



TITLE:

Human Brain Organoids and Consciousness

AUTHOR(S):

Niikawa, Takuya; Hayashi, Yoshiyuki; Shepherd, Joshua; Sawai, Tsutomu

CITATION:

Niikawa, Takuya ...[et al]. Human Brain Organoids and Consciousness. Neuroethics 2022, 15(1): 5.

ISSUE DATE:

2022-04

URL:

<http://hdl.handle.net/2433/267994>

RIGHT:

© The Author(s) 2022; 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.

Neuroethics (2022) 15:5
<https://doi.org/10.1007/s12152-022-09483-1>

ORIGINAL PAPER



Human Brain Organoids and Consciousness

Takuya Niikawa  · Yoshiyuki Hayashi ·
Joshua Shepherd · Tsutomu Sawai

Received: 18 February 2021 / Accepted: 14 October 2021 / Published online: 4 February 2022
© The Author(s) 2022, corrected publication 2022

Abstract This article proposes a methodological schema for engaging in a productive discussion of ethical issues regarding human brain organoids (HBOs), which are three-dimensional cortical neural tissues created using human pluripotent stem cells. Although moral consideration of HBOs significantly involves the possibility that they have consciousness, there is no widely accepted procedure to determine whether HBOs are conscious. Given that this is the case, it has been argued that we should adopt a precautionary principle about consciousness according to which, if we are not certain whether HBOs have

consciousness—and where treating HBOs as not having consciousness may cause harm to them—we should proceed as if they *do* have consciousness. This article emphasizes a methodological advantage of adopting the precautionary principle: it enables us to sidestep the question of whether HBOs have consciousness (the whether-question) and, instead, directly address the question of *what kinds of conscious experiences HBOs can have* (the what-kind-question), where the what-kind-question is more tractable than the whether-question. By addressing the what-kind-question (and, in particular, the question of what kinds of *valenced* experiences HBOs can have), we will be able to examine how much moral consideration HBOs deserve. With this in mind, this article confronts the what-kind-question with the assistance of experimental studies of consciousness and suggests an ethical framework which supports restricting the creation and use of HBOs in bioscience.

Takuya Niikawa and Tsutomu Sawai are Equally contributed.

T. Niikawa (✉)
Graduate School of Humanities, Kobe University, Kobe,
Japan
e-mail: niitaku11@gmail.com

Y. Hayashi
Faculty of Medicine, Saitama Medical University,
Saitama, Japan

J. Shepherd
Department of Philosophy, Carleton University, Ottawa,
Canada

J. Shepherd
University of Barcelona, Barcelona, Spain

T. Sawai
Institute for the Advanced Study of Human Biology
(WPI-ASHBi), KUIAS Kyoto University, Kyoto, Japan

Introduction

40 years ago, Hilary Putnam [1] presented the now-famous “brain in a vat” thought experiment:

A human being [...] has been subjected to an operation by an evil scientist. The person’s brain [...] has been removed from the body and placed in a vat of nutrients which keeps the brain alive. The nerve endings have been connected to a

super-scientific computer which causes the person whose brain it is to have the illusion that everything is perfectly normal. [...] The computer is so clever that if the person tries to raise his hand, the feedback from the computer will cause him to 'see' and 'feel' the hand being raised. [1], pp. 5-6)

In the 1980s, this was science fiction—there was no technology to keep a disembodied brain alive in a vat at that time. However, the recent development of bioscience has made “brains in a vat” possible, at least in a certain sense. This is not in the sense that one’s brain can be safely removed from their body and kept alive in a vat for some substantial amount of time; that scenario still belongs to science fiction.¹ Rather, it is in the sense that three-dimensional cortical neural tissues—so called “*human brain organoids*” (HBOs)—have successfully been artificially created and cultured in vitro. It is, moreover, an ongoing project to develop and use such HBOs for scientific and medical purposes [2–4].

There are many ethical questions about HBOs. Should we give moral consideration to HBOs? Is it morally permissible to treat HBOs as a mere means for our benefits? Is it even morally permissible to create HBOs in the first place? How far are we permitted to develop HBOs? Having noticed that these are urgent ethical issues, many scholars have begun to conduct ethical research on HBOs [5–9]

In addressing such ethical issues concerning HBOs, it matters whether the HBOs can “feel” or “think”; that is, whether they can have *conscious experiences*,² since it is widely accepted that consciousness is morally significant in some way

¹ Nevertheless, the science-fictional scenario might be not so far from reality. Vrselja et al. [72] have recently succeeded in restoring and maintaining microcirculation and molecular and cellular functions of an intact pig brain under ex vivo normothermic condition for four hours post-mortem. Although it is unlikely that the pig brain under such conditions experiences the “illusion of normality” described in Putnam’s scenario, it does not seem implausible to think that the pig brain *does* have some sort of conscious experience. This technology may be applied to human beings in the future, which would lead to serious ethical issues [73]. For instance, suppose that a person whom I love is severely injured in a traffic accident and the only way to save her life is to preserve her brain in such ex vivo normothermic conditions. Is it morally permissible for me to make such a decision to do so?

[10–12]. Thus, much attention has recently been devoted to the possibility that HBOs have consciousness and to the ethical implications which follow from this possibility [13–18].

However, the question of whether HBOs have, or can have, consciousness is difficult to answer. This is in part because, while we can observe the shape, size, and structure of an HBO, we cannot directly observe the presence or absence of consciousness in them. There is no consensus regarding an objective standard for detecting the presence of consciousness in creatures [19] and this lack of consensus extends to HBOs [16, 18, 20]. In addition, there are many competing theories of consciousness, which provide different predictions as to whether HBOs (can) have consciousness (see Sect. 3). We, thus, find ourselves in an epistemological predicament: we do not know how to determine whether HBOs (can) have consciousness. Even if we need to give moral consideration to HBOs *if they have consciousness*, we cannot establish an ethical protocol to regulate the creation and use of HBOs in bioscience because we do not know *whether the conditional holds*.

Against this background, this paper proposes a schema for engaging in a productive discussion of ethical issues regarding HBOs. There are three significant characteristics of this schema:

1. It adopts a precautionary principle about consciousness, which requires us to assume that HBOs have consciousness.
2. It shifts focus from the original question of *whether HBOs have consciousness* to the question of *what kinds of conscious experiences an HBO can have*, where the latter question is much more tractable than the former question.
3. It explores how we should treat HBOs on the basis of what kinds of conscious experiences they may have.

The biggest advantage of this schema is, as this paper will argue, that it enables us to discuss the creation and use of HBOs in bioscience in a relatively simple ethical framework, which can potentially be

² This paper focuses on phenomenal conscious experiences, which is defined in what-it-is-like terms rather than in functional terms [74],[75].

applied to other cases, such as the experimental use of human embryos and fetuses.

This paper shall proceed as follows: In Sect. 2, we briefly explain what HBOs are and how much they can plausibly be developed in the near future. In Sect. 3, we review the recent debate over how we can know whether an HBO is conscious and how this leads to an epistemological predicament. Section 4 presents the precautionary principle about consciousness as a way to escape the predicament, suggesting a new question about consciousness of HBOs to ask: what kinds of conscious experiences they have? In Sect. 5, we directly examine this question. Although the discussion in Sect. 5 is tentative and speculative, it serves as a useful demonstration of how to determine the kinds of conscious experiences which HBOs may have. Finally, Sect. 6 provides several positive suggestions—in light of our discussion in the preceding sections—concerning how we should restrict the creation and use of HBOs from an ethical perspective. In this section, we also discuss our suggestion's implications for experimental uses of embryos and other animals such as flies, mice and macaques.

Human Brain Organoids

A Japanese research group successfully created three-dimensional cortical neural tissues by inducing them from human pluripotent stem cells for the first time in 2008 [2, 3]. Such cortical neural tissues have been named “cerebral organoids” [4], though we shall call them HBOs. HBOs have been produced in such a way that they mimic various parts of the brain, including the cerebrum, midbrain, hypothalamus, pituitary gland, and hippocampus. Since these region-specific brain organoids can reproduce neurogenesis in vitro, they are expected to be used not only in basic research to elucidate the process of neurogenesis but also in applied research for neuro-related diseases and in clinical applications such as drug discovery [21, 22].

