



The Case Against Organoid Consciousness

James Croxford · Tim Bayne

Received: 1 May 2023 / Accepted: 18 February 2024 / Published online: 14 March 2024
© The Author(s) 2024

Abstract Neural organoids are laboratory-generated entities that replicate certain structural and functional features of the human brain. Most neural organoids are disembodied—completely decoupled from sensory input and motor output. As such, questions about their potential capacity for consciousness are exceptionally difficult to answer. While not disputing the need for caution regarding certain neural organoid types, this paper appeals to two broad constraints on any adequate theory of consciousness—the first involving the dependence of consciousness on embodiment; the second involving the dependence of consciousness on representations—to argue that disembodied neural organoids are not plausible candidates for consciousness.

Keywords Neural organoids · Consciousness · Precautionary principle · Embodiment · Representation

Introduction

Derived from stem cells, organoids are laboratory-generated structures that display various features of

particular organs. Although a wide variety of organoids have been created, it is human neural organoids that are of most interest to bioethicists and those who study consciousness. Various described as ‘mini-brains’ [1–4], ‘lab-grown brains’ [5], or ‘brains in a dish’ [6, 7], neural organoids exhibit significant structural and functional commonalities with the developing human brain. In particular, neural organoids have been found to recapitulate neurogenesis and the formation of rudimentary cytoarchitecture [8, 9], exhibit synchronised spiking activity in organoid-derived neural populations [10], perhaps most strikingly, produce complex oscillatory patterns that resemble those found in developing, pre-term brains [11].

The commonalities between neural organoids and the developing brain are robust enough to raise questions about the potential for organoid consciousness [5, 12–20]. As Andrea Lavazza [21] has put it, “The fact that *in vitro* brain organoids can recreate the cytoarchitectonic development typical of a human brain and manifest a complex functional activity, including coordinated electrical activity, suggests that human cerebral organoids, if further developed for medical research and patient care ... may also develop some form of rudimentary consciousness.”

The question of organoid consciousness is important for a number of reasons. For one thing, knowing whether (certain types of) neural organoids are conscious would provide a useful data point for theories of consciousness. In the same way that a complete theory

J. Croxford · T. Bayne (✉)
School of Philosophical, Historical and International
Studies (SOPHIS), Monash University, Melbourne,
Australia VIC Monash University, Melbourne, VIC,
Australia
e-mail: timothy.bayne@monash.edu

of consciousness should explain why consciousness emerges in human development when it does and why it is distributed across the animal kingdom in the ways that it is, so too it ought to explain why certain types of organoids do—or, as the case may be, do not—possess the capacity for consciousness. Identifying organoid types with the capacity for consciousness would also provide us with a novel range of model systems on which to intervene and manipulate.

However, it's the perceived ethical consequences that have done most to drive interest in the question of organoid consciousness. The capacity for consciousness is widely taken to have normative significance, bestowing on entities that possess it a basic kind of moral status [22–24]. Although ascribing consciousness to organoids wouldn't necessarily generate moral or legal prohibitions on their production or use in research—after all, we already allow other entities that are widely regarded as conscious (such as rats) to be used for scientific research—it would introduce into debates about organoids a range of considerations that would not be appropriate absent appeals to consciousness. With this in mind, Niikawa et al. [25] have argued that we should adopt a precautionary approach to the creation and treatment of neural organoids, proceeding (pro tem) on the assumption that current neural organoids are conscious and can undergo a range of experiences. A number of other authors have suggested that although current neural organoids are not likely to be plausible candidates for consciousness, near-future advancements in organoid size, maturity, and complexity would motivate the adoption of a precautionary attitude [12, 23, 26]. Evaluating such claims is clearly both an urgent and important matter.

What is required for the justification of a precautionary approach towards some class of entities? We certainly don't require evidence beyond reasonable doubt. After all, precautionary approaches are widely accepted in a number of other domains in which questions about the presence of consciousness are highly contested, such as those that involve infants, patients suffering from severe brain damage, and various species of non-human animals [27]. At the same time, the mere fact that we can't rule out the possibility that a certain class of entities might enjoy the capacity for consciousness is insufficient to motivate a precautionary approach to their creation or treatment. If it were, the scope of the precautionary

approach would need to be extended far beyond what's plausible, for there are accounts of consciousness which take seriously the possibility that plants, large language models and even expander graphs might be conscious. Instead, what's needed is 'reasonable suspicion' that the relevant system has the capacity for consciousness. We take a precautionary approach to be appropriate only if the ascription of consciousness to the members of the relevant class is a live possibility given assumptions that command broad assent from within the community of experts. This condition is met by neonates, certain types of severely brain-damaged patients, and a wide variety of non-human animals; thus, a precautionary approach to their treatment is appropriate. The question here is whether this condition is also met by neural organoids.

The paper proceeds as follows. We begin with a more detailed look at the variety of neural organoids, focusing on disembodied neural organoids (DNOs) as our central object of investigation. We then consider and reject 'theory-first' approaches to the question of consciousness in DNOs before developing two (relatively a-theoretical) arguments against the possibility of consciousness in DNOs. The first argument appeals to the importance of embodiment and environmental embeddedness for consciousness; the second focuses on the role that representations play in underpinning consciousness.

The Varieties of Neural Organoids

Crucial to any discussion of consciousness in organoids is a recognition of the diversity of organoid types. Following Paşca et al. [28], we distinguish between regionalised neural organoids, unguided neural organoids, and assembloids.

A regionalised neural organoid is a lab-grown neural structure that recapitulates only a determinate brain-like region. To achieve this, specific signalling cues are externally introduced to the neural stem cells to guide differentiation towards particular cell fates. Since optic-cup organoids were first created in 2011 [29], regionalised organoids have been produced that approximate rudimentary features for the pituitary gland [30], forebrain [31], hippocampus [32], cerebellum [33], ganglionic eminence [34], midbrain and hypothalamus [35], thalamus [36], and the choroid

plexus [37]. The main advantage of regionalised neural organoids is their relative replicability, which is superior to that displayed by neural organoids formed by unguided protocols. However, the very features that contribute towards their replicability also prevent them from sharing the complexity of their *in vivo* counterparts, making them less-than-comprehensive models of the brain structures that they are targeting [1, 2]-b).

An unguided neural organoid is a multi-region neural structure that is promoted to grow as a whole self-organising unit [8, 9, 38]. In contrast to the guided protocols of regionalised neural organoids, unguided organoids are governed by undirected protocols, in which neural stem cells are provided with a culturing medium that encourages them to direct their own differentiation. Sometimes referred to as ‘whole-brain organoids’, unguided organoids more closely recapitulate the complexity of the developing *in vivo*-like brain than regionalised organoids do. For example, unguided organoids develop internal divisions that are akin to those found between distinct brain regions. At the same time, unguided organoids develop idiosyncratically. For instance: cell types appear in random locations and proportions; the same type of region develops more than once; and certain regions or cells may be completely missing [39, 40]. The heterogeneity of unguided organoids undermines their utility in many research contexts, which is why many labs focus on regionalised organoids.

