

Consilience, clinical validation, and global disorders of consciousness

Andrew Peterson*

Rotman Institute of Philosophy and The Brain and Mind Institute, University of Western Ontario, London, ON, Canada

*Correspondence address. Department of Philosophy, Rotman Institute of Philosophy, Western University, Stevenson Hall 2150, London, ON N6A 5B8, Canada. Tel: 519-661-2111 ext 80947; E-mail: apeter62@uwo.ca

Abstract

Behavioral diagnosis of global disorders of consciousness is difficult and errors in diagnosis occur often. Recent advances in neuroimaging may resolve this problem. However, clinical translation of neuroimaging requires clinical validation. Applying the orthodox approach of clinical validation to neuroimaging raises two critical questions: (i) What exactly is being validated? and (ii) what counts as a gold standard? I argue that confusion over these questions leads to systematic errors in the empirical literature. I propose an alternative approach to clinical validation motivated by reasoning by consilience. Consilience is a mode of reasoning that assigns a degree of plausibility to a hypothesis based on its fit with multiple pieces of evidence from independent sources. I argue that this approach resolves the questions raised by the orthodox approach and may be a useful framework for optimizing future clinical validation studies in the science of consciousness.

Key words: Global disorders of consciousness; consciousness; awareness; clinical validation; philosophy; neurology

Introduction

Improvements in intensive care have led to an increased survival rate among patients who sustain severe brain injuries. These individuals suffer lost or disordered consciousness for extended periods of time. The standard method of diagnosis is the bedside neurobehavioral exam, but misdiagnosis occurs often. A patient may be misdiagnosed due to misinterpretations of behavior, or due to an inherent insensitivity of the method of assessment to awareness, regardless of behavior. Both kinds of diagnostic errors have profound implications, and may lead to premature withdrawal of life support, insufficient pain management, or misallocation of medical resources. Accurate diagnosis is imperative, yet assessment of severe brain injury remains one of the most challenging obstacles of modern medicine.

Recent findings in clinical neuroscience offer a potential solution. Neuroimaging can detect residual cognitive function and awareness in some brain-injured patients who appear entirely unresponsive at the bedside. Neuroimaging may provide

information ancillary to neurobehavioral exam that improves diagnostic accuracy. Yet, which neuroimaging methods have the greatest diagnostic accuracy is currently unknown.

The orthodox approach of evaluating the diagnostic accuracy of a novel diagnostic test, or “clinical validation,” involves estimating its sensitivity and specificity against a “gold standard” diagnostic test. This approach to clinical validation is common in translational research but application to the science of consciousness is not straightforward and raises two critical questions. First, what exactly is being validated? And second, what counts as a gold standard?

In this article, I explicate these questions. I argue that confusion over these questions leads to systematic errors in the empirical literature. The central problem with the orthodox approach when applied to the science of consciousness is that the validity of the gold standard cannot be taken for granted. I propose an alternative approach to clinical validation based on reasoning by consilience. According to this approach, the validity of a novel test is determined by the degree to which its

Received: 16 February 2016; Revised: 8 June 2016. Accepted: 9 July 2016

© The Author 2016. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

findings are conciliate with a broad range of evidence from independent sources. I argue that the concisience approach resolves the critical questions raised by the orthodox approach, and may serve as a useful framework for optimizing future clinical validation studies in the science of consciousness.

Neuroimaging Global Disorders of Consciousness

Severe brain injury is a leading cause of death and disability (Thurman et al., 1999; Greenwald et al., 2003). Improvements in critical care have prolonged the lives of individuals who sustain severe brain injury, including those who are comatose. Coma is characterized by unarousable unawareness that lasts approximately 2 weeks (Young, 2009). The pathophysiology of coma is damage to the cortex, underlying white matter, or bilateral thalamus (Giacino et al., 2014). Damage to the ascending reticular activating system prevents comatose patients from breathing on their own (Young, 2009). Artificial hydration, nutrition, and ventilation are required.

Following a period of coma some patients may enter into a vegetative state (also known as unresponsive wakefulness syndrome; Laureys et al., 2010). Clinically, this transition coincides with recovery of spontaneous respiration (Demertzi et al., 2013). Mechanical ventilation can be safely removed; however, artificial hydration and nutrition are still required. Behaviorally, the vegetative state is characterized by dissociation between wakefulness and awareness. Vegetative patients will exhibit sleep/wake cycles yet evidence no concomitant awareness of visual, auditory, or tactile stimuli. The vegetative state is often referred to as “wakeful unresponsiveness” (Plum and Posner, 1982; Multi-Society Task Force, 1994).