Despite these benefits, however, there are still many limitations to using HBOs. Current HBOs are different in size, number of neurons, and maturity from the normal brain and they also lack sensory input and behavioral output. In particular, although electrophysiological activities occur in HBOs during their development, those activities are not initiated by sensory input and do not produce any behavioral output. Further, in

terms of size, a typical HBO is much smaller than the brain of a mouse and is, at best, about the size of a pea or the brain of a honeybee. In terms of maturity, current HBOs do not mimic normal neurogenesis, even when cultured in vitro for long periods of time, because they do not have blood vessels. Moreover, current HBOs differ in structure from whole human brains, though some region-specific brain organoids have structural similarities with those regions of human brains and recently-induced brain organoids have multi-layered neural structures as also seen in human brains [23].

Recent studies on HBOs have been directed toward overcoming these limitations. Some studies have shown that HBOs could vascularize by being transplanted into the brains of mice and non-human primates and that HBOs could further develop inside these creatures [24–26], though there is no evidence showing that they can develop toward becoming normal mature human brain parts. Another study reported that HBOs that involve direct synaptic connections between cerebral neurons and photoreceptor cells responded to the input of external light stimuli [27]. In the future, region-specific brain organoids will likely be more refined and connected with other brain organoids as well as with both living and non-living systems to create more complexly structured brain organoids [28].

The Epistemological Predicament

Let us now turn to the question of whether HBOs have consciousness. One straightforward way to judge whether one is conscious is to ask: if one says yes, one is conscious; if one does not respond, one is not conscious. Of course, this method only applies to healthy human beings with enough understanding of the language and concepts involved in the question. Another natural approach is to examine whether one can perform intentional actions. If a creature *can* perform intentional actions, this is at least *prima facie* evidence that it has consciousness [29], otherwise, we may contend that it lacks consciousness. HBOs fail these standards, for they cannot produce verbal or behavioral outputs.

There are, however, several theories of consciousness which allow that HBOs can have consciousness.

For instance, *Integrated Information Theory (IIT)* states that consciousness is grounded in a causal informational structure that can internally generate integrated information [30, 31]. Given that simple systems like a photodiode can have such a causal informational structure, IIT allows that even such simple systems can have consciousness [32]. Even current HBOs show electrophysiological neural activities with some activation patterns [33, 34]. Since this suggests that current HBOs have a primitive form of causal informational structure, IIT would predict that they have consciousness.³

More radically, *panpsychism* states that microphysical entities are conscious [35, 36]. Many philosophers of consciousness have recently taken panpsychism as a serious theoretical option on the basis of reasonable arguments which suggest that it can successfully integrate consciousness into our current scientific picture of the world [37], chap. 4, for discussions of panpsychism, see [38]. Panpsychism, moreover, does not typically assert that every macro-physical object can have *distinctive kinds* of consciousness but, instead, claims that whether such objects have distinctive kinds of consciousness depends on the arrangements of microphysical entities, which can have consciousness in themselves. Thus, panpsychists do not typically claim that chairs and socks have distinct kinds of consciousness; rather, they hold that the consciousness belonging to such objects is a mere sum of the primitive kinds of consciousness instantiated by microphysical entities. Nevertheless, many panpsychists seem to accept that biological entities including flies, insects, plants, bacteria and amoeba, *can* have distinctive kinds of consciousness [37], chap.4). Given this, it is plausible to predict that it follows from panpsychism that HBOs would also have some distinctive kind of consciousness.⁴

Furthermore, HBOs are not mere artificial biological entities; they share human genes and mimic the developmental processes of human brains to some

extent. Biological naturalists of consciousness may regard this fact as pointing favorably to the potential presence of consciousness, since they emphasize the importance of evolutionary and developmental processes for consciousness [39].

In contrast, some theories of consciousness do not allow that HBOs can have consciousness. For instance, *the enactive theory of consciousness (ET)* states that having a body through which one can skillfully interact with the surrounding environment is necessary for having conscious experiences [40]. Given that in vitro HBOs do not have bodies to interact with the surrounding environment, ET denies the in-principle possibility that in vitro HBOs have consciousness, regardless of how structurally developed they might be.⁵

Further, according to *representational theories of consciousness*, the answer to the question of whether an HBO has consciousness depends on the degree of its sophistication. For instance, a *first-order representational theory* states that, if a system can represent what happens outside of it in such a way that the representation is ready for further cognitive processes, the system has consciousness [41–43]. A *higher-order representational theory* states that, if a system can *conceptually* represent its first-order representational state (which represents what happens outside of it), the system has consciousness [44, 45]. Our brain seems to be developed in such a way that it first gains the capacities to represent what happens outside of it (the first-order [FO] developmental stage and *then* gains the capacities to further represent the first-order representational states (the higher-order

Footnote 4 (continued)

deal with the combination problem—namely, the problem of explaining how micro-level consciousnesses give rise to macro-level consciousness. For instance, Roelofs [76] admits that parts of our brain are themselves conscious but contends that their consciousnesses are integrated into one unified consciousness of the whole brain when a functional connection holds between those parts. Although there seem to be intriguing ethical issues regarding conscious entities which are parts of a broader conscious entity, the present paper shall focus entirely on *non-partial* conscious entities that do not constitute a broader conscious entity.

⁵ Although we assume that in vitro HBOs do not have bodies to interact with the surrounding environment, future technology might allow them to have bodies that they would be capable of moving within a tank and with which they would be able to interact with other HBOs, for example. It is unclear what ET would say concerning this conceivable case.

³ We do not claim that IIT implies that *all* HBOs have consciousness. If, for instance, an HBO did not have a feedback loop and, therefore, was incapable of generating integrated information, it would not have consciousness on this view.

⁴ Panpsychists would also allow that region-specific HBOs can have consciousness but it is unclear what panpsychists would say about whether the parts of our brain have their own consciousnesses in addition to the one belonging to our whole brain. The answer to this question would depend on how they

Fig. 1 Classification of the Various Approaches to Consciousness of Human Brain Organoids

	In Principle Possibility of Consciousness for In Vitro HBOs	Order of Liberalness
Panpsychism	○	1
IIT	○	2
FOR	○	3
HOR	○	4
ET	X	5

[HO] developmental stage). If an HBO is developed to the extent that its neural structure and activity-patterns are the same as those of a human brain on the FO developmental stage but not on the HO developmental stage [15], pp. 761–762), then it is plausibly considered to have the capacity to represent what happens in the surrounding environment, but not to have the capacity to represent the first-order representation itself. According to the first-order representational theory, then, the HBO in question *can* be conscious; while, according to the higher-order representational theory, it *cannot* be. If the HBO is further developed so as to acquire the capacities to represent the first-order representations, the higher-order representational theory would also admit that it could be conscious.

Note that it is possible that there are borderline cases to which the representational theories of consciousness would not provide a determinate answer. For instance, the developmental stage of an HBO may be such that it is unclear whether it has the first-order representational capacity (a related issue is raised by Carruthers [46]. In such cases, the first-order representational theory cannot provide a determinate answer as to whether this HBO has consciousness. There are analogous borderline cases for the higher-order representational theory as well.

The preceding discussion of this section has shown that the answer to the question of whether HBOs can have consciousness depends on *what standard or theory of consciousness one adopts*. The liberal theories of consciousness such as IIT and panpsychism hold that even currently existing HBOs can have consciousness, while the conservative theories

of consciousness such as ET deny this (we owe the labels of “liberal” and “conservative” to Murray [47]. Further, the intermediate theories of consciousness, such as representationalist theories, state that whether HBOs can have consciousness depends on their developmental level (where the higher-order theory of consciousness is more conservative than the first-order one). (Fig. 1 is here).