Neural assembloids (hereafter: “assembloids”) are created by combining separately generated regionalised organoids with each other or with various kinds of specialised cell types [41, 42]. For example, a dorsal forebrain-like organoid can be fused with a separately generated ventral forebrain-like organoid to create a more complete forebrain-like assembloid [34]. Assembloid protocols offer almost unlimited potential for variation and augmentation, allowing neural organoids to be integrated with non-neural organoids—such as spinal organoids [43], vascular organoids [44], and retinal organoids [45]—thus equipping neural organoids with sensorimotor capacities. The potential of assembloid protocols extends to the possibility of integrating neural organoids with robotic or computer interfaces, thus generating cyborg assembloids of various kinds.

Although questions about the possibility of consciousness in organoid-involving systems with

sensorimotor capacities are important, the issues that they raise are fundamentally different from those raised by the possibility of consciousness in regionalised and unguided neural organoids, and we will set them to one side in order to focus on the possibility of consciousness in *disembodied neural organoids* (DNOs). Consciousness in a DNO would be utterly unlike any kind of consciousness with which we are currently familiar. It would constitute an ‘island of awareness’—a stream of consciousness “whose contents are not shaped by sensory input from either the external world or the body and which cannot be expressed via motor output” [46]. Are islands of awareness possible? Indeed, could we ever have grounds for thinking that an organoid supported an island of awareness?

How should we approach the question of DNO consciousness?

Although the question of organoid consciousness is in some sense as new as organoids themselves, it is a version of a very old problem: the problem of other minds: How can we determine which kinds of systems have an experiential point of view and which kinds of systems don’t? Ascriptions of consciousness to other human beings are typically based on their intentional behaviour—especially (but not exclusively) their verbal behaviour. We rely on people’s actions to figure out not just *what* they are conscious of but *whether* they are conscious at all. However, appeals to intentional behaviour are of little help when it comes to the question of DNO consciousness, for DNOs lack the capacity for behavioural output. Given the DNO predicament, how might we figure out if certain types of DNOs have the capacity for consciousness?

One response to this question would be to take a theory-first approach. According to Giulio Tononi and Christof Koch, in order to address questions about the distribution of consciousness “we need not only more data but also a theory of consciousness—one that says what experience is and what type of physical systems can have it” [47]: 1). Peter Carruthers makes a similar claim in the context of debates about animal consciousness: “one needs to know what consciousness is before one can ask about its distribution across the animal kingdom” [48].

It might be argued that the theory-first approach puts the cart before the horse, and that we need to first identify the distribution of consciousness before we can develop an adequate theory of it.¹ Applied to DNOs, the worry here is that unless we already know whether DNOs are conscious, we won't be able to develop a comprehensive theory of consciousness—or at least, a theory that applies not just to human beings but to DNOs. Instead of starting with a theory of consciousness, we need to settle the question of DNO consciousness on independent grounds and then use that answer to constrain our account of consciousness. Call this the 'cart/horse objection'.

In response to the cart/horse objection, one might argue that it is possible to justify a theory of consciousness without making assumptions about the distribution of consciousness. The account of consciousness defended by Tononi and Koch—the Integrated Information Theory (IIT) [49]—adopts precisely that approach. Tononi and Koch argue for IIT on the basis of what they call 'phenomenological axioms'—claims about the structure of experience that are introspectively accessible and self-evidently true of all possible forms of consciousness. As one of us has previously argued [50], this approach is unlikely to succeed, for the alleged examples of phenomenological axioms to which IIT appeals are not plausibly treated as axiomatic. More generally, it seems unlikely that an account of consciousness could be justified along axiomatic lines, for that's simply not how accounts of biological or psychological phenomena are justified.

In our view, the cart/horse objection goes wrong not because it is possible to develop a comprehensive theory of consciousness independently of claims about the distribution of consciousness, but because it fails to recognise that developing a theory of consciousness and addressing the distribution question go hand-in-glove: our account of how consciousness is distributed ought to be informed by our best account of its nature, and our best account of the nature of consciousness ought, in turn, to be informed by our best account of its distribution. There is a circularity

here, but it is virtuous rather than vicious. In Hasok Chang's [51] useful phrase, what's required here is a process of 'epistemic iteration'. We begin with a tentative account of the distribution of consciousness ('these entities are conscious', 'these entities are not conscious') and revise that account as we gain a better understanding of the nature of consciousness (and thus of its markers or indicators). Revising our conception of the distribution of consciousness leads, in turn, to revising our account of its nature, and so on [52].

A further problem with the theory-first approach is that no single theory—or even theoretical framework—commands general assent. A recent review identified 22 neurobiological theories of consciousness [53, 54]. Although some of these theories have significant overlap and might have the same implications with respect to the distribution of consciousness, other theories have distinct implications for which kinds of organoids (if any) would have the capacity for consciousness. The embarrassment of theoretical riches that we face here might not be so problematic if the trend in consciousness science was towards integration, cohesion and convergence. Unfortunately, that seems not to be the case; indeed, if anything, the field is witnessing a proliferation in theories of consciousness [55].

Of course, there is nothing to prevent individual researchers from appealing to their favoured theory of consciousness in order to motivate an account of DNO consciousness (see [16, 20], but because no theory of consciousness commands widespread endorsement, appeals to theory won't provide the kind of well-grounded suspicion that is needed to trigger the application of the precautionary principle.

Another approach to figuring out whether DNOs fall within the distribution of consciousness focuses on the (putative) non-behavioural markers of consciousness that have been developed in recent years, and which have been applied to questions of consciousness in infants, brain-damaged individuals, and non-human animals. Unfortunately, few of these markers can be applied to DNOs. For example, the capacity to produce visual imagery [56] or allocate attention [57] on command has been employed as a marker of consciousness in brain-damaged individuals, but DNOs lack the capacity to comprehend commands. Similarly,

¹ As Michael Tye puts it, "a condition of adequacy on any theory of the nature of experience is that it cover all and only creatures that have experiences. But if we don't already agree on which creatures these are, how are we to construct the theory in the first place?" [118]: 11).

the presence of a reliable ERP response (the so-called ‘P300’ or ‘P3b’ response) to second-order auditory oddballs has been employed as a marker of consciousness in severely brain-damaged patients [58] and infants/foetuses [59, 60], but again, this test cannot be applied to DNOs which (by definition lack the machinery required for processing sensory input.

There are, however, avenues for probing consciousness in DNOs that are potentially more promising. Perhaps the most prominent of these is the perturbational complexity index (PCI), developed by Marcello Massimini and his colleagues [61]. Here, the cortex is directly perturbed with transcranial magnetic stimulation (TMS), and the complexity of the neural response to that perturbation is measured with electroencephalography (EEG) and then mapped onto a perturbational complexity index. Results to date indicate that PCI responses correlate well with pre-theoretical judgments regarding the distribution of consciousness across a wide range of conditions [62, 63]. Because the PCI doesn’t require any kind of perceptual contact with an organism, it could, in principle, be applied to DNOs [20, 46].

