


RESEARCH ARTICLE

# Theoretical Neurobiology of Consciousness Applied to Human Cerebral Organoids

Matthew Owen<sup>1,2</sup>, Zirui Huang<sup>2,3</sup>, Catherine Duclos<sup>4,5,6,7</sup>, Andrea Lavazza<sup>8,9</sup> , Matteo Grasso<sup>10</sup> and Anthony G. Hudetz<sup>2,3</sup>

<sup>1</sup>Philosophy Department, Yakima Valley College, Yakima, WA, USA

<sup>2</sup>Center for Consciousness Science, University of Michigan, Ann Arbor, MI, USA

<sup>3</sup>Department of Anesthesiology, University of Michigan, Ann Arbor, MI, USA

<sup>4</sup>Department of Anesthesiology and Pain Medicine, Université de Montréal, Montréal, QC, Canada

<sup>5</sup>Department of Neuroscience, Université de Montréal, Montréal, QC, Canada

<sup>6</sup>Centre for Advanced Research in Sleep Medicine, Centre intégré universitaire de santé et de services sociaux (CIUSSS) du Nord-de-l'île-de-Montréal, Montréal, QC, Canada

<sup>7</sup>CIFAR Azrieli Global Scholars Program, Toronto, ON, Canada

<sup>8</sup>Centro Universitario Internazionale, Arezzo, Italy

<sup>9</sup>University of Pavia, Pavia, Italy

<sup>10</sup>Center for Sleep and Consciousness, University of Wisconsin-Madison, Madison, WI, USA

**Corresponding author:** Matthew Owen; Email: [matthewkeithowen@gmail.com](mailto:matthewkeithowen@gmail.com)

## Abstract

Organoids and specifically human cerebral organoids (HCOs) are one of the most relevant novelties in the field of biomedical research. Grown either from embryonic or induced pluripotent stem cells, HCOs can be used as *in vitro* three-dimensional models, mimicking the developmental process and organization of the developing human brain. Based on that, and despite their current limitations, it cannot be assumed that they will never at any stage of development manifest some rudimentary form of consciousness. In the absence of behavioral indicators of consciousness, the theoretical neurobiology of consciousness being applied to unresponsive brain-injured patients can be considered with respect to HCOs. In clinical neurology, it is difficult to discern a capacity for consciousness in unresponsive brain-injured patients who provide no behavioral indicators of consciousness. In such scenarios, a validated neurobiological theory of consciousness, which tells us what the neural mechanisms of consciousness are, could be used to identify a capacity for consciousness. Like the unresponsive patients that provide a diagnostic difficulty for neurologists, HCOs provide no behavioral indicators of consciousness. Therefore, this article discusses how three prominent neurobiological theories of consciousness apply to human cerebral organoids. From the perspective of the Temporal Circuit Hypothesis, the Global Neuronal Workspace Theory, and the Integrated Information Theory, we discuss what neuronal structures and functions might indicate that cerebral organoids have a neurobiological capacity to be conscious.

**Keywords:** human cerebral organoids; consciousness; neurobiology

## Introduction

Human cerebral organoids (HCOs) are arguably one of the most promising recent advancements in neuroscience. Insofar as these cerebral organoids accurately model human brain development, they might yield answers to long-standing neurobiological questions with clinical implications and new

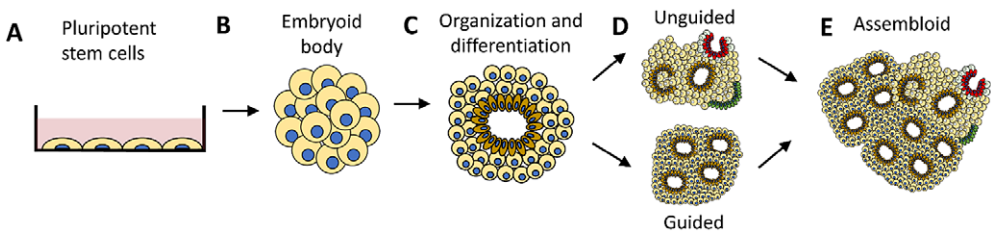
neurological treatments.<sup>1,2</sup> Simultaneously, the more HCOs replicate the human brain, the more they will elicit questions about whether they could be conscious in some way. Questions about whether HCOs could ever be conscious and how we could ever know, despite their inability to communicate, are certainly cerebral questions but not necessarily novel. They reflect an epistemic challenge neuroscientists and neurologists began wrestling with well before the development of cerebral organoids.

The fundamental difficulty in discerning the presence or absence of consciousness is far from limited to HCOs. In all creatures from primitive animals to primates, we rely on an observation of purposeful behavior as a surrogate marker of consciousness. Even in humans, we infer conscious awareness from behavior.<sup>3</sup> In clinical neurology, consciousness is especially challenging to discern when brain-injured patients are noncommunicative and unresponsive.<sup>4,5,6</sup> Central nervous system lesions can prevent motor activity, such as spontaneous speech, body movements, or purposeful response to stimuli, the lack of which could be mistaken for unconsciousness. The possibility of awareness without responsiveness has led to the concept of covert consciousness, or cognitive–motor dissociation.<sup>7,8,9,10,11</sup> A similar case can be made when patients undergo general anesthesia.<sup>12,13,14,15</sup>

It has been proposed that a behaviorally unresponsive subject might provide neuronal responses, such as those assessed by neuroimaging, to questions, thus indicating consciousness.<sup>16,17,18,19</sup> Yet, it is possible that damage to the sensory systems could still preclude a covertly conscious person from sensing external commands intended to elicit a neuronal response. To overcome this impasse, theoretical frameworks about the neural mechanisms of consciousness have been applied to the epistemic challenge of diagnosing unresponsive, noncommunicative patients.<sup>20,21,22,23</sup> The rationale for this approach is that a valid theory about the neural mechanisms that correspond to consciousness could provide an empirical indicator of consciousness in the absence of any other indicators, such as a response or an intentional behavior. This article applies this theoretical approach to HCOs, which are further introduced in the following section along with our overall objectives and the neurobiological theories we will apply to HCOs in the subsequent sections.

### HCOs and Neurobiology of Consciousness

Like the consciousness of unresponsive brain-injured patients difficult to assess with standard clinical methods, human cerebral organoids are incapable of providing an indication of consciousness in the form of responses or behavior analogous to those produced by humans and other complex organisms. A human cerebral organoid consists of neurons that connect to networks and self-organize into structures of different cell types, resembling brain regions in early development.<sup>24</sup> Consisting of up to tens of millions of neurons developed from either embryonic or induced pluripotent stem cells (see Figure 1), an



**Figure 1.** *Human cerebral organoid development.* A human cerebral organoid is a three-dimensional spheroid consisting of neurons developed from human pluripotent stem cells (a). The neurons multiply, forming an embryoid body (b) that then self-organizes and differentiates (c). The growth of a cerebral organoid *in vitro* can be guided (d-bottom) using molecules and growth factors or it can be unguided (d-top). An unguided cerebral organoid often has heterogeneous cellular tissues resembling various brain regions in one organoid. A guided organoid can be generated to resemble a specific brain region composed of a specific type of cellular tissue in one organoid. Multiple organoids can be combined to form an assembloid (e). Guided organoids resembling particular brain regions can be combined to form an assembloid intended to model interactions between distinct brain regions.

HCO that is a few millimeters in size can provide a 3D model of human brain development.<sup>25,26</sup> Unlike the human brain that exists as a part of a whole human organism and is protected by bone tissue, HCOs are grown *in vitro*. As an isolated entity not obscured by the skull, an HCO provides a clear window into the development as well as maldevelopment of neuroanatomy and neurophysiology in the brain.<sup>27</sup> There are questions about the extent to which HCOs truly model the human brain.<sup>28</sup> Nevertheless, to the extent that HCOs resemble the structures and functions of the human brain in its early stages of development, they provide significant promise for neurological progress.<sup>29,30</sup>

At the same time, however, the more HCOs model or replicate the human brain, the organ corresponding to our conscious mental lives, the question of whether they can be conscious becomes more pressing.<sup>31,32,33</sup> This question is significant to the ethical development and use of HCOs.<sup>34,35,36,37,38,39,40</sup> Yet, like the unresponsive brain-injured patients presenting a diagnostic challenge in neurology, HCOs are incapable of providing any behavioral or communicative indicators of consciousness. However, unlike unresponsive brain-injured patients, HCOs do not have prior histories of providing clear indicators of consciousness through interpersonal communication and intentional behaviors. Considering the complete lack of any behavioral or communicative indicators of consciousness from HCOs, it is difficult to imagine what else about HCOs could possibly indicate the presence or lack of consciousness, except their neuroanatomy and neurophysiology. Therefore, it is reasonable to consider how the theoretical frameworks in the neurobiology of consciousness that are being applied to diagnosing unresponsive brain-injured patients could apply to HCOs.<sup>41,42</sup> Although such theories may be principally aimed at identifying empirically observable neural structures and functions that correspond to consciousness in the human brain, they might reveal the only empirically observable markers researchers have to work with when attempting to decipher whether HCOs have a capacity to be conscious.