A number of vegetative patients may enter into a minimally conscious state (Giacino et al., 2002). The minimally conscious state is characterized by sleep/wake cycles and intermittent behavioral evidence of awareness. Distinguishing features of the minimally conscious state include inconsistent command following, purposeful yet minimal language use, visual pursuit, object recognition, and localization of painful stimuli. The pathophysiology of the minimally conscious state is diffuse axonal injury with variable thalamic involvement. The severity of impairment to the thalamus and specific cortico-thalamic connections is usually greater in vegetative than minimally conscious patients. This may explain why vegetative and minimally conscious patients differ in cognitive processing and behavior (Giacino et al., 2014).

These conditions are often referred to as “global disorders of consciousness” (Schiff, 2007; Bayne and Hohwy, 2014). Global disorders of consciousness are distinct from focal disorders of consciousness, such as blind-sight, in that the disturbance to consciousness is global rather than in one kind of conscious content. Of those who fully recover from global disorders of consciousness, nearly half remain disabled and in need of continued specialized care (Dikmen et al., 2003).

The standard method for diagnosis of global disorders of consciousness is the bedside neurobehavioral exam. These exams assess behavioral evidence of awareness (Multi-society Task Force, 1994; Giacino et al., 2004). The neurobehavioral exam—especially the Coma Recovery Scale-Revised (CRS-R)—is widely deemed the most effective method of assessment, yet several studies reveal errors and ambiguities in diagnostic accuracy (see Seel et al., 2010 for extensive review). Childs et al. (1993) and Andrews et al. (1996) demonstrate that neurobehavioral

diagnoses of the vegetative state from referring clinicians are often inaccurate when compared to assessments from specialized neurorehabilitation centers. Likewise, Schnakers et al. (2009) show significant error rates in clinical consensus diagnoses of the vegetative state when compared to the CRS-R. Discordance in diagnoses from different neurobehavioral methods is consistently between 30% and 40%. In most cases, a patient is misdiagnosed as being in the vegetative state when, in fact, she is minimally conscious.

Recent advances in neuroimaging provide information ancillary to neurobehavioral examination and may improve diagnostic accuracy. Several neuroimaging methods have been refined in clinical populations for this purpose. First, functional magnetic resonance imaging (fMRI) mental imagery measures hemodynamics as participants imagine activities. While lying in the scanner, patients are instructed to imagine playing tennis or navigating through their home for sustained and repeated 30-s intervals. Studies involving healthy participants demonstrate that imagining these activities reliably activates the supplementary motor area and parahippocampal gyrus, respectively (Boly et al., 2007). Task-appropriate and sustained activation is interpreted as volitional response to command. This technique has been effective in assessing command following in vegetative and minimally conscious patients (Owen et al., 2006; Monti et al., 2010; Fernández-Espejo and Owen, 2013). Indeed, in one large-scale study, it was found that 4 of 24 clinically vegetative patients were able to modulate their brain activity to command (Monti et al., 2010).

Second, fluorodeoxyglucose positron emission tomography (^{18}F -FDG PET) measures brain metabolism. Patients are injected with a synthetic compound, ^{18}F -fluorodeoxyglucose, containing a glucose molecule with the radioactive isotope, fluorine-18. Tissue uptake of ^{18}F -fluorodeoxyglucose is a marker of neural glucose uptake and metabolism. Measurement of isotope emission allows estimation of glucose metabolism at the whole brain level.

Several studies have applied this method in global disorders of consciousness (Laureys et al., 2004; Nakayama et al., 2006; Thibaut et al., 2012; Stender et al., 2014, 2015, 2016). Quantitative rates of glucose metabolism in brain regions hypothesized to support awareness—including the frontoparietal associative cortices, cingulate gyrus, precuneus, and thalamus—are markedly different in patients when compared to healthy participants at rest. In one recent study, it was demonstrated that the resting-state whole brain metabolic rate for vegetative patients is 42% of normal (Stender et al., 2015). Differences in metabolism between patients in the vegetative and minimally conscious states are most pronounced in the frontoparietal associative cortices.

Third, functional MRI can be used to measure brain activity during passive exposure to a variety of auditory or visual stimuli. Owen et al. (2002) demonstrated characteristic activation of the fusiform face area in several vegetative patients upon exposure to familiar faces. Coleman et al. (2009) found evidence of preserved language processing in vegetative and minimally conscious patients when exposed to sentences containing semantically ambiguous words (e.g. “there were a pair of dates in the fruit bowl”) versus semantically unambiguous words (e.g. “there is a beer in the refrigerator”). Naci et al. (2014) recently extended this approach by exposing patients to a highly engaging Alfred Hitchcock film. One vegetative patient showed brain activity similar to that observed in healthy participants. The time course of brain activity also matched independent behavioral ratings of executive demand and suspense. This suggests that

the patient likely had similar experiences to those of healthy participants as he watched the film (Naci et al., 2014, p. 14279).