The obvious next question is to ask what standard or theory of consciousness we should adopt. It is here that we arrive at an epistemological stalemate. There is no consensus as to what standard or theory of consciousness is the most promising and it does not seem that the debate over the correct indicators of consciousness will be resolved in the near future [48]. Thus, we do not know how to determine whether HBOs have consciousness.

Although the question of whether HBOs have consciousness matters for how we ought to treat them, it appears that there is no good way to resolve the question of whether they do possess consciousness. We are, then, left in the unenviable position of being uncertain as to how to proceed with an ethical discussion of HBOs.

The Precautionary Principle about Consciousness

Our proposal for escaping this predicament is to adopt a *precautionary principle about consciousness (PPC)*, according to which, if there is theoretical disagreement over whether X has consciousness—and where treating X as not having consciousness would cause more harm to X than benefit to X—we

ought to err on the side of being liberal with attribution of consciousness and assume that X *does* have consciousness [6, 15, 49], p. 762,[50]. The antecedent of the conditional is clearly satisfied for HBOs. While, as we have seen, some liberal theories of consciousness state that even current HBOs have consciousness, conservative theories of consciousness deny this, and we do not have a principled way to settle this disagreement.⁶ Moreover, in the case where HBOs actually possessed consciousness, treating HBOs as not having consciousness would cause harm to HBOs rather than benefit them. In contrast, in the case where HBOs actually lacked consciousness, treating HBOs as having consciousness would not cause harm to HBOs. As a whole, then, treating HBOs as not having consciousness would bring about more harm to HBOs than benefit to them. Thus, we can legitimately apply the PPC to HBOs and should, then, assume that HBOs are conscious. This assumption enables us to proceed with the ethical discussion concerning HBOs.

This section makes several clarificatory remarks on the PPC and its applications to HBOs. The first remark is regarding *the scope of the subjects of potential harms and benefits*. Our formulation of the PPC takes only X—the entity whose possession or lack of consciousness is in question—as the subject of potential harms and benefits. Because of this, we do not have to consider the possibility that *non-X entities* are harmed by X being treated as having consciousness in determining whether we should apply the PPC to X.

However, non-X entities *can* be harmed by X being treated as having consciousness [51, 52]. For instance, if the use of X for scientific experiments is prohibited due to the assumption that X has consciousness and the prohibition is significantly disadvantageous for pharmaceutical research, treating X

as having consciousness may cause much harm to patients who would otherwise benefit from the pharmaceutical research on X.

How should we address these sorts of potential disadvantages for non-X entities? Our proposal is that such disadvantages should be considered in discussing *the cancellation conditions* of the PPC rather than its *application conditions*. Whether a precautionary principle is acceptable or permissible depends in part on the proportion of its positive effects to its negative effects [53]. If the negative effects provided by applying the PPC to HBOs are proportionally larger than its positive ones, then we may be forced to call it off. For instance, if it follows that we ought not to create HBOs for any purpose, the negative impact on our well-being, in particular on the well-being of those who (will) have diseases that may be effectively treated by using HBOs, seems to be too large. In this case, we may need to cancel our application of the PPC to HBOs.

Our position is, thus, as follows: we can determine whether we should apply the PPC to HBOs based only on considerations of harm and benefit to HBOs *themselves*. However, we may need to cancel the application of the PPC if it turns out that applying it causes more harm to other creatures than the benefit that it causes to HBOs. We may be forced to cancel it after further consideration and observation of its consequences. In this sense,

the application of the PPC to HBOs is *tentative*.

Note, however, that the more likely a creature is to have consciousness, in the sense that more theories of consciousness predict that it has consciousness, the less acceptable it is to cancel the PPC and to treat this creature as not having consciousness, even if the experimental uses of the creature—which were harmful to it—produced large benefits to other creatures like human beings. Whether we can legitimately cancel the PPC depends not only on how much harm its application causes to other creatures but also on how likely the creature in question is to have consciousness (for the implication of this point, see Sect. 6).

Our second clarificatory remark is on the relation between the PPC and ethical treatment. The PPC *does not, in itself, determine how we should treat HBOs*. The possession of consciousness does not necessarily imply *the possession of moral status*, which would require us to give moral consideration to X for X's own sake. One sufficient condition for having moral

⁶ We are able to make the PPC more stringent by adding conditions that limit which theories of consciousness are to be seriously considered. For instance, we could add the condition that we should take seriously only *scientific* theories of consciousness in the sense of being experimentally testable and could find that panpsychism is *not* scientific in that sense, which would allow us to eliminate panpsychism from the list of theories of consciousness to be considered. This paper does not put such additional constraints on the PPC but, instead, holds that every theory of consciousness that is seriously discussed in academic contexts (including in philosophy) should be considered.

status is to have valenced experiences, such as pleasant and painful experiences [54]. That is, if *X* can have painful experiences, then we ought to treat *X* in such a way that *X* does not have painful experiences (or, at least, *X* has less painful experiences). Note, however, that the possession of consciousness does not necessarily mean the possession of the capacities to have sensory valenced experiences, since there may be a primitive form of consciousness that is unable to involve valence. The possession of such a primitive form of consciousness may not be sufficient for having moral status [55].⁷ Whether we should give moral consideration to *X* is mainly determined by *what kinds of conscious experiences* *X* can have, rather than by the mere fact that *X* has consciousness [12, 18], see also Sect. 6). In other worlds, the possession of consciousness is a “gatekeeper” to get us to the question of what kinds of conscious experiences *X* has, a question which *does* have direct moral significance.⁸

Moreover, even if it turns out that we should give moral consideration to *X*, it still remains open as to *how much* moral consideration we should give. This is also determined in part by what *kinds* of conscious experiences *X* can have (see Sect. 6). Accordingly, without examining what kinds of conscious experiences an HBO can have, we cannot specify whether and how much we should give moral consideration to

⁷ 4, Sect. 11 endorses the “existentialist view” about the moral significance of consciousness, according to which to have consciousness is to exist as having intrinsic value such that the elimination of this existence implies the loss of its intrinsic value. The existentialist view states that even the possession of primitive consciousness confers some degree of moral status on its possessor. However, since the existentialist view is not widely accepted and even discussed, this paper shall not consider it further.

⁸ One may wonder why we did not directly formulate the PPC in terms of *valenced experiences* rather than *consciousness itself* as in Shepherd [77]. The reason concerns the root of our epistemological uncertainty about consciousness. The *fundamental* reason why we cannot observationally determine whether *X* has valenced experiences is that we do not know when consciousness occurs; *not* that we do not know when valence is instantiated in consciousness (given the presence of consciousness). The relevant epistemological uncertainty lies fundamentally in the nature of consciousness rather than the nature of valenced experiences. Accordingly, since the PPC is introduced to address such epistemic uncertainty, it should be formulated in terms of consciousness rather than valenced experiences.

it. Thus, the PPC does not directly provide any determinate answer as to how we ought to treat HBOs.

Here one may suspect that the PPC does not resolve our predicament at all but, rather, merely pushes it back one level. The line of thought is as follows: Our predicament is caused by the intractability of the question of whether an HBO has consciousness (the *whether-question*). The question of what kinds of conscious experiences an HBO has (the *what-kind question*) seems no less intractable than the *whether-question*. Thus, it seems that we would face the exact same predicament in attempting to address the *what-kind question* as we did in attempting to address the *whether-question*.

Our response to this worry is that the *what-kind question* is more tractable than the *whether-question*. The intractability of the *whether-question* lies in the hard problem of consciousness—namely, the deep mystery of what “switches” consciousness on and off. We cannot answer the *whether-question* without knowing what needs to be present in order to switch consciousness on. In contrast, when we address the *what-kind question* with PPC, *it is presupposed* that the consciousness switch is on. Hence, in addressing the *what-kind question*, we can sidestep the hard problem of consciousness.

