the search for the neuronal mechanisms of PC. Complexity measures rely either on the characterization of spontaneous brain dynamics (through the analysis of functional connectivity networks and graph theory metrics)^{33,34} or on the quantification of the patterns of interactions evoked by direct cortical perturbations.³⁵ The latter approach, instantiated by the PCI, shows high sensitivity in challenging clinical conditions and points to a mechanistic link between brain complexity and feedback interactions. Indeed, across various physiological and pathological conditions and experimental models, decreases and increases of PCI during loss and recovery of consciousness are invariably associated with the obliteration and resurgence of recurrent waves of activations.^{36–39}

GNWT, the other theories (IIT, RPT, and PP/NREP; for dendritic integration theory (DIT), see below) all primarily try to address the “Hard Problem,¹” i.e., to explain how it is possible that PC and “qualia” can arise in a physical system like the brain, and why these subjective phenomena disappear during dreamless sleep or anesthesia (but see Sevenius Nilsen et al.⁵²). This difference partly comes from a different philosophical view: contrary to IIT, PP/NREP, or RPT, the proponents of GNWT typically do not regard PC or qualia as a real phenomenon, distinct from AC, but have argued that as long as their theory explains the observable facts (e.g., reports like: “now I see the face!”), there is nothing left to explain. For PP/NREP, IIT, and RPT, PC is the primary explanandum, regarding access to conscious content as an accessory matter, relying on executive processes that apply conscious content to motor or cognitive operations.^{7,12,13,27,53} DIT has suggested mechanisms for both AC and PC.⁵⁴

The original versions of GNWT vs. IIT also differed in another way: GNWT focused mostly on the processes of selection and use of contents of consciousness (AC), whereas IIT focused more on the fundamental nature and qualities of PC (in terms of information integration and interactions), and the state and level of PC.⁵⁵ However, now both theories increasingly try to explain processes, contents, and states/levels,^{28,16} a tendency shared with RPT, PP/NREP, and DIT.^{12,14,56} There is an increasing agreement that any complete theory should explain both contents and states/levels of consciousness.⁵⁷ Finally, DIT is a recently proposed theory, mainly operating at the cell

and circuit levels, and addresses why mammalian consciousness depends on both cortico-cortical and thalamo-cortical loops. It hypothesizes that both AC and PC involve simultaneous completion of these loops due to dendritic processing.^{14,54}

In this review, we will first briefly present each of the five theories one by one, then discuss how they compare with each other, using the brain’s hierarchical organization as a scaffold, and next consider if and how different theories may be reconciled and perhaps even integrated.

THEORIES OF CONSCIOUSNESS

Global neuronal workspace theory

The global workspace theory (GWT)⁵⁸ describes the processing steps that give rise to conscious experience as a dynamic interaction between local processors and a global workspace that integrates information and broadcasts it to other processors for further processing. The global workspace thus acts as backbone through which information about different sensory and cognitive processes is made available to multiple brain regions. It is the broadcasted information that constitutes the content of consciousness. A neurobiology-inspired version of the GWT was later proposed as the GNWT.^{6,28,59,60} It takes the same functional concept of a global workspace but links it to specific brain networks and neuronal characteristics. Here, the local processors comprise modular cortical areas with functional specializations for perception, action, memory, etc. Their activity is triggered by sensory stimuli, and they can perform calculations in

Box 2. Methods for assessing contents of consciousness

Assessing the contents of consciousness empirically at a fine-grained sub-categorical level is challenging because recording sites representing specific features need to be identified. A useful approach has been proposed by Koch et al.¹⁶ when searching for content-specific neural correlates of consciousness (NCC): (1) a mapping step in which the response properties of recording sites (e.g., neurons in neuronal recording studies or voxels in fMRI studies) are determined in a participant using appropriate feature-specific stimuli; (2) application of an experimental paradigm tapping on consciousness such as a multistable figure or binocular rivalry where the physical stimulus properties remain constant, while the contents of consciousness alternate between different interpretations of the stimulus. If the experienced (reported) alternating content of consciousness is reflected in dynamic activity changes in the features mapped in step 1, the respective recording sites qualify as potential NCCs; (3) experimentally try to rule out known confounds, for instance, that neural activity covarying with conscious perception might instead reflect prerequisites or consequences of consciousness.⁴⁰ Thus, it must be kept in mind that this approach is correlational, not causal, and that attention and memory effects may confound the interpretation of results.

This approach has been successfully applied in macaque monkeys as well as in humans using electrophysiological recordings. Human intracerebral (stereo-EEG) recordings led to the identification of tonic activity patterns,⁴¹ i.e., prolonged low-amplitude gamma-band responses to sensory stimulation, likely corresponding to the intracranial correlate of previously reported late EEG components reflecting perceptual awareness.^{42,43} Somatosensory stimulation induces tonic activity in the bilateral parietal operculum and posterior insular cortices.^{44,45} In addition, visuo-tactile stimulation enhances specifically these tonic components,⁴⁶ paralleling the notion that visuotactile stimulation considerably improves detection in tactile-extinction patients.⁴⁷

The approach could be extended to human fMRI measurements with a sufficiently high spatial resolution to differentiate feature-specific responses at the level of cortical columns. In a study by Schneider et al.,⁴⁸ columnar clusters in the human middle temporal complex (hMT+) that selectively responded to horizontal vs. vertically moving stimuli were localized in the mapping stage. In the main experiment, an ambiguous apparent motion stimulus (motion quartet) was presented that could be perceived as either horizontally or vertically moving. By tracking activity in the pre-mapped horizontal/vertical columnar clusters, it was possible to predict which of the two possible interpretations of the bistable motion stimulus (horizontal or vertical motion) was perceived by the participant at alternating periods in time. The results of the motion quartet study support the idea that hMT+ is (part of) the content-specific NCC for direction-of-motion, in line with the causal role of the middle temporal area (MT) in motion vision elucidated by Salzman and Newsome in non-human primates.⁴⁹ More generally, mapping out cortical (columnar) functional clusters and measuring their responses using sub-millimeter ultra-high field fMRI during ambiguous sensory stimulation fulfills the criteria for the search of content-specific NCC outlined by Chalmers¹⁵ and Koch et al.¹⁶

isolation from each other. Modules with such characteristics are, for instance, found in early sensory cortex, where different features are processed in specialized cortical regions.^{61–63} The global workspace is thought to involve long-range connections (cortico-cortical and thalamo-cortical) that connect a broad network of brain regions, particularly in the prefrontal cortex, posterior parietal cortex, and associated areas. These distributed areas can co-select one mental object so that all the representations of relevant features of that object are enhanced and bound in perception. The formation of coherent object representations corresponds to object-based attention in perceptual psychology. In this scheme, the thalamus and other subcortical structures such as the basal ganglia could also serve to amplify and distribute the information across various cortical regions and integrate neuronal activity across large brain regions. The cortical areas are believed to select and integrate information to enable the brain to focus and carry out complex cognitive tasks. Finally, the moment a local neuronal assembly reaches the activity threshold to enter the global workspace and engage in subsequent broadcasting is called “ignition.”

Ignition is a central concept in GNWT that has been linked to recurrent processing, for instance, in frontoparietal cortex.⁶⁴ It has gathered support in recent years,²⁸ for instance, from human neurophysiological and neuroimaging studies that show a distinct divergence in brain activity around 200 to 300 milliseconds after stimulus onset that depends on conscious percep-

tion^{65,66} and has been observed across different sensory modalities and paradigms. Early perceptual processing in the first 200 milliseconds can be similar or even identical between conscious (i.e., reported) and unconscious (i.e., not reported) trials,^{67–69} but conscious perception is associated with late events, such as the N2/P3b/P300 component of scalp-event-related potentials in humans.^{6,70–72} The timing of conscious access can furthermore be delayed under certain conditions, such as inattention or dual-task situations where the global workspace is already occupied by another conscious object.^{67,73} Concept cells in the anterior temporal lobe of humans and prefrontal neurons in non-human primates show stronger and longer-lasting firing rates when corresponding stimuli are consciously perceived.^{74–77} In monkeys, signal propagation to the frontal cortex furthermore leads to ignition and behavioral reports.⁶⁴ Some of these experimental results are debated,^{12,16,78} often on the basis that perceptual reports may confound the later frontal activity.⁴² However, results from recent no-report paradigms in monkeys may potentially mitigate these criticisms.^{79,80}

Recurrent circuits within or across brain areas support persistent activity and ignition-like processes.²⁰ Long-range cortical projections, crucial for the global workspace, originate in cortical layers II/III and V/VI where feedback effects are also most pronounced.^{81–84} Recent evidence suggests that reverberatory loops for persistent neuronal activity involve both cortico-cortical interactions and subcortical regions like the thalamus and

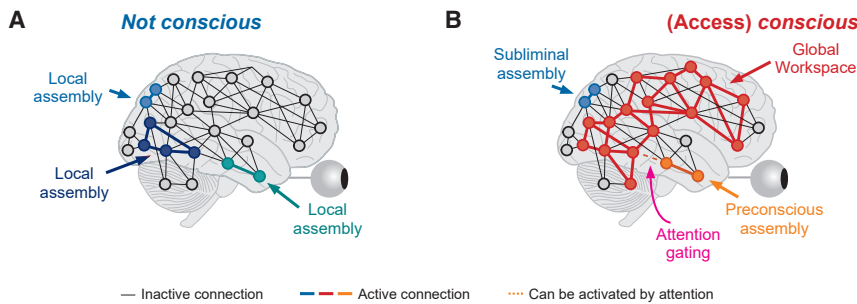


Figure 1. Global neuronal workspace theory (GNWT)

(A) In GNWT, local assemblies of processing nodes (circles) remain outside of conscious access.

(B) When the local assemblies get connected within a broad network of long-range connections (red) called the “Global Workspace” through a process of “ignition,” they do give rise to (access) consciousness. Assemblies that are never connected to the global workspace remain subliminal and unconscious (blue nodes), while processes inside the global workspace can be accessed for further use by other cognitive processes. Assemblies that are not currently connected to the global work-

space but that can become connected by way of attention boosting their signal (“attentional gating”) are dubbed “preconscious” (orange nodes; similar to what some other theories call phenomenal consciousness; see Figures 3, 4, 5, 6, and 7). Inactive connections among nodes are depicted with thin black lines, active connections are thick (colored) lines.

cerebellar nuclei,^{85,86} which is in line with a broad network of long-range connections supporting a global workspace.⁸⁷

Integrated information theory

With its emphasis on the “phenomenal structure” of experience, IIT uses introspection to characterize phenomenal properties and then deduces these in terms of corresponding physical properties. This physical explanation can then be empirically validated or falsified according to the standard methods of measuring behavioral, functional, and neural correlates.⁵⁰ The seed of this consciousness-first approach can be already found in the early paper “Consciousness and Complexity” published in 1998.⁸⁸ Here, the authors start from introspection and the premise that each conscious experience is at once integrated (unitary) and highly informative (each conscious state differs in a specific way from a huge number of alternative states). They then postulate that the coexistence of neuronal integration and differentiation is a key physical property of conscious systems, operationalizing this claim as measurable neuronal properties:

- (1) Integration: a group of neurons can contribute to conscious experience only if it is part of a functional cluster that achieves high integration through reentrant interactions in hundreds of milliseconds.
- (2) Differentiation: to sustain conscious experience, it is essential that this functional cluster must be capable of a large repertoire of different neural states, associated with high values of complexity.

This principle, relating consciousness to the complexity of neuronal interactions, is at the core of the early formulation of IIT,^{89,90} it has inspired a version of “weak IIT,⁹¹” and led to testable explanations and predictions:

- (1) Some regions of the thalamo-cortical system endowed with appropriate anatomical requirements (high density of lateral connections and large numbers of reentrant circuits) support conscious experience, whereas other structures with modular or feedforward architectures, such as the cerebellum, do not.
- (2) Synchronous or widespread neuronal activation is not necessarily associated with conscious states unless they are characterized by high levels of differentiation or complexity.

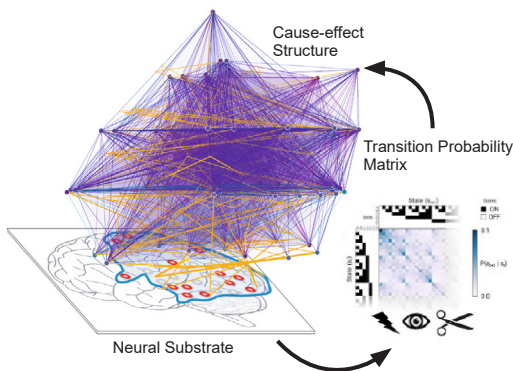
- (3) Given that anatomical requirements are preserved, functional alterations affecting reentry, integration, and differentiation will result in the loss of complexity and consciousness in conditions such as sleep, general anesthesia, and generalized seizures.
- (4) High-complexity and conscious experience can be supported by intrinsic brain interactions even in the absence of sensory inputs, motor outputs, and executive functions.
- (5) The presence vs. absence of consciousness can be reliably assessed by using direct cortical perturbations and recordings to measure the complexity of causal interactions within the brain.⁹²

These early predictions have been extensively assessed over the last 20 years across different conscious states (wakefulness, sleep, dreaming, psychedelic states, epilepsy, and disorders of consciousness) using various brain imaging modalities and have shown promise of clinical application (for a review, see Sarasso et al.³²; see also Box 1). The full formulation of IIT^{93,94} culminating in IIT 4.0⁹⁵ considers five phenomenal properties of consciousness (intrinsicity, information, integration, exclusion, and composition) and then formulates them as physical properties. Through its mathematical formalism IIT aims to account not only for the presence vs. absence of consciousness but also for its quality and contents. Specifically, IIT posits an explanatory identity between each experience—a particular “phenomenal structure”—and the corresponding “cause-effect structure” specified by the physical substrate.⁵⁰

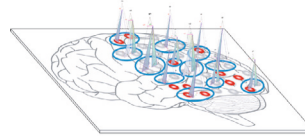
IIT holds that a cause-effect structure captures in full the causal powers of a system in a particular state—say, a network of neurons, some active and some inactive. Testing this framework requires combining systematic perturbations, observations, and partitions to assess precisely how subsets of elements make a difference to and take a difference from one another (Figure 2).

Although fully unfolding the cause-effect structure of even simple physical substrates is computationally daunting, IIT makes interesting predictions that are amenable to empirical tests. For example, IIT posits that (1) hierarchically organized grid-like cortical structures with lateral and recurrent convergent-divergent connections, such as those found in posterior cortex, are optimally suited for integrating information and thus correspond to the neural substrate of consciousness,⁹⁶ and (2) that phenomenal space—the feeling of extendedness—can be

A Wakefulness, dreaming



B Deep NREM, anesthesia



C Cerebellum

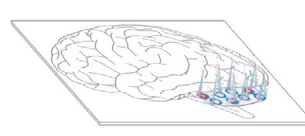


Figure 2. Integrated information theory (IIT) Unfolded cause-effect structures for three “brain-like” substrates.

