Cambridge Quarterly of Healthcare Ethics (2025), 1–22
doi:10.1017/S0963180124000446

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Neural Voices of Patients with Severe Brain Injury?

Matthew Owen[1,2], Darren Hight[3] and Anthony G. Hudetz[2,4]

[1]Department of Philosophy, Yakima Valley College, Yakima, WA, USA; [2]Center for Consciousness Science, University of Michigan, Ann Arbor, MI, USA; [3]Department of Anaesthesiology and Pain Medicine, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland and [4]Department of Anesthesiology, University of Michigan, Ann Arbor, MI, USA
Corresponding author: Matthew Owen; Email: matthewkeithowen@gmail.com

Abstract
Studies have shown that some covertly conscious brain-injured patients, who are behaviorally unresponsive, can reply to simple questions via neuronal responses. Given the possibility of such neuronal responses, Andrew Peterson et al. have argued that there is warrant for some covertly conscious patients being included in low-stakes medical decisions using neuronal responses, which could protect and enhance their autonomy. The justification for giving credence to alleged neuronal responses must be analyzed from various perspectives, including neurology, bioethics, law, and as we suggest, philosophy of mind. In this article, we analyze the warrant for giving credence to neuronal responses from two different views in philosophy of mind. We consider how nonreductive physicalism's causal exclusion problem elicits doubt about interpreting neural activity as indicating a conscious response. By contrast, such an interpretation is supported by the mind-body powers model of neural correlates of consciousness inspired by hylomorphism.

Keywords: cognitive motor dissociation; covert consciousness; hylomorphism; mental causation; neuroethics; patient autonomy; physicalism

Introduction

A patient with cognitive motor dissociation, who is covertly conscious, presents a counterexample to standard diagnostic methods.[1,2] The standard approach for identifying a patient's level of consciousness in a medical context relies on behavioral correlates of consciousness.[3,4] To the extent patients appropriately respond to stimuli, questions, or commands, their level of consciousness can be reasonably inferred, based on the patient's behavioral responses. However, if a brain-injured patient is behaviorally unresponsive due to motor impairments, the standard method is challenged. This is because there are no behavioral correlates indicating the presence of consciousness, and the lack of behavioral indicators of consciousness does not necessarily entail unconsciousness. It is possible for a patient to be covertly conscious in the sense that they behaviorally appear unconscious while they are actually conscious.[5,6]

The possibility of covertly conscious patients has motivated the development of a neuronal response paradigm. This paradigm can allow behaviorally unresponsive patients to provide neuronal responses to commands and even reply to simple "yes/no" questions by activating particular neural activity that corresponds to specific mental activity.[7,8,9] As the neural activity is observable via neural imaging, neurologists can detect the corresponding mental activity of some behaviorally unresponsive but covertly conscious patients. In light of this possibility, Andrew Peterson et al.[10] argue that covertly conscious but neuronally responsive patients could be included in low-stakes decisions, or to express preferences, via neuronal responses to protect and enhance their autonomy. Peterson et al. briefly mention the relevance of philosophy of mind to this topic of neuroethical significance, which this article demonstrates.[11]

The validity of neural responses depends on a variety of factors, from technological issues to a patient's level of arousal and neurobiology, which must be considered to determine whether a purported response accurately reflects a patient's thoughts.[12] The particular factor we focus on here pertains to philosophy of mind, and specifically how the patient's conscious intention relates to the neural activity allegedly indicative of a response. This affects how the neural activity can be justifiably interpreted, as we will demonstrate.

As Molly Cairncross et al.[13] argue, when patients have communication impairments preserving their autonomy can involve alternative forms of response, but this requires that alternative responses are reliably produced and interpreted. The meaningfulness of neuronal responses, and whether neural activity can reasonably be interpreted as indicating a consciously intended response, depends on the neural activity resulting from mental conscious intentions. After explaining the neuronal response paradigm in the following section, we highlight the critical role of mental causation in the subsequent section. Then, a problem for nonreductive physicalism regarding mental causation is applied to the neuronal response paradigm. According to the version of nonreductive physicalism we are concerned with here, conscious intentions are distinct from their physical substrate, presumably neurons. Its problem pertaining to mental causation, known as the causal exclusion problem, fosters doubt about conscious intentions being causally responsible for neural activity, thus calling into question the legitimacy of interpreting neuronal activity as indicating a consciously intended response. By contrast, the mind-body powers model of neural correlates of consciousness (NCC), considered in the fifth section, warrants interpreting neuronal activity as causally explained by a conscious intention, thus justifying the neuronal response paradigm.

Before diving into the cerebral weeds of neurology and philosophy of mind, it may be useful to mention one way our abstract reflections here can matter. Joseph Fins[14] argues for a "mosaic decision-making" approach in clinical contexts that incorporates a team of decision-makers, consisting of individuals with differing degrees of authoritative input.[15] Such an approach could be used to gradually give a severely brain-injured patient more autonomy as their agency reemerges through the use of the neuronal response paradigm discussed below. The mosaic decision-making approach provides checks and balances through the input of multiple decision-makers. One mechanism for such checks and balances could be disagreements among decision-makers. However, for such disagreements to be constructive, it can often be necessary to know the contrasting reasons prompting the disagreement.

It is possible that individual decision-makers could view alleged neuronal responses from different perspectives in philosophy of mind and consequently give such more or less decision-making authority. A decision-maker who leans toward a view that entails epiphenomenalism, meaning that mental states are causally inert, might be more apprehensive about neuronal responses. By contrast, a decision-maker who holds a view that easily accommodates the possibility of neuronal responses might readily accept an alleged neuronal response and give it significant decision-making authority.

Although some decision-makers might be aware of their philosophical presuppositions and how they influence their evaluations of neuronal responses, most will likely be unaware. Our aim is to bring awareness to the way in which different views in philosophy of mind can influence how one views the merit of neuronal responses. Such awareness can make it easier to identify and then analyze the work that such views are doing in one's reasoning process as they consider how much authority to give neuronal responses in decision-making. This can aid individual decision-makers, but also hospital committees making procedural decisions and policymakers establishing laws. With that said, let us move into the cerebral weeds, beginning with neurology that will quickly take us into philosophy of mind.

Neuronal response paradigm

A study conducted by Caroline Schnakers and her colleagues found that 41% of 44 patients in their study initially diagnosed as being in a vegetative state (VS), which is now more respectfully called unresponsive wakefulness syndrome (UWS), were actually in a minimally conscious state.[16,17] More recently, a study led by Jiahui Pan[18] implemented a machine-learning algorithm in the assessment of patients with

disorders of consciousness; and Adrian Owen[19] notes in his scientific commentary on the study that 40% of the patients clinically diagnosed as VS/UWS were able to follow commands sufficiently enough to be reclassified as consciously aware. These patients were previously considered unconscious when they were actually capable of following commands, indicating they were conscious. What such studies reveal is that covert consciousness can go undetected in clinical contexts, but it can be more reliably detected using brain imaging technology sufficient to discern a patient's neuronal response.

The neuronal response paradigm is more apt to detect covert consciousness because it does not require the proper functioning of the patient's motor systems. The patient's sensory systems that allow the perception of a command must be functional. Yet, the patient does not need to perform a behavioral task involving the neuronal motor circuits that instigate muscle contractions driving bodily movements. This is because the commands are imaginative commands such as "imagine playing tennis" or "imagine you are moving through your apartment." Although the patient must have the motivation to cooperate with the clinician's or experimenter's commands, there is no need to move a muscle in order to obey them. However, one must activate neurons in the parts of the brain that correspond to mentally imagining doing the specific task. This requires being conscious and cognizant enough to perceive the command, understand it, and consciously intend to imagine doing something specific. And of critical relevance, which will become apparent in the following sections, it requires that the conscious intention to imagine doing the task is what activates the neuronal activity in the neuronal circuits that correspond to the conscious intention. This is vitally important to how Peterson et al.[20] propose to use the neuronal response paradigm to allow behaviorally unresponsive patients to communicate decisions.

The neuronal response paradigm uniquely allows some covertly conscious patients to provide "yes" or "no" answers to questions by activating particular neural activity that corresponds to specific mental activity.[21,22,23] The groundwork for the neuronal response paradigm was laid by Mélanie Boly et al.'s study on healthy subjects, which demonstrated that when they followed commands to carry out imaginative tasks, there was specific neural activity corresponding to their specific mental activity.[24] When the individuals were asked to imagine playing tennis, there was corresponding neural activity in the supplementary motor area (SMA) involved in planning movement; and when they were asked to imagine walking through their home, different coalitions of neurons that correspond to processing spatial coordinates and memory of locations were activated in the posterior parietal cortex, the parahippocampal gyrus, and the lateral premotor cortex.[25] Once the specific neural activity that corresponds to each mental activity was mapped, a team led by Adrian Owen and Steven Laureys demonstrated that a brain-injured patient who lacked motor movement yet retained consciousness could also respond to commands to carry out imaginative tasks.[26] As was done with healthy subjects, the researchers could give the commands and observe the corresponding neural activity one would expect if the patient imagined playing tennis following the command to do so, and then, at another time, imagined walking through their home in response to the command to do so.[27,28]

Such consistency is what makes it possible to implement the neuronal response paradigm to allow patients to provide yes/no answers to simple questions.[29,30] To answer "yes" to a question such as "is your name John?" a patient can be instructed to imagine navigating through her apartment, or alternatively to imagine playing tennis to answer "no." Peterson et al. provide several examples of patients who have communicated in this way.[31] It is, however, an uncommon occurrence since there are multiple reasons why a patient who is covertly conscious might nevertheless be unable to provide a neuronal response.[32,33,34,35] In one study, Zirui Huang et al.[36] tested the presence of tennis and navigation imagery to verbal commands in healthy volunteers anesthetized with propofol titrated to the point of behavioral unresponsiveness as a pharmacological model for disorders of consciousness. Out of the 29 patients tested, Huang and colleagues found only one patient who displayed convincing neuroimaging signs of mental imagery, which suggests that either covert consciousness during anesthesia is rare or the applied task-response approach for its detection is suboptimal. The latter could be due to decreased sensory processing in secondary auditory and higher order cerebral cortex due to a suppressive effect of the anesthetic,[37,38] a diminished desire to respond, or to something else while not excluding the possibility for perhaps partial preservation of internal awareness.