That said, employing PCI to address the question of DNO consciousness is far from trivial. For one thing, new methods will need to be developed in order to apply PCI to DNOs (although see [64]. More significantly, it is unclear how the perturbational complexity that DNOs might exhibit ought to be interpreted. In the case of humans, perturbational complexity can be benchmarked by relying on our pretheoretical measures of consciousness (such as the capacity for verbal report and intentional behaviour), but given the wide-ranging and radical differences between humans and DNOs, it is unclear whether we can rely on those same benchmarking studies when it comes to identifying the level of perturbational complexity that might be indicative of consciousness in DNOs. Here, it is important to recognise that it is not just false negatives that should be avoided—we also need to avoid false positives when it comes to the ascription of consciousness. And, to avoid false positives, we would need to know what level of perturbational complexity a system might require in order to be a plausible candidate for the ascription of consciousness. To date, we do not have a well-grounded answer to that question.

DNOs and two constraints on consciousness

Rather than viewing the problem of DNO consciousness through either a theory-based lens or a marker-based lens, we suggest that at present it would be more appropriate to view it through a constraint-based lens. The idea, in other words, is to consider what broadly-accepted constraints on consciousness might suggest with respect to the prospects of consciousness in DNOs.

The remainder of this paper explores two such constraints: the first focuses on embodiment and the role of the environment in enabling consciousness; the second concerns the centrality of representations in grounding consciousness. We argue that both constraints are entailed by a wide range of theories of consciousness, and that each provides significant reason to be skeptical of the idea that DNOs could qualify as members of the ‘consciousness club’.

The embodiment constraint

One of the central debates in the study of consciousness over the last three or so decades has focused on the role played by the body and the wider environment in grounding or explaining consciousness. We can group theorists into two broad categories: internalists and externalists. Roughly speaking, internalists about consciousness downplay the relevance of the body and the wider environment in understanding consciousness and hold that the factors to which a theory of consciousness ought to appeal are purely intracranial (indeed: typically intracortical). Externalists, by contrast, emphasise the importance of the body and the wider environment, holding that consciousness is constitutively related to factors that are ‘outside the head’. The internalist/externalist debate clearly bears on the question of DNO consciousness, but what’s less clear is the precise nature of that bearing. We begin by identifying certain influential forms of externalism that don’t speak directly to the question of organoid consciousness before turning to a form of externalism—‘externalism-lite’, if you will—that does.

One strand in the externalist literature focuses on the idea that the material basis of consciousness isn’t restricted to the brain but can loop out into the body and the perceptual environment. The advocates of this view deny that the brain (or skin) is a

‘magical membrane’ (in Susan Hurley’s memorable phrase) and hold instead that ‘in principle what explains phenomenal qualities can be distributed within the brain, among brain and body, or among brain, body, and embedding environment, depending on the explanatory dynamics’ [65]: 116, see also [66, 67].

However, the relevance of this version of externalism to the DNO question is, at best, indirect, for in order to put pressure on the idea that DNO consciousness is a live possibility, we need reason to think not just that consciousness *can* extend beyond the brain, but that it *must* do so. No such argument is made by those who defend this form of phenomenal externalism.

A very different form of externalism focuses on the idea that consciousness involves a relation between an organism and its environment. Consider the following passage from Thompson and Cosmelli ([68]: 176):

“...if perceptual consciousness is a certain kind of interactive relationship between an organism and its environment, then a disembodied brain going through the same sequence of internal states as an embodied brain is like a disembodied stomach going through the same sequence of internal states as an embodied one. The disembodied stomach isn’t digesting and the disembodied brain isn’t experiencing, because the necessary contexts of the body and the environment are missing.”

This form of externalism is directly relevant to the organoid question. If true, it would appear to undermine the very possibility of consciousness in DNOs, for DNOs have no interactive relationship with their environments.

The question, of course, is whether it *is* true. It is one thing to grant that perception itself is relational and that one cannot perceive an object without being suitably related to it. It is, however, quite another thing to grant that perceptual experience—understood to include not only perception proper but also hallucination and illusion—is relational. Relational accounts of perceptual experience certainly have their advocates, but they are very far from being the only games in town (see e.g. [69, 70]). To take just one influential alternative, many hold that perceptual experience is a kind of representational state, in which the organism represents the world as being thus-and-so (e.g.

[71–73]). If representationalism is true, then perceptual experience isn’t much like digestion.

Moreover, even if *perceptual* consciousness is best understood as a relation between an organism and its environment, there are other forms of consciousness to consider, such as sensory imagery, conscious thought, and dream experience. Although relational accounts of these phenomena are sometimes offered, they lack even the *prima facie* plausibility that characterises relational accounts of perceptual experience. Thus, even if relational considerations were able to show that organoids aren’t perceptually conscious, they would leave open the possibility that DNOs might enjoy non-perceptual forms of consciousness.

So, neither of the two versions of externalism that we’ve considered thus far appropriately engage with the question of DNO consciousness in a satisfactory manner: the former leaves the question of DNO consciousness unanswered; the latter rules out the possibility of DNO consciousness, but only by appealing to a tendentious theory of perceptual experience. What’s needed, then, is a thesis which speaks directly to the question of organoid consciousness, but does so in a way that is appropriately non-partisan and would—or at least *should*—be acceptable to the vast majority of theorists.

The following constraint, we suggest, fits the bill:

Embodiment Constraint (EC): Only a brain with a history of embodiment and sensorimotor interaction with the world has a genuine chance of supporting consciousness.

EC speaks directly to the question of DNO consciousness, but is it also acceptable to the vast majority of theorists? It should certainly be embraced by externalists of all stripes, but we suggest that it should also be embraced by many versions of internalism as well. In fact, although EC is a kind of externalism, one might think of it as ‘externalism-lite’, for it is consistent with many internalist approaches to consciousness.

To see why, consider what is arguably the classical form of internalism: the mind-brain identity theory [74–76]. First advanced in the late 1950s (e.g. [77–79]), identity theorists draw inspiration from familiar *a posteriori* identities in the special sciences to argue that conscious states are identical to brain states. For example, an identity theorist might identify pain with particular patterns of activation in

the so-called ‘pain matrix’, a network involving the primary and secondary somatosensory cortices, the insula, and the anterior cingulate cortex that is typically active when people experience transient pain. On the face of things, the identity theory might seem to suggest that DNOs ought to be plausible candidates for consciousness. After all (one might think), if conscious states are brain states, and if (as seems plausible) DNOs have brain states, then they could surely have conscious states too.

The problem with this line of thought is that it avoids the key question of whether DNOs have the right *sort* of brain states. Any remotely plausible version of the identity theory should accept that many (perhaps most?) brain states don’t support mental states of any kind, let alone conscious states. The kinds of brain states seen in the context of epileptic seizures aren’t generally regarded as consciousness-supporting, nor are those which characterise dreamless sleep, deep sedation, or certain forms of severe brain damage, such as the vegetative state (also known as the ‘unresponsive wakefulness syndrome’). But given that only some brain states are consciousness-supporting, we need a reason for thinking that DNOs could have brain states of the right kinds before thinking that they might be viable candidates for consciousness.