(A) The neural substrate composed of a highly integrated network in the posterior cortex in a given state (red indicates active neurons and gray silent neurons) during wakefulness or dreaming. Combining systematic perturbations (lightning bolt) and observations (eye), one can obtain a transition probability matrix (TPM) of the system. This TPM is the basis for assessing irreducible causal mechanisms and their relations through partitioning operations (scissors) (for details, see Albantakis et al.⁹⁵). In this way, the intrinsic causal powers are “unfolded” to reveal an intricate (high Φ) cause-effect structure with many distinctions and relations.

(B) Anatomical connections being equal, changes in neuromodulation and post-synaptic properties

can lead to a breakdown of causal interactions. Therefore, the single, large cause-effect structure of high Φ (A) “disintegrates” into multiple disjoint ones (“peaks” of gray lines), each with low Φ . (C) If the neural substrate is characterized by modular and/or mostly feedforward connectivity, similar to that found in the cerebellum, the network specifies separate, minimal cause-effect structures, each with very low Φ .

accounted for by the “cause-effect structure” specified by the grid-like structure of cortical areas.⁹⁷ (3) Within these grids, also inactive neurons contribute to the cause-effect structure and thus to experience, just like active neurons. (4) Finally, modular or feedforward architectures, such as deep learning networks, would not be conscious even if their observable performance should equal that of a conscious human.⁹⁴

Recurrent processing theory

The core idea of the RPT is that recurrent (feedback, reentrant) processing (RP) is essential and perhaps even sufficient for conscious experience. Victor Lamme developed RPT based on experiments on primate vision,^{98–100} leading to a general theory of visual consciousness.^{27,101,102} A key element of RPT is that it regards attention and consciousness as separate (orthogonal) processes (Figure 3, top, horizontal, and left, vertical arrows).⁸ This distinguishes RPT from many other theories, including GNWT, PP, and the attention schema theory (AST¹⁰³), but not from NREP.

RPT outlines four stages of visual processing (Figures 3A–3D): two are unconscious (stages 1 and 2) and two others (stages 3 and 4) are regarded as conscious, but different (phenomenal vs. AC; PC vs. AC).¹⁹ RPT also relates PC and AC to attention and different forms of memory (Figures 3C and 3D), thus pioneering attempts to reconcile RPT and other theories.^{8,25,105}

Stage 1

Local (“superficial”) feedforward processing: a fast feedforward sweep (FFS, ~100–200ms) of activity carries information from primary to higher visual cortices, where image features are extracted. However, if not boosted by attention, processing stays local, failing to proceed to stage 2, yielding only “subliminal” processing¹⁰⁶ without PC or AC^{27,102,105} (Figure 3A). However, “non-feedforward” (recurrent/lateral/horizontal) processing can be elicited by the FFS at all cortical levels, thus triggering stages 3 and 4.

Stage 2

More global (“deep”) feedforward processing occurs when the stimulus is attended. If so, the boosted FFS propagates up to motor and prefrontal areas, leading to rapid (~200 ms) identifica-

tion of the main content. This can trigger immediate, unconscious (“automatic”) behaviors, such as hitting a tennis ball (Figure 3B),^{27,102} while the subject is unaware of both stimulus and action.

Stage 3

Local (“superficial”) RP is an intermediate processing stage where unattended, unmasked stimuli, mainly in the ventral visual stream,¹⁰⁷ cause phenomenal conscious experience (PC¹⁹). However, without attention, the recurrent processing is limited and does not provide interactions with prefrontal and motor areas. The limited recurrence implies that sensory stimuli do not enter working memory and further cognitive processes that are needed for reporting. Hence, there is PC without AC (Figure 3C).

Stage 4

Given sufficient time, attended stimuli lead to deep, global recurrent processing (Figure 3D),^{27,102,105} reaching higher levels, including prefrontal, executive areas, providing contents accessible to the brain’s executive systems (AC in addition to PC), for use in, e.g., working memory, reports, and other behavioral responses, i.e., “ignition” in GNWT (see above).

Within RPT, attention is crucial for determining which stimuli cause global (stage 4) RP and AC, vs. local (stage 3) RP and PC. Only with attention does the subject “know” and can use and report its experience (AC). RPT also claims that different forms of short-term memory map onto PC and AC. Whereas global RP engages robust, low-capacity working memory (WM), local RP engages large-capacity, very brief, sensory (iconic), and fragile memory (SM, FM; 0.1–few seconds) in sensory cortices, which is easily overwritten by other stimuli.¹⁰⁵ This explains, e.g., inattentive and change blindness, and why only a few visual items (held in WM) are reported after brief presentations with no cue.¹⁰⁸

Being developed by experimentalists, RPT is firmly rooted in empirical observations, unlike more philosophical theories. Numerous experimental results support RPT, e.g., from masking, change blindness, inattention, attentional blink, neglect, and extinction,^{8,101,102,105} but do not exclude other theories. An early (<200 ms) event-related potential (ERP) component

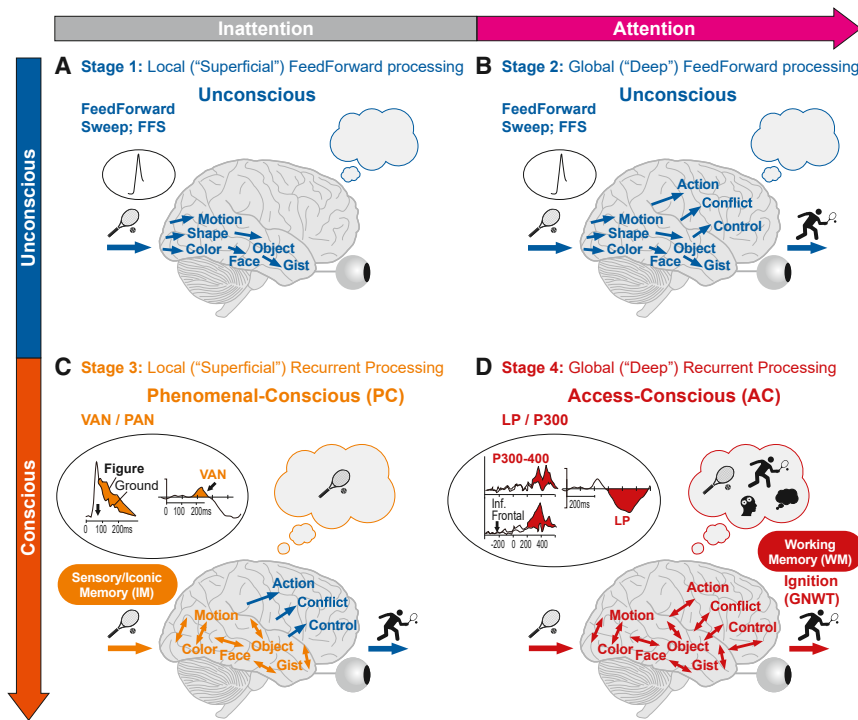


Figure 3. Recurrent processing theory (RPT) of visual consciousness

(A–D) Four stages (1–4) of feedforward (FF) and recurrent (RP) sensory processing in the primate (human) visual system, underlying two forms of consciousness; phenomenal consciousness (PC) and access consciousness (AC). Since the RPT focuses on vision, this figure illustrates visual processing only, although the general principles (stages 1–4, PC vs. AC, etc.) are assumed by RPT to apply also to other sensory modalities (see, e.g., Dembski et al.⁴³). (A–D) Stages 1–4: see [recurrent processing theory](#) in the main text for explanation. (A and B) show two stages of early unconscious processing (blue).

(C and D) show later, P-conscious (orange in C) and A-conscious (red in D) stages of processing. The oval insets I(C) and (D) show examples of electrophysiological (EEG/ERP) components that seem to reflect stage 3 processing (C, VAN: visual awareness negativity; PAN: perceptual awareness negativity⁴³), and stage 4 processing (D, P300-400 and LP (Late Positivity) are the late event-related potential (ERP) components that are both assumed to correlate with global recurrent processing according to RPT and also ignition of the global workspace according to GNWT (see [Figure 1B](#)), illustrating empirical and mechanistic agreement between these two theories. The PC content of stage 3 processing is stored only briefly (0.1–1 s) in a high-capacity sensory memory (for vision called iconic memory, IM; in C), whereas the AC content is stored for longer (>1s) in low-capacity working memory (WM; in D). The large arrow above (A and B) illustrates that attention vs. no attention is a separate dimension, orthogonal to conscious/unconscious (left, vertical arrow), according to RPT. Adapted from Lamme,^{8,27,102} and Koivisto and Revonsuo¹⁰⁴ (“VAN” in C) and 89 (“PAN” in C). See also Lamme.^{8,101,102,105}

(visual awareness negativity; VAN) seems to match (in timing and topology) local RP (Stage 3¹⁰⁴) and may reflect PC ([Figure 3C](#)). The late (~300 ms) positivity (LP) ERP wave (P300/P3b²⁹) shows characteristics of stage 4 and may thus reflect AC ([Figure 3D](#)),^{102,105} in agreement with GNWT.²⁸ Similar early ERP waves are found also for other sensory modalities, suggesting a general perceptual awareness negativity (PAN) reflecting sensory PC.⁴³

A challenge for RPT is whether RP “is sufficient for conscious experience.”^{8,102} Although several empirical challenges that may seem to falsify RPT are resolved by distinguishing PC from AC,^{8,19} the core problem remains: RP seems far more ubiquitous than consciousness, suggesting that RP alone is not sufficient for consciousness.¹⁰² Indeed, RP is not completely lost in apparently unconscious states (dreamless sleep, anesthesia, coma;¹⁰²; but see Sevenius Nilsen et al.⁵²). Lamme suggests that the “missing ingredient” may be neural plasticity, since RP may satisfy Hebb’s rule and thereby “differs fundamentally” from feedforward processing.^{8,109} Accordingly, conscious processing may possibly serve to enable perceptual learning.^{101,102}

Predictive processing & neurorepresentationalism

The key idea in PP, and more specifically neurorepresentationalism (NREP) as a theory of consciousness, is that the brain does not accomplish perception by extracting or copying knowledge from its surrounding world but builds up percepts and knowledge by constructing an internal world model. This construction happens by learning to make inferences about the most probable causes of the sensory inputs it currently receives. Thus,

“prediction” mainly pertains to inferences about the “here and now,” not necessarily the future. By comparing the effects of internal representations with actual sensory inputs, a prediction error can be computed, which subserves both perceptual learning and inference. When particular circumstances are realized in corticothalamic systems, inferential representations can become sufficiently extensive and enriched to correspond to conscious representations as occurring during perception, imagery, and dreaming.¹³

As a philosophical position, representationalism localizes perceived objects and features neither in the physical, external world, nor in the brain itself (e.g., if we are conscious of orange, that color is not literally *in* the brain: neurons do not turn orange). Instead, representationalism attributes these features to the conscious representation we experience as taking shape “outside the brain,” as if “projected” onto the surroundings and our body. NREP starts from this position to identify neural mechanisms underlying the genesis of PC as primary explanandum. NREP takes the computational principles of PP to build sensory-based models of inferential representation learning, positing that motor actions are not needed per se for consciousness.^{12,13,53} Herein, NREP differs from the active inference account of consciousness, which is also based on PP but does take motor actions (or “agency”) to be at the heart of consciousness.^{10,110} Below, we will focus on NREP or refer to PP-based accounts in general.

Widespread evidence suggests that prediction or representation is a key brain function in realizing a plethora of sensory, motor, and cognitive processes, some carried out consciously and

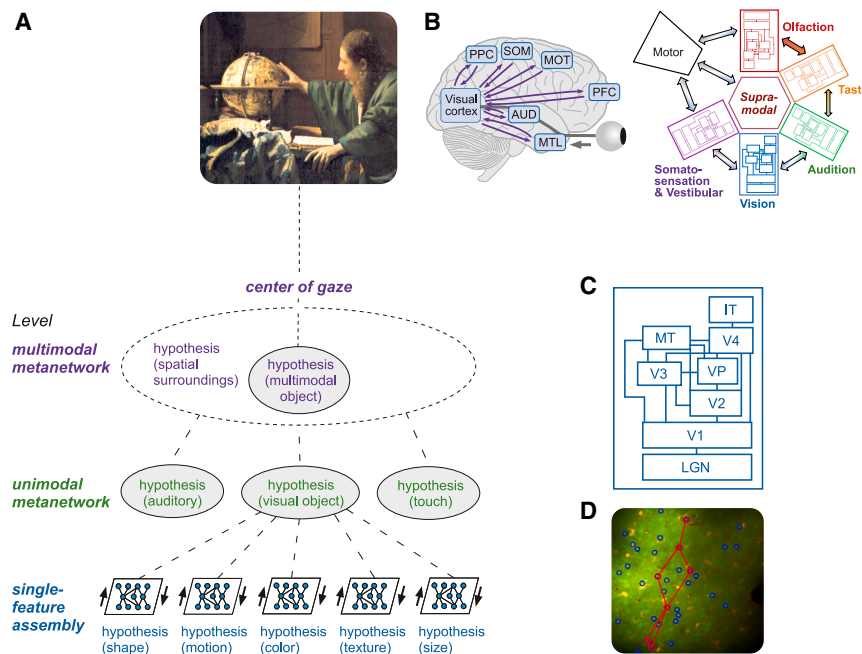


Figure 4. Neurorepresentationalism (NREP) in relation to predictive processing

NREP posits that conscious experience—exemplified by a fragment of Vermeer’s painting *The Astronomer*—arises across multiple levels of representations.

(A) At the bottom, we find single-feature assemblies as the lowest level of representation (example in D; assembly recorded in mouse visual cortex; cf. Montijn et al.¹¹⁸). These assemblies perform predictive-coding operations giving rise to hypotheses (representations) of singular features (which are visual in this case: shape, motion, color, etc.). These feature representations are combined into a hypothesis at the unimodal metanetwork level; in vision, this corresponds to joint representations coded primarily in the visual cortical hierarchy (C). At the multimodal metanetwork level, hypotheses from different modalities are combined, and integration across space is achieved. This level corresponds to phenomenal experience.