Fourth, neuroimaging can be used to assess the functional and structural integrity of networks believed to underpin awareness in the resting state. Several recent studies using fMRI demonstrate deterioration in the functional integrity of the default-mode-network—an intrinsic cortical network associated with self-awareness—in global disorders of consciousness patients (Boly et al., 2009; Vanhaudenhuyse et al., 2009; Demertzi et al., 2013). Likewise, Fernández-Espejo et al. (2011) used structural imaging to show changes of tissue integrity in long-range cortical-thalamic connections in vegetative patients. It is hypothesized that damage to these connections may inhibit voluntary control of behavior (Fernández-Espejo et al., 2015). Finally, several studies have also identified associations between electrophysiological connectivity and awareness (Chennu et al., 2013; King et al., 2013). Indeed, Chennu et al. (2013) observed that, in some patients, electrophysiological evidence of attentional abilities was associated with performance on an fMRI mental imagery task. These results have, in part, been effective for predicting recovery (cf. Di et al., 2008; Norton et al., 2012). It has also been demonstrated that changes in functional integrity track the behavioral diagnostic categories of the vegetative and minimally conscious states (Demertzi et al., 2015).

Several general conclusions can be drawn from this research. First, neuroimaging can contribute to our understanding of the kinds of misdiagnoses that occur in global disorders of consciousness patients. One kind of misdiagnosis results from errors in the interpretation of patient behavior. In principle, patients could be diagnosed correctly provided that a neurobehavioral exam is interpreted appropriately. Yet, several reasons—including suboptimal physician training or variation in patient arousal—may prevent accurate observation. Neuroimaging can provide information ancillary to neurobehavioral scales to highlight and correct this kind of misdiagnosis.

A second kind of misdiagnosis can occur because neurobehavioral scales are, in principle, insensitive to awareness unmediated by behavior. As reviewed above, a number of patients who consistently satisfy the behavioral criteria of the vegetative state can modulate their brain activity to command. These patients are covertly aware. As such, their misdiagnosis results from an error in applying the correct method of assessment, not a misinterpretation of behavior. This latter kind of misdiagnosis is intriguing as it may recast diagnostic categories in terms of the mere presence or absence of awareness, rather than behaviorally mediated evidence of awareness. Both kinds of misdiagnosis suggest that neuroimaging could be a natural compliment to bedside neurobehavioral exams in difficult cases (Giacino et al., 2014; Stender et al., 2014).

Additionally, neuroimaging and electrophysiological assessment at the bedside could contribute to more consistent diagnoses across medical institutions. Consistent neurobehavioral evaluation is challenging, and typically requires satisfaction of at least three conditions: the application of systematic method of assessment (e.g. the CRS-R), an expert evaluator capable of applying the method consistently, and repeated examinations to sufficiently account for arousal variability commonly observed in these patients. Except at specialized referral centers, one or more of these conditions are often difficult to satisfy. Neuroimaging and electrophysiological assessment can ameliorate this problem by providing compensatory information if neurobehavioral evaluation is inconsistent.

Finally, neuroimaging may—at least, in the future—improve the accuracy of prognostication in acutely comatose patients.

Evidence of preserved cognitive function may distinguish positive from poor outcomes. The accuracy of this information is important for families as it may help guide end-of-life decisions (Weijer et al., 2016).

Clinical Validation: The Orthodox Approach

The effectiveness of neuroimaging in detecting preserved cognitive function and awareness in global disorders of consciousness has motivated calls for inclusion in standard diagnostic protocol (Laureys et al., 2004; Owen, 2013). Determining which methods have the greatest diagnostic accuracy, and hence used in clinical practice, requires clinical validation.

To date, investigators have applied the orthodox approach of clinical validation to neuroimaging. The orthodox approach involves determining the diagnostic accuracy of one or more novel diagnostic tests by comparing them to a gold standard. A gold standard is a test with the highest regarded diagnostic accuracy for a target condition that is stipulated *pro tem* for a particular study. The diagnostic accuracy of a novel test is estimated by calculating the ratio of false positives and false negatives it produces as compared to the gold standard. A false positive, or type I error, occurs when an evaluated diagnostic test produces an erroneous positive result. A false negative, or type II error, occurs when an evaluated diagnostic test produces an erroneous negative result (see Table 1).

The rate of false positives and false negatives bears on the estimation of sensitivity and specificity. The sensitivity of a diagnostic test is a function of its ability to detect each and every instance of the target condition. Sensitivity is expressed formally as:

$$\text{Sensitivity} = nTP / (nTP + nFN)$$

As the rate of false negatives increase, the sensitivity of a diagnostic test is reduced. By contrast, the specificity of a diagnostic test is a measure of its ability to uniquely detect a target condition. Specificity is expressed formally as:

$$\text{Specificity} = nTN / (nTN + nFP)$$

As false positives increase, the specificity of a diagnostic test is reduced (see Table 1).