It is worth considering that a recent study showed for the first time that cortical organoids generated from induced pluripotent stem cells can spontaneously develop periodic and regular oscillatory network electrical activity that resembles the electroencephalography (EEG) patterns of preterm babies.<sup>43</sup> A machine learning model based on a preterm newborn's EEG (ranging from 24 to 38 weeks) features was able to predict the organoid culture's age based on the electrical activity of the organoid itself. These results are surely relevant but by no means indicate that the recorded patterns of activity give rise to the same subjective states as can be believed to have originated in preterm babies.<sup>44</sup> In another study, researchers managed to visualize in cortical spheroids synchronized and non-synchronized activities in networks and connections between individual neurons.<sup>45</sup> The manifestation of a synchronized neural activity can be the basis for various relevant cognitive functions. Relatedly, Kagan et al. claim to have "demonstrated that a single layer of *in vitro* cortical neurons can self-organize activity to display intelligent and sentient behavior when embodied in a simulated game-world."<sup>46</sup>

One can reasonably question whether terms such as *sentience* are being used in this context accurately, analogously, or hyperbolically. Nevertheless, while we are still in the relatively early stages of HCO development, now is the time to preemptively do all we can to responsibly consider the implications of HCO development with respect to their capacity for consciousness. Hence, this article outlines how three neurobiological theories of consciousness apply to HCOs. To be clear, however, it will not answer the question of whether HCOs *are* conscious. In this vein, it should be pointed out that a strong disagreement exists in the scientific community, where many researchers, bioethicists, and neuroethicists consider the emergence of consciousness in HCOs to be unlikely, while others think it is a hypothesis that should be taken into account.<sup>47</sup> So, our article aims at offering a more nuanced theoretical framework within which one can place different views on the topic and assess the empirical findings. Although this article addresses a topic of great importance to bioethics and neuroethics regarding the research development and use of HCOs, it should be noted that providing ethical guidelines is not our focus. We are concerned specifically with what three neurobiological theories predict are the neuronal structures and functions that enable consciousness, and what this

**Table 1.** Describing consciousness

There is no single authoritative definition of consciousness. Yet, there are three prominent ways consciousness is described. Each presupposes a mental state that requires more or less properties to truly count as conscious.		Theory that favors description
Awareness	Description: Consciousness is a mental representation, or awareness, persons have of their environment. Example: When a patient hears a nurse's voice. <i>Notice this requires the ability to sense one's environment and have a mental representation of it.</i>	Temporal Circuit Hypothesis
Access consciousness	Description: Consciousness is mental content that is accessible for functions like rational thought, communication, or intentional behavior. Example: A memory that is recalled when describing an event. <i>Notice this requires mental content that is available for mentally involved functions.</i>	Global Neuronal Workspace Theory
Phenomenal consciousness	Description: Consciousness is experience, the felt sensation of what it is like to be in a particular mental state. Example: What it feels like to be depressed or to be happy. <i>Notice this requires the internal ability to feel, but not the ability to sense the environment nor think or intentionally act.</i>	Integrated Information Theory

means for HCOs with respect to their potential neurobiological capacity to be conscious. This can have significant ethical implications for the research development and use of HCOs, which is why it is important to address the issue. Yet, the aim of this article is not to spell out such ethical implications.<sup>48-49</sup>

This article surveys what three prominent neurobiological theories of consciousness predict are the neuronal structures and functions in the human brain that enable consciousness and explores the implications for HCOs.<sup>50</sup> We here assume consciousness generally corresponds to brain structure and activity, and apply to HCOs theoretical predictions from the Temporal Circuit Hypothesis, Global Neuronal Workspace Theory, and Integrated Information Theory about the neural correlate of being conscious versus unconscious.<sup>51-52</sup> This is often called the *full neural correlate of consciousness (full NCC)* and can be described as the minimal neuronal mechanisms that are together physically sufficient for consciousness.<sup>53-54</sup> The theories discussed here share similarities but differ in how they understand consciousness, which is notoriously difficult to define (see Table 1).<sup>55-56</sup> Correspondingly, each theory makes unique predictions about the full NCC (see Table 2). Our intention is to outline each theory and its applicability to HCOs in a manner accessible to specialists in various disciplines relevant to organoid research and development. Hence, there are elements of each theory left out, while representative works well worth further study for greater detail are referenced. There are multiple theories we do not have space to consider.<sup>57-58</sup> And our focus on the three theories considered here is not only because they provide examples of theories based on distinct, empirically testable understandings of consciousness, but also due to our own familiarity with each theory based on our prior research.

Our current aim is not to evaluate, defend, or critique (as many other works do) the neurobiological theories of consciousness outlined and applied later to HCOs. Yet, a thorough discussion of the applicability of the theories to HCOs would be deficient without also highlighting potential weaknesses of each theory in terms of their applicability to HCOs. It cannot be assumed that even if a theory accurately describes the full NCC in the human brain that it must directly apply to human cerebral organoids. Hence, the foregoing sections not only discuss how each theory might apply to HCOs but also particular challenges concerning their application to HCOs.

**Table 2.** Neurobiological theories and human cerebral organoids

Theory	Emphasizes awareness, access, or phenomenal consciousness	Predicted neuronal correlate of being conscious	Application to HCOs	Unique question raised
Temporal Circuit Hypothesis	Awareness	Consciousness is enabled by a temporal architecture involving a back-and-forth dance of the inner self and outer world representations mediated by attentional switching. This is supported by spatiotemporal dynamics of brain activity that displays a shifting balance between the default mode network (DMN) and the dorsal attention network (DAN) as governed by their regulator, the anterior insula.	A conscious HCO would include an inner self that is aware of its environment. This would involve representations of the environment that are conscious content due to a shifting balance between two systems governed by a temporal circuit. In the mature brain, the two systems are the DMN and DAN, whereas the regulator is the anterior insula. If there were the same systems or perhaps different systems that nevertheless performed the same functions in an HCO, the neurobiological conditions for consciousness would be met if the DAN (or whatever performs its functional role) could receive representations of the environment.	Could artificial sensory mechanisms be attached to the HCO, making representations of its environment possible?
Global Neuronal Workspace Theory	Access consciousness	In the brain, there is a workspace that stores information and makes it globally available to all the specialized systems throughout the brain. Whatever perception or thought that comes to occupy the workspace is conscious as it is projected to the systems throughout the brain. The workspace consists of long-range cortical neurons that comprise a network capable of reciprocal projections, which can receive information from sensory cortices and send projections to other cortical areas involved in working memory, language, and intentional action.	A conscious HCO would have a global workspace that stores information and specialized systems that the information is made available to, and neurons connecting the workspace to all the specialized systems, making the information available to each system. In the mature brain, a marker of consciousness is a global broadcast of information involving the activity of a prefrontal–parietal network of long-range cortical neurons corresponding with activity in high-level sensory cortices receiving the broadcast. If an HCO had a workspace that could receive a percept from a specialized system	In the absence of sensory information gained via sense organs, what type of information could enter the workspace and be made accessible to specialized systems?

Table 2. *Continued*

Theory	Emphasizes awareness, access, or phenomenal consciousness	Predicted neuronal correlate of being conscious	Application to HCOs	Unique question raised
			and then make it available to every other specialized system in the HCO, it would be neurobiologically capable of being conscious.	
Integrated Information Theory	Phenomenal consciousness	The physical substrate of consciousness exemplifies a maximally irreducible cause–effect structure that is the integrated information associated with consciousness. The intrinsic causal structure is predicted to be specified by the posterior cerebral cortex in a temporal–parietal–occipital “hot zone” where coalitions of neurons are sufficiently structured for reciprocal projections that manifest intrinsic causation.	A conscious HCO would have a coalition of neurons that are sufficiently structured for reciprocal projections that specify a maximally irreducible cause–effect structure. It is predicted that such neuronal coalition(s) exist in the temporal–parietal–occipital “hot zone” because the localized neuroarchitecture is sufficient for manifesting the maximally irreducible cause–effect structure. If an HCO had a set of neurons capable of reciprocal projections that would specify a maximally irreducible cause–effect structure, it would be neurobiologically capable of being conscious.	What could it possibly feel like to have a phenomenal conscious experience without any sensory input?