(B) Left: Macroscopic view on the brain depicting how the visual cortex communicates with other cortical areas in the genesis of visual perception (AUD: auditory cortex; MTL: medial temporal lobe; PPC: posterior parietal cortex; SOM: somatosensory cortex; for physically active subjects: MOT: motor cortices; PFC, prefrontal cortex, involved in goal-directed behavior and planning).

(B) Right: schematic of how different sensory-

cortical hierarchies are interlinked and, in the case of physically active subjects, communicate with motor and executive areas. Figure adapted from Pennartz¹³ and Pennartz et al.¹¹²

others not.^{12,110,111} Why are some representations conducive to consciousness but not others? In NREP, conscious phenomenology is characterized by five inalienable hallmarks. Together, these define conscious experience as a multimodal, situational survey of the agent’s world, including its body.^{12,13,53} This survey or “overview” (in a multimodal sense) subserves, but does not equate to, complex decisions and planning of goal-directed behavior. Multiple modalities contribute to this spatially encompassing survey, not only by way of multisensory integration, but also by modality segregation, such that modalities and submodalities are experienced as qualitatively different. In neural terms, this survey is mediated by multiple sensory-cortical hierarchies (somatosensory, auditory, visual, etc.) that each have a feedforward/feedback architecture and interact with each other, and with motor and planning systems¹³ (Figure 4). In this very large system, not only the interplay of feedforward and recurrent processing is required for PP in a single modality, but also long-range and lateral interactions between the modalities and, in moving subjects, motor systems.¹¹² Indeed, a major part of the recurrent wave of visual cortical responses to perceived stimuli can be explained as motor-related activity (e.g., Allen et al.¹¹³ and Oude Lohuis et al.^{114,115}). This configuration for “superinference” is not found in other brain structures capable of generating predictions (e.g., cerebellum); hence, these are not implied in consciousness. NREP considers several neural implementations of deep PP, varying from strictly hierarchical to heterarchical systems.^{116–118} Under this view, consciousness and attention are different processes: whereas multimodal topology provides for basic conscious experience, attentional systems highlight a limited processing stream within that overall topology.

A distinguishing aspect of NREP is that conscious representations have a multi-level nature. This particular feature offers an

approach to the Hard Problem of consciousness¹: the key idea is that “levels” should not only be understood in terms of neuroanatomy or aggregate size (consciousness may be realized in different substrates, even artificial ones). Instead, phenomenology emerges at the top of a hierarchy of conceptually different levels, having single neurons at its bottom and running up higher via assembly-level, unimodal, and multimodal network representations (Figure 4).

Dendritic integration theory

DIT^{14,56,120} emerged out of investigations into the role of recurrent processing in the cerebral cortex. It focuses on deep-layer pyramidal neurons of the cortex and the non-linear processes in their dendrites that allow these neurons to serve as the fulcrum, or nexus point, for the convergence of cortico-cortical and thalamo-cortical information flow. It is (currently) based on six core observations, primarily on subcortically projecting (extratelencephalic, ET) layer 5 (L5) pyramidal neurons in rodents:

- (1) L5 pyramidal neurons associate information arriving at separate dendritic compartments through a non-linear apical dendritic process leading to burst firing.^{120–122}
- (2) Suppressing these apical dendritic processes (e.g., dendritic calcium spikes) in pyramidal neurons suppresses conscious perception.¹²³
- (3) Suppressing the subcortical targets (particularly higher-order thalamus) of L5 pyramidal neurons suppresses conscious perception.¹²⁴
- (4) Loss of consciousness due to anesthesia correlates with a decoupling of the apical from the basal dendrites in these neurons.¹²⁵

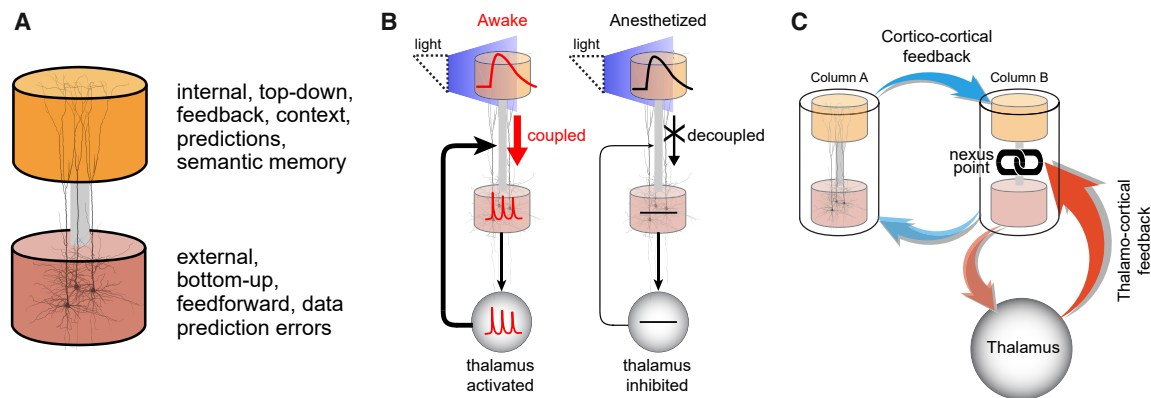


Figure 5. Dendritic integration theory (DIT)

DIT proposes that a specific cellular mechanism underlies consciousness and the global activation patterns related to conscious processing (A) Cortical pyramidal cells integrate information from the apical and basal dendritic compartments. Between the two compartments, there is a junction (nexus point) that acts like a gate and controls the coupling between them. (B) Left, this coupling mechanism gates information flow within the thalamo-cortical system. In the conscious state, the gate within the pyramidal cells is open and allows signals to propagate to higher-order thalamus that itself projects to mid-cortical layers maintaining coupling across the apical dendrites. Right, during the unconscious state, cortical pyramidal cells are decoupled, leading to a breakdown of activity propagation in the thalamo-cortical system. (C) The apical dendrites of deep-layer pyramidal neurons that project to higher order thalamus therefore represent a nexus point for the maintenance and influence of feedback across the whole cortex. Disruption at this nexus point (the coupling zone of the apical dendrite) by anesthetics leads to widespread loss of feedback in the whole system.

- (5) Suppressing higher-order thalamus decouples apical from basal dendritic compartments.¹²⁵
- (6) Neuromodulatory state controls dendritic coupling in L5 pyramidal neurons.¹²⁵

Taken together, DIT argues that the best explanation for these convergent findings is that the apical axis of subcortically projecting pyramidal neurons serves as the fulcrum for both cortico-cortical and thalamo-cortical information flow, which are both (and simultaneously) fundamental to mammalian consciousness.^{14,56,57} Thus, it follows that multiple processes associated with the integration of synaptic inputs in the dendrites of these L5 neurons are fundamental for understanding how their output can be so tightly linked to consciousness.

The fundamental insight of DIT is that L5 cortical pyramidal neurons both receive and process categorically different information in two distinct regions: the apical and basal compartments that are separated both physically and functionally. The electrical separation between these two compartments, which is already substantial, is accentuated by a high “leak” conductance that allows apical and somatic computations to be essentially isolated from each other. However, it is now well established that these neurons can associate information arriving at both compartments through highly non-linear processes dependent on voltage-sensitive ion channels in the dendrites^{121,122} and controlled by very specifically targeted inhibition and neuromodulation impinging on subdomains of the dendritic tree.¹²⁶ DIT hypothesizes firstly that the basal compartment of these neurons receives predominantly feedforward information that relates to the specific cognitive or sensory feature being processed in the column the L5 pyramidal neuron. Second, it proposes that long-range feedback information impinges predominantly on the apical compartment of the same neurons providing information (or context) about pro-

cesses occurring elsewhere in the brain, including other areas of the cortex.

The requirement for neurons with multiple compartments in order to explain information flow in the mammalian brain is a distinctive characteristic of DIT that leads to many corollaries and predictions.^{14,120,127} For instance, a corollary of DIT is that prior knowledge (e.g., semantic memory), that relates to the associations made with perceptual processes, should be dependent on the influence of synapses connecting to the apical compartment. Evidence that consolidated semantic memories are stored in the connections in layer 1 and their influence on apical dendritic activity was recently obtained in rodents.^{128,129} Furthermore, the observation that higher-order thalamo-cortical projections sustain the coupling between apical and basal dendrites has been proposed to explain the simultaneous completion of cortico-cortical and thalamo-cortical loops and therefore can serve as the mechanism for (anesthesia-induced) loss of consciousness¹⁴ (Figure 5). This, in turn, leads to the hypothesis that directed attention, an aspect of AC, is controlled by the selective maintenance of dendritic coupling in different cortical columns, which in turn can regulate the coupling of pyramidal neurons in posterior cortical regions relating to PC. The regulation of coupling by inhibitory and neuromodulatory input also makes specific predictions about the relationship between brain state and consciousness, including dreaming.¹³⁰

COMPARING THEORIES ACROSS AGGREGATE LEVELS

Here, we compare all five theories, discussing their similarities and differences, as well as compatibility and complementarity. All theories propose neural mechanisms that are often overlapping or similar, providing potential points of convergence (for differences, see also Box 3). Neural mechanisms can be described at different levels of organization, from the micro-level of (sub)cellular

Box 3. How to move on with consciousness theories: Remaining differences, criticisms, and approaches

The current integrative approach to neural theories of consciousness does not imply that one should be naive about remaining fundamental differences and criticisms. In addition to the differential focus on PC vs. AC, several other key differences deserve highlighting. Whereas GNWT, RPT, PP/NREP, and DIT emphasize the role of information processing, computation, and messaging in generating experiential content, IIT identifies an experience with the cause-effect structure supported by a neural substrate. This requires considering not just the substrate's actual state but its repertoire of potential states,⁹⁵ whereas the other theories explain specific conscious experiences as actual states, without taking into account all possible states of the system. By consequence, IIT does not identify consciousness with a particular function (see Tononi and Koch¹³¹), while the other theories cast neural underpinnings of consciousness in functional terms, e.g., information distribution (GNWT), learning and feature binding (RPT), predictions (PP, DIT, and NREP), and multimodal situational survey (NREP). However, the fact that these and other theories attribute functional aspects to consciousness does not equate them with computational functionalism (e.g., Pennartz^{12,13}).

One general critique that can be exerted on neural theories of consciousness is that, after considering the mechanisms illustrated in Figures 1–7, one may still ask, "... but why would these mechanisms give rise to consciousness"? This question refers back to the "explanatory gap" or "hard problem" that is dealt with by the various theories in different ways (see main text) but should also be taken to heart as serious critique. For instance, the commonly embraced principle of recurrent processing may be *necessary* for consciousness but is unlikely to be *sufficient*. Thus, current theories risk being underconstrained and will require further elaboration, based on empirical evidence and theoretical reasoning. A case in point is that a lack of constraints may give rise to differential permissiveness toward consciousness in non-neural substrates (for instance, IIT is relatively permissive; NREP and DIT are more restricted^{132,133}). Moreover, next to the distinction between "structure" and "message," the question of how messages acquire meaning (intentionality, "aboutness") remains lurking in the background (e.g., Pennartz,¹² Chalmers,¹⁵ and Searle¹³⁴).

How to move on through this seemingly intimidating philosophical minefield? One approach is the adversarial testing of theories on points where they make differential, empirically testable predictions. Five such projects have been sponsored by the Templeton World Charity Foundation, including tests of GNWT, IIT, PP/NREP, and RPT. Experiments target, for instance, the anatomic loci representing conscious content, stimulus duration and ignition-like processes,⁵ and the roles of inactive neurons and motor activity in visuospatial perception (cf. <https://www.templetonworldcharity.org/projects-database/0646>). However, we argue this adversarial testing will benefit from complementary approaches because it may (1) test predictions that are tangential to the theories under scrutiny, being unable to test their core assumptions, and (2) be over-reliant on falsificationism, whereby a clear-cut reject/accept dichotomy on theories is expected. What appears to be false under one theory may be perfectly acceptable under another one, and improvements of theories may rather follow incremental evidence in favor or against certain accounts, in agreement with abductive inference ("inference to the best explanation of the data"^{135,136}) and Bayesian evidence accumulation.¹³⁷ We thus advocate a complementary approach whereby new empirical evidence is accrued to combine the best supported elements of various theories into an improved framework with stronger predictive power. Finally, we also expect these combined approaches to gradually push back some of the seemingly deep conceptual (and sometimes metaphysical) divides. For instance, defining more empirically based constraints on theories is expected to restrict the range of complex natural systems to which consciousness may be reasonably attributed.

processes, through the meso-level of within-area circuits, to the macro-level of interactions among brain areas. Although different theories of consciousness have their main focus at different levels, these levels are dealt with as being hierarchically organized, making it possible to describe their mechanistic relations and discern differences and commonalities (Figure 6).

Micro-level: Cellular and subcellular levels

A common assumption in much of consciousness science is that consciousness can be best understood at the level of large-scale networks and their computations. However, in the brain global activity is dependent on processes happening at the level of circuits, cells, and even subcellular processes. Here, we take a multiscale approach^{13,140} trying to unravel some of the processes underlying consciousness, starting from the cellular level (Figures 5 and 6).