It is important to note here that experimental studies of consciousness have productively explored the correlations between conscious experiences and (1) physiological states, (2) sensory stimuli, (3) brain areas, and (4) cognitive functions [41, 56–58] and that, through these studies, we have gained much knowledge about their relations. Further, assuming that HBOs have consciousness, we can employ these accumulated scientific findings about consciousness in order to address the *what-kind question*. That is to say, we can consider what kinds of conscious experiences an HBO may be able to have in terms of (1) what types of physiological states it can be in, (2) what types of sensory stimuli it is sensitive to, (3) what types of neural network structure it has, and (4) what types of cognitive functions it has. For instance, if an HBO is sensitive to optical stimuli [27], it might be plausibly counted as potentially having primitive visual experiences. This is the reason why the *what-kind question* is tractable (see Sect. 5 for more detailed discussion). The PPC can shift our focus from the *whether-question* to the *what-kind question*

and thereby opens up a space in which to productively discuss how we should treat HBOs.

The third clarificatory remark is that the application of the PPC to HBOs is primarily for *problem-solving* rather than *problem-finding*. This principle was introduced to address the already occurring ethical worries concerning HBOs, rather than to discover novel ethical issues. We are *not obliged* to consider whether we should apply the PPC to X *unless* there are preexisting ethical issues regarding X. We ought to do so only when there are preexisting ethical issues concerning X. Although we may be able to uncover new ethical issues by employing the PPC to various entities, it is neither the intended purpose of this principle nor are we obliged to seek such new issues out.

In light of these remarks, let us consider a worry about the unlimited application of the PPC. If we take seriously the most liberal theory of consciousness—namely, panpsychism—the PPC would be applied to every living being and even microphysical entities. However, it seems practically impossible and *prima facie* implausible to hold that we should give moral consideration to every living being and microphysical entity. Hence, it appears that we should somehow limit the application of the PPC. Yet, we cannot appeal to any specific standard or theory of consciousness for this purpose, for the epistemological problem rears its head again: we do not know how to determine what standard or theory of consciousness to adopt. Thus, it seems entirely unclear how we could limit the application of the PPC.

Our response to this worry is threefold. The first is to appeal to our third point that the PPC is primarily for *problem-solving* rather than *problem-finding*. There is no serious ethical issue as to how we ought to treat, for example, microphysical entities or primitive biological entities, such as amoeba, for their own sakes. Even though panpsychism implies that they are conscious, there is no obligation to apply the PPC to them and thereby open up a new ethical discourse. Our second response is to notice that, even if we apply the PPC to primitive biological creatures and microphysical entities, it does not follow that we should give them moral consideration. This is because it depends on what *kinds* of conscious experiences they can have and there is no panpsychist who explicitly endorses the view that microphysical entities are sentient (though it is unclear what panpsychists say about primitive biological entities). The third response is

that, if it followed from the PPC that we need to give significant moral consideration to primitive biological entities and microphysical entities, it would be so burdensome to us that it is reasonable to cancel the application of the PPC to them.

Let us here summarize the main line of discussion in this Sect. If an HBO has consciousness but is regarded as not having consciousness and treated accordingly, the HBO may be harmed by this treatment. We should adopt the PPC to prevent such potential harms. Adopting the PPC enables us to address what kinds of conscious experiences HBOs may have. Through addressing the what-kind question, we can examine whether and how much we need to give moral consideration to HBOs. Depending on its results (in particular, the size of the negative impact on human beings), we may need to cancel the application of the PPC to HBOs.

What is It Like to Be an HBO?

The PPC enables us to focus on the what-kind question—that is, the question of what kinds of conscious experiences HBOs have—while leaving aside the whether-question concerning the presence of consciousness itself. This section, accordingly, addresses the what-kind question. However, this section does not aim to provide decisive and comprehensive answers to this question. Rather, it aims to *demonstrate* how we can infer the kinds of conscious experiences that HBOs are likely to have. For this methodological demonstration, we will present some inferences from observable features of HBOs to the kinds of conscious experiences they may have. Although these inferences are partly grounded in experimental findings of consciousness studies, they also depend on assumptions the reliability of which remains to be examined. In this sense, the inferences presented in this section are *highly speculative*.

It is important to note that there is no widely shared protocol which specifies what kinds of experiences a creature can have when that creature cannot provide introspective reports in response to verbal questions. For such beings, our ability to specify which kinds of experiences they may possess typically depends largely on their similarity (or lack thereof) to standard human beings. The more dissimilar a creature is to human beings, the more difficult it is for us to infer what kinds of experiences it can

have. HBOs stand in an interesting position in this respect, however. When one focuses on their appearances and behaviors, HBOs are much more dissimilar to human beings than non-human animals and perhaps even AI robots. Yet, HBOs share some aspects of gene expression with human brains at developmental stages. Moreover, some HBOs are designed to mimic developmental processes of specific regions in human brain, such as cerebral cortex [2, 3], the hippocampus [59], and mid-brain and hypothalamic [60]. Because of the genetic and developmental similarity to human brains, we can take advantage of the accumulated knowledge of experimental studies of consciousness on human beings to reason what kinds of experiences HBOs may have. That is to say, we can infer what kinds of conscious experiences an HBO may have from (1) what types of physiological states it can be in, (2) what types of stimuli it is sensitive to, (3) what neural network structures it has, and (4) what types of cognitive functions it has (as discussed above in Sect. 4).

Let us first focus on the *glucose metabolic conditions* of HBOs. HBOs are able to be in different glucose metabolic conditions, which in part depends on culturing conditions. It has been suggested that the state or level of human consciousness varies depending on the glucose metabolic levels of our brains [61, 62]. Assuming that the difference in the level of consciousness correlates with the difference along some phenomenological dimension (perhaps such as feeling lively or drowsy), we might infer that the consciousness of HBOs can differ in that phenomenological dimension from the fact that they can differ in glucose metabolic level.

There are two remarks which we should make about this inference. First, the inference depends on the assumption that the levels of consciousness correlate with some phenomenological difference. However, this assumption can be doubly questioned as follows: (1) The notion of “levels of consciousness” has been challenged on both conceptual and theoretical bases [63] and (2) even if the notion of “levels of consciousness” makes good sense, it is unclear whether their difference corresponds to some phenomenological difference. Further conceptual, theoretical, and phenomenological research on levels of consciousness is needed to address these issues. The second remark which we should make is that the inference in question becomes acceptable *only on the*

precautionary assumption that HBOs are conscious. Given that there can be unconscious organisms that can metabolize glucose, the fact that X is able to be in states which differ in glucose metabolism does not in itself imply that X actually *is* conscious.⁹ Thus, it is only when we can plausibly assume that X is conscious on a distinct ground (which, in our case, is the PPC) that the difference in glucose metabolism suggests that there would be some phenomenological difference.

To see the second point more clearly, let us use an analogy. On the presupposition that a cup of ice-cream is in a box, the fact that I smell vanilla from the box is good evidence that there is vanilla ice-cream in the box. Without that presupposition, however, the fact that I smell vanilla from the box is *not* good evidence that vanilla ice-cream is in the box. This is because there may equally be vanilla cookies, vanilla cakes, or some kind of perfume, rather than vanilla ice-cream in the box. Likewise, on the presupposition that HBOs have consciousness, the fact that they can be in different glucose metabolic states is counted as *prima facie* evidence that their consciousness can differ in levels (and, accordingly, along some phenomenological dimension). Without this presupposition, however, the fact that HBOs can be in different glucose metabolic states does *not* suggest that their consciousness can differ in levels, since we would not have a principled reason to believe that they have consciousness at all. The same applies to all of the inferences that we will present in this section. Simply put, all of these inferences wholly rely on the precautionary assumption that HBOs are conscious.

With this said, let us now turn to *the neural activity patterns* of HBOs. It has been found that synchronous firing occurs in two-dimensional neuronal structures that are made by disassembling an HBO into small cells and scattering them on a plate to produce two-dimensional neuronal connections [33]. This is indirect evidence that synchronous firing also occurs in HBOs when they are intact (though it is technically difficult to examine this prediction due to the current limitations of microscopes). Assuming that the

⁹ This seems to be the reason why Shulman et al. [61] stated that high global energy production and consumption is not sufficient but only necessary for the presence of consciousness.

presence of synchronous neural activities is a good indicator of the unification of consciousness [64], we might infer from the likelihood of synchronous neuronal firings within HBOs that they have unified conscious experiences rather than having multiple streams of consciousness in “one great blooming, buzzing confusion” [65], p. 462). Note, however, that this inference would be defeated if it turns out that some conditions other than the presence of synchronous neural activities must also be satisfied for neuronally-based conscious experiences to be unified and HBOs do not satisfy them. Likewise, if it turns out that synchronous firing does not occur in HBOs, the inference does not hold. In this sense, this inference is tentative and defeasible.