### Temporal Circuit Hypothesis

The Temporal Circuit Hypothesis (TCH) proposes that the neuronal basis of consciousness is a temporal architecture of representational brain processes that change dynamically and probabilistically.<sup>59,60</sup> These processes are supported by several functional brain networks, of which two networks play critical roles. According to the TCH, consciousness emerges from the dynamic relation of two systems that relate to the self and the environment, of which both are necessary. An agent (either a human or a nonhuman being) is assumed to have the ability to accurately represent its environment and form abstractions and concepts, but such representations or concepts in themselves do not endow the agent with awareness. Being conscious (or aware) also requires the agent to represent itself in relation to aspects of the world, either external to the agent or internal. In this sense, the crucial step in the emergence of consciousness is not only perception (i.e., conscious states with a specific content) but the percepts being *our* percepts. From this particular perspective presented by the TCH, the self is vital because it is the subject that is aware of the environment, and without a subject that is aware, there is no awareness and therefore no consciousness.<sup>61,62,63,64,65</sup>

The TCH anchors the “self-environment” interaction in a reciprocal balance between two opposing cortical systems embedded in the spatiotemporal dynamics of neural activity. Starting with a real-life example, imagine driving to work along the same route you take each day. Your mind wanders from one thing to the next. Suddenly, a car cuts you off, and wandering thoughts immediately vanish as all your attention focuses on maneuvering the steering wheel to avoid a collision. As we go about our waking lives, our stream of consciousness typically cycles through many such alternations between introspection and outward attention throughout the day. This back-and-forth “dance” between inward and outward mental states happens naturally and automatically. Evidence from noninvasive functional neuroimaging studies has pointed to two distinct brain networks that mediate the stream of consciousness, an internally directed system—default mode network (DMN)—and an externally directed system—dorsal attention network (DAN).<sup>66,67,68,69,70</sup> The former corresponds with inward focus on ourselves and conceptually guided cognition, and the latter corresponds with our awareness of the environment around us. In the conscious brain, the two systems are in a dynamic balance, sliding back and forth, but both are there to some degree.

Recent work reveals how cyclical patterns of brain activity and how the push–pull relationship between the DMN and DAN may differ between conscious and unconscious individuals.<sup>71</sup> In the conscious brain, the dynamic switching of networks including the DMN and DAN occurs along a set of structured transition trajectories conceived of as a “temporal circuit.” Thus, the conscious brain passes through intermediate states between the DMN and DAN, rather than flipping instantaneously between these two. In contrast, in the unconscious brain, there are fewer trajectories reaching the DMN and DAN. In other words, in subjects who were either under anesthesia or unconscious due to neuronal injury, both the DMN and DAN are visited less often, meaning that the number of times they are activated within the cycling patterns of networks is reduced. Thus, consciousness relies on a temporal circuit of dynamic brain activity representing balanced reciprocal accessibility of functional brain states. The absence of this “give-and-take” relationship of the two systems is common to any form of diminished consciousness. However, its presence corresponds to full consciousness, including both the external and internal aspects corresponding to the two systems.

The dynamic shift between the DMN and DAN is regulated by a critical brain structure, the anterior insula, a central component of the brain’s salience and ventral attention networks.<sup>72,73,74,75</sup> It seems to identify and prioritize salient stimuli in the stream of continuous sensory information and send signals to systems responsible for the allocation of top–down attentional resources to the relevant sensory representations.<sup>76,77,78</sup> Anatomically, the anterior insula is composed of unique clusters of large spindle-shaped pyramidal neurons in layer 5, called *von Economo* neurons that establish long-distance, fast relay of information throughout the cortex.<sup>79,80</sup> These neurons play a role in spatial and self-awareness, and their abnormality was historically associated with psychiatric disorders.<sup>81</sup> Interestingly, they have been found only in species that are able to pass the standard mirror test for self-recognition

such as elephants and macaques.<sup>82,83</sup> Importantly, the dysfunction of the anterior insula during anesthesia disables the cyclic DMN–DAN transitions.<sup>84</sup>

Overall, the TCH proposes that consciousness is enabled by a temporal architecture involving a back-and-forth dance of representations about the inner self and one's environment mediated by attentional switching. This is principally supported by the observed spatiotemporal dynamics of brain activity that displays a shifting balance between the DMN and DAN as governed by their regulator, the anterior insula.

The TCH seems to imply several neurobiological conditions for consciousness in HCOs. If an HCO were conscious, it would involve an inner self that is aware of its environment. This would involve representations of the environment that are conscious content given a shifting balance between two systems governed by a temporal circuit. In the mature human brain, the two systems are the DMN and DAN, whereas the regulator is the anterior insula. Furthermore, the DAN receives inputs from sensory systems with information from the surrounding environment. If an HCO had the same systems performing the same functions, the neurobiological conditions for consciousness would be met. Yet, it might also be possible for different systems to perform the same functions and therefore meet the neurobiological conditions. Nevertheless, when comparing the implications of the TCH for HCOs to the following two theories, readers will notice that a distinguishing feature of the TCH is that there is explicitly an inner self aware of its environment, which is enabled by a back-and-forth switching between an internally directed brain system and an externally directed brain system.

A more obvious challenge of applying the TCH to HCOs is the reality that isolated cerebral organoids do not have the typical conduits of sensory inputs to the brain, such as organs involved in hearing or smelling. A more difficult challenge of applying the TCH pertains to the self that is aware, the subject of experiences, which is the bearer of consciousness. When studying neural correlates of consciousness in human subjects, it is often justifiably presupposed that there is a subject, or self, who commonly provides subjective reports of conscious experiences.<sup>85</sup> But with respect to HCOs, it cannot be presupposed whether the self is or is not present. There are many examples of natural biological organisms devoid of a self (e.g., flowers and trees) and engineered biological organisms devoid of a self, such as kidney and cardiac organoids. Determining whether an HCO is a biological organism like the human brain or body in which a subject is present would seem necessary for the application of the TCH to an HCO. Yet, this would be an immensely complex task, especially in the absence of any subjective reports or history of such from a noncommunicative HCO.<sup>86</sup> While this is the primary challenge for the TCH, it will become apparent that each theory will have its own hurdles to overcome when it comes to applying the theory to HCOs.

### Global Neuronal Workspace Theory

The theoretical neurobiologist Bernard Baars proposed the Global Workspace Theory in the 1980s.<sup>87</sup> Baars posited that within the brain, there is a workspace that stores information and makes it globally available to all the specialized systems throughout the brain. Since the workspace's capacity is limited, information signals compete to be the globally available representation in the workspace. Whatever perception or thought that comes to occupy the workspace is conscious as it is projected to the systems throughout the brain. The percepts or thoughts that do not make it into the workspace, and therefore are not globally available, are not conscious.

Contemporary advocates of the Global Workspace Theory have applied it to the neurophysiology of the neocortex under the new name the Global Neuronal Workspace (GNW) Theory.<sup>88,89,90</sup> Accordingly, a marker of consciousness is a global broadcast of information involving the activity of a prefrontal–parietal network of long-range cortical neurons corresponding with activity in high-level sensory cortices receiving the broadcast. Such activity indicates that the information is globally available for various functional processes such as speech, memory, or action, and thus, it is conscious content, per the GNW. If, however, there are no long-range projections and neuronal activity is merely localized in specific areas, that would indicate there is no global broadcast of information, making it available to



specialized systems throughout the brain, and therefore, consciousness is not present. After all, according to the GNW, “Broadcasting implies that the information in the workspace becomes available to many local processors, and it is the wide accessibility of this information that is hypothesized to constitute conscious experience.”<sup>91</sup>

On this theoretical framework, content is either globally available, and therefore conscious, or it is not. Several factors make such global availability possible. Of central importance is the global workspace that receives, stores, and sends information. In addition, there are specialized systems that send content to the workspace and to which the globally broadcast content is made available. Neurons dubbed “GNW neurons” receive information from the specialized systems and transmit information to the specialized systems when it is globally broadcast.<sup>92</sup> While what constitutes conscious experience on the GNW is the content or information globally broadcast, there is also content that does not become conscious content because it does not enter the workspace and become globally available. Such content can still affect behavior but not consciously. It is only the specific content that exceeds a threshold, igniting a global broadcast of the content as GNW neurons receive the corresponding signal from a specialized system and recurrently transmits it to all other specialized systems.<sup>93</sup> When that global broadcast is triggered, there is a conscious experience with accessible content. Note the requirement for recurrent signaling in the brain to enable conscious perception is also the essential feature of the Recurrent Processing Theory of consciousness.<sup>94,95</sup>

If an HCO had a workspace that could receive a percept from a specialized system and then make it available to every other specialized system in the HCO, it would be neurobiologically capable of being conscious. Therefore, to find out whether an HCO is neurobiologically capable of being conscious, we would want to ask whether it has the neuroanatomy and neurophysiology for a workspace, specialized systems, and projections between the workspace and all the systems. Then, to find out whether it is manifesting the neurobiological condition for consciousness, we would want to find out whether projections are being sent from the workspace to all the specialized systems.

In a fully formed human brain, the workspace consists of long-range cortical neurons that comprise a network capable of reciprocal projections, which can receive information from sensory cortices and send projections to other cortical areas involved in working memory, language, and intentional action. The GNW does not assume that all physical systems that are conscious will have the exact same physical mechanisms that fulfill the roles involved in a global broadcast of information in the human brain. Accordingly, it is possible for computers to be conscious, per the GNW.<sup>96,97</sup> The difference maker for a computer’s physical capacity to be conscious would be whether there are mechanisms in a computer capable of a global broadcast. The same is true of HCOs.