And here, we would argue, the identity theorist faces a difficult task, for any account of which brain states are consciousness-supporting and which aren’t must begin with the brains of the only systems that we know to be conscious—that is, organisms, with a history of sensorimotor interaction with their environment. (To the best of our knowledge, no one has been able to procure any kind of behaviour that might be indicative of consciousness from a disembodied brain.) This doesn’t entail that EC is false, but it does shift the burden of proof onto those who reject EC. An analogy with reading might be useful. Although it is conceivable (and perhaps even nomologically possible) that a creature who has no history of interacting with text could acquire the ability to read, as a matter of fact, the ability to read is found only in organisms with an extended history of interacting with words (or proto-words). Similarly, although it is conceivable (and perhaps even nomologically possible) that a disembodied brain might support consciousness, we suggest that, as a matter of fact, the capacity for consciousness is likely to arise only in the context of

embodiment and sensorimotor interaction. Organisms that have lost the capacity for sensorimotor engagement with the world might continue to support an ‘island of awareness’ [46] for a period of time, but—we suggest—without feedback from the body and wider perceptual environment the neural organisation required for consciousness would be extremely unlikely to occur.

The general point here is that the brain and the body develop together as parts of an interdependent, life-sustaining system. With the notable exception of organoids, brains are always components of organisms. Although we don’t know at what stage (if at all) the human fetus acquires the capacity for consciousness [80, 81], we do know that the fetal brain is shaped by the body and its uterine environment in fundamental ways. Cellular and regional specialisation occurs in concert with, and response to, connections and feedback from developing organs, and sensory and motor capacities [82, 83]. The genetically determined neural connectivity that occurs in embryonic development requires pruning and strengthening, and those processes occur as the brain engages in bodily, sensory, and motor interaction [84]. It is well-documented that deficiencies in sensorimotor stimulation lead to developmental deficits [85]. This point can be illustrated with reference to cataracts. Patients who have congenital binocular cataracts removed after the age of 10 experience permanent deficits in visual acuity and have difficulties perceiving shape and form. However, when cataracts develop in adulthood and are removed decades after they form, normal vision returns immediately [84, 86].

The development of a DNO, of course, is not shaped by its environment in the ways in which the development of an ordinary human brain is. The trajectory of its structure is not informed by the demands of regulating a body or by a history of sensorimotor exploration.² The thoroughgoing nature of this isolation suggests not merely that DNOs are unlikely to share the kinds of conscious states that we have, but

² This is most obviously true of unguided organoids, but we would argue that it is also true of regionalised organoids whose development is ‘guided’, for the guidance in question doesn’t provide anything like the kind of fine-grained sculpting that is provided by embodiment and sensorimotor engagement with the world.

that they are unlikely to acquire the capacity for consciousness *at all*.

Although we've developed the case for EC with reference to the identity theory, the point generalises beyond the identity theory to many other versions of internalism. It applies to those versions of functionalism that identify conscious states with 'short-arm' (that is, internally-individuated) functional roles; and to those versions of internalism which refuse to identify conscious states with internal states, but hold only that conscious states are constituted by or supervenient on internal states of some kind.³

In response to EC, a critic might point to dreaming, sedation or hallucination as evidence for the claim that neither embodiment nor environmental engagement is required for consciousness. But this line of objection is far from compelling. It is certainly true that the brain can continue to support consciousness even when largely (and perhaps wholly) insulated from input from the body or wider environment. But (to the best of our knowledge), the brains that have such capacities are brains that have also had a long history of embodiment and environmental interaction. The fact that ordinary (that is, embodied and environmentally embedded) brains might have the capacity to generate experiences that are endogenously triggered doesn't provide any reason to think that extraordinary "brains" (that is, brains that are neither embodied nor environmentally embedded) might also have that capacity.

Similar points hold with respect to the claim that EC might be false because direct stimulation of cortical areas can elicit experiences of various kinds. Activity in the fusiform face area (FFA) of an ordinary human being may be "sufficient" for a visual experience of faces [87–89], but as far as we know, FFA activity underpins visual experiences of faces only when suitably integrated with a wide range of cortical, sub-cortical and perhaps even bodily processes. We certainly have no evidence that activating a section of FFA that had been excised from a brain

and placed in a petri dish would generate any kind of experience, let alone an experience of faces.

What, then, of the fact that organoids have been observed to produce neural activity reminiscent of the *trace discontinu* seen in human preterm infants—regular periods of inactivity punctuated by synchronised, high-amplitude oscillations [11]? Is this not evidence of consciousness in the relevant class of organoids?

It is certainly a striking finding, but we would argue that it provides little evidence in favour of organoid consciousness. First, we lack reason to think that the *trace discontinu* is directly implicated in consciousness, as opposed to a more general kind of neural activity that is the precondition for cognitive development. Second, it's important to recognise that the organoids employed in the Trujillo study were cortical spheroids, a kind of highly simplified, regionalised neural organoid. Had such findings been obtained in organoids with higher degrees of structural complexity then it might be more reasonable to treat them as providing evidence of consciousness, but in the absence of such complexity it's unclear what the role of high-amplitude oscillations might be.

The more pressing question raised by the Trujillo study is whether this *kind* of finding could, in principle, provide evidence of consciousness in DNOs.⁴ We allow that it could—after all, it is difficult to say with any certainty what the study of consciousness might reveal. But we can grant that the science of consciousness could, in principle, show that DNOs have neurofunctional properties that suffice for consciousness without granting that we currently have grounds to take seriously the possibility of DNO consciousness.

We conclude this section by drawing attention to an assumption that we have implicitly granted—namely, that DNOs are brains, and that the kinds of internal states they have are legitimately treated as brain states. This assumption should not be uncritically accepted. Consider the once-influential suggestion that visual experience involves oscillations of around 40Hz in the visual cortex [90]. Clearly, only a system with a visual cortex is capable of instantiating that property. Could a DNO have a visual cortex? That is far from clear. A DNO could, of course, contain a part that is visual-cortex-like, but to be

³ We do, however, acknowledge that there are versions of internalism that would reject (or are at least not committed to) EC. As far as we can see, substance dualists would have no reasons to embrace EC, nor would advocates of the Integrated Information Theory (IIT) of consciousness. We thank a referee for raising this issue.

⁴ We are grateful to a referee for raising this point.

visual-cortex-like is not itself to be a visual cortex. Arguably, only that which has developed as part of an embodied and environmentally embedded brain could be a genuine visual cortex.

Let's take stock. Although it is evident that the brain is intimately involved in consciousness, and that some parts of the brain are more intimately involved than others, there are good reasons to hold that only neural states that have had to accommodate themselves to the task of regulating a body or tracking the contingencies of its environment have the capacity to support consciousness. By definition, DNOs face neither of these demands, and thus, there are good grounds to deny that they might be conscious. Importantly, this conclusion goes through even if we assume that consciousness supervenes solely on what's 'in the head'.