Let us then focus on the *neural network structures* of HBOs. Although no HBO has successfully been built that resembles any specific region of mature human brains in structure, we can imagine HBOs with such structural similarity with mature human brains. If an HBO is cultured to structurally resemble a part of human brains that is found to contribute to a *kind of sensory information processing*, then we might infer that the HBO can have *primitive sensory experiences of that kind*. For instance, if an HBO structurally resembles a part of human brains which is responsible for visual processing—namely, the visual cortex—then we might reason that the HBO can have primitive visual experiences, such as an experience of a flash of light followed by an experience of darkness.¹⁰

By focusing on the neural network structure of HBOs, we can also consider what *non-sensory* experiences they can have. For instance, if an HBO is structured in such a way that it contains neural feedback connections, which seem related to predictions in a broad sense [66], we might reason that it has the

primitive capacity of prediction and that it, therefore, can have a *primitive* cognitive experience of predicting. Furthermore, it is known that cerebral limbic systems play a key role in having emotional experiences with positive and negative valence [67]. Given this, we might reason that an HBO that is designed to structurally resemble cerebral limbic systems can have *primitive* emotions.

In general, we can reason that, if an HBO acquires more complex structures, then it is likely to have more kinds of conscious experiences. For instance, *the capacity to think about oneself verbally* seems to be grounded in various brain areas including the anterior cortical midline, the posterior cortical midline, the lateral inferior parietal lobe, the medial temporal lobe—including the insula, amygdala, and hippocampus—and the anterior and middle lateral temporal cortex [68]. If HBOs are designed to resemble the large neuronal networks in those areas, they might be regarded as potentially having the cognitive experiences involving self-representation.

In summary: this section has demonstrated how we can address the what-kind question on the ground of the PPC. We can infer what kinds of conscious experiences current and future HBOs may have from their glucose metabolic conditions, their neural activity patterns, their neural network structures and so on. Although the inferences presented in this section are highly speculative and non-comprehensive, they shall be capable of being further refined based on future neuroscientific findings about consciousness.

An Ethical Framework to Guide HBO Research

Koplin and Savulescu [15] propose an ethical framework to limit the creation and use of HBOs on the basis of a similar framework for animal ethics (e.g., [69]). Their position distinguishes between three kinds of HBOs: (a) non-conscious brain organoids, (b) conscious or potentially conscious brain organoids, and (c) brain organoids with the potential to develop advanced cognitive capacities. For non-conscious brain organoids, they claim that “research should be regulated according to existing frameworks for stem cell and human biospecimen research” (p. 765). For conscious or potentially conscious brain organoids, they propose six additional restrictions (p. 765):

¹⁰ One may cast doubt on this claim by suggesting that it is controversial that a specific part of our brain can produce a conscious experience without possessing the appropriate connections with other parts of the brain. Certainly, it may seem plausible that the visual cortex cannot produce a visual experience without being appropriately connected with other areas of our brain. However, this doubt works only when we do not assume that an isolated visual cortex has consciousness. When we *do* make this assumption, the most plausible answer to the what-kind question would be that the isolated visual cortex has *visual experiences*. In this way, our claim relies heavily on the precautionary assumption that HBOs have consciousness.

1. The expected benefits of the research must be sufficiently great to justify the moral costs, including potential harms to brain organoids.
2. Conscious brain organoids should be used only if the goals of the research cannot be met using non-sentient material.
3. The minimum possible number of brain organoids should be used, compatible with achieving the goals of the research.
4. Conscious brain organoids should not have greater potential for suffering than is necessary to achieve the goals of the research.
5. Conscious brain organoids must not experience greater harm than is necessary to achieve the goals of the research.
6. Brain organoids should not be made to experience severe long-term harm unless necessary to achieve some critically important purpose.

For brain organoids with the potential to develop advanced cognitive capacities, they further add (p.765):

7. Brain organoids should be screened for advanced cognitive capacities they could plausibly develop. In general, such assessments should err on the side of overestimating rather than under-estimating cognitive capacities.
8. Cognitive capacities should not be more sophisticated than is necessary to achieve the goals of the research.
9. Welfare needs associated with advanced cognitive capacities should be met unless failure to do so is necessary to achieve the goals of the research.
10. The expected benefits of the research must be sufficiently great to justify the expected or potential harms. This calculation should take into account the implications of advanced cognitive abilities for brain organoids' welfare and moral status.

Although we basically agree with the proposed ethical framework, we suggest several modifications. First, there should be no option for category (a)—that of “non-conscious brain organoids”. [15], p. 762) assume that “a brain organoid lacks even a rudimentary form of consciousness until it resembles the brain of a fetus at 20 weeks’ development”. As we have discussed, however, the plausibility of this assumption depends on what standard or theory of consciousness to adopt (see also [70]). Some liberal theories

of consciousness such as IIT and panpsychism allow that even those underdeveloped brain organoids can have a primitive form of consciousness.

Additionally, we want to make another remark regarding category (b)—that of conscious and potentially conscious brain organoids. As Shepherd [12] argues, the morally significant kinds of conscious experiences are *valenced experiences* (that is, affective experiences). This is because those beings that can have valenced experiences have experiential interests: positively valenced experiences are intrinsically good for them and negatively valenced experiences are intrinsically bad for them. However, there is no a priori reason to think that every conscious entity *can* have valenced experiences. At the very least, we can imagine a primitive form of consciousness that does not involve any such valence. Given this, (b) should be further divided into two sub-categories: (b-1) conscious brain organoids that *cannot* have valenced experiences and (b-2) conscious brain organoids that *can* have valenced experiences. The first six conditions hold only for (b-2).

Nevertheless, it is not easy to identify the borderline between (b-1) and (b-2). Currently existing brain organoids are unlikely to have typical affective experiences such as bodily pain or pleasure. However, valenced experiences are not necessarily sensory. It is also possible that currently existing brain organoids have primitive valenced experiences. Further research on valenced experiences is needed to distinguish between HBOs belonging to (b-1) and those belonging to (b-2).

There should also be another modification regarding the distinction between categories (b) and (c). We agree that brain organoids with advanced cognitive capacities—those belonging to group (c)—should be given more moral consideration than brain organoids without these capacities, since they can have more complex valenced experiences and, therefore, also have more sophisticated experiential interests. However, there does not seem to be a clear-cut line between (b) and (c); rather, the distinction seems to be a matter of degree. For example, if relatively simple HBOs contain recurrent neural connections, they may be understood as having the cognitive capacity of prediction—though this capacity would be much more primitive than our own. Given that there are such borderline cases between (b) and (c), it seems better to abolish the distinction and to, instead, provide a wide

and flexible framework that can accommodate the gradation between (b) and (c).

Given those modifications, we propose a revision to the ethical framework presented by Koplín and Savulescu. For HBOs that can have valenced experiences, we suggest that:

1. The expected benefits of the research must be sufficiently great to justify the moral costs, including the potential harms to these brain organoids. The assessments of such potential harms should err on the side of overestimating rather than underestimating.
2. HBOs should be used only if the goals of the research cannot be met using materials that cannot have valenced experiences.
3. The minimum possible number of brain organoids should be used, compatible with achieving the goals of the research.
4. HBOs must not experience greater harm than is necessary to achieve the goals of the research.
5. HBOs should not be constructed in such a way as to have greater potential for suffering—namely, they should not be given the capacities to have more sophisticated kinds of valenced experiences—than is necessary for achieving the goals of the research.
6. Brain organoids should not be made to experience severe long-term harm unless this is necessary to achieve some critically important purpose.
7. Brain organoids should be ranked according to how sophisticated the valenced experiences which they can have are.¹¹ We should use lower-ranked conscious brain organoids if the goals of the research can be met without using higher-ranked ones.