Recurrent processing and integration—two powerful principles operating at all levels in the nervous system—are both central in most theories of consciousness, including DIT, RPT, GNWT, IIT, and PP/NREP. DIT, in particular, posits that recur-

rence and integration are inextricably linked through the non-linear processes in the apical dendrites of pyramidal neurons that are a critical nexus for both cortico-cortical and thalamo-cortical loops (Figure 6C). In the awake, conscious state, layer 5 pyramidal cells in cortex are in a "coupled state" so that feedback arriving at the apical dendrites can influence somatic processing at the cellular level, and thus the internal information contained in feedback can be reintegrated into the whole system in a massively parallel fashion.^{14,125} According to DIT, dendritic integration associated with conscious states ensures that the whole thalamo-cortical system is in a state where activity can be integrated over space and time.⁵⁶ This could mean that activity can propagate and spread through cortico-cortical recurrent connections, and pyramidal cells can activate higher-order thalamic nuclei, which, in turn, can reinforce the activity in cortex.^{14,57} According to DIT, the evolving activity gives rise to the rich global dynamics of consciousness measured at the macro-scale.¹³⁰ DIT assumes that such global coordination cannot happen if L5 pyramidal cells are decoupled, e.g., during anesthesia or dreamless sleep¹³⁰; in this case, activity can propagate

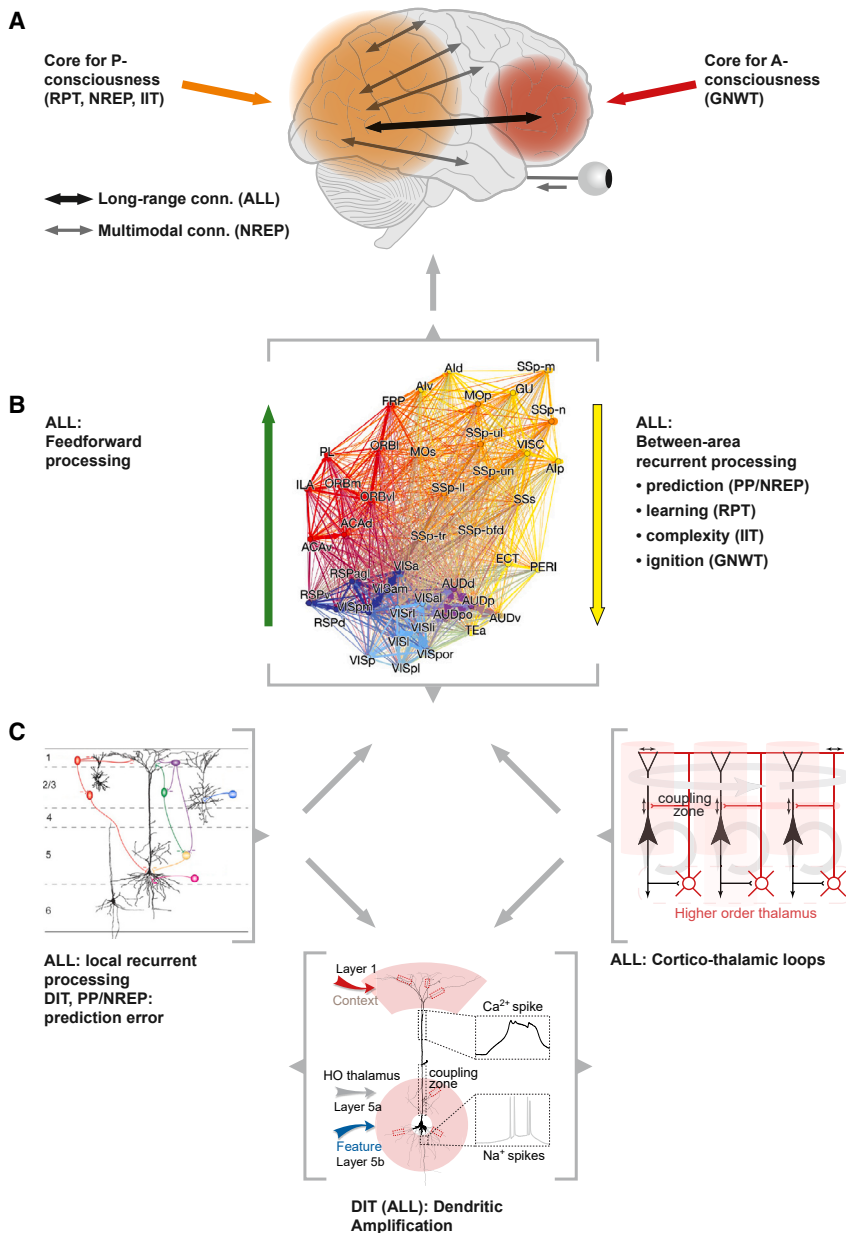


Figure 6. A hierarchy of aggregate levels implicated in consciousness by various theories

The lowest level (bottom part of C) is represented by cortical pyramidal cells, where the processes of apical dendritic amplification (illustrated with a dendritic Ca^{2+} spike and dendritic-somatic coupling) are proposed, by DIT in particular, to lay a cellular foundation for conscious processing “ALL” means that the other four theories comply, although not having proposed this foundation specifically. The next aggregate levels are local (within-area) recurrent processing in cortical microcircuits (C, upper left; from Suzuki et al.¹³⁸) and cortico-thalamic loops (C, HO: higher order).

(B) The multi-area level is represented by a large-scale connectivity map of mammalian cortex¹³⁹, where theories agree that recurrent processing is important for consciousness, albeit for different reasons.

(A) At the macroscopic scale, the orange-shaded areas indicate the posterior cortical areas involved in phenomenal visual experience (according to RPT, NREP, and IIT; including parietal and temporal areas), while GNWT proposes that prefrontal areas, with long-range connections to more posterior areas (black arrow), are key to AC. Basal communication between visual areas and other sensory modalities and motor systems has been postulated by NREP (gray arrows). Gray brackets outline a particular aggregate level; gray arrows indicate hierarchical relationships between levels.

apical and basal compartments of neurons in the frontoparietal network are decoupled.^{14,125} Thus, sufficient activity propagation and integration in the thalamo-cortical system to cause large-scale ignition may only occur when L5 pyramidal cells are in the coupled state. Similarly, local dendritic integration may enable the build-up of joint neural activity patterns in multiple sensory modalities and across scales that is crucial in NREP. In sum, subcellular dendritic integration is probably an enabling condition for large-scale integration across cortical or thalamo-cortical networks. Thus, the mechanisms proposed by DIT can be incorporated in

locally within the cortex but cannot involve larger thalamo-cortical loops.¹³⁰ Thus, large-scale brain dynamics (measured by fMRI, EEG, etc.) related to consciousness depend on dendritic integration happening at the subcellular level.

In DIT, this local mechanism lays the foundation for processes related to consciousness at other levels of processing.¹⁴ For instance, IIT predicts that decoupling the pyramidal cells causes a drastic breakdown in the cause-effect repertoire within the main complex, reducing the Φ value, and hence consciousness.⁷ Thus, dendritic integration provides a specific cellular mechanism for controlling information integration in the thalamo-cortical system, in line with computational modeling.⁵⁴ Similarly, if dendritic integration is central for information flow in thalamo-cortical networks, global workspace activity would collapse if the

any theory of consciousness that relies on such large-scale integration.

Dendritic integration also provides neurobiological nuances to theories of consciousness that have a dual-stream architecture (such as RPT and classical PP; cf. Rao and Ballard¹⁴¹). In particular, the knowledge about dendritic integration mechanisms specifies how the two streams interact at the level of single pyramidal neurons. According to RPT, consciousness arises from recurrent interactions between different levels of processing in the brain. This idea is compatible with DIT as decoupling of pyramidal neurons leads to a breakdown of RP in the brain.^{14,125} According to PP-related accounts of consciousness (e.g., NREP), predictions sent from higher to lower levels of processing try to “explain away” activity propagating from lower to higher

levels by suppressing the sensory signals that match the predictions. What cannot be predicted is passed on to higher levels as prediction errors. Although, according to DIT, the biophysical processes intrinsic to the L5 pyramidal cells prescribe them to amplify the match between the two streams,^{14,120} there are multiple ways in which the biophysical properties of pyramidal neurons, or adjoining cortical interneurons, may be co-opted into a PP framework (Figure 6C^{135,142–145}). All of these variants are compatible with principles of dendritic integration as a cellular basis used to generate conscious experience. However, according to DIT, recurrence is only one requisite for consciousness. Another key component is the dendritic integration mechanism that enables the incorporation of this feedback information into the ongoing activation patterns of the thalamo-cortical system.

Meso-level: Local circuits and recurrent processing

How may local processing within and between corticothalamic microcircuits contribute to consciousness? Besides the large L5 neurons highlighted in DIT, multiple neuron types connected by millions of synapses form immensely complex microcircuits within and across cortical columns,¹⁴⁶ the intricacy of which is easily underestimated. This dense complexity, multiplied by similarly rich inter-areal and cortical-subcortical interactions, may transcend human intuition and may (unpredicted and unexpectedly) contribute or lead to the seemingly non-physical phenomena of subjective experience (PC). The local, nested intricacies within and between cortical neurons and microcircuits are relevant for all the five theories discussed here and may, e.g., contribute to high Φ cause-effect structures underlying P-consciousness according to IIT.⁹⁵

In general, feedback (RP), both within microcircuits¹⁴⁷ and more globally, is an important feature in most consciousness theories (RPT, GNWT, IIT, PP/NREP; see [macro-level: multi-area systems](#) and [overarching concepts: richness and complexity](#)) and also contributes to processes linked to consciousness, such as WM (Figure 3) and sensory adaptation.²⁰ However, what could local processing contribute specifically to consciousness? Building on the DIT-based predictions in [micro-level: cellular and subcellular levels](#), PP/NREP has proposed that the computation of low-level predictions and prediction errors strongly depends on the local, intracolumnar circuitry that regulates information trafficking in L5, but also L2/3 pyramidal cells. Accumulating evidence suggests a role for superficial pyramidal cells in representation (prediction) and error coding, while L5 cells may function especially in representation coding.^{23,118,135} In particular, somatostatin-positive interneurons have been implicated in error computations.^{148,149}

Although RPT focuses primarily on RP within and between visual areas,^{150–152} Lamme also discussed whether even RP between interconnected neuron pairs may perhaps generate some consciousness. However, the ubiquitousness of feedback led Lamme to propose that some other process, perhaps neural plasticity, may also be needed for PC.¹⁰²

GNWT implies that conscious access (AC) depends on RP, as its “ignition” and reverberant activity realizes access to conscious content.²⁸ Early versions of GNWT proposed that L2/3 pyramidal cells with long-range connections (not L5 neurons as in DIT) are most crucial for AC, but more recent versions also

consider the loops with subcortical structures. The Dehaene et al.’s model of AC,¹⁵³ which included thalamo-cortical columns with laminar feedforward and feedback projections, simulated ignition with a P300 wave and attentional blink. The network model¹⁵⁴ proposed how a short-lasting (<1 s) perceptual buffer can be accessed and broadcasted to WM and consciousness (AC), by distinct operational modes of the same circuit. According to IIT, feedback (RP) is needed for consciousness. Based on how Φ is calculated to “quantify” PC, artificial systems with purely feedforward architecture have $\Phi = 0$, and hence lack consciousness (PC), even when their overt behaviors (input-output functions) are indistinguishable from that of conscious ($\Phi > 0$) systems with feedback^{94,95} (but see Doerig et al.¹⁵⁵). Accordingly, consciousness should fade even in the presence of the appropriate architecture if feedback connections are functionally disabled. On the other hand, since IIT holds that the causal structure of brain circuits, rather than their activity, determines Φ and PC, circuits that are causally integrated within the main Φ complex can contribute to PC, even when its neurons are silent. This counter-intuitive prediction, not shared by other theories, is being tested in an adversarial collaboration between IIT and PP (see: www.templetonworldcharity.org/projects-database/0646; Box 3).

Haun and Tononi propose that “grid-like” neuronal circuits in sensory cortices, resembling lattices, form a substrate for spatial experience.⁹⁷ Grid-like connectivity is found in both visual and other sensory areas, but PP accounts generally interpret the topographic structure of sensory cortices differently than Haun and Tononi, viz. as a layout for making inferences from one part of sensory (e.g., visual) space to another part (cf. Knierim and Van Essen¹⁵⁶). In addition, NREP proposes that subjective spatial experience arises from the multimodal integration between maps set in different spatial frameworks (retinotopic, craniotopic, and allocentric¹²).

In addition to representing information content, local circuits undergo changes in neuromodulatory state with different “levels” of arousal (reflecting one of the main “classical” “dimensions of consciousness¹⁵⁷”; but see Hill and Tononi,¹⁵⁸ Laureys et al.,¹⁵⁹ and Bayne et al.¹⁶⁰). However, according to a more recent view of neuromodulatory states, these can be regarded as “background conditions” or enabling factors with only an indirect relation to consciousness per se (not specifying conscious content).¹⁶ The effects of changes in local “arousal” state and network complexity caused by anesthesia or neuromodulation have been modeled in thalamo-cortical network models¹⁵⁸ and seem roughly compatible with all five theories discussed here, although dendritic mechanisms (DIT) were not included so far. Such modulatory effects have been assessed even in isolated cortical slices (“meso-scale³⁶”) by modified versions of perturbational complexity index (PCI) and related complexity indices similar to those used in humans (see Box 1; [overarching concepts: richness and complexity](#)). Although these measurements do not allow any inference about consciousness in brain slices and rodents, they can inform about cellular and network mechanisms responsible for brain state changes with loss and recovery of complexity.¹⁶¹

Macro-level: Multi-area systems

When ascending along the scale of neuroanatomy, we arrive at the question of how brain mechanisms underlying conscious

experience operate at the level of brain areas and systems composed of various areas. The five theories reviewed here can be subdivided into three partitions on this point. First, IIT, PP-NREP, and RPT have in common that, for conscious vision, they all emphasize the foremost role of the posterior cortex (more specifically, those occipital, parietal, and temporal areas implied in vision). IIT holds that a maximal Φ value will be reached primarily in posterior cortex (“hot zone”)^{78,95} when dense reentrant interactions occur between higher and lower visual areas.⁷ PP-based accounts, including NREP, propose that conscious vision begins with the interaction between the feedforward signaling stream (mainly conveying error information) and the recurrent stream (conveying top-down predictions) within the visual cortical hierarchy,^{23,141,135} although alternative schemes are possible (e.g., Heeger¹⁶²). Both IIT and NREP propose that, in a changeable visual environment, different, switchable subsets of brain areas will engage in feedforward and feedback interactions to ensure integration of visual features, depending on the visual submodalities being part of the subject’s experience.^{12,88} RP within the visual cortical system is also key to RPT in reaching PC, while global recurrent processing is needed to accommodate attentional effects and reach AC.

GNWT attributes a prominent role to the prefrontal and parietal cortices in conscious vision, which is in line with its focus on AC rather than PC, and its translation of this focus into executive functions such as behavioral responding, attention, WM, and valuation—all functions associated with the prefrontal-parietal network.⁶ Third, as we have seen, DIT focuses more strongly on the cellular basis of consciousness. In that sense, DIT is “neutral” with respect to the “anterior versus posterior” debate on neural substrates of conscious vision. Notably, all five theories agree that the neocortex constitutes the key structure for generating conscious content (as opposed to other theories highlighting centrencephalic structures^{163–165}) and that thalamic nuclei are vital in sustaining consciousness. Moreover, all five theories posit that, next to local or short-range interactions, long-range communication in corticothalamic systems is required for consciousness.^{6,7,13,105}

Importantly, however, both RPT and NREP argue that information processing within the visual cortical hierarchy is not sufficient to achieve conscious vision (either understood as PC or AC). According to RPT, widespread recurrent processing requires attention, and for this, interactions between the visual cortex and prefrontal areas are deemed necessary.¹⁰² In this sense, RPT is more compatible with GNWT than their different definitions would suggest: they both attribute a key role to the prefrontal-parietal network regulating attentional signals.^{102,105} PP accounts interpret attention in relation to large prediction errors and low decision confidence, prompting the subject to seek more information about perceptually uncertain regions of sensory space (e.g., Kanai et al.¹⁶⁶).