The goal of clinical validation is to determine which diagnostic test has the greatest diagnostic accuracy. Tests with the greatest diagnostic accuracy are ultimately translated into clinical practice. Importantly, there is no single cut off of diagnostic accuracy required for any diagnostic test to be accepted in diagnostic protocol. Ideally, a novel diagnostic test should optimize a clinician's ability to detect the majority of cases of a target condition. Nevertheless, in practice, clinicians must be aware of the possibility of misdiagnosis and strike a balance between the risks of different diagnostic errors when formulating a general picture of a patient's condition (Peterson et al., 2015).

The orthodox approach is fruitful in many areas of medicine. In these domains, the science has sufficiently matured and there is a clear relationship between gold standard diagnostic methods and the mechanism of disease. In cardiology, for example, extensive knowledge of anatomical and electromechanical properties that contribute to cardiac function underlie interpretations of diagnostic tests. This allows for a clear understanding of the mechanisms that contribute to heart failure or cardiac ischemia, and how the results of different diagnostic tests corroborate each other. By contrast, theory and

Table 1. Standard formulae for estimating diagnostic accuracy

Test outcome	Positive result on gold standard	Negative result on gold standard
Positive result on novel test	True positive	False positive
Negative result on novel test	False negative	True negative
Calculations	Sensitivity = $nTP/(nTP + nFN)$	Specificity = $nTN/(nTN + nFP)$

measurement in the science of consciousness are still in their infancy. There is no consensus as to the conceptual nature of consciousness, nor the essential measurable phenomenon that contribute to the realization of consciousness. This hinders attempts to clinically validate neuroimaging methods in global disorders of consciousness. As we shall see, an alternative approach is needed to determine which diagnostic methods are most accurate.

Two Critical Questions Raised by the Orthodox Approach

The orthodox approach to clinical validation raises several perennial philosophical questions about the nature and accuracy of measurement. How do we know our instruments are accurately measuring a target phenomenon? And, when validating an instrument, how do we know that the gold standard is itself valid? (for detailed discussion, see [Chang, 2004](#); [Chang and Cartwright, 2008](#)). These questions cast an important, critical light on the design of clinical validation studies in the science of consciousness. In what follows, I explicate two critical questions raised by the orthodox approach when applied to neuroimaging: (i) What exactly is being validated? and (ii) What counts as a gold standard? I then outline an alternative approach to clinical validation based on reasoning by consilience.

What exactly is being validated?

A first question raised by application of the orthodox approach to neuroimaging is, what exactly is being validated? Each neuroimaging method used to evaluate global disorders of consciousness involves a complex combination of imaging modality, analytic technique, and task design. Of these components, each requires a particular kind of evidence to produce positive results. fMRI and PET, for example, track different neural phenomena. Likewise, different task designs may place varying demands on a patient; fMRI mental imagery involves a cognitively demanding task while other neuroimaging methods do not require task compliance. Such components—or combinations thereof—bear directly on a method's diagnostic accuracy. Without explicating how these components lend themselves to diagnostic accuracy, clinical validation may produce misleading results.

To date, little work has been done to track the relevant differences between components of neuroimaging methods and their diagnostic accuracy. This has, in part, led to systematic errors in clinical validation in the science of consciousness. For example, in the first large-scale clinical validation study of its kind, [Stender et al. \(2014\)](#) compared ^{18}F -FDG PET measurement of brain metabolism against fMRI mental imagery, using the CRS-R as a gold standard. [Stender et al.](#) reported that ^{18}F -FDG PET was 93% sensitive to the CRS-R diagnosis of the minimally conscious state, while the sensitivity of fMRI mental imagery was a mere 45%. Based on these findings, [Stender et al.](#) concluded that, “mental imagery fMRI is less reliable for differential

diagnostic purposes than ... ^{18}F -FDG PET” ([Stender et al., 2014](#), p. 519).

Two meta-analyses published the following year drew similar comparisons. [Bender et al. \(2015\)](#) compared fMRI-based methods and electroencephalographic-based methods. The sensitivity and specificity of all reviewed fMRI-based methods, including those assessing mental imagery ([Monti et al., 2010](#)), passive speech processing ([Coleman et al., 2009; 2007](#)), and resting-state networks ([Demertzi et al., 2014](#)), were aggregated to estimate overall diagnostic accuracy. [Bender et al.](#) concluded that, “the sensitivity and specificity of functional MRI-based techniques are 44% and 67%, respectively; [while] those of quantitative [electroencephalography] are 90% and 80%, respectively” ([Bender et al., 2015](#), p. 235). Similarly, [Kondziella et al., \(2016\)](#) compared both neuroimaging and electroencephalographic-based methods according to task design. They concluded that methods which did not require the patient to perform a task “... suggested preserved consciousness more often than active paradigms” ([Kondziella et al., 2016](#), p. 1).