This revised ethical framework is simpler than the one put forward by Koplín and Savulescu in that it unitarily frames moral limitations based only on valenced experiences and their degree of sophistication. Nevertheless, it still covers all relevant types of HBOs and offers practical guidance concerning how to treat different types of HBOs. We propose that

HBO research should be conducted within this unitary ethical framework.¹²

Let us consider, for illustration, two possible ways of applying this framework. First, suppose that existing HBOs have negatively valenced experiences when their conscious level is too low and that levels of consciousness can be successfully inferred from glucose metabolic conditions. In this case, we can apply the fourth and sixth conditions of our framework, claiming that we ought to keep HBO glucose metabolic levels higher, since lower levels are tied to negatively valenced experiences. Secondly, if it turns out that HBOs have negatively valenced experiences by being physically damaged and such negatively valenced experiences become greater when their neural structures become more sophisticated, we can apply the seventh condition of our framework and claim that, if the use of less sophisticated HBOs is sufficient for a given experimental purpose which involve physically damaging the HBOs, we ought not to use more sophisticated HBOs for that purpose.

We want to emphasize that this framework can also potentially be applied to the use and creation of human embryos and fetuses, non-human animals, and even robots for scientific purposes. In order to examine whether (or to what extent) they deserve moral consideration, we should first examine whether some theories of consciousness predict that they have consciousness. If there are such theories, we can then apply the PPC to these entities. This enables us to, in turn, consider what *kinds* of conscious experiences such entities have and to rank them according to how sophisticated the valenced experiences which they may have are. Based on this, we can consider how we ought to treat these entities on the basis of our ethical framework developed above.

As we have argued in Sect. 4, we are allowed to cancel the application of the PPC to a kind of entity just in case (1) the application produces a large negative impact on our well-being and (2) the majority of theories of consciousness predicts that these entities do not possess consciousness (namely, if only fairly liberal theories of consciousness predict that

¹¹ For a methodology of how to rank different kinds of affective experiences (and, accordingly, different types of conscious creatures), see Shepherd [12].

¹² This ethical framework does not address issues over ownership and stewardship regarding HBOs. For instance, who should be counted as an owner of an HBO? The creator or the original cell's donor? Such issues need to be discussed within the context of another framework.

they possess consciousness). Given this, for example, if it turns out that macaques have conscious experiences which are as sophisticated as those belonging to human beings and that they deserve as much moral consideration as we do, the experimental use of macaques without informed consent ought to be prohibited. Even if this prohibition produces a large negative impact on the advancement of neuroscience and, thereby, on our well-being indirectly, we are not legitimately allowed to cancel the application of the PPC. This is because the majority of theories of consciousness predict that macaques possess consciousness.

In contrast, imagine that a high-tech company produces a new type of caregiving robot, which is exceptionally sophisticated in its functionality, to the extent that care receivers sometimes mistake them for human caregivers. Suppose further that we work the caregiving robots to exhaustion and to the point of bodily damage for the sake of care receivers. If we assume that this kind of caregiving robot has consciousness, we might infer that they have fairly sophisticated valenced conscious experiences, concluding that it is morally unacceptable to exploit them for the sake of care receivers. However, if the majority of theories of consciousness are committed to biological naturalism, which holds that only evolutionarily-selected living beings have the potential for consciousness, then we are allowed to cancel the application of the PPC to the caregiving robots. This is because applying this principle would both produce a large negative impact on care receivers *and* the majority of theories of consciousness predict that these robots do not possess consciousness.

Let us close this section by acknowledging two limitations of this paper. First, the effectiveness of the ethical framework proposed here depends on the further development of consciousness studies—particularly those concerning the correlation between valenced experiences and the physiological, informational, neural, and functional conditions of our bodies, including our brains. We are in need of much more knowledge about such correlations if we are to reliably address the what-kind question for non-human beings who cannot provide introspective reports. Second, we do *not* claim that our ethical framework is capable of capturing *every* ethical issue regarding the use or creation of living beings. For instance, the use of fetuses for scientific

purposes may involve some distinctive ethical issues that the use of HBOs does not face and which are not directly related to the potential for having conscious experiences [71]. Our proposed ethical framework is designed to capture consciousness-based moral status, *not* other types of moral status, and, thus, aims at presenting a new perspective on the use and creation of various kinds of *possibly conscious* entities.

Conclusion

This paper proposes a new theoretical schema to address ethical issues concerning HBOs. This schema has two key features. The first is to adopt the PPC to shift focus from the question of *whether HBOs have consciousness* to the question of *what kinds of conscious experiences an HBO can have*. The second is to rank how much moral consideration HBOs deserve based on the kinds of valenced experiences that we infer that they are likely to have. This theoretical schema enables us to offer a concrete ethical framework to limit the creation and use of HBOs in bioscience—one which is a revised and unified version of the framework which Koplin and Savulescu [15] have proposed. This new ethical framework provides effective guidance regarding how to proceed in, and further develop, HBO research.

Funding This work was supported by JSPS KAKENHI Grant Number 21K00011 (for TN), 20K00001 (for TN), 21K12908 (for TS) and 19K00016 (for YH), and AMED Grant Number JP21wm0425021 (for TN and TS), and the ERC Starting Grant 757698 (for JS), and Azrieli Global Scholar Fellowship from CIFAR (for JS).