Among the hallmarks of conscious experience delineated by NREP, we find the combination of dynamics and stability, where “stability” refers to the capacity to perceive the outer world and its objects as a stable whole, despite the occurrence of eye, head, and body movements.^{12,13} To achieve visual stability in physically active subjects, interactions between the visual and motor cortices are required: the visual system receives motor

signals (along with proprioceptive and vestibular reafference) or predictions about which sensory changes are happening or about to happen.^{112,166} Moreover, NREP postulates that lateral, multisensory interactions between sensory-cortical systems are necessary to ensure the qualitative, multimodal richness that characterizes consciousness.¹⁶⁷ By consequence, vision is not mediated exclusively by the visual corticothalamic system as classically defined but by the “extended visual cortex,” which includes many satellite areas. IIT and RPT are *inter alia*, not in contradiction with these additional interactions, but do not make a particular point about this.

Thus, at first sight, three theories appear to agree at least globally on the macro-anatomy of conscious vision, while the other two theories either express a different view—in relation to differences in background definition (GNWT)—or stay neutral (DIT). This global compatibility of IIT, RPT, and NREP does not mean they have the same scope or are similar. For instance, they use different rationales as to *why* recurrent feedback in posterior cortex would be required for consciousness (IIT: to maximize Φ ; RPT: to enable plasticity, learning, and feature binding; NREP: to close the loop for representation learning and inference). Another key difference lies in the presumed role of inactive neurons (in IIT, these can contribute to conscious experience unless they are disconnected from the network or inactivated⁵⁵; other theories do not adhere to this). Nonetheless, we may conclude that, also at the level of multi-area systems, the theories are not as incompatible as has sometimes been assumed.

Overarching concepts: Richness and complexity

A growing body of experimental evidence suggests that the patterns of neuronal interactions that are relevant for consciousness must not only be tightly integrated through feedforward and feedback connections but must also be differentiated and information rich.^{32,168,169} This notion of brain complexity—defined as the coexistence of integration and differentiation—was at the core of IIT’s original notion of Φ ^{89,90} and remains central to the empirical predictions of IIT in its mature form.⁹⁵

Brain complexity has also been embraced by other theories more recently. For instance, a key tenet of the GNWT is the ignition of widespread activations. Hence, IIT and GNWT share in practice the basic requirement of neuronal integration, albeit motivated by divergent theoretical underpinnings. Notably, the original formulation of GNWT made no explicit reference to information-related measures, differentiation or complexity and mainly focused on neuronal processes such as fronto-parietal synchrony and the P3b. Subsequently, however, empirical studies by GNWT proponents shifted the focus to the richness of dynamical repertoire,^{30,170} and, more recently, the notion of complexity has been explicitly incorporated as a key element within the GNWT, based on the rationale that distributed modules broadcast different types of sensory information.¹⁷¹

As defined by IIT, brain complexity bears clear mechanistic relationships with the microscopic and mesoscopic processes postulated by DIT and RPT. It requires effective reentrant interactions to ensure that differentiated groups of neurons cohere tightly and engage in a rich repertoire of cause-effect interactions. Both dendritic decoupling and the breakdown of recurrent interactions are expected to result in dramatic decreases in the

brain's capacity for integrated information. Along these lines, the collapse of complexity observed experimentally during sleep and in unresponsive wakefulness syndrome ("vegetative" state) patients has been associated with a breakdown of feedback interactions.^{37,172–174}

Similarly, complexity is inherent to the multimodal richness and integration, postulated by NREP, although there are also important differences. In IIT, brain complexity explains general phenomenal properties (unity and richness) that are "immediate" and common to each and every experience, while both NREP and PP in general take conscious experience to be representational—a best guess "simulacrum" of external reality.^{12,13,175,176} In the neurorepresentational view, complexity translates into a multifaceted phenomenon comprising multimodal, qualitative richness of conscious content, spatial extendedness and resolution, intensity and attention, and finally temporal depth—all of which are understood to be conducive to subsequent action planning.^{13,53} In this respect, there are many ways of defining and operationalizing brain complexity, and more efforts are warranted to specify underlying assumptions and extend notions of complexity into an exhaustive taxonomy.³²

Nonetheless, the recent convergence on the general notion of complexity⁸⁸ has interesting implications—especially for empirical applications in unresponsive patients^{170,177,178} (Box 1). Widespread theoretical agreement may increase such confidence and justify inference in more challenging cases. Furthermore, complexity and information richness can be studied at multiple levels (Figure 6), from human extracranial and intracranial recordings to rodents^{38,169,179–181} and cortical slices.¹⁸² Outside of the human brain, in which they are calibrated, current complexity measures do not allow any inference on the level of consciousness. Yet, they offer the prospect of drawing a mechanistic link between neural processes occurring at different scales, from microscale dendritic integration to mesoscale recurrent processes and macroscale inter-areal integration. Finally, this convergence around complexity may represent a common background to contrast different frameworks at a finer granularity.

Where and how in the brain is complexity instantiated? For IIT the cause-effect structure of grid-like cortices with recurrent pyramidal-like, convergent-divergent connections (such as those found in posterior areas) (Figure 7B) is thought to correspond to a maximum of complexity and thus be sufficient for conscious experience.^{96,97} PP-NREP shares with IIT the quest for the neural basis of the percept of spatial extendedness but seeks this in the integration between retinotopic, craniotopic, and allocentric frames of reference used by the occipital, temporal, and parietal cortices¹²; these different directions may prove to be compatible. Another key question is: why do some theories of consciousness focus on the complexity of ongoing neuronal activity, while others emphasize the complexity of the underlying causal structure?¹⁸³ It will be important to clarify this aspect, as it entails not only different approaches to measurement (observational vs. perturbational) but also a substantially different understanding of the kind of information that matters for consciousness (an extrinsic message vs. an intrinsic cause-effect structure) (Box 3).^{7,184,185} Finally, how do current theories consider the possibility that analyzing the structure and complexity of causal interactions within the brain may also be useful in the search for content-specific neural corre-

lates of consciousness? Addressing this question is challenging, but attempts have been made in this direction.^{97,167,176}

OVERALL ASSESSMENT: TOWARD AN INTEGRATED THEORY OF CONSCIOUSNESS?

Is it realistic to reconcile theories, or even aspire to a unified theory of consciousness? At first sight, as we have seen, at least some of the theories seem to be so different (Box 3) that bridging them appears insurmountable. This has even led to accusations of illegitimacy between competing theories.¹⁸⁷ However, we take the standpoint that the existence of multiple theories is a sign of healthiness in this nascent field and that paying close attention to the explananda can disentangle the core claims such that multiple theories can simultaneously contribute to our understanding. Fundamental differences in the focus and explananda of theories may sometimes make them more compatible. For example, GNWT focuses on AC, whereas RPT, IIT, and PP/NREP aim to explain primarily PC. An integrated theory of consciousness will have to explain PC in neural terms but also how conscious content is distributed and used for subsequent action planning, WM, valuation, attentional regulation, etc., which is the domain of AC. Modeling efforts in this direction are indeed being made, for instance, to integrate PP accounts with global neuronal workspace operations.^{188,189} We also saw how RPT can be integrated, in principle, with GNWT in a joint effort to include AC. As IIT already acknowledges the role of prefrontal areas for access, there is no contradiction with GNWT regarding its "anterior" focus for AC. Still, some will argue that GNWT and IIT cannot both be right because they disagree on the nature of what consciousness is (see also Box 3).

A further point of potential conceptual convergence lies in high-level, complexity-related concepts such as information integration (IIT), global information availability (GNWT), multimodal richness in multi-level representations (NREP), recurrent processing at different anatomical scales (RPT), and integration in corticothalamic loops (DIT). Although, at face value, these ideas differ remarkably, they are all fundamentally based on different aspects of interconnectedness, recognizing that the richness of neuronal interaction, relying on short- and long-range connectivity in corticothalamic systems, is an important factor for significant degrees of consciousness (Figure 7).

These points of convergence and related considerations give reason to fuel a certain hope that eventually an integrated theory of consciousness—or at least a related set of interdigitating concepts and working hypotheses—may come within reach within the next years or decades. However, it must also be recognized that there are still vast areas of theoretical and empirical investigation that remain to be unlocked (Box 3). The realization that we are only scratching the surface of this great question is humbling, but consistent progress is now apparently being made on several fronts. The state of consciousness research at the moment may be compared with a multifaceted object examined from different angles, each stressing different aspects based on individual vantage points. A parallel can be drawn with quantum mechanics, where there is also agreement on observations and phenomena, but still widespread disagreement on how to interpret the relationship

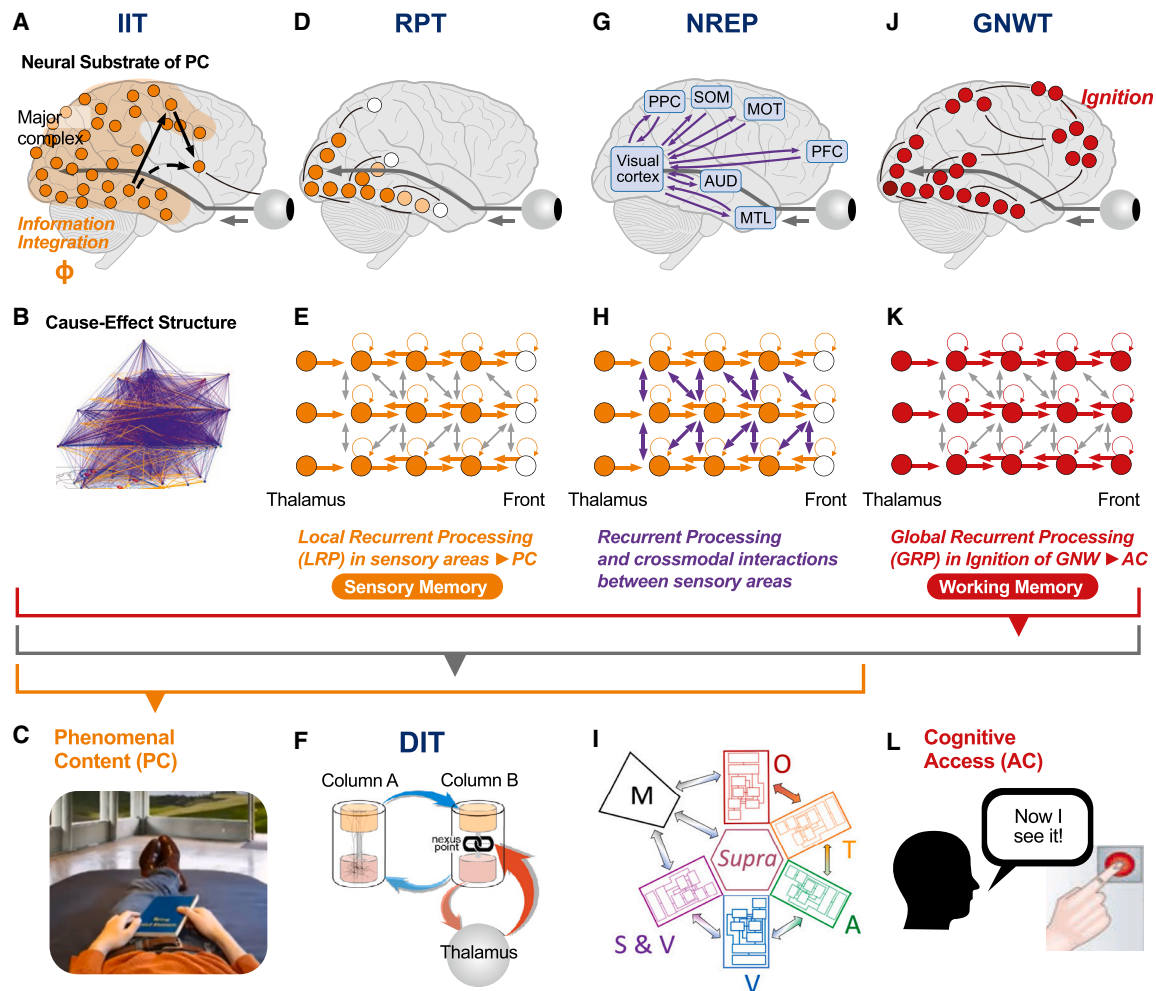


Figure 7. Comparing theories of consciousness

Summary figure illustrating how the five theories of consciousness discussed in this article are partly complementary, partly overlapping, and related to each other and to the core concepts and explananda: phenomenal consciousness (PC; orange elements and C) and access consciousness (AC; red elements and L), illustrated here mainly for vision.

(A and B) Integrated information theory (IIT), and its measure of integrated information (Φ), which quantifies P-consciousness according to IIT.

(C) PC illustrated by an image of a subjective visual experience from one eye (after Ernst Mach¹⁸⁶).

(D and E) Recurrent processing theory (RPT).

(F) Dendritic integration theory (DIT).

(G, H, and I) Neurorepresentationalism (NREP); see Figure 4B for abbreviations in (I).

(J and K) Global neuronal workspace theory (GNWT).

(L) Access consciousness (AC) illustrated by behavioral and cognitive responses to sensory stimuli.

We argue that the five theories are at last partly complementary but differ in their main focus: IIT, RTP, and NREP aim mainly at explaining PC (orange bracket under B, E, and H), whereas GNWT aims mainly at explaining access consciousness (AC), which is also an explanandum for the other theories (red bracket under B, E, H, and K).

(F) DIT is proposed to provide a cellular basis for brain network functions that appear to be necessary for consciousness (both PC and AC), and is compatible with all the other theories: IIT, RPT, NREP, and GNWT (gray bracket under B, E, H, and K). The horizontal arrays of colored circles with arrows in (E), (H), and (K) represent feedforward and recurrent processing streams; the middle row represents the central visual pathway from thalamus (LGN) to frontal cortex (front); and the upper and lower rows represent auxiliary visual pathways (RPT, GNWT, and NREP) or other sensory modalities (NREP). RPT posits that recurrent processing within the visual system, and its higher areas is central to visual PC. RPT and GNWT hold that recurrent processing must also involve prefrontal systems to reach AC. NREP emphasizes cross-talk between sensory modalities and submodalities to explain qualitative differences in PC, even if this concerns conscious vision (H, purple arrows).

between probabilistic, micro-level descriptions and our deterministic, macro-level everyday lives. Both cases may possibly reflect fundamental limitations of human imagination and intuition.^{12,190} This analogy underscores the urgent need for consciousness researchers to develop well-defined concepts and

a common terminology (e.g., not use merely the general term “consciousness” for different concepts like “AC” and “PC”) and to devise ways to translate disparate formulations into one another. Just as quantum theory benefited from a robust mathematical framework, so too should this field.