The representationalist constraint

Although there is much disagreement between contemporary theories of consciousness, a surprisingly wide range of theories have a shared commitment to what we will call a 'representationalist' conception of consciousness. Although representationalism in the study of consciousness is typically understood as the thesis that conscious mental states have no conscious properties other than those that are fixed by their representational properties [91, 92], that is not how we are using that term here. Instead, we use 'representationalism' to refer to the claim that conscious states (experiences) are representations, and (thus) that consciousness is restricted to representational systems.⁵ This notion of representationalism is weaker than the standard notion of representationalism, for one could (and indeed many do) hold that conscious states are representations but deny that their conscious properties are fixed by their representational properties.

The widespread endorsement of representationalism (in our sense of the term) is partly obscured by the fact many of the most high-profile debates concern the question of how representations are implicated in consciousness, and the participants in these disputes have shared (albeit rarely articulated)

commitment to the idea that consciousness is fundamentally representational. The main divide here is between first-order representationalists, who hold that conscious states are representations of a certain kind, and higher-order representationalists, who hold that a mental state is conscious in virtue of being represented (in a certain kind of way). Some versions of first-order representationalism focus on the ways in which representations are available to guide cognition and behaviour (e.g., Dretske [93], Tye [94], Baars [95] and Dehaene & Changeux [96]). Others focus on the ways in which representations are integrated with each other. For example, Lamme's [97] local recurrency theory holds that visual experiences are generated when feature-specific representations are bound together to form representations of integrated visual objects by recurrent processing within perceptual cortices, while Merker's [98] account appeals to the role of mid-brain structures in supporting the integration of representations that capture both an organism's physiological state and motivations with representation of its perceptual environment. Another version of first-order representationalism is the Attended Intermediate Representation (AIR) theory of consciousness. First advanced by Jackendoff [99] and defended at length by Prinz [100], the AIR theory holds that consciousness arises when intermediate-level representations are modulated by attention. Again, the view treats representations as crucial to consciousness.⁶

Higher-order representationalism also comes in a variety of forms. Some higher-order representationalists hold that consciousness arises when mental states are monitored by a quasi-perceptual process [101,

⁵ Strictly speaking, all that this objection really requires is the claim that conscious systems are representational systems, rather than the claim that conscious states are representations.

⁶ The representational nature of some of these accounts is not always fully appreciated. For example, Birch and Browning [12] suggest that Merker's mid-brain account of consciousness would give rise to the reasonable suspicion that a functional midbrain DNO is conscious, since Merker's theory (they suggest) posits a functioning midbrain as sufficient for consciousness. In our view, this is a misreading of Merker's position, and treats his view as a kind of identity theory rather than as a form of representationalism. On our reading, Merker holds that midbrain processing is sufficient for consciousness only insofar as it integrates interoceptive and exteroceptive information into a unified model. Disembodied 'mid-brain structures' (if indeed they deserve that label) might support neural activity, but—we suggest—that activity would not realise the kinds of representations that Merker's account takes to be required for consciousness.

102], others argue that it arises when mental states are monitored by a thought-like process [103], and still others argue that consciousness involves information-processing forms of meta-representation that are arguably neither perception-like nor thought-like [104, 105]. Orthogonal to the debate about the nature of the monitoring process required for consciousness is a debate about how the monitoring state and the monitored state can (or must) be related. Although most higher-order representationalism requires that these states are distinct, self-representationalist versions of the view hold that they must be identical, and that consciousness arises when a mental state takes itself as its own intentional object [106–108].

Although first-order and higher-order representationalists disagree about the kinds of representational structure that are required for consciousness, they are all implicitly committed to the following:

Representationalist Constraint (RC): Only representational systems are candidates for being conscious.

We suggest that RC grounds a powerful argument for denying consciousness to DNOs, for it is implausible to think that DNOs could qualify as representational systems.

There are two ways of arguing that DNOs are unlikely to qualify as representational systems. One line of argument focuses on the conditions required for representations to be contentful. According to content externalists, the contents of a representation are fixed by a certain subset of an organism's relations to its environment (e.g. [93, 109, 110]). Given the extremely coarse-grained nature of the relations that obtain between a DNO and its environment, content externalism would seem to imply that whatever representations a DNO might have, they could not have any content. But surely contentless representations—if indeed such states are possible—could not support consciousness by the lights of any version of representationalism.

While we are sympathetic to the line of argument just sketched, it won't convince those who harbour doubts about the tenability of content externalism. A more powerful line of argument, we suggest, focuses not on what fixes a representation's content but on what makes it the case that a particular feature of the physical world qualifies as a representation in the first place.

The key point here is that physical states qualify as representational only insofar as they (are poised to) play a representational role. The point is perhaps most evident with respect to public representations. An arrangement of seaweed on the beach might spell out the word “dragon”, but it won't be an instance of that word unless it is incorporated into a system that treats it as an instance of that word. A pattern of sounds produced by the wind knocking one branch against another might share the acoustic properties of a particular Morse code message, but those sounds won't be representational in the way that the message in Morse code is unless they are incorporated into a representational system. In effect, something qualifies as a representation only if there is something that treats it as a representation. Representations need not only to be produced, they must also be consumed.

Although we know of no reason to rule out the possibility that DNOs could acquire the capacity to consume representations, we think it exceedingly unlikely that current generation DNOs have such capacities. Further, we doubt that future generations of DNOs could acquire such capacities in the absence of the kind of input from the body or the environment that, by definition, DNOs lack. Representationalists should be more concerned about the possibility of consciousness in systems that are clearly representational—such as laptops or self-driving cars—than they are about the possibility of consciousness in DNOs.

In response, it might be argued that DNOs have representational capacities not in virtue of their direct relational features (as it were), but in virtue of the fact that homologous structures are representational when instantiated in us (or other species). Consider the kinds of somatotopic maps found in the primary somatosensory and motor cortices. Here, one might argue that neural activity in the ‘primary somatosensory and motor cortices’ of a DNO is representational in virtue of the fact that activity in homologous structures of an ordinary (embodied, embedded) is representational. (Activity in the somatotopic map of a DNO might, of course, be misrepresentational, but there is nothing in RC which requires that the kinds of representations that underwrite consciousness are veridical.) The idea, in other words, is that the representational properties of DNOs are inherited from the

representational properties of the systems from which they derive—that is, human beings.⁷

This proposal is an intriguing one, but we don't find it convincing. For one thing, we doubt that a DNO might develop neural structures that are truly isomorphic to somatotopic maps. While it is true that somatotopic “protomaps” are initially directed by genetic mechanisms, these protomaps require connection and functional activity from the body to appropriately develop and form mature functional networks [111]. Moreover, given the complete absence of somatic stimulation seen in the case of DNOs, we would expect synaptic and axonal pruning to disrupt any prepotent drive there might be towards the appropriate maturation of somatotopic maps. More fundamentally, the proposal doesn't engage with the requirement that genuine representations must be usable as such, but merely side-steps that constraint. A model of the human primary somatosensory and motor cortices produced by a 3D printer would share the structure of a somatotopic map, but it wouldn't support representations of any kind—let alone representations of body parts—unless it was used in an appropriate manner. We conclude that those sympathetic to representationalist conceptions of consciousness have good reasons for scepticism about the prospects of DNO consciousness.