These studies rest on at least two conceptual errors. First, there is ambiguity in the target of clinical validation. Are these studies comparing the diagnostic accuracy of imaging modality, data analysis, task design, or some combination thereof? [Bender et al. \(2015\)](#) individuate methods according to imaging modality. Yet, in doing so, they overlook differences in task designs that are distributed across and within imaging modality categories. This suggests that the sensitivity and specificity calculated reflect—at most—the diagnostic accuracy of a particular imaging modality rather than the broader neuroimaging or electroencephalographic-based method.

A second, related error is confusion over the inferential grounds of different methods. The task designs deployed by different methods are relevant for clinical validation, as each elicits evidence of more or less strength for the attribution of awareness. [Kondziella et al., \(2016\)](#) meta-analysis is an improvement in this regard, as it individuates neuroimaging and electroencephalographic-based methods according to canonical task categories—namely, active paradigms, passive paradigms, and resting-state paradigms (cf. [Laureys and Schiff, 2012](#)). Yet, there is a lingering confusion regarding inferential grounds. Active paradigms produce very strong evidence of awareness while passive paradigms produce evidence that is relatively weaker. This difference is due, in part, to the role that awareness plays in task performance (cf. [Bayne, 2013](#)); volitional task performance requires awareness while passive sensory processing, in most cases, does not (cf. [Shea and Bayne, 2010](#); [Bayne and Hohwy, 2014](#)). Confusion arises for [Kondziella et al.](#) when they claim that, “... passive paradigms suggested preserved consciousness more often than active paradigms” ([Kondziella et al., 2016](#), p. 1, emphasis added). Since the passive paradigms assessed in [Kondziella et al.](#)'s study do not necessarily warrant the attribution of awareness (or consciousness) as active paradigms do, it is unclear how they arrive at this conclusion.

This confusion regarding inferential grounds is sharpened by several critiques following the publication of [Stender et al.](#)'s

(2014) study. Sleigh and Warnaby, for example, take issue with the assertion that awareness can be inferred from ^{18}F -FDG PET measurement of metabolic activity. They argue that:

We do not, and perhaps cannot, know if the presence of [metabolic activity] is both a sufficient and necessary cause of consciousness. The converse question is: are all patients with unresponsive wakefulness syndrome who have depressed brain metabolism actually unconscious? (Sleigh and Warnaby, 2014, p. 476)

Additionally, Owen argues further that:

... the techniques employed by Stender *et al.* have fundamental differences that render any direct comparison, in terms of diagnostic utility, inappropriate. ^{18}F -FDG PET directly measures the metabolic integrity of cortical networks believed to underpin consciousness, while fMRI mental imagery indirectly demonstrates consciousness by defining awareness as intentional neural modulation (or neural 'command following'). Crucially, metabolic integrity of cortical networks is necessary for consciousness, but does not guarantee it. By contrast, intentional neural modulation ... is ... sufficient to confirm consciousness in the absence of overt behavioural command following. (Owen, 2014, p. 370)

Problems arise for clinical validation if neuroimaging methods are compared without acknowledging these possible errors. Either there is ambiguity in the target of clinical validation, or there is confusion regarding the inferential grounds of the compared neuroimaging methods. In either case, diagnostic accuracy is obfuscated. Methods that are simply easier to satisfy may be thought to have superior diagnostic accuracy. Yet, on close inspection, such methods may produce evidence that is relatively weak—so weak, in fact, that attribution of awareness may not be warranted.

What counts as a gold standard?

A second question raised by the orthodox approach is determining what counts as a gold standard. Recall that, according to the orthodox approach, a gold standard is a test with the highest regarded diagnostic accuracy for a target condition that is stipulated *pro tem* for a particular study. The selected gold standard must, therefore, be reasonably regarded as more sensitive and specific than the evaluated diagnostic tests. If it is not, false positive and false negative rates cannot be accurately estimated.

It is, however, possible to discover that a novel diagnostic test has greater diagnostic accuracy than a selected gold standard. After all, this is how diagnostic precision evolves. Nevertheless, it is mysterious how this occurs in the context of the orthodox approach. If it is discovered that the diagnostic accuracy of a novel test is greater than the stipulated gold standard, then the novel diagnostic test is being compared to something else. But what is this something else? This raises deep philosophical questions about how the validity of the gold standard is determined in the first place.