While acknowledging that the details of the neuronal response paradigm need refinement, the remainder of this article will presuppose that the technical problems will be resolved with time in clinical neurology. Our interest here pertains to what the expected and consistent neural activity following a command would imply, and whether it can be reasonably interpreted as a conscious intentional response. More specifically, we are interested in this topic in light of philosophy of mind, and the implications that views about the metaphysics of mind have for the neuronal response paradigm. The following section discusses the problems raised by physicalism for interpreting the expected neural activity following a command as a neuronal response. By contrast, the subsequent section presents a hylomorphic view that strongly implies that such expected neural activity should be understood as a neuronal, consciously intended, response.

Neural responses and mental causation

Peterson et al. propose that neuronal responses can be used to involve the covertly conscious patient in low-stakes decisions, thereby preserving some degree of autonomy.[39] It is important to understand that the response utilized is a particular type of response. It is not an unconscious reflexive response, but rather a consciously intended response, which is precisely why it is thought to indicate covert consciousness. As suggested by pioneering articles on the neuronal response paradigm, such as "When thoughts become action"[40] and "Willful Modulation of Brain Activity in Disorders of Consciousness,"[41] empirically detecting covert conscious intentions is possible when they willfully affect brain activity.

This is evident from descriptions of a paradigmatic case, involving Mr. R, that was initially reported by Davinia Fernández-Espejo and Adrian Owen[42] and recounted at the beginning of Peterson et al.[43] and Peterson et al.[44] In December of 1999, Mr. R endured a traumatic brain injury that left him behaviorally unresponsive. For over a decade, he was regularly diagnosed via clinical exams as VS/UWS. However, that would change, as recounted by Peterson and colleagues.

> In February 2012, Mr. R's parents enrolled him in a study at the University of Western Ontario to discover whether functional neuroimaging could detect awareness in persons with catastrophic brain injuries. While lying in the scanner he was instructed to imagine one of two activities—playing tennis or visiting the rooms of his home—for repeated 30-second intervals…Remarkably, Mr. R's brain activity matched that of fully conscious persons, suggesting that he was aware of his surroundings. Investigators then adapted this method to assess whether Mr. R could communicate. They asked a series of yes/no questions. To answer "yes," Mr. was asked to imagine playing tennis. To answer "no," he was asked to imagine visiting the rooms of his home. Mr. R correctly answered questions regarding his name, the name of his support worker (whom he had met following his injury), the date and his location…He was even asked twice whether he was in physical pain, to which, each time, he replied "no." These results were revolutionary. They demonstrated that, despite being consistently diagnosed as being in the vegetative state for 12 years, Mr. R was conscious.[45]

After discussing this case, Peterson et al. describe why it was fitting to conclude that Mr. R was covertly conscious: "Neuroimaging demonstrated that despite satisfying the behavioral criteria of the VS, Mr. R could willfully modulate his brain activity to command, indicating he was aware."[46] In other words, and perhaps more precisely, Mr. R could consciously intend to imagine doing one activity or another, and his conscious intention brought about particular neural activity.

The observed neural activity in the neuronal response paradigm can be understood as indicating a genuine response because it is the result of the patient's conscious intention—or, to use the terminology of Peterson et al., because the neural activity is willfully modulated. This implies that there is an explanatory relation in a particular direction: the conscious intention explains the neural activity. When a patient is providing a meaningful contemplated (as opposed to automated or reflexive) neuronal response, the expected neural activity can reasonably be interpreted as a genuine response to

a question only if the patient's conscious intention to respond in a particular way is what explains the neural activity.

It is worth asking what type of relation the explanatory relation from the conscious intention to the neural activity is. When considering this issue, cognitive neuroscientists and philosophers would do well to take into account Fernández-Espejo and Owen's description of neuronal response tasks reported by Tristan Bekinschtiein et al.,[47] Martin Monti et al.,[48] and Monti et al.[49] Fernández-Espejo and Owen write:

> Importantly, the neural responses that characterize all of these tasks are not automatically produced by the eliciting stimulus but rather *depend on* the will or the *intention* of the participant to *generate and sustain* a response to the given instruction. Such (neural) behavior provides a proxy for a motor action and is therefore an appropriate vehicle for reportable awareness.[50] Indeed, given the complexity of the tasks used and the specificity of the responses measured, one can draw far more elaborate conclusions about the mental state of these patients than the fact that they are merely 'conscious.' For example, at the very least, sustained attention (required to maintain focus through each task), language comprehension (required to understand the task instructions), response selection (required to switch between alternative tasks or conditions) and working memory (required to remember which task to perform when instructed) must all be substantially preserved. These are all aspects of *'top-down' cognitive control* that are typically associated with normal levels of conscious awareness.[51]

It is difficult to avoid interpreting this description of neuronal responses as a description of top-down mental causation, in which a conscious intention causes neural activity. Such causation, however, has been notoriously difficult to make sense of in contemporary philosophy of mind, as it has been dominated by physicalism.

Nonreductive physicalism's causal exclusion problem

The most well-known problem of mental causation was a subject of debate between René Descartes and Princess Elisabeth of Bohemia, who asked Descartes to explain how a nonphysical mind could causally interact with a physical body. Physicalist proponent, Jaegwon Kim, developed the issue the Princess raised into what he called the causal pairing problem.[52] However, the most notorious contemporary problem of mental causation—the causal exclusion problem—was referred to by Kim as "Descartes's revenge" because, rather than raising trouble for Descartes's substance dualism, it strikes "at the very heart of physicalism."[53] In addition to the hard problem of consciousness, physicalism's own difficulties with mental causation ignited (ongoing) reconsiderations of non-physicalist, dualist views.[54,55,56,57,58,59,60,61] Yet, many philosophers still maintain physicalism, holding to a nonreductive physicalism that admits the existence of irreducible conscious states that supervene on physical states.

As we will see, the causal exclusion problem threatens to exclude irreducible consciousness from producing physical effects. This could have profound implications for the neuronal response paradigm if conscious intentions are irreducible to their physical substrate. In essence, the conscious intentions of covertly conscious patients would be excluded from being the cause of neural activity that is interpreted as the result of a conscious intention to respond. If the conscious intention of the patient is not the cause of the neural activity, it would be difficult to see why the neural activity should be understood as indicating the patient's intentional response.

Before explicating the causal exclusion problem and its implications, it will be helpful to entertain a reason to think that the conscious intentions of covertly conscious patients thought to explain neural activity are not reducible to (i.e., identical to) the neuronal activity. Let us consider an imaginary patient, named Connie, who is being studied by a team of neuroscientists interested in the extent to which covertly conscious patients can answer philosophical questions. The researchers have asked Connie to answer the question "Do you think you have direct, first-person, private epistemic access to your own conscious mental state?" Since Connie is unable to give a behavioral response, she must provide a

neuronal response to the philosophical question. She has been instructed to imagine playing tennis if her answer is "yes" and to imagine navigating through her home if her answer is "no." The team of neuroscientists observing her brain is therefore expecting to see robust neural activity instigated in the SMA associated with imagining playing tennis or robust neural activity in regions that correspond to real or imagined spatial navigation, namely the parahippocampal cortices, the lateral premotor cortices, and the posterior parietal lobe.[62]

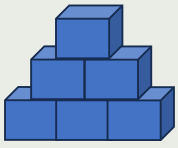
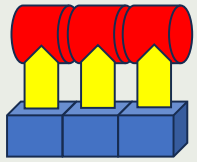
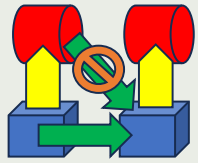
Just as Connie consciously decides to answer the question by consciously imagining one of the tasks, the neuroimaging being used to observe the neural activity in Connie's brain stops working. The neuroscientists are consequently unable to observe any neural activity in Connie's brain at the time she is answering the question. At that point, does anyone know what Connie's answer is? Surely, Connie knows, but the neuroscientists do not. Her conscious intention to answer "yes" or "no" by consciously intending to imagine one of the tasks is directly and privately accessible to her from her first-person perspective. (It is this type of epistemic access to consciousness that makes it possible for consciousness to be covert in the absence of behavioral responses that indirectly make it known to outside observers.)

In contrast to her conscious mental intention, the neural activity in Connie's brain cannot be known by her in the same way. Apart from neuroscientists using neuroimaging technology to observe the specific neural activity in her brain, she would never know what neural activity corresponds to her conscious intentions. This reveals a fundamental difference between her conscious mental intention and the corresponding neural activity. The conscious intention is directly and privately knowable to the subject from her first-person perspective, but the neural activity is not. Therefore, it is difficult to avoid the conclusion that the conscious intention is *not* the same as the neural activity. Because if they were the same, then what is true of the conscious intention would likewise be true of the neural activity, which is not the case. Connie has direct, first-person, epistemic access to her conscious intention, but she lacks such epistemic access to the neural activity; therefore, her conscious intention and the neural activity cannot be identical.

However, if the conscious intention is not reducible to its physical substrate, the causal exclusion problem is prompted from a broadly physicalist framework that many philosophers have found attractive.[63] To see why, it will be important to understand physicalism's basic tenets presented in Table 1.