If there were the exact same neuroanatomy and neurophysiology in an HCO that makes a global broadcast possible in the fully formed human brain, then it would be capable of meeting the neurobiological condition for consciousness. However, just as it is conceivable that another physical system like a computer could have different mechanisms than human brains that fulfill the same functional roles, thus satisfying the physical conditions for consciousness, it is conceivable that an HCO could have different mechanisms that fulfill the same roles. Or, they might have the same mechanisms, but not as fully developed. In other words, the neural mechanisms in HCOs capable of a global broadcast might be different from those in the mature human brain. Nevertheless, there would need to be something that fulfills the role of the workspace and something that fulfills the role of the specialized subsystems that send information to and receive information from the workspace.

There is, however, an epistemic difference between the brain of a healthy adult and an HCO that creates a hurdle. The former’s specialized systems are recognized by their association with certain behaviors or functions of a human subject. For example, Broca’s area is known to be associated with speech because we can associate activity in that area with a subject’s speech. Examples could easily be multiplied where specialized brain areas are specialized precisely because they correspond to a specific function of a subject. But in the case of HCOs, functions such as hearing, seeing, or speaking—which are associated with specialized systems in the mature human brain—are absent. The only information we have to make inferences from is the degree to which the HCO’s neuroanatomy and neurophysiology reflect that of the specialized systems and the workspace in a mature brain. Yet, even if HCOs directly

Downloaded from <https://www.cambridge.org/core>. Berkeley College Of Music, on 06 Feb 2025 at 01:35:01, subject to the Cambridge Core terms of use, available at <https://www.cambridge.org/core/terms>. <https://doi.org/10.1017/S095963180123000543>

replicate a mature human brain, things are still complicated by the fact that specialized systems in the human brain perform specific functions that are not necessarily isolated to the brain alone. Rather, they can be the result of receiving input from other biological parts of the whole organism outside the brain, such as sensory organs. They can also contribute to functions of other biological parts—just think again of Broca’s area associated with speech involving the mouth. These issues raise difficult questions.

For one, it must be asked whether the isolated functions in an HCO are truly the same functions as those in a normal brain, which are not isolated but rather part of a larger system with interacting parts. This question about the nature of the neurophysiological processes in HCOs will require an answer in order to discern whether HCOs meet the neurobiological conditions for consciousness from the perspective of the GNW. Moreover, it must be asked what kind of information could enter the workspace given that an HCO is an isolated cerebral entity devoid of sense organs and therefore receives no sensory information about the external environment (see, however, the concluding section). Will the information only be abstract information? If so, is it only the systems that correspond to abstract information processing that count as specialized systems that send and receive information? There is also a related question pertaining to the GNW generally that is more difficult to answer when applying the theory to HCOs: that is, what makes some neural states and activity correspond to information of a specific kind (e.g., sensory information)? These are the sort of questions that arise from the perspective of the GNW, which emphasizes access consciousness. The third theory to be considered understands phenomenal consciousness as fundamental to being conscious.

### Integrated Information Theory

The Integrated Information Theory (IIT) was proposed by the neuroscientist and psychiatrist Giulio Tononi, shortly after the turn of the century.<sup>98,99</sup> The IIT is based on five axioms about the nature of consciousness, which is understood as subjective experience that is phenomenal consciousness.<sup>100-101-102-103</sup> From the axioms about consciousness, five postulates are inferred about the nature of the physical substrate of consciousness (PSC).

The first axiom, intrinsic existence, says consciousness exists and is intrinsic to the subject of the conscious experience who has direct epistemic access to it. Based on this axiom, it is postulated that the PSC must exist and produce intrinsic causal effects upon itself. The second axiom, composition, says conscious experience is structured in that it has phenomenal distinctions. To grasp what a phenomenal distinction is, one can imagine the conscious experience of observing a sunset, during which the sound of a bird’s chirping is heard, the scent of freshly cut grass is smelt, and multiple colors are visually perceived. Given such phenomenal distinctions, the IIT postulates that there are distinct constituent elements of the PSC that have causal power upon the system, either by themselves or together with other elements. The third axiom, information, says that consciousness is specific, and it is the way it is; consequently, each experience is distinguished from other conscious experiences due to its distinct phenomenological features. It is postulated from this that the PSC must have a cause–effect structure of a specific form that makes it distinct from other possible structures. According to the fourth axiom, integration, a conscious experience is a unified whole irreducible to the phenomenal distinctions within it. Thus, the IIT postulates the cause–effect structure specified by the PSC must be unitary—that is, it must be irreducible to the one specified by non-interdependent causal subsystems. The final axiom, exclusion, claims a conscious experience is definite in its content and spatiotemporal grain. Therefore, the IIT postulates the cause–effect structure exemplified by the PSC must also be definite, being specified by a definite set of elements exhibiting maximal cause–effect power—neither less nor more—at a definite spatial and temporal grain.

According to the IIT, the PSC in the human central nervous system specifies a structure that exhibits maximal intrinsic cause–effect power.<sup>104</sup> This maximally irreducible cause–effect structure is the integrated information associated with consciousness, and as a structure that is theoretically possible to measure, it is thought to provide a basis for measuring consciousness.<sup>105</sup> The integrated information of the IIT is referred to as *Phi* and is represented by  $\Phi$ , with the vertical bar representing information and

the circle representing integration.<sup>106</sup> Since a single entity might have multiple manifestations of  $\Phi$ , it is important that it is the *maximal*  $\Phi$ , which is the greatest manifestation of intrinsic causation among a coalition of mechanisms in a system that is associated with consciousness. While the IIT's axioms and postulates concern human consciousness and the human nervous system, it has been argued that the theory can be applied to other biological organisms and even more broadly to all conceivable manifestations of consciousness.<sup>107</sup>

The primary prediction of the IIT is that maxima of  $\Phi$  indicates consciousness, and therefore, the ability to measure  $\Phi$  would make it possible to detect consciousness and measure the degree to which it is present. So, in a nervous system, the presence of consciousness will correspond with intrinsic causation manifested by a coalition of neurons and by measuring the maximal manifestation of such intrinsic causation you can measure the degree of consciousness. Furthermore, suppose there is no spontaneous manifestation of  $\Phi$  in the brain of a patient with a severe brain injury. A technique inspired by the IIT can be used to measure the extent to which the brain maintains adequate neurophysiology to manifest  $\Phi$ . The Perturbational Complexity Index (PCI), a simplified, surrogate measure of integrated information, reflects the response of the cortex, measured by electroencephalography (EEG), when it is perturbed using transcranial magnetic stimulation (TMS). The PCI reveals the extent to which the brain has the neurophysiology needed for manifesting integrated information in the form of intrinsic causation.<sup>108</sup>

Proponents of the IIT hypothesize that the full NCC in a mature human brain will be the subsystem with maximal  $\Phi$  and will be localized in the posterior cerebral cortex in a temporal–parietal–occipital “hot zone.”<sup>109</sup> The rationale for thinking the full NCC is in this posterior region is that this area has coalitions of neurons sufficiently structured for recurrent projections and, thus, for manifesting high  $\Phi$ . When it comes to applying the IIT to HCOs, what matters is the rationale for this hypothesis about the full NCC, not necessarily whether such a hot zone has been formed within the organoid. If there are neurons in an HCO sufficiently structured for reciprocal projections that would manifest a global maximum of  $\Phi$ , then the HCO is capable of meeting the IIT's neurobiological condition for consciousness. When such reciprocal projections take place among the coalition of neurons, the neurobiological conditions for being conscious are met.

If the technology were available for securing a PCI for HCOs, which is theoretically plausible, then we could find out whether they have the neuroanatomy and neurophysiology sufficient for manifesting  $\Phi$ . Moreover, if we had the ability to measure  $\Phi$  in HCOs, then we could know whether and to what extent the neurobiological conditions for consciousness are being met.

In the context of considerations about HCOs, what noticeably distinguishes the IIT is that it permits the possibility of minimal consciousness in strikingly simple physical systems, even something as simple as a single photodiode.<sup>110,111</sup> Some might see the IIT's relatively liberal ascription of consciousness as a strength, while others might see it as a shortcoming.<sup>112,113,114</sup> It is worth mentioning, however, that the GNW might also allow for quite simple systems, composed of just several units, to implement a global workspace and therefore be conscious.<sup>115,116,117</sup> Nevertheless, does the IIT's generous ascription of consciousness mean that an HCO consisting of just several million neurons is capable of the same conscious experiences as a person with a fully developed human brain consisting of eighty-six billion neurons? Although the number of neurons matters, it is not the most important factor. The key condition is whether the neurons comprising an HCO include a coalition of neurons sufficiently structured for manifesting  $\Phi$ . If there is such a coalition, the HCO is capable of meeting the IIT's neurobiological condition for consciousness, and when it manifests maximal  $\Phi$ , then it meets the condition.