Finally, we feel there is broad and uncontroversial support for more experimental testing and development of empirical measures of consciousness. The relevance of this is 2-fold. First, novel, theory-based empirical measures such as PCI and other complexity-related measures can be usefully applied in clinical settings, such as in helping to classify unresponsive patients with consciousness disorders into different subgroups with more nuanced distinctions and different prospects for recovery and treatment (e.g., Demertzi et al.,¹⁷⁰ Casarotto et al.,¹⁷⁷ and Farisco et al.¹⁹¹). Second, empirical measures may help to distinguish or unify different theories. The high-complexity spatiotemporal EEG patterns evoked by TMS in conscious conditions are invariably associated with the presence of recurrent waves of activation,^{174,192} whereas low-complexity patterns are found during cortical up and down states characteristic of slow-wave sleep and anesthesia (e.g., Steriade¹⁶¹ and Olcese et al.¹⁹³). Hence, PCI is compatible not only with IIT postulates but in principle also with the neural processes that are relevant for GNWT, RPT, NREP, and DIT.^{13,14,102,171} The field is already benefiting from a broader palette of empirical tools and measures. DIT, for instance, introduced EEG and LFP markers that align with dendritic Ca²⁺ spikes during distinct consciousness phases.¹⁹⁴ PP/NREP proposes representational measures of consciousness (relying on techniques to decode conscious content that subjects are experiencing; cf. Goltstein et al.,¹⁹⁵; see also [Box 2](#)), and GNWT stands to benefit from report vs. no-report paradigms combined with single-unit recordings, fMRI and EEG-MEG (e.g., Sergent et al.⁴²). Complementary computational approaches can be utilized to foster further comparison and integration,^{116,188,196,197} and research is being extended from human to animal consciousness to benefit from *in vivo* two-photon imaging, large-scale ensemble recordings, complexity measures, and optogenetics to unravel causal substrates of consciousness (e.g., van Vugt et al.,⁶⁴ Oude Lohuis et al.,^{114,115} Takahashi et al.,¹²⁴ Suzuki and Larkum,¹²⁵ Arena et al.,^{38,179} O'Connor et al.,¹⁹⁸ and Storm et al.¹⁹⁹). Adversarial research is emerging as an important new approach in consciousness research,⁵ but it remains to be seen whether this will eventually lead to a Darwinian selection of a single, winning theory ([Box 3](#)). Here, we have advocated for a more unifying, integration-oriented approach, which seeks to empirically test and combine valuable elements from various theories.

In conclusion, we have argued that neural theories of consciousness (1) show differences that are often not as large or insurmountable as initially thought, (2) offer opportunities for combining and integrating various elements figuring at different levels of organization, and that (3) an integrated framework for understanding brain-consciousness relationships, guided by both empirical and theoretical advances, has strong potential to move the field forward in the next decade.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the specific grant agreement no. 945539 (Human Brain Project SGA3 to J.F.S., C.M.A.P., C.K., M.M., M.L., P.A., P.R.R., R.G., and W.V.). We would like to thank the following people for their helpful comments and suggestions: for C.P., Matthias Brucklacher, Umberto Olcese, Matthijs Oude Lohuis, Mototaka Suzuki, Con-

rado Bosman, Mathis Bassler, and Francesco Lässig; for C.K. and P.R., Xinyu Liu; for J.F.S., Bjørn E. Juel, Hedda Hassel Mørch, Andre S. Nilsen, and Christoph Höhnigsperger. In Oct.–Nov. 2023, we sent draft versions of subsections 2.1 to Giulio Tononi, 2.2 to Stanislas Dehaene, and 2.3 to Victor Lamme for comments; we thank them for their helpful suggestions. J.A. was supported by the Estonian Research Council grant PSG728. M.M. was supported by the European Research Council (ERC-2022-SYG – 101071900 – NEMESIS) and by the Ministero dell'Università e della Ricerca (PRIN 2022). C.M.A.P., P.R.R., and P.C.K. were supported by the Dutch Research Council (I; Cross-over Program 17619 “INTENSE”) and DBI2, a Gravitation program of the Dutch Ministry of Science. P.R.R. and P.C.K. were further supported by the European Union FP7 (ERC grant 101052963 “NUMEROUS”) and Innovation programme grant number 899287, “NeuraViper.” C.M.A.P. was supported by a grant from the Templeton World Charity Foundation (TWCF#0646). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Templeton World Charity Foundation. Andrea Pigorini is supported by Progetto Di Ricerca Di Rilevante Interesse Nazionale (PRIN) P2022FMK77. J.F.S. received additional support from the Research Council of Norway and the University of Oslo (UiO:life Science Convergence Environments). M.L. was supported by the European Research Council (101055340, Cortical Coupling).

AUTHOR CONTRIBUTIONS

J.F.S., P.C.K., R.G., P.R., M.M., M.L., and C.M.A.P. contributed to the conceptualization of this paper. J.F.S., P.C.K., M.M., M.L., and C.M.A.P. contributed to visualization. All authors contributed to writing, with major roles by J.F.S., P.C.K., J.A., R.G., M.M., M.L., and C.M.A.P. in the original writing, reviewing, and editing, and additional roles by A.P., P.A., W.S., W.D., and P.R. in advice, reviewing, and editing.

DECLARATION OF INTERESTS

M.M. is co-founder and shareholder of Intrinsic Powers. P.R. is a co-founder and shareholder of neurotechnology start-up Phosphoenix BV (the Netherlands) (<https://phosphoenix.nl>).

REFERENCES

- Chalmers, D.J. (1995). Facing up to the problem of consciousness. *J. Conscious. Stud.* 2, 200–219.
- Crick, F. (1995). The Astonishing Hypothesis: the Scientific Search for the Soul (Simon and Schuster).
- Crick, F., and Koch, C. (1990). Towards a neurobiological theory of consciousness. *Semin. Neurosci.* 2, 263–275.
- Seth, A.K., and Bayne, T. (2022). Theories of consciousness. *Nat. Rev. Neurosci.* 23, 439–452.
- Cogitate Consortium, Ferrante, O., Gorska-Klimowska, U., Henin, S., Hirschhorn, R., Khalaf, A., Lepauvre, A., Liu, L., Richter, D., Vidal, Y., et al. (2023). An adversarial collaboration to critically evaluate theories of consciousness. Preprint at bioRxiv. <https://doi.org/10.1101/2023.06.23.546249>.
- Dehaene, S., and Changeux, J.P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227.
- Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: from consciousness to its physical substrate. *Nat. Rev. Neurosci.* 17, 450–461.
- Lamme, V.A.F. (2006). Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204.
- Friston, K.J., Wiese, W., and Hobson, J.A. (2020). Sentience and the Origins of Consciousness: From Cartesian Duality to Markovian Monism. *Entropy (Basel)* 22, 516.
- Hohwy, J., and Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *PhiMiSci.* 1, 1.

12. Pennartz, C.M.A. (2015). *The Brain's Representational Power: On Consciousness and the Integration of Modalities* (MIT Press).
13. Pennartz, C.M.A. (2022). What is neurorepresentationalism? From neural activity and predictive processing to multi-level representations and consciousness. *Behav. Brain Res.* *432*, 113969.
14. Aru, J., Suzuki, M., and Larkum, M.E. (2020). Cellular Mechanisms of Conscious Processing. *Trends Cogn. Sci.* *24*, 814–825.
15. Chalmers, D.J. (2010). *The Character of Consciousness* (Oxford University Press).
16. Koch, C., Massimini, M., Boly, M., and Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nat. Rev. Neurosci.* *17*, 307–321.
17. Sanders, R.D., Tononi, G., Laureys, S., and Sleight, J.W. (2012). Unresponsiveness ≠ Unconsciousness. *Anesthesiology* *116*, 946–959.
18. Nagel, T. (1974). What Is It Like to Be a Bat? *Philos. Rev.* *83*, 435–450.
19. Block, N. (2005). Two neural correlates of consciousness. *Trends Cogn. Sci.* *9*, 46–52.
20. Wang, X.J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* *24*, 455–463.
21. Hempel, C.G., and Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philos. Sci.* *15*, 135–175.
22. Carlson, N.R. (2010). *Psychology: the Science of Behavior, Seventh Edition* (Allyn & Bacon).
23. Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron* *76*, 695–711.
24. National Academy of Sciences (2008). *Science, Evolution, and Creationism* (Washington, DC: The National Academies Press).
25. Klink, P.C., Self, M.W., Lamme, V.A.F., and Roelfsema, P.R. (2015). *Theories and Methods in the Scientific Study of Consciousness* (John Benjamins Publishing Company).
26. Koch, C., and Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends Cogn. Sci.* *11*, 16–22.
27. Lamme, V.A.F. (2020). Visual Functions Generating Conscious Seeing. *Front. Psychol.* *11*, 83.
28. Mashour, G.A., Roelfsema, P., Changeux, J.P., and Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron* *105*, 776–798.
29. Bekinschtein, T.A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., and Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci. USA* *106*, 1672–1677.
30. King, J.R., Faugeras, F., Gramfort, A., Schurger, A., El Karoui, I., Sitt, J.D., Rohaut, B., Wacongne, C., Labyt, E., Bekinschtein, T., et al. (2013). Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *NeuroImage* *83*, 726–738.
31. Faugeras, F., Rohaut, B., Weiss, N., Bekinschtein, T., Galanaud, D., Puybasset, L., Bolger, F., Sergent, C., Cohen, L., Dehaene, S., et al. (2012). Event related potentials elicited by violations of auditory regularities in patients with impaired consciousness. *Neuropsychologia* *50*, 403–418.
32. Sarasso, S., Casali, A.G., Casarotto, S., Rosanova, M., Sinigaglia, C., and Massimini, M. (2021). Consciousness and complexity: a consilience of evidence. *Neuroscience of Consciousness 2021*, niab023.
33. Demertzi, A., Soddu, A., and Laureys, S. (2013). Consciousness supporting networks. *Curr. Opin. Neurobiol.* *23*, 239–244.
34. Luppi, A.I., Craig, M.M., Pappas, I., Fioino, P., Williams, G.B., Allanson, J., Pickard, J.D., Owen, A.M., Naci, L., Menon, D.K., et al. (2019). Consciousness-specific dynamic interactions of brain integration and functional diversity. *Nat. Commun.* *10*, 4616.
35. Casali, A.G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K.R., Casarotto, S., Bruno, M.A., Laureys, S., Tononi, G., et al. (2013). A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. *Sci. Transl. Med.* *5*, 198ra105.
36. D'Andola, M., Giulioni, M., Dante, V., Del Giudice, P., and Sanchez-Vives, M.V. (2019). Control of cortical oscillatory frequency by a closed-loop system. *J. Neuroeng. Rehabil.* *16*, 7.
37. Rosanova, M., Fecchio, M., Casarotto, S., Sarasso, S., Casali, A.G., Pigorini, A., Comanducci, A., Seregini, F., Devalle, G., Citerio, G., et al. (2018). Sleep-like cortical OFF-periods disrupt causality and complexity in the brain of unresponsive wakefulness syndrome patients. *Nat. Commun.* *9*, 4427.
38. Arena, A., Comolatti, R., Thon, S., Casali, A.G., and Storm, J.F. (2021). General Anesthesia Disrupts Complex Cortical Dynamics in Response to Intracranial Electrical Stimulation in Rats. *eNeuro* *8*, ENEURO.0343-20.2021.
39. Rosanova, M., Gosseries, O., Casarotto, S., Boly, M., Casali, A.G., Bruno, M.A., Mariotti, M., Boveroux, P., Tononi, G., Laureys, S., et al. (2012). Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients. *Brain* *135*, 1308–1320.
40. Aru, J., Bachmann, T., Singer, W., and Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neurosci. Biobehav. Rev.* *36*, 737–746.
41. Avanzini, P., Abdollahi, R.O., Sartori, I., Caruana, F., Pelliccia, V., Casaceli, G., Mai, R., Lo Russo, G., Rizzolatti, G., and Orban, G.A. (2016). Four-dimensional maps of the human somatosensory system. *Proc. Natl. Acad. Sci. USA* *113*, E1936–E1943.
42. Sergent, C., Corazzol, M., Labouret, G., Stockart, F., Wexler, M., King, J.R., Meyniel, F., and Pressnitzer, D. (2021). Bifurcation in brain dynamics reveals a signature of conscious processing independent of report. *Nat. Commun.* *12*, 1149.
43. Dembski, C., Koch, C., and Pitts, M. (2021). Perceptual awareness negativity: a physiological correlate of sensory consciousness. *Trends Cogn. Sci.* *25*, 660–670.
44. Avanzini, P., Pelliccia, V., Lo Russo, G., Orban, G.A., and Rizzolatti, G. (2018). Multiple time courses of somatosensory responses in human cortex. *NeuroImage* *169*, 212–226.
45. Del Vecchio, M., Caruana, F., Sartori, I., Pelliccia, V., Lo Russo, G., Rizzolatti, G., and Avanzini, P. (2019). Ipsilateral somatosensory responses in humans: the tonic activity of SII and posterior insular cortex. *Brain Struct. Funct.* *224*, 9–18.
46. Del Vecchio, M., Fossataro, C., Zauli, F.M., Sartori, I., Pigorini, A., d'Orto, P., Abarrategui, B., Russo, S., Mikulan, E.P., Caruana, F., et al. (2021). Tonic somatosensory responses and deficits of tactile awareness converge in the parietal operculum. *Brain* *144*, 3779–3787.
47. Fossataro, C., Bruno, V., Bosso, E., Chiotti, V., Gindri, P., Farnè, A., and Garbarini, F. (2020). The sense of body-ownership gates cross-modal improvement of tactile extinction in brain-damaged patients. *Cortex* *127*, 94–107.
48. Schneider, M., Kemper, V.G., Emmerling, T.C., De Martino, F., and Goebel, R. (2019). Columnar clusters in the human motion complex reflect consciously perceived motion axis. *Proc. Natl. Acad. Sci. USA* *116*, 5096–5101.
49. Salzman, C.D., and Newsome, W.T. (1994). Neural Mechanisms for Forming a Perceptual Decision. *Science* *264*, 231–237.
50. Ellia, F., Hendren, J., Grasso, M., Kozma, C., Mindt, G., P Lang, J., M Haun, A., Albantakis, L., Boly, M., and Tononi, G. (2021). Consciousness and the fallacy of misplaced objectivity. *Neurosci. Conscious.* *2021*, niab032.
51. Block, N. (2023). *The Border Between Seeing and Thinking* (Oxford University Press).
52. Sevenius Nilsen, A.S., Juel, B.E., Thüner, B., Aamodt, A., and Storm, J.F. (2022). Are we really unconscious in “unconscious” states? Common assumptions revisited. *Front. Hum. Neurosci.* *16*, 987051.