There are various views that can be classified as physicalist views. Yet, there are some essential principles that arguably any version of physicalism must maintain in order to genuinely count as a version of physicalism.[64] Most fundamentally is that physics is fundamental and all explanations of everything in the physical world can ultimately be traced back to fundamental physical explanations that appeal to fundamental physical entities.[65]

Table 1. Physicalism's essential tenets in logical order from left to right

Physics fundamental	Mind-body supervenience	Physical causal closure
Whether physical entities are the only entities or there are also nonphysical mental entities, the physical is fundamental	Whether the mental is identical to or distinct from its physical substrate, it is determined by its physical substrate	Whether or not there are nonphysical mental entities, they cannot cause physical effects, which have sufficient physical causes
		
World's fundamental building blocks are physical entities (blue), even if there are nonphysical mental entities	Even if mental entities (red) are distinct from physical entities (blue), they supervene (yellow) on physical entities	Nonphysical mental entities (red) cannot cause (green) physical entities (blue) which have physical causes

Mind-body supervenience flows out of the fundamentality of physics. Kim describes supervenience in the following technical way while letting *t* stand for a particular time: "Mental properties supervene on physical properties in the sense that if something instantiates any mental property *M* at *t*, there is a physical base property *P* such that the thing has *P* at *t*, and necessarily anything with *P* at a time has *M* at that time."[66] More simply, supervenience says every mental property (or state) is fixed by a physical property that is its base. Supervenience is thought by many philosophers to entail that if an entity has mental properties, such as conscious states, they are determined by and depend on their physical base properties.[67] Such dependence on the physical by the mental is an outworking of the ontological primacy and priority of the physical.[68]

The causal closure of the physical domain is a third principle often considered essential to physicalism, as it safeguards the fundamentality of the physical.[69] If there could be nonphysical causes of physical effects, then it would not be the case that everything in the physical world has a physical explanation. However, according to causal closure, there are no nonphysical causes of physical effects, nor causal chains that include nonphysical causes of physical effects; because all physical effects have purely physical causes.[70] The causal closure principle seems to exclude the possibility of any mental state that is not reducible to its physical substrate from causing physical effects. After all, it says only physical entities can cause physical effects. So insofar as a conscious intention is not ontologically identical to something physical, most plausibly its neural correlates, it cannot causally produce physical effects.

To apply this to the neuronal response paradigm, let us suppose Connie answered the philosophical question above with a firm "yes" by imagining playing tennis, thus instigating neural activity in the SMA. So, Connie's conscious intention to imagine playing tennis (which we will label CT) caused specific neural activity in the SMA (which we will label NS). That is, CT caused NS. This is why NS can reasonably be interpreted as a conscious response; for its explanation is that CT caused it. But, according to causal closure, CT could be the cause of NS only if CT is reducible to something physical, the most plausible candidate of which would be the neural activity that correlates with CT, which is NS.[71,72] This seems to result in an odd potential dilemma from the standpoint of physicalism. If CT is identical to NS, then NS is the cause of NS, and it is difficult to see how neural activity that is the cause of itself could be indicative of a conscious intention. On the other hand, if CT is not identical to something physical, then it could not be the cause of NS because causal closure would exclude CT from having any causal effect in the physical world. Above we entertained one reason to think CT is not identical to NS, and if such a conclusion is justified then causal closure would seem to imply that CT is causally impotent and excluded from being the cause of NS.

Some philosophers have thought that mind-body supervenience could save mental causation by allowing irreducible mental states to cause physical effects via the causal efficacy of their underlying physical base property. However, Kim's "supervenience argument" makes a convincing case that supervenience actually reinforces the causal exclusion problem.[73] The argument, formally presented in Box 1, purportedly demonstrates how ideas at the core of physicalism, such as mind-body supervenience, exclude any irreducible mental state from causing physical effects. The upshot is that irreducible mental states are epiphenomenal (i.e., causally impotent). Although Kim himself reluctantly accepted epiphenomenalism regarding irreducible conscious states, specifically qualia,[74] he acknowledged that "…the idea that our thoughts, wants, and intentions might lack causal efficacy of any kind is deeply troubling, going as it does against everything we believe about ourselves as agents and cognizers."[75] The implications are not safely sequestered to philosophical debates, but logically spill over to neurology.

The causal exclusion problem presents a considerable problem for the neuronal response paradigm. Because if the patient's conscious intention to answer a question by imagining a task is irreducible to its physical substrate, then it is excluded from causing any physical effect. This would exclude the patient's conscious intention from being the cause of the neural activity that is supposed to be indicative of the conscious intention. And if the patient's conscious intention is not and cannot be the cause of the neural activity, then why does the neural activity indicate the conscious intention? It is difficult to see why it would be indicative of the conscious intention. That is the causal exclusion problem for the neuronal

Box 1. The supervenience argument for the causal exclusion problem

The supervenience argument for the causal exclusion problem can be formulated in various ways. Below is one way to formally present the argument. *MBS* represents the physicalist principle of mind-body supervenience, while *CCP* represents the physicalist principle of the causal closure of the physical domain. *M* represents an irreducible mental state, and *PB* represents its subvenient physical base property. *PE* represents a physical effect that has been caused.

1. According to *CCP*, every physical effect has a sufficient physical cause and it has no cause that is not identical to a physical entity.
2. According to *MBS*, every mental property has a physical property necessarily sufficient for its existence that determines it.
3. Given *MBS* and *CCP*, any mental state *M* necessarily supervenes on its subvenient physical base property *PB*; and given that *M* is not identical to a physical entity, *M* cannot cause a physical effect *PE*.
4. Therefore, for *M* to cause *PE* it must do so via its subvenient physical base property *PB* sufficiently causing *PE*. However, if *M* supervenes on *PB* as *PB* sufficiently causes *PE* then *PB* (not *M*) causes *PE*. Thus, *PB* preempts *M* as the cause of *PE*, assuming that effects are not causally overdetermined.
5. Therefore, *M* is excluded from causing *PE*.

response paradigm, formally presented in Box 2. In summary, ideas central to physicalism render the patient's conscious intention epiphenomenal or causally impotent, so it cannot be the cause of the neural activity thought to indicate a response. And if the patient's conscious intention is not the cause of the neural activity, it is difficult to see how it indicates a response.

There are several ways physicalists might try to deal with this causal exclusion problem for the neuronal response paradigm. Some might attempt to counter the arguments for the causal exclusion problem while still holding to nonreductive physicalism.[76,77] Others might argue that a patient's conscious intention is reducible to something physical that is, therefore, causally efficacious. Another option is to simply accept the epiphenomenalism, despite it entailing that covert conscious intentions cannot be discerned through neuronal responses. Either way, nonreductive physicalism's causal exclusion problem does cast doubt on the idea that neural activity can be justifiably interpreted as indicating a consciously intended response.

Box 2. The causal exclusion problem applied to a neuronal response

The argument for the causal exclusion problem presented in Box 1 has been refurbished below to illustrate the problem for the neuronal response paradigm. *MBS* represents the physicalist principle of mind-body supervenience, while *CCP* represents the physicalist principle of the causal closure of the physical domain. *CI-Tennis* represents the irreducible mental property of the patient's conscious intention to answer "yes" by imagining playing tennis, and *PB* represents its subvenient physical base property. *SMA* represents increased and robust neural activity in the SMA that correlates with *CI-Tennis*.

1. If, and only if, *CI-Tennis* causes *SMA*, then *SMA* indicates a response.
2. Given *MBS* and *CCP*, any mental state, including *CI-Tennis*, necessarily supervenes on its subvenient physical base property *PB*; and given that *CI-Tennis* is not identical to a physical entity, *CI-Tennis* cannot cause a physical effect, including *SMA*.
3. Therefore, for *CI-Tennis* to cause *SMA* it must do so via its subvenient physical base property *PB* sufficiently causing *SMA*. However, if *CI-Tennis* supervenes on *PB* as *PB* sufficiently causes *SMA* then *PB* (not *CI-Tennis*) causes *SMA*. Thus, *PB* preempts *CI-Tennis* as the cause of *SMA*, assuming that effects are not causally overdetermined.
4. Therefore, *CI-Tennis* is excluded from causing *SMA*.
5. Therefore, given premises 1 and 4, *SMA* does not indicate a response.

However, it could be argued that the empirical validation and reliability of the neuronal response paradigm, given its low percentage of false positives,[78] provides empirical support for such an interpretation regardless of what any philosophical view implies. Moreover, one might argue that the empirical support for the neuronal response paradigm is empirical evidence against any view that implies neural responses are unfeasible and empirical evidence for views that imply neural responses are feasible. In any event, we will next consider a view that implies the latter.

Hylomorphism and neural voices

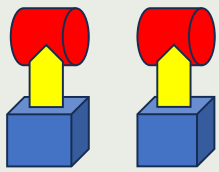
In contemporary philosophy, there is an ongoing Aristotelian revival reigniting interest in hylomorphism.[79,80,81,82,83,84,85,86,87,88,89,90] Given this modern revival, along with hylomorphism's rich historical pedigree, it is worth considering how hylomorphism applies to the neuronal response paradigm. In this section, we will demonstrate how a model inspired by hylomorphism supports Peterson et al.'s[91] interpretation of neural activity as indicating a consciously intended response. However, before presenting the model, it will be helpful to provide a brief introduction to its hylomorphic roots.[92]

Hylomorphism is a view about the constitution of ordinary objects that Aristotle proposed in *Metaphysics* and applied to human persons in *De Anima*. The word "hylomorphism" is a compound of the Greek "hyle," which means matter, and "morphe," which means form. As the compound of "hyle" and "morphe" together constitute the term "hylomorphism," the view says that form and matter together constitute real material objects. In *Metaphysics* Z 17, Aristotle asks why some matter composes a house while other matter composes a body. Aristotle's answer to his question is that each material (or physical) entity is composed of its matter plus a form in virtue of which the matter is unified into one thing of a particular kind.[93]

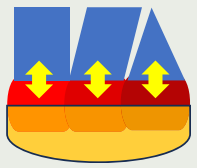
In *De Anima* Aristotle argues that the form of a natural living body such as Socrates's body is its soul.[94] The medieval philosopher Thomas Aquinas followed Aristotle and likewise thought that the form that unifies a human body is the human soul.[95] Hylomorphists make a distinction between accidental forms that externally unify material objects versus substantial forms that internally unify material substances. Like all natural living bodies, on this Aristotelian-Thomistic framework, the human body has a soul that is its substantial form, internally grounding the body's unity and essence.[96]

The soul that is the form of a biological substance is the source of its capacities/powers—such as self-nutrition, locomotion, sensation, or rationality—that determine or characterize its kind.[97] On nonreductive physicalism, the physical substrate of consciousness is ontologically fundamental, and consciousness supervenes upon it. On hylomorphism, there is a grounding relation that goes in the opposite direction (see Table 2). The physical parts of a living body exist, or develop, so the capacities of its substantial form can be manifested.[98,99] The hylomorphic explanation for why the neural mechanisms of consciousness

Table 2. Contrasting ontological foundations of physicalism and hylomorphism



Physicalism says the physical constituents (blue) of all material substances are ontologically fundamental. So, if humans have irreducible mental properties (red), they supervene (yellow) on the physical entities (blue). Physicalism does not specify what unifies the mental properties nor their physical base into one substance.