There are, of course, non-representationalist conceptions of consciousness. We have already mentioned one such view: the mind-brain identity theory. Other non-representational approaches to consciousness include the Integrated Information Theory [49], quantum accounts [112, 113], and various forms of dualism [114] and Russellian Monism [115, 116]. As best we can tell, each of these approaches leaves open the possibility that certain kinds of DNOs might be conscious. Indeed, Christof Koch—one of IIT's leading advocates—has suggested that certain DNOs would be likely to have a measure of informational complexity (or ‘phi’) that is indicative of consciousness [117].

However, the implications of this for precautionary approaches to consciousness are far from straightforward, even from the perspective of these approaches to consciousness. Crucially, these approaches also leave open the possibility that consciousness might

characterise all manner of systems. For example, IIT is explicitly committed to the possibility of consciousness in thermostats and expander graphs [47], while Russellian Monists ascribe consciousness—or ‘proto-consciousness’, as they sometimes put it—to the fundamental components of the physical world. If the proponents of these accounts are prepared to adopt a precautionary approach to DNOs, then—we would argue—they should also adopt a precautionary approach to a wide range of other entities, such as thermostats, expander graphs and quarks.

At this point, we are clearly a long way from the domains in which precautionary approaches are most compelling, such as those involving the treatment of non-human animals, human infants, and individuals who have suffered from severe brain damage. The mere fact that serious theories of consciousness allow for the possibility of DNO consciousness isn't enough to motivate the adoption of a precautionary approach to them, for if it were, then we would have to take a precautionary approach to a bewildering array of non-living entities. One person's *modus ponens* is, of course, another's *modus tollens*, but in this case the *tollens* strikes us as singularly unattractive.

Conclusion

The main aim of this paper has been to pour some cold water on speculations regarding consciousness in DNOs. Although we don't claim to have ruled out the possibility of DNO consciousness, we do think that the considerations outlined here provide strong reasons for treating such suggestions with a great deal of suspicion. Not only is there little reason to ascribe consciousness to the current generations of DNOs, there are also grounds for thinking that DNOs *as such* are not viable candidates for consciousness.

It is important to recognise that these conclusions apply only to *disembodied* neural organoids, and not to other types of organoid-involving systems, such as animals into which organoids have been implanted or cyborg assembloids that integrate neural organoids with some form of robotic or computer interface. Systems of these kinds need not flout either the embodiment or representationalist constraints, and there may be a strong case for adopting a precautionary approach to their creation and use. (Of course, determining whether such systems actually possess the

⁷ We are indebted to a referee for this point.

capacity for consciousness would be a further question.) The development of neural organoids may well lead to the creation of the first synthetically conscious systems, but the systems in question are unlikely to be islands of awareness.

Acknowledgements T.B. acknowledges support from the Canadian Institute for Advanced Research (CIFAR) and the Australian Research Council (FT 150100266). We are grateful to the organizers of a workshop on detecting unusual forms of consciousness held at the University of Bonn (September 2023) for the opportunity to present this paper, and to participants at that workshop (in particular Jonathan Birch, Charlotte Gauvry, and Elizabeth Schechter) for their feedback on it. We are also very grateful to two referees for this journal for their very helpful comments on earlier versions of this paper.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

Conflicts of Interest The authors have no conflicts of interest to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Kelava, I., and M.A. Lancaster. 2016a. Dishing out mini-brains: Current progress and future prospects in brain organoid research. *Developmental Biology* 420 (2): 199–209.
- Kelava, I., and M.A. Lancaster. 2016b. Stem cell models of human brain development. *Cell Stem Cell* 18 (6): 736–748.
- Chapman, A.R. 2019. Brain models in a dish: Ethical issues in developing brain organoids. *AJOB Neuroscience* 10 (3): 113–115.
- Lavazza, A., and M. Massimini. 2018. Cerebral organoids: Ethical issues and consciousness assessment. *Journal of Medical Ethics* 44: 606–610.
- Reardon, S. 2020. Can lab-grown brains become conscious? *Nature* 586: 658–661.
- DiSalvo, D. 2019. The real story about organoids: What you should know about “brains in a dish”. *Forbes*.
- Lavazza, A. 2021. Potential ethical problems with human cerebral organoids: Consciousness and moral status of future brains in a dish. *Brain Research*, 1750. <https://doi.org/10.1016/j.brainres.2020.147146>.
- Lancaster, M.A., et al. 2013. Cerebral organoids model human brain development and microcephaly. *Nature* 501: 373–379.
- Lancaster, M.A., et al. 2017. Guided self-organization and cortical plate formation in human brain organoids. *Nature Biotechnology* 35 (7): 659–666.
- Sakaguchi, H., et al. 2019. Self-organized synchronous calcium transients in a cultured human neural network derived from cerebral organoids. *Stem Cell Reports* 13: 458–473.
- Trujillo, C., et al. 2019. Complex oscillatory waves emerging from cortical organoids model early human brain network development. *Cell Stem Cell* 25 (4): 558–569.e7.
- Birch, J., and H. Browning. 2021. Neural organoids and the precautionary principle. *The American Journal of Bioethics* 21 (1): 56–58.
- Greely, H.T.H. 2021. Human brain surrogates research: The onrushing ethical dilemma. *The American Journal of Bioethics* 21 (1): 34–45.
- Greely, H.T.H., and K.V. Kreitmair. 2021. Should cerebral organoids be used for research if they have the capacity for consciousness? *Cambridge Quarterly Healthcare Ethics* 30 (4): 575–584.
- Koplin, J., O. Carter, and J. Savulescu. 2021. Moral status of brain organoids. In *Rethinking Moral Status*, ed. S. Clarke, H. Zohny, and J. Savulescu, 250–268. Oxford: OUP.
- Owen, M., et al. 2023. Theoretical neurobiology of consciousness applied to human cerebral organoids. *Cambridge Quarterly of Healthcare Ethics*. <https://doi.org/10.1017/S0963180123000543>.
- Sawai, T., et al. 2019. The ethics of cerebral organoid research: Being conscious of consciousness. *Stem Cell Reports* 13: 440–447.
- Sawai, T. et al. 2021. Mapping the ethical issues of brain organoid research and application. *AJOB Neuroscience*: 1–14.
- Shepherd, J. 2018a. Ethical (and epistemological) issues regarding consciousness in cerebral organoids. *Journal of Medical Ethics* 44 (9): 611–612.
- Zilio, F., and A. Lavazza. 2023. Consciousness in a rotor? Science and ethics of potentially conscious human cerebral organoids. *AJOB Neuroscience* 14 (2): 178–196.
- Lavazza, A. 2020. Human cerebral organoids and consciousness: A double-edged sword. *Monash bioethics review* 38 (2): 105–128.
- Clarke, S., H. Zohny, and J. Savulescu, eds. 2021. *Rethinking Moral Status*. Oxford: Oxford University Press.
- Koplin, J., and J. Savulescu. 2019. Moral limits of brain organoid research. *The Journal of Law, Medicine & Ethics* 47: 760–767.