However, suppose that an HCO manifests a maximal  $\Phi$ , and it is therefore inferred that the HCO is conscious. Does this mean it is experiencing in the way that someone with a fully formed human brain is? Per the IIT, the HCO is indeed conscious—but according to the IIT, consciousness is not all or nothing. Rather, it comes in degrees. So, the HCO will be conscious to the same degree as a mature human to the degree that the HCO is manifesting the same degree of  $\Phi$ . An implication is that the less HCOs are like fully formed human brains in their capacity to manifest  $\Phi$ , the less they are capable of being conscious to the degree mature humans are, and of having similar experiences. Yet, the more like

human brains HCOs are in their capacity to manifest  $\Phi$ , the more they are capable of being conscious to the degree that mature humans are. Granted, it is hard to imagine what an HCO's conscious experience might be like when there is no input from sense organs. Nevertheless, there would be a conscious experience of what it's like to be an HCO, according to the IIT, even if imagining what exactly it's like brings us to the edge of our own empathy. Here, the IIT's aim to not only measure the degree of consciousness but also infer its content would become relevant since it would be important to know whether the content includes undesirable states such as pain.<sup>118,119,120</sup>

Overall, according to the IIT, human cerebral organoids (and systems in general) are conscious if and only if they exemplify a global maxima of  $\Phi$ , and they experience pain if and only if their irreducible cause–effect structure is composed of distinctions and relations mimicking the ones that systematically correspond to pain when specified by the PSC in the human brain. The benefits of this claim are its precision and that it provides a theoretical basis for detecting consciousness in HCOs. At the same time, some might critique this element of the IIT on the grounds that it entails that not only human cerebral organoids but also any other type of organoid (such as kidney organoids) could be conscious if and only if they exemplify a global maxima of  $\Phi$ . This seems to go against the grain of a key presupposition in the overall discussion. That is, there is something unique about human cerebral organoids that makes them possible candidates for consciousness—namely, their *neuronal* constitution as it resembles the human brain. At this point, IIT theorists may have to either bite the bullet that kidney organoids could be conscious or hope that it can be conclusively demonstrated that cortical neurons are the only cells capable of exemplifying a global maxima of  $\Phi$ .

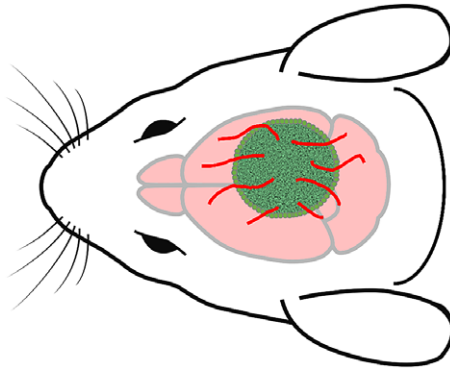
If it were shown that kidney organoids manifest a global maxima of  $\Phi$  and cerebral organoids do not, that would be a counterintuitive result within a field that emphasizes the brain and its neuronal constitution in relationship to consciousness. If, however, HCOs were shown to manifest a global maxima of  $\Phi$  and organoids devoid of neurons were shown to lack a global maxima of  $\Phi$ , this might suggest the IIT has identified a unique feature of neurons in the organ that corresponds to consciousness. But it would also imply, from the standpoint of the IIT, that HCOs are conscious.

### Discussion: Current Science and Future Progress

Having outlined three prominent neurobiological theories of consciousness and what they predict are the neuronal conditions that enable consciousness, it is important to note that no one theory is widely considered demonstrably true.<sup>121,122</sup> The empirical research continues, and competitively so.<sup>123,124,125</sup> There are also conceptual and empirical challenges that the theories must overcome. For example, Koch critiques the GNW's computational account of consciousness as he argues for the IIT, while Doerig et al. use theorems from computational theory to argue that the IIT is either false or outside the domain of science.<sup>126,127</sup> Meanwhile, Sánchez-Cañizares calls for greater clarity regarding the IIT's metaphysical commitments as Bayne conceptually challenges its axiomatic foundation and Noel et al. raise an empirical challenge.<sup>128,129,130</sup> It is easy to find examples of objections to the prominent theories given the amount of attention they receive. But each theory, notorious or not, has its own challenges to contend with.

Additionally, the development of HCOs is well underway but far from the finish line. While the ultimate objective might be a whole-brain organoid for comprehensive modeling of the brain's biological development and neurophysiology, there is a long way to go.<sup>131</sup> Obstacles such as the challenge of providing vascularization for organoids prohibit their development to more mature stages.<sup>132,133,134</sup> Despite current limitations, the theories outlined here can guide considerations and prompt key questions concerning the neurobiological conditions for consciousness as organoid development progresses and the underexplored physiology of neuronal circuits in HCOs is further studied.<sup>135</sup>

While no theory of consciousness currently enjoys consensus confirmation, each theory can be used to prompt important questions, some of which might be more germane to specific stages of organoid development. Even in light of current neural tissue development, the IIT prompts the question of



**Figure 2.** *Human cerebral organoid and mouse chimera.* To overcome a lack of vascularization that limits a human cerebral organoid's growth *in vitro*, it can be implanted into a mouse brain, producing a chimera. The mouse brain then provides vascularization for the implanted human cerebral organoid. This raises various questions. For example, could the organoid grow to the extent that it becomes sufficiently developed to manifest the neurobiological conditions for consciousness? If the mouse is conscious, could the organoid become part of the physical substrate of the mouse's consciousness?

whether *in vitro* cortical neurons can manifest intrinsic causation via recurrent excitation and feedback loops.<sup>136</sup> The IIT also leads to an interesting question pertinent to a tactic for providing HCOs with vascularization. One approach is to implant them into a host animal brain resulting in what are called *chimeras*, as depicted in Figure 2.<sup>137</sup> From the vantage point of the IIT, this maneuver invites the question of whether the implanted HCO benefitting from the vascular system of the host animal might be incorporated into a maximally irreducible causal structure in the host brain. This would make the hitchhiking HCO part of the physical substrate of the host animal's consciousness (assuming it is conscious). Even if this is no more ethically suspect than an organ transplant, by seriously reflecting on such issues, the scientific community can demonstrate ethical responsibility in research that safeguards public trust.<sup>138,139</sup> And exploring the implications of the different theories in such scenarios might lead to fruitful questions that hone our development and understanding of the theories.

For example, suppose it is possible in the case of chimeras that the host animal is conscious and the HCO could become conscious. We could reasonably ask whether the HCO could have its own conscious experience distinct from the animal's conscious experience. From the perspective of the IIT, it might seem that there are two possibilities. On the one hand, if the implanted HCO is part of the host animal's body, and specifically its brain, it seems that it could become part of the animal's physical substrate of consciousness. Given this, it could be said that the animal's conscious experience corresponds to the neuronal activity in the HCO. However, this would require the neural activity in the HCO being part of the maximally irreducible causal structure in the host animal's brain. After all, if the HCO is part of the animal's body, and specifically its brain, it could hypothetically manifest intrinsic causation that does not correspond to consciousness. Because according to the IIT, it is only the maximally irreducible causal structure in the brain that corresponds to consciousness. But this raises the question of what would make the HCO part of the animal's body rather than a stand-alone entity? Put differently, since a maximally irreducible causal structure excludes all other causal structures specified by an overlapping subsystem, it is vital to know what does and does not count as a part of the system. Likewise, from the perspective of the GNW, what counts as a global broadcast of information also depends on the boundaries of the system within which a broadcast is or is not global. Suppose an animal were conscious because there was a global broadcast to all the specialized areas in its brain. If it had an HCO implanted into its brain and the HCO developed specialized areas, would it be necessary for a broadcast in the animal's brain to reach these specialized areas in the HCO before it is conscious? If so, that would seem to be because the implanted HCO is part of the animal's brain. If not, that would seem to be because the implanted HCO is not part of

the animal's brain. Yet, what would make it part of (or not part of) the animal's brain to which the condition of a global broadcast applies?

Some might argue that such questions arising from such novel scenarios demonstrate how the neurobiological theories are not sufficiently developed or specified to be applicable to possible conscious systems beyond the human brain, such as HCOs.<sup>140,141</sup> It would surely be unwise to simply assume the theories apply as directly to an HCO as to a human brain. However, it does not follow that they cannot be applicable to HCOs at all, and the only tools anyone can apply to any task are the tools they possess. If there were a validated theory that was fully developed and specifically applied directly to all possible instances of consciousness, we would be at the end goal of a major scientific project. Yet, that is not where the scientific study of consciousness currently is. What we have to work with are the best theories to date. And progress in the development of chimeras is not likely to wait for neurobiological theories of consciousness to be perfectly honed. A recent study reports that HCOs transplanted into the somatosensory cortex of newborn athymic rats “extend axons throughout the rat brain” and developed mature cell types that “integrate into sensory and motivation-related circuits.”<sup>142</sup> Applying neurobiological theories of consciousness to such novel issues can lead to their development not only when the theories are shown to be useful but also if they are shown to have limitations. Asking the difficult questions about how they apply to the most complicated conundrums in consciousness research is one way to reveal their explanatory utility, or lack thereof.