Hylomorphism says each material substance is unified by an ontologically fundamental substantial form (orange). A human is a material substance unified by a substantial form with distinct mental powers (shades of red) that determine (yellow) the composition of the substance's physical parts (blue), which they rely on (yellow) for their manifestation.

exist in a human's brain (or an animal's brain) is that the substantial form of the human body has the capacity to be conscious, and the form utilizes the neural mechanisms to manifest the capacity.

Like animals, humans have the powers of self-nutrition and sensation, but also the higher capacity of rational thought. Hence, the substantial form of a human body is referred to as a *rational soul.* The rational intention to do something or to deliberate how to make known "interior speech, from which comes external speech" are actions of the rational human soul's intellect, according to Aquinas.[100] The human rational soul, like all substantial forms of biological bodies, exercises its capacities via the body it unifies. When dealing with a question about whether a human rational soul ought to be united to a body such as the human body, Aquinas clarifies that the human soul is united to the human body specifically so that it can manifest its rational capacity of understanding.[101] "And consequently," Aquinas argued, "it is necessary that the body which is united to a rational soul be the kind of body that is best fitted for being of service to the soul and whatever understanding demands."[102]

In the same context, he claims that human rational understanding relies on information gained from sensing the world through bodily organs.[103] "Consequently," he infers, "the body to which a rational soul is united must be arranged in the best possible way for the operation of sensing."[104] "Consequently," he further concludes, "because a well ordered brain is necessary for the effective condition of the internal sense powers, for example, the imagination, the memory and the cogitative power, a human being was made in such a way that in proportion to his size he has a brain that is larger than that of any other animal."[105] The cogitative power to which Aquinas here refers is the human capacity to form rational intentions after comparing options.[106] Aquinas clarifies that the cogitative power "…is also called *particular reason* (to which physicians have assigned a definite organ, the middle part of the head), because it compares individual intentions just as intellective reason compares universal intentions."[107]

Centuries before the advent of neurobiology, Aquinas embraced the textbook idea in cognitive neuroscience that there is a correlation between structure and function, but he thought about it at the level of organs.[108,109] "For just as the whole organic body is constituted in order to assist those operations of the soul which are carried out through it, so also a single organ is so constituted that it may assist in a particular operation."[110] Aquinas's claim that the cogitative power to form an intention after comparing options relies on a "well ordered brain"[111] is pertinent to the neuronal response paradigm in which intentional responses are evidenced by activity in specific brain structures. However, to become directly relevant, Aquinas's line of reasoning needs to be applied to the cellular level of neuronal responses. The mind-body powers model of NCC brings Aquinas's hylomorphic reasoning regarding organs down to the cellular and subcellular levels of neurobiology.[112,113,114]

Mind-body powers model of NCC

The mind-body powers model of NCC does not make empirical predictions about the physical characteristics of neural correlates. Rather, its purpose is to provide a metaphysical explanation of why NCC exist.[115] NCC are the minimal neuronal mechanisms that are physically sufficient for consciousness.[116,117,118] Neurobiologists make a distinction between the *full NCC*, which corresponds to being conscious irrespective of the contents of one's conscious experience, and *content specific NCC* that correspond to specific conscious states with particular content, such as seeing a face.[119] It has been previously argued that the mind-body powers model provides philosophical justification for a passive neural paradigm for detecting covert consciousness via the full NCC, which requires no response from the patient.[120,121] Yet, the model also applies to the neuronal response paradigm and provides philosophical justification for interpreting specific neural activity as a patient's response, as we will demonstrate in due course.

While the mind-body powers model is built on the ontological foundation of hylomorphism, and the mind-body dependence evident in Aristotle and Aquinas's thought, it also appeals to Aristotle's ontology of causal powers. Nature is replete, according to Aristotle, with interdependent causal powers, both active powers to produce change and passive powers to undergo change.[122] The sun has the active power to heat a pool of water, which has the passive power to be heated. The sperm has the active power to fertilize the egg,

which has the passive power to be fertilized. The presynaptic neuron has the active power to send an electrochemical pulse to postsynaptic neurons that have the passive power to receive it. These are just several examples of natural causal powers that highlight the need for both active and passive powers to be co-activated, or manifested, in concert. The sun's active power to heat the water and the water's passive power to be heated must be co-manifested for either one to be manifested. They are "mutual partner powers," which due to their mutual dependence for activation realize their natures in activities that are "co-determined, co-varying, and co-extensive in time."[123]

The distinction between active and passive powers easily becomes complex. Imagine, for example, a fire manifesting its active power to heat a metal pot, which simultaneously manifests not only its passive power to be heated but also its active power to heat the water within it, which manifests its passive power to be heated while also acting on the noodles within it. Despite the complexity, there is nevertheless active and passive powers being co-manifested as interdependent partner powers. The mind-body powers model of NCC applies the Aristotelian concept of interdependent active and passive powers to conscious biological organisms with sensory or rational substantial forms. Such forms, according to the model, have active and passive mental powers that naturally depend on the co-manifestation of active and passive partner powers in the biological bodies they enform. Consciousness is considered a mental power of sensory and rational substantial forms that naturally co-manifests with powers of the bodies that the substantial forms enform. Although not reduced to its physical substrate, consciousness relies on it. Consciousness generally and each specific conscious state relies on the co-manifestation of particular powers of specific physical parts in the biological organism the substantial form enforms. These powers are called body/bodily powers, and their manifestations require that the biological parts that manifest them have sufficient physical properties regarding their structures and constitutions.

Every conscious state, on this framework, is the manifestation of a mental power that naturally co-manifests with a corresponding bodily power, or bodily powers, requiring sufficient physical structures and constitutions to be manifested. Consequently, each conscious state naturally corresponds to physical properties—such as physical structures, functions, or states of composition—of its substrate manifesting its partner bodily power(s). The model does not say what those physical properties of the substrate are, but rather provides philosophical warrant for the neurobiological search for the physical correlates of consciousness, such as neural correlates at the cellular level.

Let us consider an example involving a content-specific NCC, which corresponds to a particular conscious state with specific content.[124] As portrayed in Diagram 1, the conscious state of seeing a face is the manifestation of a human person's mental power that will correspond to the co-manifestations of its corresponding partner bodily powers. These bodily powers require sufficient physical properties for their manifestation that specific organs such as the eyes and the brain have, and neuroscience is revealing the neuronal structures and functions manifesting the corresponding bodily powers in brain structures, such as the optic chiasm, the optic tract, the visual cortex, and more specifically, the fusiform face area.[125,126,127]

The conscious state of seeing a face is just one of countless conscious states someone could be in. Regardless of the conscious state—whether it is the conscious recall of an autobiographical episodic memory, consciously considering logical entailments, or consciously making a decision—it will correspond to particular bodily powers manifested by particular neural mechanism, according to the model.[128] This is because each conscious state is the manifestation of a mental power that relies on the co-manifestation of a bodily power(s).

On the mind-body relational level, sometimes the bodily powers will be active, activating the co-manifestation of the passive conscious state, as is the case in our first example (see Diagram 1). However, the mental power can also be the active power, instigating the manifestation of corresponding bodily power(s). For instance, if someone is in the conscious state of intending to sign a check, the mental power will be active, instigating the corresponding co-manifestation of bodily powers in the premotor cortex,[129] which in turn trigger the manifestation of bodily powers in the central nervous system, and peripheral nervous system (see Diagram 2). Since the mental power is active, one might infer that the bodily powers are passive, according to the model. This is correct insofar as the co-manifestation of the bodily powers are triggered by the active mental power, on the level of mind-body causation. On the level of neuronal causation, however, the complex neural processes in the premotor cortex and beyond will