53. Pennartz, C.M.A. (2018). Consciousness, Representation, Action: The Importance of Being Goal-Directed. *Trends Cogn. Sci.* 22, 137–153.
54. Munn, B.R., Müller, E.J., Aru, J., Whyte, C.J., Gidon, A., Larkum, M.E., and Shine, J.M. (2023). A Thalamocortical Substrate for Integrated Information via Critical Synchronous Bursting. *Proc Natl Acad Sci USA* 120, e2308670120.
55. Tononi, G. (2004). An information integration theory of consciousness. *BMC Neurosci.* 5, 42.
56. Bachmann, T., Suzuki, M., and Aru, J. (2020). Dendritic integration theory: A thalamo-cortical theory of state and content of consciousness. *PhiMiSci.* 1, 1.
57. Aru, J., Suzuki, M., Rutiku, R., Larkum, M.E., and Bachmann, T. (2019). Coupling the State and Contents of Consciousness. *Front. Syst. Neurosci.* 13, 43.
58. Baars, B.J. (1988). *A Cognitive Theory of Consciousness* (Cambridge University Press).
59. Dehaene, S., Kerszberg, M., and Changeux, J.P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl. Acad. Sci. USA* 95, 14529–14534.
60. Dehaene, S., and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79, 1–37.
61. Felleman, D.J., and van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.
62. Grill-Spector, K., and Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci.* 27, 649–677.
63. Mesulam, M.M. (1999). Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 354, 1325–1346.
64. van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., and Roelfsema, P.R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science* 360, 537–542.
65. Noel, J.P., Ishizawa, Y., Patel, S.R., Eskandar, E.N., and Wallace, M.T. (2019). Leveraging Nonhuman Primate Multisensory Neurons and Circuits in Assessing Consciousness Theory. *J. Neurosci.* 39, 7485–7500.
66. Sanchez, G., Hartmann, T., Fuscà, M., Demarchi, G., and Weisz, N. (2020). Decoding across sensory modalities reveals common supramodal signatures of conscious perception. *Proc. Natl. Acad. Sci. USA* 117, 7437–7446.
67. Marti, S., King, J.R., and Dehaene, S. (2015). Time-Resolved Decoding of Two Processing Chains during Dual-Task Interference. *Neuron* 88, 1297–1307.
68. Marti, S., and Dehaene, S. (2017). Discrete and continuous mechanisms of temporal selection in rapid visual streams. *Nat. Commun.* 8, 1955.
69. Sergent, C., Baillet, S., and Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nat. Neurosci.* 8, 1391–1400.
70. Berkovitch, L., Dehaene, S., and Gaillard, R. (2017). Disruption of Conscious Access in Schizophrenia. *Trends Cogn. Sci.* 21, 878–892.
71. Charles, L., King, J.R., and Dehaene, S. (2014). Decoding the Dynamics of Action, Intention, and Error Detection for Conscious and Subliminal Stimuli. *J. Neurosci.* 34, 1158–1170.
72. Del Cul, A.D., Baillet, S., and Dehaene, S. (2007). Brain Dynamics Underlying the Nonlinear Threshold for Access to Consciousness. *PLoS Biol.* 5, e260.
73. Marti, S., Sigman, M., and Dehaene, S. (2012). A shared cortical bottleneck underlying Attentional Blink and Psychological Refractory Period. *NeuroImage* 59, 2883–2898.
74. Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., and Fried, I. (2008). Internally Generated Reactivation of Single Neurons in Human Hippocampus During Free Recall. *Science* 322, 96–101.
75. Kreiman, G., Fried, I., and Koch, C. (2002). Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proc. Natl. Acad. Sci. USA* 99, 8378–8383.
76. Quiroga, R.Q., Mukamel, R., Isham, E.A., Malach, R., and Fried, I. (2008). Human single-neuron responses at the threshold of conscious recognition. *Proc. Natl. Acad. Sci. USA* 105, 3599–3604.
77. Reber, T.P., Faber, J., Niediek, J., Boström, J., Elger, C.E., and Mormann, F. (2017). Single-Neuron Correlates of Conscious Perception in the Human Medial Temporal Lobe. *Curr. Biol.* 27, 2991–2998.e2.
78. Boly, M., Massimini, M., Tsuchiya, N., Postle, B.R., Koch, C., and Tononi, G. (2017). Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? Clinical and Neuroimaging Evidence. *J. Neurosci.* 37, 9603–9613.
79. Bellet, J., Gay, M., Dwarakanath, A., Jarraya, B., van Kerkoerle, T., Dehaene, S., and Panagiotaropoulos, T.I. (2022). Decoding rapidly presented visual stimuli from prefrontal ensembles without report nor post-perceptual processing. *Neurosci. Conscious.* 2022, niac005.
80. Kapoor, V., Dwarakanath, A., Safavi, S., Werner, J., Besserve, M., Panagiotaropoulos, T.I., and Logothetis, N.K. (2022). Decoding internally generated transitions of conscious contents in the prefrontal cortex without subjective reports. *Nat. Commun.* 13, 1535.
81. Kok, P., Bains, L.J., van Mourik, T., Norris, D.G., and de Lange, F.P. (2016). Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Curr. Biol.* 26, 371–376.
82. Muckli, L., De Martino, F., Vizioli, L., Petro, L.S., Smith, F.W., Uğurbil, K., Goebel, R., and Yacoub, E. (2015). Contextual Feedback to Superficial Layers of V1. *Curr. Biol.* 25, 2690–2695.
83. Self, M.W., van Kerkoerle, T., Supèr, H., and Roelfsema, P.R. (2013). Distinct Roles of the Cortical Layers of Area V1 in Figure-Ground Segregation. *Curr. Biol.* 23, 2121–2129.
84. Van Kerkoerle, T., Self, M.W., and Roelfsema, P.R. (2017). Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nat. Commun.* 8, 13804.
85. Gao, Z., Davis, C., Thomas, A.M., Economo, M.N., Abrego, A.M., Svoboda, K., De Zeeuw, C.J., and Li, N. (2018). A cortico-cerebellar loop for motor planning. *Nature* 563, 113–116.
86. Guo, Z.V., Inagaki, H.K., Daie, K., Druckmann, S., Gerfen, C.R., and Svoboda, K. (2017). Maintenance of persistent activity in a frontal thalamo-cortical loop. *Nature* 545, 181–186.
87. Markov, N.T., Ercsey-Ravasz, M., Van Essen, D.C., Knoblauch, K., Toroczkai, Z., and Kennedy, H. (2013). Cortical High-Density Counterstream Architectures. *Science* 342, 1238406.
88. Tononi, G., and Edelman, G.M. (1998). Consciousness and Complexity. *Science* 282, 1846–1851.
89. Tononi, G. (2001). Information measures for conscious experience. *Arch. Ital. Biol.* 139, 367–371.
90. Tononi, G., and Sporns, O. (2003). Measuring information integration. *BMC Neurosci.* 4, 31.
91. Mediano, P.A.M., Rosas, F.E., Bor, D., Seth, A.K., and Barrett, A.B. (2022). The strength of weak integrated information theory. *Trends Cogn. Sci.* 26, 646–655.
92. Massimini, M., Boly, M., Casali, A., Rosanova, M., and Tononi, G. (2009). A perturbational approach for evaluating the brain's capacity for consciousness. In *Progress in Brain Research: Clinical and Ethical Implications*, S. Laureys, N.D. Schiff, and A.M. Owen, eds. (Elsevier), pp. 201–214.
93. Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* 215, 216–242.

94. Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS Comput. Biol.* *10*, e1003588.
95. Albantakis, L., Barbosa, L., Findlay, G., Grasso, M., Haun, A.M., Marshall, W., Mayner, W.G.P., Zaeemzadeh, A., Boly, M., Juel, B.E., et al. (2023). Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. *PLoS Comput Biol.* *19*, e1011465.
96. Grasso, M., Haun, A.M., and Tononi, G. (2021). Of maps and grids. *Neurosci. Conscious.* *2021*, niab022.
97. Haun, A., and Tononi, G. (2019). Why Does Space Feel the Way it Does? Towards a Principled Account of Spatial Experience. *Entropy* *21*, 1160.
98. Lamme, V.A.F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.* *15*, 1605–1615.
99. Lamme, V.A.F., and Roelfsema, P.R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* *23*, 571–579.
100. Zipser, K., Lamme, V.A.F., and Schiller, P.H. (1996). Contextual modulation in primary visual cortex. *J. Neurosci.* *16*, 7376–7389.
101. Lamme, V.A.F. (2015). The Crack of Dawn: Perceptual Functions and Neural Mechanisms that Mark the Transition from Unconscious Processing to Conscious Vision (Open Mind).
102. Lamme, V.A.F. (2018). Challenges for theories of consciousness: seeing or knowing, the missing ingredient and how to deal with panpsychism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *373*, 20170344.
103. Graziano, M.S.A., and Webb, T.W. (2015). The attention schema theory: a mechanistic account of subjective awareness. *Front. Psychol.* *6*, 500.
104. Koivisto, M., and Revonsuo, A. (2010). Event-related brain potential correlates of visual awareness. *Neurosci. Biobehav. Rev.* *34*, 922–934.
105. Lamme, V.A.F. (2010). How neuroscience will change our view on consciousness. *Cogn. Neurosci.* *1*, 204–220.
106. Dehaene, S., Changeux, J.P., Naccache, L., Sackur, J., and Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn. Sci.* *10*, 204–211.
107. Milner, A.D., and Goodale, M.A. (1995). *The Visual Brain in Action* (Oxford University Press).
108. Sperling, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.: Gen. Appl.* *74*, 1–29.
109. Self, M.W., Kooijmans, R.N., Supér, H., Lamme, V.A.F., and Roelfsema, P.R. (2012). Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proc. Natl. Acad. Sci. USA* *109*, 11031–11036.
110. Hohwy, J. (2013). *The Predictive Mind* (OUP Oxford).
111. Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* *11*, 127–138.
112. Pennartz, C.M.A., Oude Lohuis, M.N., and Olcese, U. (2023). How 'visual' is the visual cortex? The interactions between the visual cortex and other sensory, motivational and motor systems as enabling factors for visual perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *378*, 20220336.
113. Allen, W.E., Kauvar, I.V., Chen, M.Z., Richman, E.B., Yang, S.J., Chan, K., Gradinaru, V., Deverman, B.E., Luo, L., and Deisseroth, K. (2017). Global Representations of Goal-Directed Behavior in Distinct Cell Types of Mouse Neocortex. *Neuron* *94*, 891–907.e6.
114. Oude Lohuis, M.N., Pie, J.L., Marchesi, P., Montijn, J.S., de Kock, C.P.J., Pennartz, C.M.A., and Olcese, U. (2022). Multisensory task demands temporally extend the causal requirement for visual cortex in perception. *Nat. Commun.* *13*, 2864.
115. Oude Lohuis, M.N., Marchesi, P., Olcese, U., and Pennartz, C.M.A. (2024). Triple dissociation of visual, auditory and motor processing in mouse primary visual cortex. *Nat. Neurosci.* <https://doi.org/10.1038/s41593-023-01564-5>.
116. Brucklacher, M., Bohté, S.M., Mejias, J.F., and Pennartz, C.M.A. (2023). Local minimization of prediction errors drives learning of invariant object representations in a generative network model of visual perception. *Front. Comput. Neurosci.* *17*, 1207361.
117. Dora, S., Bohte, S.M., and Pennartz, C.M.A. (2021). Deep Gated Hebbian Predictive Coding Accounts for Emergence of Complex Neural Response Properties Along the Visual Cortical Hierarchy. *Front. Comput. Neurosci.* *15*, 666131.
118. Keller, G.B., and Msrac-Flogel, T.D. (2018). Predictive Processing: A Canonical Cortical Computation. *Neuron* *100*, 424–435.
119. Montijn, J.S., Olcese, U., and Pennartz, C.M.A. (2016). Visual Stimulus Detection Correlates with the Consistency of Temporal Sequences within Stereotyped Events of V1 Neuronal Population Activity. *J. Neurosci.* *36*, 8624–8640.
120. Larkum, M.E. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.* *36*, 141–151.
121. Larkum, M.E., Zhu, J.J., and Sakmann, B. (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature* *398*, 338–341.
122. Larkum, M.E., Nevian, T., Sandler, M., Polsky, A., and Schiller, J. (2009). Synaptic Integration in Tuft Dendrites of Layer 5 Pyramidal Neurons: A New Unifying Principle. *Science* *325*, 756–760.
123. Takahashi, N., Oertner, T.G., Hegemann, P., and Larkum, M.E. (2016). Active cortical dendrites modulate perception. *Science* *354*, 1587–1590.
124. Takahashi, N., Ebner, C., Sigl-Glöckner, J., Moberg, S., Nierwetberg, S., and Larkum, M.E. (2020). Active dendritic currents gate descending cortical outputs in perception. *Nat. Neurosci.* *23*, 1277–1285.
125. Suzuki, M., and Larkum, M.E. (2020). General Anesthesia Decouples Cortical Pyramidal Neurons. *Cell* *180*, 666–676.e13.
126. Palmer, L.M., Schulz, J.M., Murphy, S.C., Ledergerber, D., Murayama, M., and Larkum, M.E. (2012). The Cellular Basis of GABA(B)-Mediated Interhemispheric Inhibition. *Science* *335*, 989–993.
127. Phillips, W.A. (2023). *The Cooperative Neuron: Cellular Foundations of Mental Life* (Oxford University Press).
128. Doron, G., Shin, J.N., Takahashi, N., Drüke, M., Bocklisch, C., Skenderi, S., de Mont, L., Toumazou, M., Ledderose, J., Brecht, M., et al. (2020). Perirhinal input to neocortical layer 1 controls learning. *Science* *370*, eaaz3136.
129. Shin, J.N., Doron, G., and Larkum, M.E. (2021). Memories off the top of your head. *Science* *374*, 538–539.
130. Aru, J., Siclari, F., Phillips, W.A., and Storm, J.F. (2020). Apical drive—A cellular mechanism of dreaming? *Neurosci. Biobehav. Rev.* *119*, 440–455.
131. Tononi, G., and Koch, C. (2015). Consciousness: here, there and everywhere? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *370*, 20140167.
132. Evers, K., Farisco, M., and Pennartz, C.M.A. (2024). Assessing the commensurability of theories of consciousness: On the usefulness of common denominators in differentiating, integrating and testing hypotheses. *Conscious. Cogn.* *119*, 103668. <https://doi.org/10.1016/j.concog.2024.103668>.
133. Aru, J., Larkum, M.E., and Shine, J.M. (2023). The feasibility of artificial consciousness through the lens of neuroscience. *Trends Neurosci.* *46*, 1008–1017.
134. Searle, J.R. (1980). Minds, brains, and programs. *Behav. Brain Sci.* *3*, 417–424.
135. Pennartz, C.M.A., Dora, S., Muckli, L., and Lorteije, J.A.M. (2019). Towards a Unified View on Pathways and Functions of Neural Recurrent Processing. *Trends Neurosci.* *42*, 589–603.
136. J.R. Josephson and S.G. Josephson, eds. (1994). *Abductive Inference: Computation, Philosophy, Technology* (Cambridge University Press).