How, then, is a gold standard selected according to the orthodox approach when applied to neuroimaging methods? To ensure accurate estimation of diagnostic accuracy, investigators should endeavor to avoid selecting a gold standard that is already regarded as less sensitive or specific than the evaluated diagnostic tests. Nevertheless, a commonly selected gold standard is the CRS-R, primarily on the grounds of inter-rater reliability and criterion validity. Stender *et al.*, for example, argue that:

The only broadly accepted test for ... awareness is behavioural responsiveness. However, the reliability of the behavioural reference

is a key issue. The diagnostic inter-rater agreement of one CRS-R assessment ranges between 89% and 100%, and inter-rater reliability on two subsequent days is roughly 95%. Thus, CRS-R is a robust method, even for one assessment. Serial CRS-R assessments by several experienced raters ensured a highly reliable clinical diagnosis. (Stender *et al.*, 2014, pp. 6–7)

Despite this justification, use of the CRS-R as a gold standard for evaluating neuroimaging is problematic. A number of carefully conducted studies have shown that a proportion of brain injured patients who consistently satisfy the CRS-R criteria of the vegetative state are able to volitionally modulate their brain activity to command (see, among other studies, Owen *et al.*, 2006; Monti *et al.*, 2010; Goldfine *et al.*, 2011; Cruse *et al.*, 2012; Fernández-Espejo and Owen, 2013; Naci and Owen, 2013). This suggests that the CRS-R, as with all neurobehavioral exams, are in principle insensitive to covert awareness. The central problem is a conflict in the way diagnostic categories are articulated according to neurobehavioral scales versus neuroimaging. Neurobehavioral scales individuate diagnoses according to behaviorally mediated evidence of awareness. Yet, neuroimaging individuates diagnoses according to the mere presence or absence of awareness, unmediated by behavior. While it may be convention to use the CRS-R for validation, there is a conceptual error in using it to validate neuroimaging.

Since neuroimaging methods appear, in some cases, to be more sensitive to awareness than neurobehavioral evaluation, it may be argued that a task-driven neuroimaging method, such as fMRI mental imagery, could serve as a gold standard to validate both novel neuroimaging methods and neurobehavioral scales. After all, if a participant can satisfy a neurobehavioral scale, there seems to be no reason, in principle, why she could not also perform mental imagery. But this approach is also problematic. Some brain injured patients who satisfy neurobehavioral criteria of the minimally conscious state are unable to perform fMRI mental imagery. Monti *et al.* (2010) found only 1 of 31 minimally conscious patients could perform fMRI mental imagery upon instruction. Likewise, Stender *et al.* (2014) found similar difficulties in eliciting mental imagery from a group of minimally conscious patients. These results may be explained by patient fatigue, inhibition in sustained attention, deficits in “high-level cognition” required for mental imagery, or technical challenges inherent to neuroimaging, such as artifacts generated by uncontrolled movement (Naci *et al.*, 2012, p. 316; Stender *et al.*, 2014, p. 519). Nonetheless, these confounding factors demonstrate that a task-driven neuroimaging method is also inappropriate to use as a gold standard.

A third alternative for a gold standard is to use a neuroimaging method that measures a neural correlate of awareness. Promising work on the functional and structural integrity of intrinsic cortical networks and metabolic rates of glucose may reveal a neural correlate of awareness with high diagnostic accuracy (Laureys *et al.*, 2004; Boly *et al.*, 2009; Vanhaudenhuyse *et al.*, 2009; Demertzi *et al.*, 2013; Fernández-Espejo *et al.*, 2015; Stender *et al.*, 2015). This approach is favorable because it avoids problems generated by neurobehavioral exams and task-driven neuroimaging methods; it can identify awareness in patients regardless of motor or cognitive deficits, and is not contingent upon the subjective interpretation of neurobehavioral examination. Yet, this approach also engenders methodological problems.

To date, methods for detecting a neural correlate of awareness are, in some cases, insensitive to patients known to be aware. Stender *et al.* (2014), for example, found that ^{18}F -FDG PET measurement of glucose metabolism was not sensitive to all

study participants clinically diagnosed as minimally conscious. Likewise, [Fernández-Espejo et al. \(2011\)](#) and [Demertzi et al. \(2015\)](#) have shown that methods assessing functional and structural correlates of awareness, while highly accurate, are less than 100% sensitive to awareness in clinical populations. Such methods may, in the future, be the best alternative for a gold standard, yet it is unclear precisely when—or how—they will be optimized for this purpose.

The general difficulty in identifying a gold standard in the science of consciousness is that each method of assessment systematically overlooks a proportion of patients in whom we have good reason to believe are aware. Neurobehavioral examination is, in principle, insensitive to awareness unmediated by behavior, while task-driven neuroimaging methods are insensitive to some patients presumed to be aware according to neurobehavioral evaluation. Meanwhile, methods for evaluating a neural correlate of awareness may resolve this problem, but such methods still require optimization.