24. Shepherd, J. 2018b. Consciousness and Moral Status. *Routledge*. <https://doi.org/10.4324/9781315396347>.
25. Niikawa, T. et al. 2022. Human brain organoids and consciousness. *Neuroethics*, 15/5: <https://doi.org/10.1007/s12152-022-09483-1>.
26. Birch, J. 2023. When is a brain organoid a sentience candidate? *Molecular Psychology: Brain, Behavior, and Society* 2: 22.
27. Birch, J. 2017. Animal sentience and the precautionary principle. *Animal Sentience*. 2 (16): 1.
28. Paşca, S.P., et al. 2022. A nomenclature consensus for nervous system organoids and assembloids. *Nature* 609: 907–910.
29. Eiraku, M., et al. 2011. Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature* 472 (7341): 51.
30. Suga, H., et al. 2011. Self-formation of functional adenohypophysis in three-dimensional culture. *Nature* 480 (7375): 57.
31. Kadoshima, T., et al. 2013. Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proceedings of the National Academy of Sciences (USA)* 110 (50): 20284.
32. Sakaguchi, H. et al. 2015. Generation of functional hippocampal neurons from self-organizing human embryonic stem cell-derived dorsomedial telencephalic tissue. *Nature Communications* 6 (1). <https://doi.org/10.1038/ncomms9896>.
33. Muguruma, Keiko, et al. 2015. Self-organization of polarized cerebellar tissue in 3D culture of human pluripotent stem cells. *Cell Reports* 10 (4): 537–550.
34. Xiang, Yangfei, et al. 2017. Fusion of regionally specified hPSC-derived organoids models human brain development and interneuron migration. *Cell Stem Cell* 21 (3): 383–398.e7.
35. Qian, X., F. Jacob, and M. Song. 2018. Generation of human brain region-specific organoids using a miniaturized spinning bioreactor. *Nature Protocols* 13 (3): 565–580.
36. Xiang, Y., et al. 2019. hESC-derived thalamic organoids form reciprocal projections when fused with cortical organoids. *Cell Stem Cell* 24 (3): 487–497.e7.
37. Pellegrini, L., et al. 2020. Human CNS barrier-forming organoids with cerebrospinal fluid production. *Science* 369 (6500). <https://doi.org/10.1126/science.aaz5626>.
38. Giandomenico, S.L., et al. 2019. Cerebral organoids at the air-liquid interface generate diverse nerve tracts with functional output. *Nature Neuroscience* 22 (4): 669–679.
39. Bhaduri, A., et al. 2020. Cell stress in cortical organoids impairs molecular subtype specification. *Nature* 578 (7793): 142.
40. Weiler, N. 2020. Not 'Brains in a Dish': Cerebral organoids flunk comparison to developing nervous system. UCSF News. <https://www.ucsf.edu/news/2020/01/416526/not-brains-dish-cerebral-organoids-flunk-comparison-developing-nervous-system>.
41. Marton, R., and S.P. Paşca. 2020. Organoid and assembloid technologies for investigating cellular crosstalk in human brain development and disease. *Trends in Cell Biology* 30 (2): 133–143.
42. Paşca, S.P. 2019. Assembling human brain organoids. *Science* 363 (6423): 126–127.
43. Andersen, J., et al. 2020. Generation of functional human 3D cortico-motor assembloids. *Cell* 183 (7): 1913–1929.e26.
44. Song, L., et al. 2019. Assembly of human stem cell-derived cortical spheroids and vascular spheroids to model 3-D brain-like tissues. *Scientific Reports* 9 (1): 5977–6016.
45. Fligor, C.M., et al. 2021. Extension of retinofugal projections in an assembled model of human pluripotent stem cell-derived organoids. *Stem Cell Reports* 16 (9): 2228–2241.
46. Bayne, T., A. Seth, and M. Massimini. 2020. Are there islands of awareness? *Trends in Neurosciences* 43 (1): 6–16.
47. Tononi, G., and C. Koch. 2015. Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society B* 370: 20140167.
48. Carruthers, P. 2018. Comparative psychology without consciousness. *Consciousness and Cognition*. 63: 47–60.
49. Tononi, G., et al. 2016. Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience* 17 (7): 450–461. <https://doi.org/10.1038/nrn.2016.4>.
50. Bayne, T. 2018. On the axiomatic foundations of the integrated information theory of consciousness. *Neuroscience of Consciousness*, 1/1: niy007. <https://doi.org/10.1093/nc/niy007>.
51. Chang, H. 2008. *Inventing Temperature*. Oxford: Oxford University Press.
52. Bayne, T., et al. In press. Tests for consciousness in humans and beyond. *Trends in Cognitive Sciences*.
53. Francken, J.C., L. Beerendonk, D. Molenaar, et al. 2022. An academic survey on theoretical foundations, common assumptions and the current state of consciousness science. *Neuroscience of Consciousness* 1: 1–13.
54. Seth, A., and T. Bayne. 2022. Theories of consciousness. *Nature Reviews Neuroscience* 23: 439–452. <https://doi.org/10.1038/s41583-022-00587-4>.
55. Yaron, I., et al. 2022. The ConTraSt database for analysing and comparing empirical studies of consciousness theories. *Nature Human Behaviour* 6 (4): 593–604.
56. Owen, A., et al. 2006. Detecting awareness in the vegetative state. *Science* 313 (5792): 1402.
57. Naci, L., and A.M. Owen. 2013. Making every word count for nonresponsive patients. *JAMA Neurology* 70 (10): 1235–1241.
58. Bekinschtein, T., et al. 2009. Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences (USA)* 106 (5): 1672–1677.
59. Moser, J., et al. 2020. Magnetoencephalographic signatures of hierarchical rule learning in newborns. *Developmental Cognitive Neuroscience* 46: 100871.
60. Moser, J., et al. 2021. Magnetoencephalographic signatures of conscious processing before birth. *Developmental Cognitive Neuroscience* 49: 100964.
61. Massimini, M., et al. 2009. A perturbational approach for evaluating the brain's capacity for consciousness. *Progress in Brain Research* 177: 201–214.