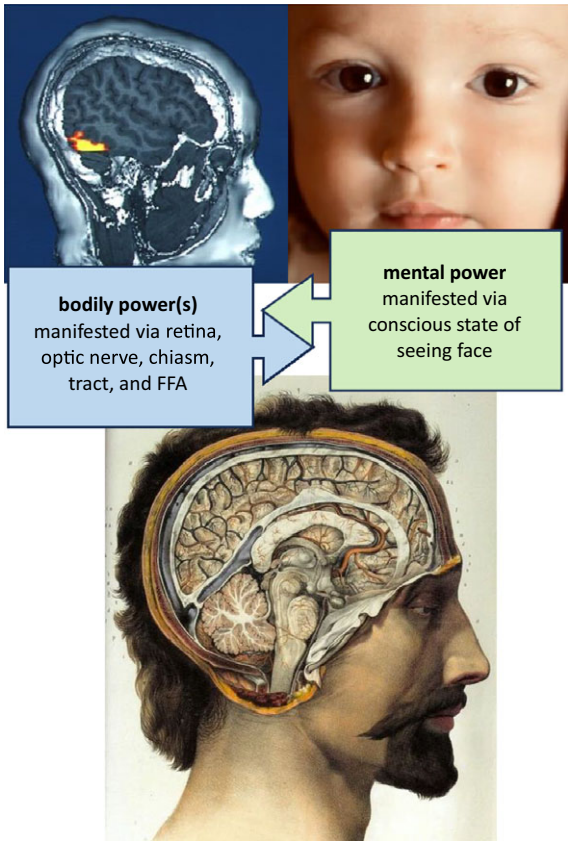


Diagram 1. *Mind-body powers and neuronal correlates of consciously seeing a face*. A human person with the mind-body power to consciously see a face is depicted. The person's bodily powers (blue) to receive light signals, translate them into electrical signals, and transmit the signals are manifested via the retina, optic nerve, optic chiasm, and optic tract as the signal is relayed to the fusiform face area in the visual cortex, which is activated in the image on the left. The bodily powers activate the natural co-manifestation of the person's mental power (green) to consciously experience seeing a face, which is manifested via the person's conscious state of seeing a face depicted on the right. The person's bodily and mental powers are mutual partner powers that naturally co-manifest as constituents of the mind-body power to consciously see a face. Therefore, the conscious state and neuronal correlates manifesting the powers consistently correspond. Top left image: Modified computer-enhanced fMRI scan of a person's brain when asked to look at a face. Source: National Institute of Health. Public domain via Wikimedia commons: https://commons.wikimedia.org/wiki/File:Fusiform_face_area_face_recognition.jpg. Top right image: By Matthew Owen. Bottom image: By J.M. Bourgery (1831–1854) via Erald Mecani, Wikimedia Commons: https://commons.wikimedia.org/wiki/File:Human_brain.jpg (License: https://creativecommons.org/licenses/by-sa/3.0/deed.en).

inevitably involve the manifestation of a complexity of active and passive powers, as signals are sent via the central and peripheral nervous systems to muscles in the arm and hand (see Diagram 4).

What matters most for the neuronal response paradigm is the relation between an active mental power manifested via a conscious state and the bodily powers co-manifested via neuronal activity in the central nervous system, and specifically, particular brain areas. The mind-body powers model was formulated for the purpose of providing a metaphysical explanation of why there are neuronal correlates of consciousness. It was not designed to provide justification for interpreting specific neural activity as indicative of a conscious response from behaviorally unresponsive patients. Nonetheless, it does just that, as we will discuss next.

Interpreting neural activity as a conscious response

Instead of casting doubt about whether specific neural activity can indicate a covert conscious intentional thought, the mind-body model inspired by hylomorphism does the opposite. Of surprising

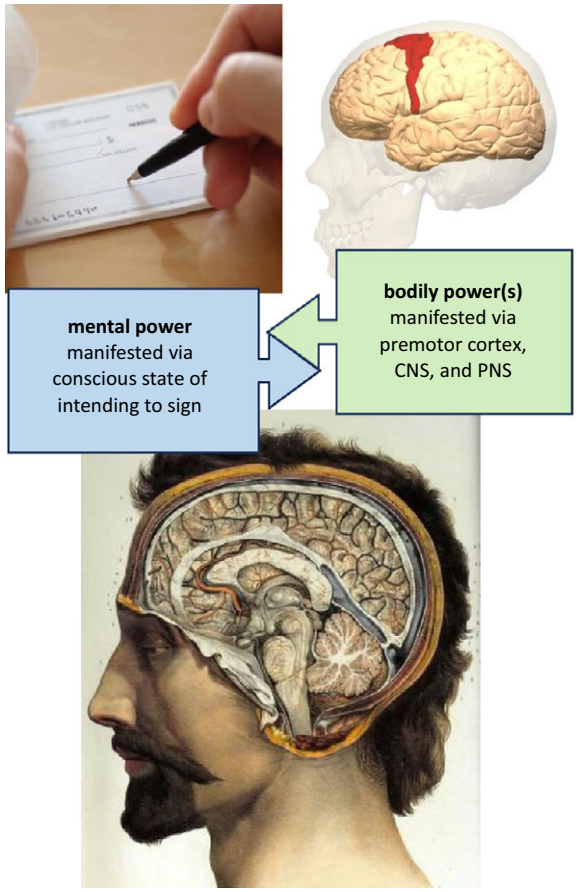


Diagram 2. *Mind-body powers and neuronal correlates of consciously intending to sign a check*. A human person with the mind-body power to consciously intend to sign a check is depicted. The person's mental power (blue) to consciously intend to sign a check is manifested via the person's conscious state of intending to sign a check, represented by the picture on the left. The manifestation of the mental power activates the natural co-manifestation of the bodily powers (green) to signal hand muscles that are manifested via synaptic transmission in the premotor cortex, through the central nervous system and peripheral nervous system, partly represented by the picture on the right. The person's mental and bodily powers are mutual partner powers that naturally co-manifest as constituents of the mind-body power to consciously intend to sign a check. Therefore, the conscious state and neuronal correlates manifesting the powers consistently correspond. Top right image: Brodmann area 6, premotor cortex. Source: Anatomography via Wikimedia Commons: https://commons.wikimedia.org/wiki/File:Brodmann_area_6_lateral.jpg (License: https://creativecommons.org/licenses/by-sa/2.1/jp/deed.en). Top left image: IStock by Getty Images (License: https://www.istockphoto.com/legal/license-agreement). Bottom image: By J.M. Bourgery (1831–1854) via Erald Mecani, Wikimedia Commons: https://commons.wikimedia.org/wiki/File:Human_brain.jpg (License: https://creativecommons.org/licenses/by-sa/3.0/deed.en).

relevance, Aquinas once consider how angels could have third-person epistemic access to "secret thoughts" of humans. This might sound like a strange question to modern ears, but the way he dealt with the question is similar to asking how outside observers such as physicians could discern conscious intentional thoughts of a behaviorally unresponsive patient. Aquinas argued that a secret thought could be known by its effects, not just bodily actions like speech, but also subtle physiological effects. "In this way it [i.e., the thought] can be known not only by an angel, but also by man; and with so much the greater subtlety according as the effect is the more hidden," wrote Aquinas.[130] "For thought," he reasoned, "is sometimes discovered not merely by outward act, but also by change of countenance; and doctors can tell some passions of the soul by the mere pulse."[131]

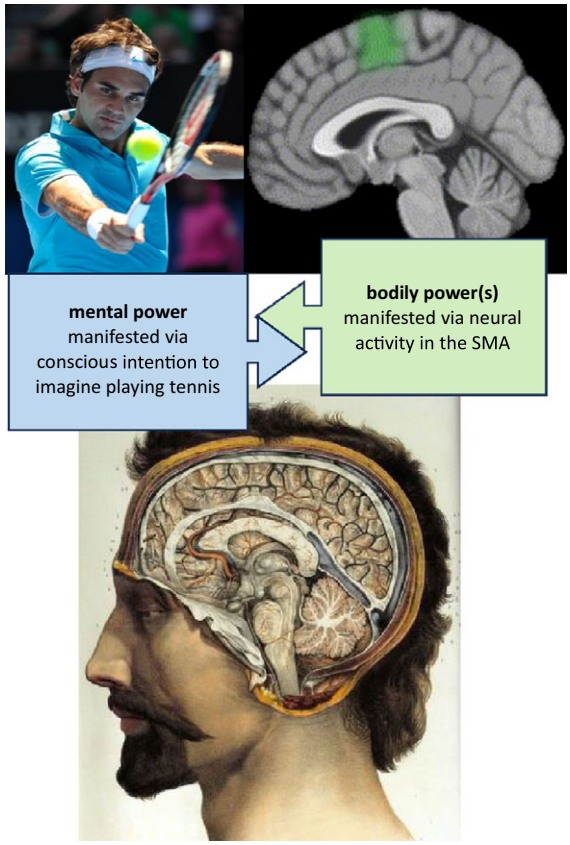


Diagram 3. *Mind-body powers and neuronal correlates of consciously imagining playing tennis*. A human person with the mind-body power to consciously imagine playing tennis is depicted. The person's mental power (blue) to consciously imagine playing tennis is manifested via the conscious state of imagining playing tennis, depicted by the image on the left. The manifestation of the mental power activates the natural co-manifestation of the bodily powers (green) to send a signal to hand muscles via the supplementary motor area, depicted by the image on the right. The person's mental and bodily powers are mutual partner powers that naturally co-manifest as constituents of the mind-body power to consciously imagine playing tennis. Therefore, the conscious state and neuronal correlates manifesting the powers consistently correspond. As a result, when the bodily power is manifested via the neuronal activity that indicates the corresponding mental power is simultaneously manifested via the conscious intention. Consequently, it is most reasonable to infer the presence of the conscious intention given the neural activity. And if the person was commanded to consciously imagine playing tennis to answer yes, then the neural activity is reasonably seen as indicating a conscious intention to imagine playing tennis in order to answer yes. Top right image: Modified version of Supplementary Motor Cortex Sagittal Sections by Daniel Sabinasz via Wikimedia Commons: https://commons.wikimedia.org/wiki/File:Supplementary_motor_cortex_sagittal_sections.gif (License: https://creativecommons.org/licenses/by-sa/4.0/deed.en). Top left image: Roger Federer at the 2010 Australian Open by Esther Lim via Wikimedia Commons: https://commons.wikimedia.org/wiki/File:R_federer.jpg (License: https://creativecommons.org/licenses/by-sa/2.0/deed.en). Bottom image: By J.M. Bourgery (1831-1854) via Erald Mecani, Wikimedia Commons: https://commons.wikimedia.org/wiki/File:Human_brain.jpg (License: https://creativecommons.org/licenses/by-sa/3.0/deed.en).