These problems generate a significant challenge for clinical validation. Indeed, there appears to be no single method—and no single imaging modality—that could be reasonably regarded as a gold standard. This problem is strongly felt within the empirical literature. [Stender et al.](#), for example, admit that:

Because no gold standard exists for absence of consciousness, sensitivity to unresponsive wakefulness syndrome or specificity to minimally conscious states seem like meaningless measures. ([Stender et al., 2014](#), p. 7)

Likewise, [Kondizella et al.](#) observe that:

In the absence of a gold standard for consciousness, precise estimates of sensitivity and specificity of active and passive paradigms are futile. ([Kondziella et al., 2016](#), p. 5)

Additionally, even if there were such a method, there are deeper philosophical problems regarding the validity of the gold standard itself. How do we know that the gold standard is valid? And, according to what is its validity confirmed? This line of questioning threatens a viscous epistemic regress.

Finally, there is lingering tension over the compatibility of diagnostic categories as understood by neurobehavioral evaluation versus neuroimaging. Neurobehavioral evaluation individuates diagnoses according to behavioral evidence of awareness, while neuroimaging individuates diagnoses according to the mere presence or absence of awareness. There is an inherent problem, then, in evaluating the diagnostic accuracy of neuroimaging on the basis of neurobehavioral examination. Any evaluation of the validity of novel method on the basis of neurobehavioral examination will generate a systematic bias toward behaviorally individuated diagnostic categories.

The Consilience Approach

One way to avoid the forgoing problems is to reject the orthodox approach in favor of an alternative approach motivated by reasoning by consilience. Consilience is a mode of reasoning that assigns a degree of plausibility to a hypothesis based on support by a diverse set of evidence from independent sources (cf. [Vezér, 2015](#)). This approach has been instrumental for determining the status of hypotheses with little or no independent means of confirmation, including hypotheses in climate science ([Oreskes, 2007](#); [Lloyd, 2015](#)), cosmology ([Harper, 1989](#)), and evolutionary biology ([Gould, 2002](#)). Applied in the science of consciousness, reasoning by consilience may allow investigators to assign a degree of plausibility to the results of a novel method

by comparison to other, methodologically distinct tests. No single test would function as a gold standard. Rather, the degree to which results are conciliate or discordant with a patient's broader assessment would serve as a starting point for clinical validation.

This approach has been applied—at least, implicitly—in several studies evaluating the diagnostic accuracy of neuroimaging and EEG methods. [Forgacs et al. \(2014\)](#) recently compared performance on fMRI mental imagery with preservation of metabolic activity and electrophysiological background organization during periods of wakefulness and sleep. They found that, in global disorders of consciousness patients, performance of fMRI mental imagery was highly associated with background organization during wakefulness, sleep spindle activity, and relatively normal metabolism. Likewise, [Gibson et al. \(2014\)](#) applied a variety of fMRI and EEG-based methods in a group of patients to determine where performance overlapped and diverged. Additionally, [Sitt et al. \(2014\)](#) compared a variety of electrophysiological markers of awareness in a single patient cohort to develop an EEG-based screening technique that differentiates the vegetative from the minimally conscious state. Finally, [Di Perri et al. \(2016\)](#) recently compared resting-state functional connectivity of the default mode network with brain metabolism in patients with global disorders of consciousness. They found that brain metabolism was correlated with functional connectivity, and that partial preservation of between-network anti-correlations characterized the recovery of consciousness. These studies suggest that a variety of evidence from independent sources may provide mutual support for a diagnostic hypothesis, or the accuracy of a novel neuroimaging or EEG-based method.

The consilience approach also resolves critical questions raised by the orthodox approach. First, there is no ambiguity regarding the target of validation. Because the consilience approach compares a novel method against a variety of evidence, there is no need to isolate a particular component of a method for validation, nor explicitly match the inferential grounds of tasks used in a method. While it may be organizationally helpful to specify particular kinds of consilience—for example, consilience between active paradigms; consilience across active, passive, and resting-state paradigms; or consilience across imaging modalities—this does not bear directly on prospects for validating a novel method according to the consilience approach. Simply put, variation in underlying components of neuroimaging methods is irrelevant. It is only the degree to which a method's results are conciliate or discordant with a broader set of evidence from multiple, independent sources that counts.

Additionally, the consilience approach resolves questions surrounding the potential bias of a gold standard. Recall that there appears to be no single diagnostic test for global disorders of consciousness that reasonably satisfies the epistemic requirements of a gold standard. All tests systematically overlook a group of patients for whom we have good reason to believe are aware. The consilience approach cuts through this problem by deriving the epistemic strength of a novel method from its fit with a broader set of evidence. This accounts for and corrects potential biases in clinical validation. Neurobehavioral examination, for example, is biased due to its inherent insensitivity to awareness unmediated by behavior. Likewise, the tasks involved in different neuroimaging methods are biased due to variations in their cognitive demand. Importantly, these are biases of a different kind. When aggregated, the results of different methods naturally compliment each other by counterbalancing respective biases. This prevents the exclusion of relevant diagnostic evidence that can result from a biased gold standard.