62. Casali, A., et al. 2013. A theoretically based index of consciousness independent of sensory processing and behaviour. *Science Translation Medicine* 5 (198): 198ra105.
63. Casarotto, S., et al. 2016. Stratification of unresponsive patients by an independently validated index of brain complexity. *Annals of Neurology* 80: 718–729.
64. Colombi, I., et al. 2021. Spontaneous and perturbational complexity in cortical cultures. *Brain Sciences* 11 (11): 1453.
65. Hurley, S.L. 2010. Varieties of externalism. In *The extended mind*, ed. R. Menary, 101–153. Cambridge, MA: MIT Press.
66. Ward, D. 2012. Enjoying the spread: Conscious externalism reconsidered. *Mind* 121 (483): 731–751.
67. Vold, K. 2020. Can consciousness extend? *Philosophical Topics* 48 (1): 243–264.
68. Thompson, E., and D. Cosmelli. 2011. Brain in a vat or body in a world? Brainbound versus enactive views of experience. *Philosophical Topics* 39 (1): 163–180.
69. Papineau, D. 2021. *The Metaphysics of Sensory Experience*. Oxford: Oxford University Press.
70. Pautz, A. 2021. *Perception*. Routledge.
71. Dretske, F. 1993. Conscious experience. *Mind* 102 (406): 263–283.
72. Schellenberg, S. 2011. Perceptual content defended. *Noûs* 45 (4): 714–750.
73. Siegel, S. 2010. *The contents of visual experience*. New York: Oxford University Press.
74. Gozzano, S., and C. Hill. 2012. *New Perspectives on Type Identity: The Mental and the Physical*. Cambridge: Cambridge University Press.
75. Macdonald, C. 1989. *Mind-Brain Identity Theories*. Routledge.
76. Polger, Thomas W. 2009. Identity theories. *Philosophy Compass* 4 (5): 822–834.
77. Feigl, H. 1958. The ‘mental’ and the ‘physical.’ *Minnesota Studies in the Philosophy of Science* 2: 370–497.
78. Place, U.T. 1956. Is consciousness a brain process? *British Journal of Psychology* 47: 44–50.
79. Smart, J.J.C. 1959. Sensations and brain processes. *Philosophical Review* 68 (2): 141–156.
80. Bayne, T., J. Frohlich, R. Cusack, J. Moser, and L. Naci. 2023. Consciousness in the cradle: on the emergence of infant experience. *Trends in Cognitive Sciences* 27 (12): 1135–1149.
81. Frohlich, J., et al. 2023. Not with a “zap” but with a “beep”: measuring the origins of perinatal experience. *NeuroImage* 273: 1200537.
82. Fagard, J., et al. 2018. Fetal origin of sensorimotor behavior. *Frontiers in Neuroinformatics* 12: 23–23.
83. Khazipov, R., and M. Milh. 2018. Early patterns of activity in the developing cortex: Focus on the sensorimotor system. *Seminars in Cell & Developmental Biology* 76: 120–129.
84. Sanes, J.R. 2021. Experience and the refinement of synaptic connections. In *Principles of Neural Science*, ed. E.R. Kandel et al. New York: McGraw Hill Medical.
85. Cascio, C. 2010. Somatosensory processing in neurodevelopmental disorders. *Journal of Neurodevelopmental Disorders* 2 (2): 62–69.
86. Vogelsang, L., et al. 2018. Potential downside of high initial visual acuity. *Proceedings of the National Academy of Sciences of the United States of America* 115 (44): 11333–11338.
87. Corrow, S.L., K.A. Dalrymple, and J.J. Barton. 2016. Prosopagnosia: Current perspectives. *Eye Brain* 8: 165–175.
88. Rangarajan, V., et al. 2014. Electrical stimulation of the left and right human fusiform gyrus causes different effects in conscious face perception. *Journal of Neuroscience* 34 (38): 12828–12836.
89. Tong, F., and e al. 1998. Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21 (4): 753–759.
90. Crick, F., and C. Koch. 1990. Towards a neurobiological theory of consciousness. *Seminars in Neuroscience* 2: 263–275.
91. Bayne, T. 2021. *Philosophy of Mind: An Introduction*. Milton: Taylor & Francis Group.
92. Van Gulick, R. 2022. *Consciousness*. The Stanford Encyclopedia of Philosophy (Winter 2022 Edition). <https://plato.stanford.edu/entries/consciousness/>.
93. Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.
94. Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.
95. Baars, B. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
96. Dehaene, S., and J.P. Changeux. 2011. Experimental and theoretical approaches to conscious processing. *Neuron* 70: 200–227.
97. Lamme, V. 2006. Towards a true neural stance on consciousness. *Trends in Cognitive Sciences* 10 (11): 494–501.
98. Merker, B. 2007. Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences* 30 (1): 63–81.
99. Jackendoff, R. 1987. *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.
100. Prinz, J. 2012. *The Conscious Brain*. New York: Oxford University Press.
101. Armstrong, D. 1968. *A Materialist Theory of Mind*. London: Routledge and Kegan Paul.
102. Lycan, W. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
103. Rosenthal, D.M. 2005. *Consciousness and Mind*. Oxford: Clarendon.
104. Brown, R., H. Lau, and J.E. LeDoux. 2019. Understanding the higher-order approach to consciousness. *Trends in Cognitive Sciences* 23 (9): 754–768.
105. Lau, H. 2008. A higher-order Bayesian decision theory of consciousness. *Progress in Brain Research* 168: 35–48.
106. Kriegel, U. 2009. *Subjective Consciousness: A Self-Representational Theory*. Oxford: OUP.
107. Kriegel, U., and K. Williford. 2006. *Self-Representational Approaches to Consciousness*. Cambridge, MA: MIT Press.
108. Thomasson, A.L. 2000. After Brentano: A one-level theory of consciousness. *European Journal of Philosophy* 8 (2): 190–210.

109. Fodor, J. 1990. *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.
110. Millikan, R. 1984. *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.
111. Dall’Orso, S., et al. 2018. Somatotopic mapping of the developing sensorimotor cortex in the preterm human brain. *Cerebral Cortex* 28: 2507–2515.
112. Chalmers, D.J., and K. McQueen. 2021. Consciousness and the collapse of the wave function. In *Quantum mechanics and consciousness*, ed. S. Gao. Oxford: Oxford University Press.
113. Hameroff, S., and R. Penrose. 2014. Consciousness in the universe: A review of the “Orch OR” theory. *Physics of Life Reviews* 11 (1): 39–78. <https://doi.org/10.1016/j.plrev.2013.08.002>.
114. Loose, J. L., Menuge, A., and (Eds) Moreland, J. 2018. *The Blackwell Companion to Substance Dualism*. Wiley-Blackwell. <https://doi.org/10.1002/9781119468004>.
115. Coleman, S., and P. Goff. 2020. Russellian monism. In *Oxford Handbook of the Philosophy of Consciousness*, ed. U. Kriegel. Oxford: Oxford University Press.
116. Brüntrup, G., and L. Jaskolla, eds. 2017. *Panpsychism: Contemporary Perspectives*. New York: Oxford University Press.
117. Koch, C. 2019. *The Feeling of Life Itself: Why consciousness is Widespread but Can’t be Computed*. Cambridge, MA: The MIT Press.
118. Tye, M. 2017. *Tense Bees and Shell-Shocked Crabs: Are Animals Conscious?* New York: Oxford University Press.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.