Aquinas reasoned that thought can be known not only by bodily behavioral acts but also subtle effects in the body, which the thought causally explains. This fits well with Aquinas's view that the rational soul, which is the substantial form of the body, can rule it very strictly:

> For a power is called despotic whereby a man rules his slaves, who have not the right to resist in any way the orders of the one that commands them…And so, the soul is said to rule the body by a despotic power, because the members of the body cannot in any way resist the sway of the soul, but at the soul's command both hand and foot, and whatever member is naturally moved by voluntary movement, are moved at once.[132]

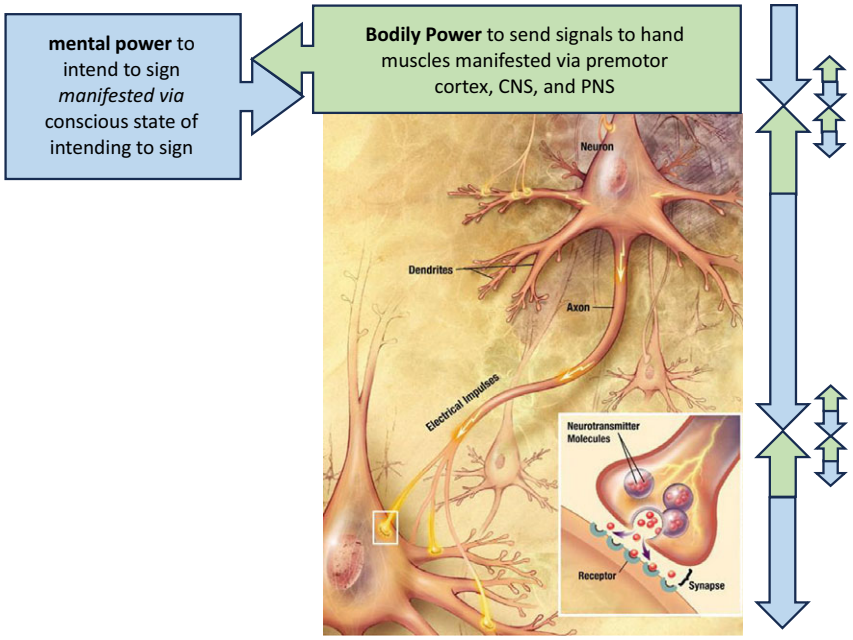


Diagram 4. *Mind-body causation and neuronal causation*. On the mind-body level, which is horizontally represented, the active (blue) mental power on the left activates the passive (green) bodily powers manifested via neural activity on the right, as in Diagram 2. On the neuronal level and the subcellular level, which is vertically represented, the sending of signals via synaptic transmission, involves a complex matrix of active (blue) and passive (green) powers that would be impossible to fully illustrate. Image credit: Drawing by Christy Krames illustrating the process of synaptic transmission in neurons, cropped from original in a National Institute of Health, National Institute of Aging brochure. Source: Common Domain via Wikipedia Commons: https://commons.wikimedia.org/wiki/File:Chemical_synapse_schema_cropped.jpg

Although Aquinas was not explicitly referring to effects at the neuronal level, the mind-body powers model can be used to apply his thirteenth century line of reasoning to the contemporary neural response paradigm. Suppose a physician asks a patient named Eliud if his name is Eliud, and commands him to imagine navigating through his home if the answer is *no*, or to imagine playing tennis if the answer is *yes*. Suppose further that directly following the command there is neural activity in the SMA, which would be expected if Eliud imagined playing tennis. The key question is: What explains the neural activity? From the philosophical framework of the mind-body powers model, the most natural explanation is that Eliud is consciously intending to answer "yes" by imagining playing tennis (see Diagram 3).

The natural function of Eliud's mental power to consciously intend to imagine playing tennis would always be accompanied by the co-manifestation of its partner bodily powers manifested by neural activity in the SMA. According to the model, the mental and bodily powers are always co-manifested, and therefore, the conscious state and the neural activity that are the manifestations of the powers naturally coincide. This is the model's explanation of why there are NCC, but it also explains why the neural activity should be understood as indicating the presence of the conscious intention. After all, the manifestation of the two powers is naturally coupled; consequently, the manifestation of one implies the manifestation of the other. In other words, the manifestation of the mental power via the conscious intention naturally corresponds to the manifestation of the bodily powers via the neural activity in the SMA.

The model does permit the possibility that the same bodily powers manifested by the same neural activity could also correspond to other conscious states. For example, the model does not rule out the possibility of the neural activity in the SMA also corresponding to the conscious state of imagining playing basketball, instead of tennis.[133] However, this has no consequences for how the activity should be interpreted. Given the clinical context in which the patient has just been commanded to imagine playing tennis to answer yes, if the neural activity in the SMA follows, then the best explanation for it is that the

patient is consciously imagining playing tennis. And on the mind-body powers model, the conscious intention not only can be but naturally is the causal explanation of the neural activity, as the conscious intention and neural activity co-manifest corresponding causal powers. As Aquinas once said "the soul is the part producing motion" whereas "the ensouled body is the part that is moved,"[134] we can say that the conscious intention produces activity while the neurons in the SMA are activated.

In sum, the natural correspondence between the conscious intention to imagine playing tennis and the neural activity in the SMA, grounded in the co-manifestation of interdependent mental and bodily powers, justifies interpreting the neural activity as indicating the conscious intention. And if a patient is cognizant enough to follow the command to imagine playing tennis or to imagine navigating through their home, it is probable that they are cognizant enough to answer simple questions by imagining one of the tasks. If the neuronally indicated answers to factual questions are consistently correct, there is no reason, from the hylomorphic vantage point presented here, not to interpret the neural activity following specific questions as a consciously intended neural response.

Conclusion

Olivia Gosseries et al.[135] have pointed out that the utility of neuroimaging approaches to diagnosing disorders of consciousness depends on understanding how consciousness relates to the brain. We have tried to demonstrate how this is likewise true when interpreting neural activity that is purportedly indicative of a conscious response. Nonreductive physicalism's causal exclusion problem elicits reservations about the neuronal response paradigm and whether neural activity could be causally explained by a conscious intention. By contrast, the mind-body powers model justifies interpreting neural activity as indicating a conscious intention that causally explains it. That would be, given the mind-body powers model, the most natural interpretation.[136] So, from the perspective of the model, there is more warrant for giving credence to neural responses of behaviorally unresponsive patients, which could preserve their autonomy. That said, it must be acknowledged that a responsible approach to interpreting alleged neuronal responses must weigh various factors.

In addition to clinical factors (e.g., a patient's level of arousal) and neurological questions about the efficacy of the task-response approach already alluded to,[137] there are also sociological factors. Physicalism is more widely embraced among academic philosophers, with a recent survey suggesting that 51.9% lean toward physicalism while 32.1% lean toward non-physicalism and 15.9% lean toward something else.[138] Although it has been argued that there is a prevalent "closet dualism"[139] in cognitive neuroscience, it is likely that most neuroscientists and neurologists would claim to lean toward physicalism. However, if the opinions of professional academics are not the only opinions that should inform ethical clinical practice, and public opinion should also be considered, then physicalism might not provide the only philosophical lens through which possible neuronal responses should be considered.[140]

According to sociologist John Evans, in American culture physicalism is not the most common view of human nature relied on to form bioethical opinions, but a theological view informed by Christianity "has had enormous influence in the West, such as being the origin of secular human rights, and is widely held by Americans."[141] If that is correct, the general public in the United States could be unlikely to view the topic of neuronal responses from a physicalist perspective and more likely to view it from a perspective akin to the hylomorphism held by Aquinas (who many Christians consider a preeminent theological authority) that inspired the mind-body powers model. Something similar might be said of cultures influenced by Buddhism, insofar as its view of consciousness is incompatible with physicalism.[142] And where there is profound ideological diversity on these matters, it is possible that individual decision-makers might consider alleged neuronal responses from diverging philosophical frameworks and, consequently, arrive at differing conclusions about their validity. Our hope is that the reflections in this article will make the role of such philosophical frameworks more apparent and therefore easier to critically analyze.

Acknowledgments. MO thanks Christof Koch and Mihretu Guta for many helpful discussions relevant to this article.

Competing interest. The authors declare none.

Author contribution. MO: Conceptualization; visualization; writing—original draft; writing—review, revisions, and editing. DH: Review, revisions, editing. AGH: Writing—review, revisions, editing.