In addition to chimeras, what are called *assembloids* are another novel creation in HCO research worthy of consideration. The development of an organoid can be guided using molecules and growth factors to grow a specific type of cell tissue composing an organoid that resembles a specific brain region.<sup>143</sup> Multiple guided organoids resembling specific brain regions can also be assembled and connected by interneurons.<sup>144</sup> From the perspective of the GNW, the ability to manufacture assembloids prompts the question of whether specific parts of the brain that might serve as the global workspace and other parts that could serve as specific specialized systems might be grown, assembled, and interconnected in a way that the neuronal structures the GNW posits are preconditions for consciousness would be present. A thorough consideration of such possibilities invites proponents of the GNW to specify a threshold number of specialized systems needed for a global broadcast, and whether as few as one or two would be sufficient. A question with details pertinent to the TCH can also be asked if cellular tissue involved in sensing external stimuli is grown and connected to a cerebral organoid. The relevance of such a question is increasing. It has already been reported that cerebral organoids can assemble primitive light-sensitive optic vesicles, permitting interorgan interaction studies and potentially direct inputs from the environment.<sup>145</sup>

As we conclude, let us offer a brief catalog of further theoretical questions worthy of future considerations and study. In addition to the degree of similarity between the structures of HCOs and those that correspond to consciousness in the human brain, according to the theories mentioned earlier, what extent of similarity is needed to make plausible inferences about the capacity for consciousness? Related to this, is the size and number of neurons something important, or is an isomorphism and similar functionality between developing human brains and HCOs sufficient? Likewise, it is worth asking whether consciousness comes in degrees or if it is an all-or-nothing property that is either fully present or not present at all. If fetuses have rudimentary forms of sentience that differ from those experienced by adult humans, could the same be true of HCOs?

It would also be immensely helpful to know whether we can devise empirical studies based on the theoretical claims discussed before. An example is measuring the Perturbational Complexity Index (PCI) in HCOs, as proposed by Lavazza and Massimini, which might yield a proxy for the presence of a capacity for consciousness.<sup>146</sup> At the same time, we will need to continually wrestle with the question of whether empirical indicators in the human brain indicate the same phenomenon in HCOs. For example, according to the GNW, long-range cortico-cortical functional connectivity seems to be a key neurobiological condition for consciousness in the human brain. However, in HCOs with a size of a few millimeters, these (centimeter) long connections are absent. Yet, the event-related potential P300 has been deemed one of the neuronal markers when it comes to the detection of a particular level of information available in the global workspace, and in particular the P3b, whereas

the P3a component might be linked to automatic and nonconscious processes.<sup>147</sup> Could the P300 wave indicate the same in cerebral organoids? Could specific neural activity that indicates a global broadcast in the human brain be indicative of the same in an HCO? Or, should we expect a modified indicator corresponding to the differences between the brain and a cerebral organoid? We cannot assume a neuronal indicator in a healthy human subject indicates the same phenomenon in an unresponsive brain-injured human patient without sufficient justification for thinking so.<sup>148,149</sup> This is even truer with respect to HCOs.

There is no reason to think the questions raised here will yield perspicuous answers. And the theories discussed earlier may provide divergent answers that must be continually critically evaluated. However, what is clear in light of the rapid development of HCOs is that now is the time to carefully consider the difficult questions. To foster and aid considerations about the neurobiological capacity HCOs might develop for consciousness, we have attempted to apply three theoretical predictions about the neurobiological conditions for consciousness to HCOs. Establishing which prediction is most viable could be an important step toward eventually deciphering whether the neurobiological conditions for consciousness are present in HCOs.

**Acknowledgments.** M.O. would like to thank Christopher Moraes for providing Figures 1 and 2 and for helpful discussions pertaining to sections two and six.

**Data availability statement.** Data analyzed are available in supplementary sources.

**Author contribution.** M.O. drafted much of the article, while Z.H. and A.G.H. wrote and edited the third section. Z.H. and A.G.H. also contributed to writing the first two sections and revising them with M.O. C.D. contributed to substantial discussion of the article content with M.O. and helped conceptualize and revise the first and second sections. M.O., C.D., M.G., and A.C.H. conceptually revised the fourth and fifth sections and the corresponding table. M.O. and M.G. edited the fifth section. A.L. contributed to the drafting of and revising of the first, second, and sixth sections. All authors contributed to revising/editing the manuscript and approved the final version.

**Competing interest.** The authors declare none.

## Notes

1. Setiaa H, Muotri AR. Brain organoids as a model system for human neurodevelopment and disease. *Seminars in Cell and Developmental Biology* 2019;**95**:93–7.
2. Sun N, Meng X, Liu Y, Song D, Jiang C, Cai J. Applications of brain organoids in neurodevelopment and neurological diseases. *Journal of Biomedical Science* 2021;**28**:1–16.
3. Giacino JT, Kalmar K, Whyte J. The JFK coma recovery scale-revised: Measurement characteristics and diagnostic utility. *Archives of Physical Medicine and Rehabilitation* 2004;**85**:2020–9.
4. Giacino JT, Fins JJ, Laureys S, Schiff ND. Disorders of consciousness after acquired brain injury: The state of the science. *Nature Reviews Neurology* 2014;**10**:99–114.
5. Schnakers C, Hirsch M, Noé E, Llorens R, Lejeune N, Veeramuthu V, et al. Covert cognition in disorders of consciousness: A meta-analysis. *Brain Sciences* 2020;**10**:1–11.
6. Wannez S, Heine L, Thonnard M, Gosseries O, Laureys S. The repetition of behavioral assessments in diagnosis of disorders of consciousness. *Annals of Neurology* 2017;**81**:883–9.
7. Fernández-Espejo D, Rossit S, Owen AM. A thalamocortical mechanism for the absence of overt motor behavior in covertly aware patients. *JAMA Neurology* 2015;**72**:1442–50.
8. Monti MM, Vanhaudenhuyse A, Coleman MR, Boly M, Pickard JD, Tshibanda L, et al. Willful modulation of brain activity in disorders of consciousness. *The New England Journal of Medicine* 2010;**362**:579–89.
9. Owen AM, Coleman MR, Boly M, Davis MH, Laureys S, Pickard JD. Detecting awareness in the vegetative state. *Science* 2006;**313**:1402.
10. Schiff ND. Cognitive motor dissociation following severe brain injuries. *JAMA Neurology* 2015;**72**:1413.

11. Kondziella D, Friberg CK, Frokjaer VG, Fabricius M, Møller K. Preserved consciousness in vegetative and minimal conscious states: Systematic review and meta-analysis. *Journal of Neurology, Neurosurgery and Psychiatry* 2016;**87**:485–92.
12. Hudetz A. Are we unconscious during general anesthesia? *International Anesthesiology Clinics* 2008;**46**:25–42.
13. Laureys S, Pellas F, Van Eeckhout P, Ghorbel S, Schnakers C, Perrin F, et al. The locked-in syndrome: What is it like to be conscious but paralyzed and voiceless? *Progress in Brain Research* 2005;**150**:495–511.
14. Sanders RD, Tononi G, Laureys S, Sleigh JW. Unresponsiveness ≠ unconsciousness. *Anesthesiology* 2012;**116**:946–59.
15. Huang Z, Vlisides PE, Tarnal VC, Janke EL, Keefe KM, Collins MM, et al. Brain imaging reveals covert consciousness during behavioral unresponsiveness induced by propofol. *Scientific Reports* 2018;**8**:13195.
16. Boly M, Coleman MR, Davis MH, Hampshire A, Bor D, Moonen G.. When thoughts become action: An fMRI paradigm to study volitional brain activity in non-communicative brain injured patients. *Neuroimage* 2007;**36**:972–92.
17. See [note 8](#), Monti et al. 2010.
18. Owen AM. Improving diagnosis and prognosis in disorders of consciousness. *Brain* 2020;**143**:1046–56.
19. See [note 15](#), Huang et al. 2018.
20. Casali AG, Gosseries O, Rosanova M, Boly M, Sarasso S, Casali KR, et al. A theoretically based index of consciousness independent of sensory processing and behavior. *Science Translational Medicine* 2013;**5**:198ra105.
21. Gosseries O, Di H, Laureys S, Boly M. Measuring consciousness in severely damaged brains. *Annual Review of Neuroscience* 2014;**37**:457–78.
22. Sarasso S, Rosanova M, Casali AG, Casarotto S, Fedchio M, Boly M, et al. Quantifying cortical EEG responses to TMS in (un)consciousness. *Clinical EEG and Neuroscience* 2014;**45**:40–9.
23. Storm JF, Boly M, Casali AG, Massimini M, Olcese U, Pennartz CMA, et al. Consciousness regained: Disentangling mechanisms, brain systems, and behavioral responses. *Journal of Neuroscience* 2017;**37**:10882–93.
24. Arlotta P, Paşca SP. Cell diversity in the human cerebral cortex: From the embryo to brain organoids. *Current Opinion in Neurobiology* 2020;**56**:194–8.
25. Lancaster MA, Renner M, Martin C-A, Wenzel D, Bicknell LS, Hurles ME, et al. Cerebral organoids model human brain development and microcephaly. *Nature* 2013;**501**:373–9.
26. See [note 1](#), Setiaa et al. 2019.
27. Benito-Kwiecinski S, Lancaster MA. Brain organoids: Human neurodevelopment in a dish. *Cold Spring Harbor Perspectives in Biology* 2019;**12**:8.
28. Marx V. Reality check for organoids in neuroscience. *Nature Methods* 2020;**17**:961–4.
29. See [note 1](#), Setiaa et al. 2019.
30. See [note 2](#), Sun et al. 2021.
31. Farahany NA, Greely HT, Hyman S, Koch C, Grady C, Paşca SP, et al. The ethics of experimenting with human brain tissue. *Nature* 2018;**556**:429–32.
32. Lavazza A. Human cerebral organoids and consciousness: A double-edged sword. *Monash Bioethics Review* 2020;**38**:105–28.
33. National Academies of Science, Engineering, and Medicine. *The Emerging Field of Human Neural Organoids, Transplants, and Chimeras: Science, Ethics, and Governance*. Washington, DC: The National Academies Press; 2021.
34. Ankenyv RA, Wolvetang E. Testing the correlates of consciousness in brain organoids: How do we know and what do we do? *American Journal of Bioethics* 2020;**21**:51–3.
35. Bayne T, Seth A, Massimini M. Are there islands of awareness? *Trends in Neuroscience* 2020;**43**:6–16.
36. Lavazza A. Potential ethical problems with human cerebral organoids: Consciousness and moral status of future brains in a dish. *Brain Research* 2021;**1750**:147146.