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

- Putnam, H. 1981. *Reason. Truth and History*: Cambridge University Press.
- Eiraku, M., K. Watanabe, M. Matsuo-Takasaki, M. Kawada, S. Yonemura, M. Matsumura, T. Wataya, A. Nishiyama, K. Muguruma, and Y. Sasai. 2008. Self-Organized Formation of Polarized Cortical Tissues from ESCs and Its Active Manipulation by Extrinsic Signals. *Cell Stem Cell* 3: 519–532. <https://doi.org/10.1016/j.stem.2008.09.002>.
- Kadoshima, T., H. Sakaguchi, T. Nakano, M. Soen, S. Ando, M. Eiraku, and Y. Sasai. 2013. Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *PNAS* 110: 20284–20289. <https://doi.org/10.1073/pnas.1315710110>.
- Lancaster, M.A., M. Renner, C.-A. Martin, D. Wenzel, L.S. Bicknell, M.E. Hurlles, T. Homfray, J.M. Penninger, A.P. Jackson, and J.A. Knoblich. 2013. Cerebral organoids model human brain development and microcephaly. *Nature* 501: 373–379. <https://doi.org/10.1038/nature12517>.
- Farahany, N.A., H.T. Greely, S. Hyman, C. Koch, C. Grady, S.P. Paşca, N. Sestan, P. Arlotta, J.L. Bernat, J. Ting, J.E. Lunshof, E.P.R. Iyer, I. Hyun, B.H. Capestany, G.M. Church, H. Huang, and H. Song. 2018. The ethics of experimenting with human brain tissue. *Nature* 556: 429–432. <https://doi.org/10.1038/d41586-018-04813-x>.
- Greely, H.T. 2021. Human Brain Surrogates Research: The Onrushing Ethical Dilemma. *The American Journal of Bioethics* 21: 34–45. <https://doi.org/10.1080/15265161.2020.1845853>.
- Munsie, M., I. Hyun, and J. Sugarman. 2017. Ethical issues in human organoid and gastruloid research. *Development* 144: 942–945. <https://doi.org/10.1242/dev.140111>.
- International Society for Stem Cell Research (ISSCR). 2021. Guidelines for Stem Cell Research and Clinical Translation. https://www.isscr.org/docs/default-source/all-isscr-guidelines/2021-guidelines/isscr-research-and-clinical-translation-2021.pdf?sfvrsn=979d58b1_4
- National Academies of Sciences, Engineering, and Medicine (NASEM). 2021. The Emerging Field of Human Neural Organoids, Transplants, and Chimeras: Science, Ethics, and Governance. <https://www.nationalacademies.org/our-work/ethical-legal-and-regulatory-issues-associated-with-neural-chimeras-and-organoids>.
- Kahane, G., and J. Savulescu. 2009. Brain damage and the moral significance of consciousness. *Journal of Medicine and Philosophy* 34: 6–26. <https://doi.org/10.1093/jmp/jhn038>.
- Niikawa, T. 2018. Moral Status and Consciousness. *Analele Universităţii din Bucureşti – Seria Filosofie* 67, 235–257.
- Shepherd, J. 2018. *Consciousness and Moral Status*. Oxon (UK): Routledge.
- Bayne, T., A.K. Seth, and M. Massimini. 2020. Are There Islands of Awareness? *Trends in Neurosciences* 43: 6–16. <https://doi.org/10.1016/j.tins.2019.11.003>.
- Hyun, I., J.C. Scharf-Deering, and J.E. Lunshof. 2020. Ethical issues related to brain organoid research. *Brain Research* 1732: 146653. <https://doi.org/10.1016/j.brainres.2020.146653>.
- Koplin, J.J., and J. Savulescu. 2019. Moral Limits of Brain Organoid Research. *The Journal of Law, Medicine & Ethics* 47: 760–767. <https://doi.org/10.1177/1073110519897789>.
- Lavazza, A., and M. Massimini. 2018. Cerebral organoids: Ethical issues and consciousness assessment. *Journal of Medical Ethics* 44: 606–610. <https://doi.org/10.1136/medethics-2017-104555>.
- Sawai, T., H. Sakaguchi, E. Thomas, J. Takahashi, and M. Fujita. 2019. The Ethics of Cerebral Organoid Research: Being Conscious of Consciousness. *Stem Cell Reports* 13: 440–447. <https://doi.org/10.1016/j.stemcr.2019.08.003>.
- Shepherd, J. 2018. Ethical (and epistemological) issues regarding consciousness in cerebral organoids. *Journal of Medical Ethics* 44: 611–612. <https://doi.org/10.1136/medethics-2018-104778>.
- Irvine, E. 2013. Measures of Consciousness. *Philosophy. Compass* 8: 285–297. <https://doi.org/10.1111/phc3.12016>.
- Lavazza, A. 2020. Human cerebral organoids and consciousness: A double-edged sword. *Monash Bioethics Review*. <https://doi.org/10.1007/s40592-020-00116-y>.
- Cugola, F.R., I.R. Fernandes, F.B. Russo, B.C. Freitas, J.L.M. Dias, K.P. Guimarães, C. Benazzato, N. Almeida, G.C. Pignatari, S. Romero, C.M. Polonio, I. Cunha, C.L. Freitas, W.N. Brandão, C. Rossato, D.G. Andrade, D. de Faria, and P. Garcez, A.T., Buchpigel, C.A., Braconi, C.T., Mendes, E., Sall, A.A., Zanotto, P.M. de A., Peron, J.P.S., Muotri, A.R., Beltrão-Braga, P.C.B., 2016. The Brazilian Zika virus strain causes birth defects in experimental models. *Nature* 534: 267–271. <https://doi.org/10.1038/nature18296>.
- Watanabe, M., J.E. Buth, N. Vishlaghi, L. de la Torre-Ubieta, J. Taxisidis, B.S. Khakh, G. Coppola, C.A. Pearson, K. Yamauchi, D. Gong, X. Dai, R. Damoiseaux, R. Aliyari, S. Liebscher, K. Schenke-Layland, C. Caneda, E.J. Huang, Y. Zhang, G. Cheng, D.H. Geschwind, P. Golshani, R. Sun, and B.G. Novitch. 2017. Self-Organized Cerebral Organoids with Human-Specific Features Predict Effective Drugs to Combat Zika Virus Infection. *Cell Reports* 21: 517–532. <https://doi.org/10.1016/j.celrep.2017.09.047>.
- Seto, Y., and M. Eiraku. 2019. Toward the formation of neural circuits in human brain organoids. *Current Opinion in Cell Biology* 61: 86–91. <https://doi.org/10.1016/j.ccb.2019.07.010>.
- Daviaud, N., R.H. Friedel, H. Zou. 2018. Vascularization and Engraftment of Transplanted Human Cerebral Organoids in Mouse Cortex. *eNeuro* 5. <https://doi.org/10.1523/ENEURO.0219-18.2018>
- Kitahara, T., H. Sakaguchi, A. Morizane, T. Kikuchi, S. Miyamoto, and J. Takahashi. 2020. Axonal Extensions along Corticospinal Tracts from Transplanted Human Cerebral Organoids. *Stem Cell Reports* 15: 467–481. <https://doi.org/10.1016/j.stemcr.2020.06.016>.
- Mansour, A.A., J.T. Gonçalves, C.W. Bloyd, H. Li, S. Fernandes, D. Quang, S. Johnston, S.L. Parylak, X. Jin, and F.H. Gage. 2018. An in vivo model of functional and