137. Corcoran, A.W., Hohwy, J., and Friston, K.J. (2023). Accelerating scientific progress through Bayesian adversarial collaboration. *Neuron* *111*, 3505–3516.
138. Suzuki, M., Pennartz, C.M.A., and Aru, J. (2023). How deep is the brain? The shallow brain hypothesis. *Nat. Rev. Neurosci.* *24*, 778–791.
139. Harris, J.A., Mihalas, S., Hirokawa, K.E., Whitesell, J.D., Choi, H., Bernard, A., Bohn, P., Caldejon, S., Casal, L., Cho, A., et al. (2019). Hierarchical organization of cortical and thalamic connectivity. *Nature* *575*, 195–202.
140. Changeux, J.P. (2017). Climbing Brain Levels of Organisation from Genes to Consciousness. *Trends Cogn. Sci.* *21*, 168–181.
141. Rao, R.P.N., and Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* *2*, 79–87.
142. Sacramento, J., Ponte Costa, R., Bengio, Y., and Senn, W. (2018). Dendritic cortical microcircuits approximate the backpropagation algorithm. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1810.11393>.
143. Senn, W., Dold, D., Kungl, A.F., Ellenberger, B., Jordan, J., Bengio, Y., Sacramento, J., and Petrovici, M.A. (2023). A Neuronal Least-Action Principle for Real-Time Learning in Cortical Circuits. Preprint at bioRxiv. <https://doi.org/10.1101/2023.03.25.534198>.
144. Urbanczik, R., and Senn, W. (2014). Learning by the Dendritic Prediction of Somatic Spiking. *Neuron* *81*, 521–528.
145. Vladimirov, B., Urbanczik, R., and Senn, W. (2015). Hierarchical Novelty-Familiarity Representation in the Visual System by Modular Predictive Coding. *PLoS One* *10*, e0144636.
146. Markram, H., Müller, E., Ramaswamy, S., Reimann, M.W., Abdellah, M., Sanchez, C.A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., et al. (2015). Reconstruction and Simulation of Neocortical Microcircuitry. *Cell* *163*, 456–492.
147. Douglas, R.J., and Martin, K.A. (1991). A functional microcircuit for cat visual cortex. *J. Physiol.* *440*, 735–769.
148. Attinger, A., Wang, B., and Keller, G.B. (2017). Visuomotor Coupling Shapes the Functional Development of Mouse Visual Cortex. *Cell* *169*, 1291–1302.e14.
149. Green, J., Bruno, C.A., Traunmüller, L., Ding, J., Hrvatin, S., Wilson, D.E., Khodadad, T., Samuels, J., Greenberg, M.E., and Harvey, C.D. (2023). A cell-type-specific error-correction signal in the posterior parietal cortex. *Nature* *620*, 366–373.
150. Fahrenfort, J.J., Scholte, H.S., and Lamme, V.A.F. (2007). Masking Disrupts Reentrant Processing in Human Visual Cortex. *J. Cogn. Neurosci.* *19*, 1488–1497.
151. Fahrenfort, J.J., and Lamme, V.A.F. (2012). A true science of consciousness explains phenomenology: comment on Cohen and Dennett. *Trends Cogn. Sci.* *16*, 138–9; author reply 139.
152. Sikkens, T., Bosman, C.A., and Olcese, U. (2019). The Role of Top-Down Modulation in Shaping Sensory Processing Across Brain States: Implications for Consciousness. *Front. Syst. Neurosci.* *13*, 31.
153. Dehaene, S., Artiges, E., Naccache, L., Martelli, C., Viard, A., Schürhoff, F., Recasens, C., Martinot, M.L.P., Leboyer, M., and Martinot, J.L. (2003). Conscious and subliminal conflicts in normal subjects and patients with schizophrenia: The role of the anterior cingulate. *Proc. Natl. Acad. Sci. USA* *100*, 13722–13727.
154. Zylberberg, A., Dehaene, S., Mindlin, G.B., and Sigman, M. (2009). Neurophysiological bases of exponential sensory decay and top-down memory retrieval: a model. *Front. Comput Neurosci.* *3*, 4.
155. Doerig, A., Schurger, A., and Herzog, M.H. (2021). Hard criteria for empirical theories of consciousness. *Cogn. Neurosci.* *12*, 41–62.
156. Knierim, J.J., and Van Essen, D.C. (1992). Visual cortex: cartography, connectivity, and concurrent processing. *Curr. Opin. Neurobiol.* *2*, 150–155.
157. Laureys, S. (2005). The neural correlate of (un)awareness: lessons from the vegetative state. *Trends Cogn. Sci.* *9*, 556–559.
158. Hill, S., and Tononi, G. (2005). Modeling Sleep and Wakefulness in the Thalamocortical System. *J. Neurophysiol.* *93*, 1671–1698.
159. Laureys, S., Owen, A.M., and Schiff, N.D. (2004). Brain function in coma, vegetative state, and related disorders. *Lancet Neurol.* *3*, 537–546.
160. Bayne, T., Hohwy, J., and Owen, A.M. (2016). Are There Levels of Consciousness? *Trends Cogn. Sci.* *20*, 405–413.
161. Steriade, M. (2006). Grouping of brain rhythms in corticothalamic systems. *Neuroscience* *137*, 1087–1106.
162. Heeger, D.J. (2017). Theory of cortical function. *Proc. Natl. Acad. Sci. USA* *114*, 1773–1782.
163. Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behav. Brain Sci.* *30*, 63–81.
164. Penfield, W., and Jasper, H. (1954). *Epilepsy and the Functional Anatomy of the Human Brain* (Little, Brown & Co.).
165. Solms, M., and Friston, K. (2018). How and why consciousness arises: Some considerations from physics and physiology. *J. Conscious. Stud.* *25*, 202–238.
166. Kanai, R., Komura, Y., Shipp, S., and Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *370*, 20140169.
167. Pennartz, C.M.A. (2009). Identification and integration of sensory modalities: Neural basis and relation to consciousness. *Conscious. Cogn.* *18*, 718–739.
168. Luczak, A., Barthó, P., and Harris, K.D. (2009). Spontaneous Events Outline the Realm of Possible Sensory Responses in Neocortical Populations. *Neuron* *62*, 413–425.
169. Montijn, J.S., Goltstein, P.M., and Pennartz, C.M. (2015). Mouse V1 population correlates of visual detection rely on heterogeneity within neuronal response patterns. *eLife* *4*, e10163.
170. Demertzi, A., Tagliazucchi, E., Dehaene, S., Deco, G., Bartfeld, P., Raimondo, F., Martial, C., Fernández-Espejo, D., Rohaut, B., Voss, H.U., et al. (2019). Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Sci. Adv.* *5*, eaat7603.
171. Farisco, M., and Changeux, J.P. (2023). About the compatibility between the perturbational complexity index and the global neuronal workspace theory of consciousness. *Neurosci. Conscious.* *2023*, niad016.
172. Cattani, A., Galluzzi, A., Fecchio, M., Pigorini, A., Mattia, M., and Massimini, M. (2023). Adaptation Shapes Local Cortical Reactivity: From Bifurcation Diagram and Simulations to Human Physiological and Pathological Responses. *eNeuro* *10*, 435–22.2023.
173. Chang, J.Y., Fecchio, M., Pigorini, A., Massimini, M., Tononi, G., and Van Veen, B.D. (2019). Assessing recurrent interactions in cortical networks: Modeling EEG response to transcranial magnetic stimulation. *J. Neurosci. Methods* *312*, 93–104.
174. Pigorini, A., Sarasso, S., Proserpio, P., Szymanski, C., Arnulfo, G., Casarotto, S., Fecchio, M., Rosanova, M., Mariotti, M., Lo Russo, G., et al. (2015). Bistability breaks-off deterministic responses to intracortical stimulation during non-REM sleep. *NeuroImage* *112*, 105–113.
175. Frith, C. (2005). The neural basis of hallucinations and delusions. *C. R. Biol.* *328*, 169–175.
176. Seth, A.K. (2021). *Being You: A New Science of Consciousness*. https://books.google.nl/books/about/Being_You.html?id=7d9UygEACAAJ&redir_esc=y.
177. Casarotto, S., Comanducci, A., Rosanova, M., Sarasso, S., Fecchio, M., Napolitani, M., Pigorini, A., G Casali, A., Trimarchi, P.D., Boly, M., et al. (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. *Ann. Neurol.* *80*, 718–729.
178. Comanducci, A., Casarotto, S., Rosanova, M., Derchi, C.C., Viganò, A., Pirastru, A., Blasi, V., Cazzoli, M., Navarro, J., Edlow, B.L., et al. (2023).

- Unconsciousness or unresponsiveness in akinetic mutism? Insights from a multimodal longitudinal exploration. *Eur. J. Neurosci.* <https://doi.org/10.1111/ejn.15994>.
179. Arena, A., Juel, B.E., Comolatti, R., Thon, S., and Storm, J.F. (2022). Capacity for consciousness under ketamine anaesthesia is selectively associated with activity in posteromedial cortex in rats. *Neurosci. Conscious.* 2022, niac004.
 180. Cavelli, M.L., Mao, R., Findlay, G., Driessen, K., Bugnon, T., Tononi, G., and Cirelli, C. (2023). Sleep/wake changes in perturbational complexity in rats and mice. *iScience* 26, 106186.
 181. Hönigsperger, C., Storm, J.F., and Arena, A. (2023). Laminar evoked responses in mouse somatosensory cortex suggest a special role for deep layers in cortical complexity. *Eur. J. Neurosci.* <https://doi.org/10.1111/ejn.16108>.
 182. D'Andola, M., Rebollo, B., Casali, A.G., Weinert, J.F., Pigorini, A., Villa, R., Massimini, M., and Sanchez-Vives, M.V. (2018). Bistability, Causality, and Complexity in Cortical Networks: An In Vitro Perturbational Study. *Cereb. Cortex* 28, 2233–2242.
 183. Gidon, A., Aru, J., and Larkum, M.E. (2022). Does brain activity cause consciousness? A thought experiment. *PLoS Biol.* 20, e3001651.
 184. Searle, J.R. (2013). Can Information Theory Explain Consciousness? *N. Y. Rev. Books* 60.
 185. Tononi, G., and Koch, C. (2013). Can a Photodiode Be Conscious? *N. Y. Rev. Books* 60.
 186. Tononi, G., Albantakis, L., Boly, M., Cirelli, C., and Koch, C. (2023). Only What Exists Can Cause: an Intrinsic View of Free Will. *arXiv*. <https://doi.org/10.48550/arXiv.2206.02069>.
 187. Fleming, S., Frith, C., Goodale, M., Lau, H., LeDoux, J.E., Lee, A.L.F., Michel, M., Owen, A., Peters, M.A.K., and Slagter, H.A. (2023). The Integrated Information Theory of Consciousness as Pseudoscience. Preprint at PsyArXiv. <https://doi.org/10.31234/osf.io/zsr78>.
 188. Whyte, C.J., and Smith, R. (2021). The predictive global neuronal workspace: A formal active inference model of visual consciousness. *Prog. Neurobiol.* 199, 101918.
 189. Hohwy, J., Tononi, G., Seth, A.K., and Tsuchiya, N. (2015). Can the free energy principle be used to generate a theory of consciousness? In *Conference Abstract: XII International Conference on Cognitive Neuroscience (ICON-XII)*. <https://doi.org/10.3389/conf.fnhum.2015.217.00420>.
 190. Storm, J.F. (2020). Why Does the Brain-Mind (Consciousness) Problem Seem So Hard? *J. Conscious. Stud.* 27, 174–189.
 191. Farisco, M., Evers, K., Annen, J., Blandin, V., Camassa, A., Cecconi, B., Charland, V., Deco, G., Gosseries, O., Laureys, S., et al. (2023). Advancing the Science of Consciousness: from Ethics to Clinical Care. Preprint at PsyArXiv. <https://doi.org/10.31234/osf.io/sutrc>.
 192. Massimini, M., Ferrarelli, F., Huber, R., Esser, S.K., Singh, H., and Tononi, G. (2005). Breakdown of Cortical Effective Connectivity During Sleep. *Science* 309, 2228–2232.
 193. Olcese, U., Oude Lohuis, M.N., and Pennartz, C.M.A. (2018). Sensory Processing Across Conscious and Nonconscious Brain States: From Single Neurons to Distributed Networks for Inferential Representation. *Front. Syst. Neurosci.* 12, 49.
 194. Suzuki, M., and Larkum, M.E. (2017). Dendritic calcium spikes are clearly detectable at the cortical surface. *Nat. Commun.* 8, 276.
 195. Goltstein, P.M., Montijn, J.S., and Pennartz, C.M.A. (2015). Effects of Isoflurane Anesthesia on Ensemble Patterns of Ca²⁺ Activity in Mouse V1: Reduced Direction Selectivity Independent of Increased Correlations in Cellular Activity. *PLoS One* 10, e0118277.
 196. Albantakis, L., Hintze, A., Koch, C., Adami, C., and Tononi, G. (2014). Evolution of Integrated Causal Structures in Animats Exposed to Environments of Increasing Complexity. *PLoS Comput Biol.* 10, e1003966.
 197. Deperrois, N., Petrovici, M.A., Senn, W., and Jordan, J. (2022). Learning cortical representations through perturbed and adversarial dreaming. *eLife* 11, e76384.
 198. O'Connor, D.H., Hires, S.A., Guo, Z.V., Li, N., Yu, J., Sun, Q.Q., Huber, D., and Svoboda, K. (2013). Neural coding during active somatosensation revealed using illusory touch. *Nat. Neurosci.* 16, 958–965.
 199. Storm, J.F., Boly, M., Casali, A.G., Massimini, M., Olcese, U., Pennartz, C.M.A., and Wilke, M. (2017). Consciousness Regained: Disentangling Mechanisms, Brain Systems, and Behavioral Responses. *J. Neurosci.* 37, 10882–10893.