Notes

1. Edlow BL, Fins JJ. Assessment of covert consciousness in the intensive care unit: Clinical and ethical considerations. *The Journal of Head Trauma Rehabilitation* 2018;**33**(6):424–34.
2. Kazazian K, Edlow BL, Owen AM. Detecting awareness after acute brain injury. *The Lancet Neurology* 2024;**23**(8):836–44.
3. Giacino JT, Kalmar K, Whyte J. The JFK coma recovery scale–revised: Measurement characteristics and diagnostic utility. *Archives of Physical Medicine and Rehabilitation* 2004;**85**(12):2020–9.
4. Mat B, Sanz LRD, Arzi A, Boly M, Laureys S, Gosseries O. New behavioral signs of consciousness in patients with severe brain injuries. *Seminars in Neurology* 2022;**42**(03):259–72.
5. Edlow BL. Covert consciousness: Searching for volitional brain activity in the unresponsive. *Current Biology* 2018;**28**(23):R1345–8.
6. For a succinct historical overview of covert consciousness, amidst ethical considerations about the use of AI to aid patient communication, see Fins JJ, Shulman KS. Neuroethics, covert consciousness, and disability rights: What happens when artificial intelligence meets cognitive motor dissociation? *Journal of Cognitive Neuroscience* 2024;**36**(8):1667–74.
7. Owen AM, Coleman MR, Boly M, Davis MH, Laureys S, Pickard JD. Detecting awareness in the vegetative state. *Science* 2006;**313**(5792):1402.
8. Boly M, Coleman M, Davis M, et al. When thoughts become action: An fMRI paradigm to study volitional brain activity in non-communicative brain injured patients. *NeuroImage* 2007;**36** (3):979–92.
9. Monti MM, Vanhaudenhuyse A, Coleman MR, et al. Willful modulation of brain activity in disorders of consciousness. *The New England Journal of Medicine* 2010;**362**:579–89.
10. Peterson A, Mintz K, Owen AM. Unlocking the voices of patients with severe brain injury. *Neuroethics* 2022;**15**(9).
11. See note 10, Peterson, Mintz, Owen 2022, at 2.
12. Fins JJ. Mosaic decisionmaking and reemergent agency after severe brain injury. *Cambridge Quarterly of Healthcare Ethics* 2018;**27**(1):163–74, at 166–7.
13. Cairncross M, Peterson A, Lazosky A, Gofton T, Weijer C. Assessing decision-making capacity in patients with communication impairments: A case study. *Cambridge Quarterly of Healthcare Ethics* 2016; **25**(4): 691–9, at 695.
14. See note 12, Fins 2018.
15. For a response to Finns, see Peterson A. A critical analysis of Joseph Fins' mosaic decisionmaking: A response to "mosaic decisionmaking and reemergent agency after severe brain injury". *Cambridge Quarterly of Healthcare Ethics* 2019; **28**: 725–36.
16. Schnakers C, Edlow BL, Chatelle C, Giacino JT. Minimally conscious state. In: Laureys S, Gosseries O, Tononi G, eds. *The Neurology of Consciousness*. 2nd ed. San Diego, CA: Elsevier; 2016:167–81, at 171.
17. Schnakers C, Vanhaudenhuyse A, Giacino J, et al. Diagnostic accuracy of the vegetative and minimally conscious state: Clinical consensus versus standardized neurobehavioral assessment. *BMC Neurology* 2009;**9**(35).
18. Pan J, Xie Q, Qin P, et al. Prognosis for patients with cognitive motor dissociation identified by brain-computer interface. *Brain: A Journal of Neurology* 2020;**143**:1177–89.
19. Owen AM. Improving diagnosis and prognosis in disorders of consciousness. *Brain* 2020;**143**: 1046–56, at 1052.
20. See note 10, Peterson, Mintz, Owen 2022.
21. See note 7, Owen, Coleman, Boly, et al. 2006.
22. See note 8, Boly, Coleman, Davis, et al. 2007.

23. See note 9, Monti, Vanhaudenhuyse, Coleman, et al. 2010.
24. See note 8, Boly, Coleman, Davis, et al. 2007.
25. See Massimini M, Tononi G. *Sizing Up Consciousness: Towards an Objective Measure of the Capacity for Experience*. New York, NY: Oxford University Press; 2018, at 37.
26. See note 7, Owen, Coleman, Boly, et al. 2006.
27. Cf. note 18, Pan, Xie, Qin, et al. 2020.
28. Cf. Naci L, Cusack R, Anello M, Owen AM. A common neural code for similar conscious experiences in different individuals. *Proceedings of the National Academy of Sciences of the USA* 2014;**111**(39):14277–82.
29. See note 9, Monti, Vanhaudenhuyse, Coleman, et al. 2010.
30. See note 19, Owen 2020, at 1051.
31. See note 10, Peterson, Mintz, Owen 2022, at 4.
32. Boly M, Seth AK, Wilke M, et al. Consciousness in humans and non-human animals: recent advances and future directions. *Frontiers in Psychology* 2013;**4**:1–20, at 3.
33. Giacino JT, Fins JJ, Laureys S, Schiff ND. Disorders of consciousness after acquired brain injury: the state of the science. *Nature Reviews Neurology* 2014;**10**(2):99–114, at 105.
34. Owen M, Owen AD, Hudetz AG. An Aristotelian-thomistic framework for detecting covert consciousness in unresponsive persons. In: Guta MP, Rae SB, eds. *Taking Persons Seriously: Where Philosophy and Bioethics Intersect*. Eugene, OR: Pickwick Publications; 2024:93–125, at 103–4.
35. See note 25, Massimini, Tononi 2018, at 38–9.
36. Huang Z, Vlisides PE, Tarnal VC, et al. Brain imaging reveals covert consciousness during behavioral unresponsiveness induced by propofol. *Nature Scientific Reports* 2018;**8**(1):13195.
37. Gross WL, Lauer KK, Liu X, et al. Propofol sedation alters perceptual and cognitive functions in healthy volunteers as revealed by functional magnetic resonance imaging. *Anesthesiology* 2019;**131**(2):254–65.
38. Liu X, Lauer KK, Ward BD, Rao SM, Li S-J, Hudetz AG. Propofol disrupts functional interactions between sensory and high-order processing of auditory verbal memory. *Human Brain Mapping* 2012;**33**(10):2487–98.
39. See note 10, Peterson, Mintz, Owen 2022.
40. See note 8, Boly, Coleman, Davis, et al. 2007.
41. See note 9, Monti, Vanhaudenhuyse, Coleman, et al. 2010.
42. Fernández-Espejo D, Owen AM. Detecting awareness after severe brain injury. *Nature Reviews Neuroscience* 2013;**14**(11):801–9.
43. Peterson A, Owen AM, Karlawish J. Alive inside. *Bioethics* 2020;34(3):295–305.
44. See note 10, Peterson, Mintz, Owen 2022.
45. See note 43, Peterson, Owen, Karlawish 2020, at 296.
46. See note 10, Peterson, Mintz, Owen 2022, at 2.
47. Bekinschtein T, Manes F, Villarreal M, Owen A, Della Maggiore V. Functional imaging reveals movement preparatory activity in the vegetative state. *Frontiers in Human Neuroscience* 2011;**5**.
48. Monti MM, Coleman MR, Owen AM. Executive functions in the absence of behavior: Functional imaging of the minimally conscious state. *Progress in Brain Research* 2009;**177**:249–60.
49. Monti MM, Pickard JD, Owen AM. Visual cognition in disorders of consciousness: From V1 to top-down attention. *Human Brain Mapping* 2013;**34**(6):1245–53.
50. Zeman A. The problem of unreportable awareness. In: Laureys S, Schiff ND, Owen AM, eds. *Progress in Brain Research*. Vol **177**. Elsevier; 2009:1–9.
51. See note 42, Fernández-Espejo, Owen 2013, at 802. Italics added.
52. Kim J. *Physicalism, Or Something Near Enough*. Princeton, New Jersey: Princeton University Press; 2005, at ch. 3.
53. Kim J. *Mind in a Physical World*. Cambridge, MA: MIT Press; 2000, at 30, 38.
54. See note 52, Kim 2005, at 70.
55. Koons RC, Bealer G, eds. *The Waning of Materialism*. New York, NY: Oxford University Press; 2010.

56. Göcke BP, ed. *After Physicalism*. Notre Dame, Indiana: University of Notre Dame Press; 2012.
57. Koch C. Reflections of a natural scientist on panpsychism. *Journal of Consciousness Studies* 2021;**28** (9–10):65–75, at 73.
58. Koch C. Re-Enchanting the World. *The American Journal of Psychology* 2020;**133**(4):519–22, at 519.
59. Loose JJ, Menuge AJL, Moreland JP, eds. *The Blackwell Companion to Substance Dualism*. Oxford, UK: Wiley Blackwell; 2018.
60. Rickabaugh B, Moreland JP. *The Substance of Consciousness: A Comprehensive Defense of Contemporary Substance Dualism*. Oxford, UK: Wiley-Blackwell; 2023.
61. Weir RS. *The Mind-Body Problem and Metaphysics: An Argument From Consciousness to Mental Substance*. New York, NY: Routledge; 2024.
62. See note 42, Fernández-Espejo, Owen 2013, at 802.
63. See note 53, Kim 2000, at 30.
64. As with many views, there is debate about what ideas are essential to physicalism and how those ideas should be articulated. Our goal is not to provide a survey of the range of options and articulations. Therefore, we rely here on Jaegwon Kim's thought as a representative example of physicalism and the rationale for the causal exclusion problem, since Kim was arguably the most influential architect of arguments demonstrating the problem.
65. Kim J. The non-reductivist's troubles with mental causation. In: Heil J, Mele A, eds. *Mental Causation*. New York, NY: Oxford University Press; 1993:189–210, at 209–10.
66. See note 53, Kim 2000, at 39.
67. Kim J. *Philosophy of Mind*. Third ed. Boulder, CO: Westview Press; 2011, at 12.
68. See note 67, Kim 2011, at 12.
69. See note 65, Kim 1993, at 209–10.
70. See note 53, Kim 2000, at 40.
71. See Smart JJC. Sensations and Brain Processes. *The Philosophical Review* 1959;**68**(2):141–56.
72. Alternatively, one might identify the conscious intention with the neural activity that slightly precedes NS, perhaps even in a different brain region, to avoid the dilemma raised here. It has been argued that a neural correlate of a conscious state should be more broadly conceived to include prior neural activity, which makes one wonder if identifying the conscious intention with a specific phase of broadly corresponding neural activity could be another alternative. See Northoff G, Zilio F, Zhang J. Beyond task response—Pre-stimulus activity modulates contents of consciousness. *Physics of Life Reviews* 2024;**49**:19–37.
73. See note 53, Kim 2000, at 39–46.
74. Qualia are the felt sensations of what-it's-like to experience particular conscious states. For example, the quale of tasting your cup of coffee in the morning is the feeling of what-it's-like for you to taste the coffee.
75. See note 52, Kim 2005, at 70–1, see also fn. 2 and ch. 6.
76. See Roche M. Causal Overdetermination and Kim's Exclusion Argument. *Philosophia* 2014;**42** (3):809–26.
77. See Won C. Mental causation as joint causation. *Synthese* 2019; **198**:4917–37.
78. See Owen, A. Adrian Owen discusses detecting awareness after acute brain injury. *The Lancet Neurology in conversation with.* The Lancet Group. 17 July, 2024; available at https://open.spotify.com/episode/6UPscQVG0n90RG4CsQRtjT?si=9b03f1420d8f42fd (last accessed 24 July 2007).
79. Simpson WMR. *Hylomorphism*. Cambridge: Cambridge University Press; 2023.
80. Simpson WMR, Koons RC, Teh NJ, eds. *Neo-Aristotelian Perspectives on Contemporary Science*. New York, NY: Routledge; 2018. Routledge Studies in the Philosophy of Science.
81. Marmodoro A, Paolini Paoletti M. Introduction to the special issue on form, structure and hylomorphism. *Synthese* 2021;**198**(11):2647–56.
82. Inman RD. *Substances and the Fundamentality of the Familiar: A Neo-Aristotelian Mereology*. New York, NY: Routledge; 2018.