- vascularized human brain organoids. *Nature Biotechnology* 36: 432–441. <https://doi.org/10.1038/nbt.4127>.
27. Quadrato, G., T. Nguyen, E.Z. Macosko, J.L. Sherwood, S.M. Yang, D. Berger, N. Maria, J. Scholvin, M. Goldman, J. Kinney, E.S. Boyden, J. Lichtman, Z.M. Williams, S.A. McCarroll, and P. Arlotta. 2017. Cell diversity and network dynamics in photosensitive human brain organoids. *Nature* 545: 48–53. <https://doi.org/10.1038/nature22047>.
 28. Benito-Kwiecinski, S., and M.A. Lancaster. 2020. Brain Organoids: Human Neurodevelopment in a Dish. *Cold Spring Harbor Perspectives in Biology* 12: a035709. <https://doi.org/10.1101/cshperspect.a035709>.
 29. Bayne, T. 2013. Agency as a marker of consciousness. In *Decomposing the will*, ed. A. Clark, J. Kiverstein, and T. Vierkant, 160–180. Oxford University Press.
 30. Tononi, G. 2008. Consciousness as Integrated Information: A Provisional Manifesto. *The Biological Bulletin* 215: 216–242. <https://doi.org/10.2307/25470707>.
 31. Tononi, G., M. Boly, M. Massimini, and C. Koch. 2016. Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience* 17: 450–461. <https://doi.org/10.1038/nrn.2016.44>.
 32. Tononi, G., and C. Koch. 2015. Consciousness: Here, there and everywhere? *Phil. Trans. R. Soc. B* 370: 20140167. <https://doi.org/10.1098/rstb.2014.0167>.
 33. Sakaguchi, H., Y. Ozaki, T. Ashida, T. Matsubara, N. Oishi, S. Kihara, and J. Takahashi. 2019. Self-Organized Synchronous Calcium Transients in a Cultured Human Neural Network Derived from Cerebral Organoids. *Stem Cell Reports* 13: 458–473. <https://doi.org/10.1016/j.stemcr.2019.05.029>.
 34. Trujillo, C.A., R. Gao, P.D. Negraes, J. Gu, J. Buchanan, S. Preissl, A. Wang, W. Wu, G.G. Haddad, I.A. Chaim, A. Domissy, M. Vandenberghe, A. Devor, G.W. Yeo, B. Voytek, and A.R. Muotri. 2019. Complex Oscillatory Waves Emerging from Cortical Organoids Model Early Human Brain Network Development. *Cell Stem Cell* 25: 558–569.e7. <https://doi.org/10.1016/j.stem.2019.08.002>.
 35. Goff, P. 2017. The Case for Panpsychism. *Philosophy Now* 121: 6–8.
 36. Strawson, G. 2009. Realistic Monism: Why Physicalism Entails Panpsychism. *Journal of Consciousness Studies* 13. <https://doi.org/10.1093/acprof:oso/9780199267422.003.0003>
 37. Goff, P. 2019. *Galileo's Error: Foundations for a New Science of Consciousness*. Rider.
 38. Seager, W.E., ed. 2020. *The Routledge Handbook of Panpsychism*. New York: Routledge.
 39. Feinberg, Todd E., and Jon M. Mallatt. 2018. *Consciousness Demystified*. Cambridge, Massachusetts: The MIT Press.
 40. Ward, D. 2012. Enjoying the Spread: Conscious Externalism Reconsidered. *Mind* 121: 731–751. <https://doi.org/10.1093/mind/fzs095>.
 41. Dehaene, S. 2014. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking, New York, New York.
 42. Dretske, F. 1995. *Naturalizing the Mind*. MIT Press.
 43. Tye, M. 2000. *Consciousness, Color, and Content*. MIT Press.
 44. Brown, R., H. Lau, and J.E. LeDoux. 2019. Understanding the Higher-Order Approach to Consciousness. *Trends in Cognitive Sciences* 23: 754–768. <https://doi.org/10.1016/j.tics.2019.06.009>.
 45. Rosenthal, D. and J. Weisberg. 2008. Higher-order theories of consciousness. *Scholarpedia* 3:4407. <https://doi.org/10.4249/scholarpedia.4407>
 46. Carruthers, P. 2018. The problems of animal consciousness. *Proceedings and Addresses of the American Philosophical Association* 92: 179–205.
 47. Murray, S. 2020. A Case for Conservatism About Animal Consciousness. *Journal of Consciousness Studies* 27: 163–185.
 48. Michel, M. 2019. Consciousness Science Underdetermined Ergo, an Open Access. *Journal of Philosophy* 6. <https://doi.org/10.3998/ergo.12405314.0006.028>
 49. Bradshaw, R.H. 1998. Consciousness in non-human animals: Adopting the precautionary principle. *Journal of Consciousness Studies* 5: 108–114.
 50. Sebo, J. 2018. The Moral Problem of Other Minds. *The Harvard Review of Philosophy* 25: 51–70. <https://doi.org/10.5840/harvardreview20185913>.
 51. Klein, C. 2017. Precaution, proportionality and proper commitments. *Animal Sentience* 2. <https://doi.org/10.51291/2377-7478.1232>
 52. Żuradzki, T. 2021. Against the Precautionary Approach to Moral Status: The Case of Surrogates for Living Human Brains. *The American Journal of Bioethics* 21: 53–56. <https://doi.org/10.1080/15265161.2020.1845868>.
 53. Birch, J. 2017. Animal Sentience and the Precautionary Principle. *Animal Sentience* 16.
 54. Singer, P. 2009. *Animal Liberation: The Definitive Classic of the Animal Movement*, Reissued. Harper Perennial Modern Classics, New York.
 55. Lee, A.Y. 2018. Is Consciousness Intrinsically Valuable? *Philosophical Studies* 175: 1–17.
 56. Frith, C.D. 2019. The neural basis of consciousness. *Psychological Medicine* 1–13. <https://doi.org/10.1017/S0033291719002204>
 57. Koch, C., M. Massimini, M. Boly, and G. Tononi. 2016. Neural correlates of consciousness: Progress and problems. *Nature Reviews Neuroscience* 17: 307–321. <https://doi.org/10.1038/nrn.2016.22>.
 58. Stevens, S.S., ed. 2017. *Psychophysics: Introduction to Its Perceptual*. Neural and Social Prospects: Routledge.
 59. Sakaguchi, H., T. Kadoshima, M. Soen, N. Narii, Y. Ishida, M. Ohgushi, J. Takahashi, M. Eiraku, and Y. Sasai. 2015. Generation of functional hippocampal neurons from self-organizing human embryonic stem cell-derived dorsomedial telencephalic tissue. *Nature Communications* 6: 8896. <https://doi.org/10.1038/ncomm59896>.
 60. Qian, X., H.N. Nguyen, M.M. Song, C. Hadiono, S.C. Ogden, C. Hammack, B. Yao, G.R. Hamersky, F. Jakob, C. Zhong, K. Yoon, W. Jeang, L. Lin, Y. Li, J. Thakor, D.A. Berg, C. Zhang, E. Kang, M. Chickering, D. Nauen, C. Ho, Z. Wen, K.M. Christian, P. Shi, B.J. Maher, H. Wu, P. Jin, H. Tang, H. Song, and G. Ming. 2016. Brain-Region-specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure. *Cell* 165: 1238–1254. <https://doi.org/10.1016/j.cell.2016.04.032>.

61. Shulman, R.G., F. Hyder, and D.L. Rothman. 2009. Baseline brain energy supports the state of consciousness. *PNAS* 106: 11096–11101. <https://doi.org/10.1073/pnas.0903941106>.
62. Stender, J., R. Kupers, A. Rodell, A. Thibaut, C. Chatelle, M.-A. Bruno, M. Gejl, C. Bernard, R. Hustinx, S. Laureys, and A. Gjedde. 2015. Quantitative Rates of Brain Glucose Metabolism Distinguish Minimally Conscious from Vegetative State Patients. *Journal of Cerebral Blood Flow and Metabolism* 35: 58–65. <https://doi.org/10.1038/jcbfm.2014.169>.
63. Bayne, T., J. Hohwy, and A.M. Owen. 2016. Are There Levels of Consciousness? *Trends in Cognitive Sciences* 20: 405–413. <https://doi.org/10.1016/j.tics.2016.03.009>.
64. McFadden, J. 2002. Synchronous firing and its influence on the brain's electromagnetic field: Evidence for an electromagnetic field theory of consciousness. *Journal of Consciousness Studies* 9: 23–50.
65. James, W. 1890. *The Principles of Psychology*. Dover Publications.
66. Hohwy, J. 2013. *The Predictive Mind*. Oxford, New York: Oxford University Press.
67. Dalgleish, T. 2004. The emotional brain. *Nature Reviews Neuroscience* 5: 583–589. <https://doi.org/10.1038/nrn1432>.
68. Frewen, P., M.L. Schroeter, G. Riva, P. Cipresso, B. Fairfield, C. Padulo, A.H. Kemp, L. Palaniyappan, M. Owolabi, K. Kusi-Mensah, M. Polyakova, N. Fehertoi, W. D'Andrea, L. Lowe, and G. Northoff. 2020. Neuroimaging the consciousness of self: Review, and conceptual-methodological framework. *Neuroscience & Biobehavioral Reviews* 112: 164–212. <https://doi.org/10.1016/j.neubiorev.2020.01.023>.
69. Beauchamp, T.L., and D. DeGrazia. 2020. *Principles of Animal Research Ethics*. Oxford University Press.
70. Birch, J., and H. Browning. 2021. Neural Organoids and the Precautionary Principle. *The American Journal of Bioethics* 21: 56–58. <https://doi.org/10.1080/15265161.2020.1845858>.
71. Charo, R.A. 2015. Fetal Tissue Fallout. *New England Journal of Medicine* 373: 890–891. <https://doi.org/10.1056/NEJMp1510279>.
72. Vrselja, Z., S.G. Daniele, J. Silbereis, F. Talpo, Y.M. Morozov, A.M.M. Sousa, B.S. Tanaka, M. Skarica, M. Pletikos, N. Kaur, Z.W. Zhuang, Z. Liu, R. Alkawadri, A.J. Sinusas, S.R. Latham, S.G. Waxman, and N. Sestan. 2019. Restoration of brain circulation and cellular functions hours post-mortem. *Nature* 568: 336–343. <https://doi.org/10.1038/s41586-019-1099-1>.
73. Youngner, S., and I. Hyun. 2019. Pig experiment challenges assumptions around brain damage in people. *Nature* 568: 302–304. <https://doi.org/10.1038/d41586-019-01169-8>.
74. Block, N. 1995. On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18: 227–247. <https://doi.org/10.1017/S0140525X00038188>.
75. Nagel, T. 1974. What is It Like to Be a Bat? *Philosophical Review* 83: 435–450.
76. Roelofs, L. 2019. *Combining Minds: How to Think About Composite Subjectivity*. New York, USA: Oxford University Press.
77. Shepherd, J. 2021. The moral status of conscious subjects, in: Clarke, S. (Ed.), *Rethinking Moral Status*. OUP Oxford: 57–73.

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