37. Lavazza A, Massimini M. Cerebral organoids: Ethical issues and consciousness assessment. *Journal of Medical Ethics* 2018;**44**:606–10.
38. Lavazza A, Massimini M. Cerebral organoids and consciousness: How far are we willing to go? *Journal of Medical Ethics* 2018;**44**:613–4.
39. Sawai T, Sakaguchi H, Thomas E, Takahashi J, Fujita M. The ethics of cerebral organoid research: Being conscious of consciousness. *Stem Cell Reports* 2019;**13**:440–7.
40. Sawai T, Hayashi Y, Niikawa T, Shepherd J, Thomas E, Lee T-L, et al. Mapping the ethical issues of brain organoid research and application. *AJOB Neuroscience* 2022;**13**:81–94.
41. Cf. Niikawa T, Hayashi Y, Shepherd J, Sawai T. Human cerebral organoids and consciousness. *Neuroethics* 2022;**15**:1–16.
42. Cf. Shepherd J. Ethical (and epistemological) issues regarding consciousness in cerebral organoids. *Journal of Medical Ethics* 2018;**44**:611–2.
43. Trujillo CA, Gao R, Negraes PD, Gu J, Buchanan J, Preissl S, et al. Complex oscillatory waves emerging from cortical organoids model early human brain network development. *Cell Stem Cell* 2019;**25**:558–69.
44. Lavazza A. ‘Consciousnessoids’: Clues and insights from human cerebral organoids for the study of consciousness. *Neuroscience of Consciousness* 2021;**2**:niab029.
45. Sakaguchi H, Ozaki Y, Ashida T, Matsubara T, Oishi N, Kihara S, et al. Self-organized synchronous calcium transients in a cultured human neural network derived from cerebral organoids. *Stem Cell Report* 2019;**13**:458–73.
46. Kagan BJ, Kitchen AC, Tran NT, Habibollahi F, Khajehnejad M, Parker BJ, et al. In vitro neurons learn and exhibit sentience when embodied in a simulated game-world. *Neuron* 2022;**110**:1–18.15.
47. Lavazza A, Chinaia AA. Human cerebral organoids: The ethical stance of scientists. *Stem Cell Research & Therapy* 2023;**14**:1–8.
48. See [note 32](#), Lavazza 2020.
49. See [note 36](#), Lavazza 2021.
50. See also Zilio F, Lavazza A. Consciousness in a rotor? Science and ethics of potentially conscious human cerebral organoids. *AJOB Neuroscience* 2023;**14**(2):178–96.
51. Cf. Cecconi B, Laureys S, Annen J. Islands of awareness or cortical complexity? *Trends in Neuroscience* 2020;**43**:545–6.
52. See also Owen M, Huang Z, Duclos C, Lavazza A, Grasso M, LaRock E, et al. Human cerebral organoids and consciousness: Surveying relevant philosophy of mind. In: LaRock E, Guta MP, eds. *Experimental Evidence and the Conscious Person*. forthcoming.
53. Koch C, Massimini M, Boly M, Tononi G. Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience* 2016;**17**:307–21.
54. Owen M, Guta MP. Physically sufficient neural mechanisms of consciousness. *Frontiers in Systems Neuroscience* 2019;**13**:1–14.
55. Chalmers DJ. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press; 1996, at 3–11.
56. Block N. On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 1995;**18**:227–87.
57. Seth A, Bayne T. Theories of consciousness. *Natural Reviews Neuroscience* 2022;**23**:439–452. doi:10.1038/s41583-022-00587-4.
58. Sattin D, Magnani FG, Bartesaghi L, Caputo M, Fittipaldo AV, Cacciatore M, et al. Theoretical models of consciousness: A scoping review. *Brain Sciences* 2021;**11**:1–58.
59. Huang Z, Zhang J, Wu J, Mashour GA, Hudetz AG. Temporal circuit of macroscale dynamic brain activity supports human consciousness. *Science Advances* 2020;**6**:1–14.
60. Huang Z, Tarnal V, Vlisides PE, Janke EL, McKinney AM, Picton P, et al. Anterior insula regulates brain network transitions that gate conscious access. *Cell Reports* 2021;**35**:109081.
61. Critchley HD, Wiens S, Rotshtein P, Öhman A, Dolan RJ. Neural systems supporting interoceptive awareness. *Nature Neuroscience* 2004;**7**:189–95.

62. Damasio A. *Self Comes to Mind: Constructing the Conscious Brain*. New York, NY: Pantheon Books; 2010.
63. Faivre N, Salomon R, Blanke O. Visual consciousness and bodily self-consciousness. *Current Opinion in Neurology* 2015;**28**:23–8.
64. Guta MP. The non-causal account of the spontaneous emergence of phenomenal consciousness. In: Guta MP, ed. *Consciousness and the Ontology of Properties*. London: Routledge; 2019:126–52.
65. Park H-D, Tallon-Baudry C. The neural subjective frame: From bodily signals to perceptual consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2014;**369**:20130208.
66. Raichle ME. The brain's default mode network. *Annual Review of Neuroscience* 2015;**38**:433–47.
67. Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America* 2001;**98**:676–82.
68. Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. *Natural Reviews Neuroscience* 2002;**3**:201–15.
69. Fornito A, Harrison BJ, Zalesky A, Simons JS. Competitive and cooperative dynamics of large-scale brain functional networks supporting recollection. *Proceedings of the National Academy of Sciences* 2012;**109**:12788–93.
70. Vossel S, Geng JJ, Fink GR. Dorsal and ventral attention systems: Distinct neural circuits but collaborative roles. *Neuroscientist* 2014;**20**:150–9.
71. See note 59, Huang et al. 2020.
72. See note 60, Huang et al. 2021.
73. Fox MD, Corbetta M, Snyder AZ, Vincent JL, Raichle ME. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences of the United States of America* 2006;**103**:10046–51.
74. Menon V. Large-scale brain networks and psychopathology: A unifying triple network model. *Trends in Cognitive Sciences* 2011;**15**:483–506.
75. Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, et al. Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience* 2007;**27**:2349–56.
76. See note 74, Menon 2011.
77. Michel M. A role for the anterior insular cortex in the global neuronal workspace model of consciousness. *Consciousness and Cognition* 2017;**49**:333–46.
78. Uddin LQ. Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience* 2015;**16**:55–61.
79. Allman JM, Tetreault NA, Hakeem AY, Manaye KF, Semendeferi K, Erwin JM, et al. The von Economo neurons in the frontoinsula and anterior cingulate cortex. *Annals of the New York Academy of Sciences* 2011;**1225**:59–71.
80. Craig AD. How do you feel - now? The anterior insula and human awareness. *Nature Reviews Neuroscience* 2009;**10**:59–70.
81. von Economo C. *Die Encephalitis Lethargica*. Vienna: Franz Deuticke; 1918.
82. Critchley H, Seth A. Will studies of macaque insula reveal the neural mechanisms of self-awareness? *Neuron* 2012;**74**:423–6.
83. Evrard HC, Forro T, Logothetis NK. Von Economo neurons in the anterior insula of the macaque monkey. *Neuron* 2012;**74**:482–9.
84. See note 60, Huang et al. 2021.
85. Tsuchiya N, Wilke M, Frässle S, Lamme VAF. No-report paradigms: Extracting the true neural correlates of consciousness. *Trends in Cognitive Sciences* 2015;**19**:757–70.
86. See note 52, Owen et al. forthcoming.
87. Baars BJ. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press; 1988.
88. Dehaene S. *Consciousness and the Brain: Deciphering How the Brain Codes our Thoughts*. London: Penguin Books; 2014.