83. Koons RC. *Is St. Thomas's Aristotelian Philosophy of Nature Obsolete?* South Bend, Indiana: St. Augustine Press; 2022.
84. Haldane J. A return to form in the philosophy of mind. *Ratio* 1998;**11**(3):253–77.
85. Marmodoro A. *Aristotle On Perceiving Objects*. New York, NY: Oxford University Press; 2014.
86. Owen M. *Measuring the Immeasurable Mind: Where Contemporary Neuroscience Meets the Aristotelian Tradition*. Lanham, MD: Lexington Books; 2021.
87. Owen M. Exploring common ground between integrated information theory and Aristotelian metaphysics. *Journal of Consciousness Studies* 2019;**26**(1–2):163–87.
88. Sánchez-Cañizares J. Integrated information theory as testing ground for causation: why nested hylomorphism overcomes physicalism and panpsychism. *Journal of Consciousness Studies* 2022;**29**(1).
89. Sánchez-Cañizares J. Formal causation in integrated information theory: An answer to the intrinsicality problem. *Foundations of Science* 2022;**27**(1):77–94.
90. Jaworski W. *Structure and the Metaphysics of Mind: How Hylomorphism Solves the Mind-Body Problem*. New York, NY: Oxford University Press; 2016.
91. See note 10, Peterson, Mintz, Owen 2022.
92. For further details, see note 86, Owen 2021.
93. Aristotle. Metaphysics. In: Barnes J, ed. Ross WD, trans. *The Complete Works of Aristotle.* Vol Two. The Revised Oxford Translation ed. Princeton, NJ: Princeton University Press; 1984:1552–728, at 1644.
94. Aristotle. On the Soul. In: Barnes J, ed. Smith JA, trans. *The Complete Works of Aristotle.* Vol One. The Revised Oxford Translation ed. Princeton, NJ: Princeton University Press; 1984:641–92, at 656–9.
95. See Aquinas T. *The Treatise on Human Nature: Summa Theologiae 1a 75-89*. Translated by Pasnau R. Indianapolis, IN: Hackett Publishing; 2002, at ST 1a 76.1c, 76.3c, 76.8c. When citing Aquinas, 'ST' refers to the Summa Theologiae and '1a' means prima pars (first part) followed by the question number and subsequently the article number. The 'c' means main reply whereas 'sc' refers to the said contra, and 'ad' refers to a responding argument.
96. See note 95, Aquinas 2002, at ST 1a 76.1sc, 76.4sc.
97. See note 94, Aristotle 1984, at 658 (413b11-12).
98. See, e.g., Aquinas T. *Questions on the Soul.* Translated by Robb JH. Milwaukee: Marquette University Press; 1984, at 121 (QDA 8 ad 16). 'QDA' refers to Questions on the Soul and the first number refers to the particular question number whereas 'ad' refers to a numbered responding argument.
99. This is not to rule out efficient physical causes (e.g., genetics) of the body's biological parts. See note 86, Owen 2021, at 153–5.
100. See note 95, Aquinas 2002, at ST 1a 79.10 ad 3.
101. See note 98, Aquinas 1984, at 120 (QDA 8 ad 15).
102. See note 98, Aquinas 1984, at 120 (QDA 8 ad 15).
103. See note 98, Aquinas 1984, at 115 (QDA 8c).
104. See note 98, Aquinas 1984, at 115 (QDA 8c).
105. See note 98, Aquinas 1984, at 116 (QDA 8c).
106. See note 95, Aquinas 2002, at ST 1a 78.4c.
107. See note 95, Aquinas 2002, at ST 1a 78.4c.
108. See Bear MF, Connors BW, Paradiso MA, eds. *Neuroscience: Exploring the Brain*. Fourth Edition ed. Philadelphia, PA: Wolters Kluwer; 2016, at 5.
109. See note 86, Owen 2021, at 151–2.
110. See note 98, Aquinas 1984, at 142 (QDA 10 ad 9).
111. See note 98, Aquinas 1984, at 116 (QDA 8c).
112. Owen M. Aristotelian Causation and Neural Correlates of Consciousness. *Topoi* 2020;**39**(5):1113–24.
113. See note 86, Owen 2021, at ch. 7.
114. Here we apply the model to human consciousness, but it could also be adapted and applied to consciousness in animals.

115. See note 112, Owen 2020.
116. Crick F, Koch C. What are the neural correlates of consciousness? In: Hemmen Lv, Sejnowski TJ, eds. **23** *Problems in Systems Neuroscience*. Oxford University Press; 2006:474–91.
117. Chalmers DJ. What is a neural correlate of consciousness? In: Metzinger T, ed. *Neural Correlates of Consciousness*. Cambridge, Massachussetts: MIT Press; 2000:17–40.
118. Owen M, Guta MP. Physically sufficient neural mechanisms of consciousness. *Frontiers in Systems Neuroscience* 2019;**13**(24).
119. Koch C, Massimini M, Boly M, Tononi G. Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience* 2016;**17**:307–21, at 308.
120. See note 86, Owen 2021, at ch. 8.
121. See note 34, Owen, Owen, Hudetz 2024.
122. See note 93, Aristotle 1984, at 1609.
123. See note 85, Marmodoro 2014, at 32.
124. In this context, the phrase 'conscious state' can refer to one's overall or global state (such as being conscious versus unconscious) or to specific states of consciousness with particular content, as we use the phrase here.
125. See Kanwisher N, McDermott J, Chun MM. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience* 1997;**17**(11):4302–11.
126. See Duchaine B, Yovel G. A revised neural framework for face processing. *Annual Review of Vision Science* 2015;**1**(1):393–416.
127. See Bilalic M. Revisiting the role of the fusiform face area in expertise. *Journal of Cognitive Neuroscience* 2016;**28**(9):1345–57.
128. This does not mean, however, that a particular neural mechanism can only correspond to one conscious state as if it only has a single role. The same molecule (e.g., a neurotransmitter like dopamine) could do many different things in the brain. Some molecules within cells have been found to be 'moonlighting', i.e., doing multiple different jobs when not needed for their main function.
129. See Potgieser AR, van der Hoorn A, de Jong BM. Cerebral activations related to writing and drawing with each hand. *PLoS One* 2015;**10**(5):e0126723, at 16.
130. Aquinas T. *Summa Theologiae.* Vol 14. Latin/English Edition of the Works of St. Thomas Aquinas. Lander, Wyoming: The Aquinas Institute; 2012, at ST 1a 57.4c.
131. See note 130, Aquinas 2012, at ST 1a 57.4c.
132. See note 130, Aquinas 2012, at ST 1a 81.3 ad 2.
133. With more fine-grained brain imaging in the future and more advanced techniques, one may be able to make fine-grain discriminations of conscious states and their contents by imaging the activity of individual neurons.
134. See note 95, Aquinas 2002, at ST 1a 76.4 ad 2.
135. Gosseries O, Di H, Laureys S, Boly M. Measuring Consciousness in Severely Damaged Brains. *Annual Review of Neuroscience* 2014:457–78.
136. Such an interpretation would enjoy further justification if the patient were previously shown to be conscious using a no-response paradigm relying on the presence of the full NCC. See note 34, Owen, Owen, and Hudetz 2024.
137. There can also be questions raised by tangential research about how much weight to give atypical responses. In anesthesiology, for example, an unexpected state has been discovered which is called an IFT-positive state (from Isolated Forearm Test positive). In these instances, patients who have had a decent dose of anesthetic, can respond to a command—'squeeze my hand'—but do not initiate actions (even if in pain), and they do not remember it afterwards. Is this an acceptable state for surgery? One could argue that the patients seem conscious, at least in some sense, but their frontal EEG suggests they are properly anaesthetized, at least with some standard basic measures. See Gaskell AL, Hight DF, Winders J, et al. Frontal alpha-delta EEG does not preclude volitional response during anaesthesia: prospective cohort study of the isolated forearm technique. *British Journal of Anaesthesia* 2017;119(4):664–73.

138. Bourget D, Chalmers DJ. Philosophers on Philosophy: The 2020 PhilPapers Survey. *Philosophers' Imprint* 2023;**23**(11):1–53, at 7.
139. Mudrik L, Maoz U. "Me & My Brain: Exposing Neuroscience's Closet Dualism. *Journal of Cognitive Neuroscience* 2014;**27**(2):211–21.
140. Cf. Evans JH. *The History and Future of Bioethics: A Sociological View*. New York, NY: Oxford University Press; 2012, at ch. 5.
141. Evans JH. *Disembodied Brains: Understanding our Intuitions on Human-Animal Neuro-Chimeras and Human Brain Organoids*. Oxford University Press; 2024, at 16–7, 27.
142. See Lopez DS, Jr. *Buddhism & science: a guide for the perplexed.* Chicago: University of Chicago Press; 2008, at 149.

Cite this article: Owen M, Hight D and Hudetz AG (2025). Neural Voices of Patients with Severe Brain Injury? *Cambridge Quarterly of Healthcare Ethics*: 1–22, doi:10.1017/S0963180124000446