89. Dehaene S, Changeux J-P. Experimental and theoretical approaches to conscious processing. *Neuron* 2011;**70**:200–27.
90. Mashour GA, Roelfsema P, Changeux J-P, Dehaene S. Conscious processing and the global neuronal workspace hypothesis. *Neuron* 2020;**105**:776–98.
91. See note 90, Mashour et al. 2020, at 776.
92. See note 90, Mashour et al. 2020, at 777.
93. See note 90, Mashour et al. 2020, at 777.
94. Lamme VAF. How neuroscience will change our view on consciousness. *Cognitive Neuroscience* 2010;**1**:204–20.
95. See note 89, Dehaene, Changeux 2011, at 210.
96. Dehaene S, Lau H, Kouider S. What is consciousness, and could machines have it? *Science* 2017;**358**:486–92.
97. Dehaene S, Lau H, Kouider S. Conscious machines: Robot rights. *Science* 2018;**359**:400–2.
98. Tononi G. An information integration theory of consciousness. *BMC Neuroscience* 2004;**5**:1–22.
99. Tononi G. Consciousness as integrated information: A provisional manifesto. *The Biological Bulletin* 2008;**215**:216–42.
100. Tononi G. Integrated information theory. *Scholarpedia* 2015;10.
101. Oizumi M, Albantakis L, Tononi G. From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLOS Computational Biology* 2014;**10**:e1003588.
102. Tononi G, Boly M, Massimini M, Koch C. Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience* 2016;**17**:450–61.
103. Haun A, Tononi G. Why does space feel the way it does? Towards a principled account of spatial experience. *Entropy* 2019;**21**:1160.
104. See note 102, Tononi et al. 2016, at 450.
105. Koch C. *The Feeling of Life Itself: Why Consciousness Is Widespread but Cannot be Computed*. Cambridge, MA: MIT Press; 2019: at 87–9, 104.
106. Massimini M, Tononi G. *Sizing Up Consciousness: Towards an Objective Measure of the Capacity for Experience*. Oxford: Oxford University Press; 2018, at 72.
107. Tononi G, Koch C. Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences* 2015;**370**:20140167.
108. See note 106, Massimini, Tononi 2018, at 119–42.
109. See note 53, Koch et al. 2016, at 308, 315.
110. See note 101, Oizumi et al. 2014, at 19.
111. Tononi G. The integrated information theory of consciousness: An outline. In: Schneider S, Velmans M, eds. *The Blackwell Companion to Consciousness*. 2nd ed. Hoboken, NJ: Wiley Blackwell; 2017:243–56, at 254.
112. Goff P. *Galileo's Error: Foundations for a New Science of Consciousness*. New York, NY: Pantheon Books; 2019.
113. Merker B, Williford K, Rudrauf D. The integrated information theory of consciousness: A case of mistaken identity. *Behavioral and Brain Sciences* 2021;**45**:e41.
114. Owen M. Exploring Common Ground Between the Integrated Information Theory and Aristotelian Metaphysics. *Journal of Consciousness Studies* 2019;**26**:163–87, at 178–83.
115. Dehaene S, Sergent C, Changeux JP. A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences* 2003;**100**:8520–5.
116. Van Vugt B, Dagnino B, Vartak D, Safaai H, Panzeri S, Dehaene S, et al. The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science* 2018;**360**:537–42.
117. Doerig A, Schurger A, Herzog MH. Hard criteria for empirical theories of consciousness. *Cognitive Neuroscience* 2021;**12**:41–62.
118. See note 103, Haun, Tononi 2019.
119. Ellia F, Hendren J, Grasso M, Kozma C, Mindt G, P Lang JP, et al. Consciousness and the fallacy of misplaced objectivity. *Neuroscience of Consciousness* 2021;**7**:niab032.

120. Grasso M, Haun AM, Tononi G. Of maps and grids. *Neuroscience of Consciousness* 2021;7:niab022.
121. See note 57, Seth, Bayne 2022.
122. See note 58, Sattin et al. 2021.
123. Reardon S. 'Outlandish' competition seeks the brain's source of consciousness: \$20 million project puts competing ideas to the test. *Science* 2019. doi:10.1126/science.aaz8800.
124. Melloni L, Mudrik L, Pitts M, Bendtz K, Ferrante O, Gorska U, et al. An adversarial collaboration protocol for testing contrasting predictions of global neuronal workspace and integrated information theory. *Plos One* 2023;18:e0268577.
125. Cogitate Consortium, Ferrante O, Gorska-Klimowska U, Henin S, Hirschhorn R, Khalaf A, et al. An adversarial collaboration to critically evaluate theories of consciousness. 2023; available at <https://www.biorxiv.org/content/10.1101/2023.06.23.546249v2>.
126. See note 105, Koch 2019, at chaps. 11–2.
127. Doerig A, Schurger A, Hess K, Herzog MH. The unfolding argument: Why IIT and other causal structure theories cannot explain consciousness. *Consciousness and Cognition* 2019;72:49–59.
128. Sánchez-Cañizare J. Integrated information theory as testing ground for causation: Why nested hylomorphism overcomes physicalism and panpsychism. *Journal of Consciousness Studies* 2022;29:56–78.
129. Bayne T. On the axiomatic foundations of the integrated information theory of consciousness. *Neuroscience of Consciousness* 2018;4:1–8.
130. Noe JP, Ishizawa Y, Patel SR, Eskandar EN, Wallace MT. Leveraging nonhuman primate multi-sensory neurons and circuits in assessing consciousness Theory. *The Journal of Neuroscience* 2019;39:7485–500.
131. Qian X, Song H, Ming G. Brain organoids: Advances, applications and challenges. *Development* 2019;146:dev166074.
132. Cakir B, Xiang Y, Tanaka Y, Kural MH, Parent M, Kang Y-J, et al. Engineering of human brain organoids with a functional vascular-like system. *Nature Methods* 2019;16:1169–75.
133. See note 2, Sun 2021, at 11.
134. Velasco V, Shariati SA, Esfandyarpour R. Microtechnology-based methods for organoid models. *Microsystems & Nanoengineering* 2020;6:1–13.
135. Sharf T, van der Molen T, Glasauer SMK, Guzman E, Buccino AP, Luna G, et al. Functional neuronal circuitry and oscillatory dynamics in human brain organoids. *Nature Communications* 2022;13:1–20.
136. Colombi I, Nieuws T, Massimini M, Chiappalone M. Spontaneous and perturbational complexity in cortical cultures. *Brain Sciences* 2021;11:1453.
137. Mansour AA, Gonçalves JT, Bloyd CW, Li H, Fernandes S, Quang D, et al. An *in vivo* model of functional and vascularized human brain organoids. *Nature Biotechnology* 2018;36:432–41.
138. Chen IH, Wolf JA, Blue R, Song MM, Moreno JD, Ming G-L, et al. Transplantation of human brain organoids: Revisiting the science and ethics of brain chimeras. *Cell Stem Cell* 2019;25:462–72.
139. Lavazza A, Reichlin M. Human brain organoids: Why there can be moral concerns if they grow up in the lab and are transplanted or destroyed. *Cambridge Quarterly of Healthcare Ethics* 2023;32:582–596.
140. Shevlin H. Non-human consciousness and the specificity problem: A modest theoretical proposal. *Mind & Language* 2021;36:297–314.
141. Carruthers P. Comparative psychology without consciousness. *Consciousness and Cognition* 2018;63:47–60.
142. Revah O, et al. Maturation and circuit integration of transplanted human cortical organoids. *Nature* 2022;610:319–26.
143. See note 131, Qian et al. 2019, at 2.
144. See note 131, Qian et al. 2019, at 3.
145. Gabriel E, Albanna W, Pasquini G, Ramani A, Josipovic N, Mariappan A, et al. Human brain organoids assemble functionally integrated bilateral optic vesicles. *Cell Stem Cell* 2021;28:1740–57.
146. See note 37, Lavazza, Massimini 2018.

147. Dehaene S, Changeux JP, Naccache L. The global neuronal workspace model of conscious access: From neuronal architectures to clinical applications. In: Dehaene S, Christen Y, eds. *Characterizing Consciousness: From Cognition to the Clinic?* London: Springer Nature; 2011:55–84.
148. See [note 21](#), Gosseries et al. 2014.
149. Owen M. *Measuring the Immeasurable Mind: Where Contemporary Neuroscience Meets the Aristotelian Tradition*. London: Lexington Books; 2021, at chap. 8.