The consilience approach also speaks to philosophical worries regarding the presumed validity of a gold standard. The fact that consilience does not rely on a single gold standard for clinical validation resolves the problem of determining how the gold standard is validated in the first place. It may be that, in future research, the consilience approach produces a single gold standard. Neuroimaging methods that track functional and structural changes associated with the loss of consciousness may derive their epistemic strength from their consilience or discordance with a broader set of evidence. Additionally, Tomaiuolo *et al.* (2016), Noirhomme *et al.* (2015), and Stender *et al.* (2014), have argued that, among other comparisons, longitudinal evaluation of patients who recover awareness may identify which neuroimaging methods are most accurate. Such methods may, in turn, be used as a gold standard in the future, but only based on their prior consilience with, among other evidence, patient outcome. This provides a plausible story for how a gold standard could be identified as the science of consciousness matures.

Challenges to the consilience approach

The consilience approach is methodologically attractive, yet it raises a number of unique challenges. First, discordant results may be difficult to reconcile and, in turn, may undermine validation by consilience. Gibson *et al.* (2014) recently demonstrated that a single patient could return discordant results when evaluated with different modalities (fMRI vs. EEG) and tasks (motor imagery vs. spatial navigation imagery). Such results can complicate a clear understanding of how different methods are conciliate or discordant. Discordant results may be explained by practical problems occurring during assessment (Fischer and Truog, 2015). Yet, it is also possible that neurobiological deficits secondary to brain injury can cause variations in a patient's ability to satisfy different neuroimaging methods (Naci *et al.*, 2012, p. 316). This suggests there are different types of fit between evidence. What type of fit is the right fit may hinge on explaining why discordant results occur.

Another challenge is determining what information is included and excluded from the set of evidence used for comparison with a novel method. Among other methods of assessment, should the set include or exclude neurobehavioral findings, structural neuroimaging, or resting-state EEG? Moreover, on what grounds is this evidence included or excluded? Is it due to convention, or because the evidence has been rigorously shown to be diagnostically valid? The scope of evidence included may likely be based on normative rationale. Yet, from an empirical standpoint, it may remain unclear if and how this rationale is justified.

A third challenge is determining what the degree of consilience or discordance of evidence actually tells us about the results of a novel method. If, for example, the results of a novel neuroimaging method are highly conciliate with other methods of assessment, does this mean the novel method is accurately tracking a patient's "true" preserved consciousness or that there is merely a high degree of coherence between different methods of assessment? This difference is important, as medical decisions are contingent upon whether a patient is actually conscious, not the mere coherence of different methods. This problem may be resolved by comparing a set of conciliate results against patient outcome (cf. Stender *et al.*, 2014; Noirhomme *et al.*, 2015; Stender *et al.*, 2014); if patients recover, then the set of results is, in some fashion, tracking consciousness. However, this presumes that patients will recover, and in

many cases this does not occur (cf. Fernández-Espejo and Owen, 2013; Naci *et al.*, 2014). Whether it is possible to be certain of a patient's consciousness beyond the mere coherence of independent assays of brain activity is unclear. It is highly desirable, yet under the present analysis it appears difficult, if not impossible, to attain.

These challenges are difficult to resolve. However, they do not demonstrate the consilience approach should be abandoned. While the consilience approach raises its own challenges, these are potentially resolvable with future philosophical inquiry. The problem of discordant findings may be resolved by a fine-grained analysis of the aspects of consciousness that different methods of assessment probe. For example, Bayne and Hohwy (2015) and Bayne *et al.* (2016) recently proposed a multidimensional model that tracks the recovery of aspects of consciousness, rather than levels of consciousness, following brain injury. Certain aspects of consciousness may be differentially important to satisfying particular neuroimaging methods. Variation of preserved aspects of consciousness may explain why a single patient is able to satisfy some methods but not others (see also Klein and Howhy, 2015).

Additionally, the problem of including or excluding methods in a comparison set may be resolved by explicating the rationale—whether normative or empirical—for inclusion or exclusion. To date, there has been no systematic analysis of this issue. The design of future clinical validation studies may benefit from making this rationale explicit.

Finally, the problem of determining what consilience tells us about the actual recovery of consciousness may be resolved by including patient outcome in the comparison set. To be sure, outcome data may tell us more about prognosis than diagnosis. Nonetheless, comparison with standardized outcome scales, such as the Glasgow Outcome Scale, can provide compelling evidence that a novel neuroimaging method is accurately tracking the recovery of consciousness, provided a patient does not have motor deficits (cf. Stender *et al.*, 2014, 2016; 2014). This information may further corroborate diagnoses derived from conciliate evidence.

The consilience approach is—at least, provisionally—the best alternative to the orthodox approach for clinical validation in the science of consciousness. In the long run, the consilience approach may eventually yield a gold standard. Multiple lines of evidence may converge on a single test for preserved consciousness. Yet, because no single test currently satisfies the epistemic criteria of a gold standard, investigators must consider how data generated by a novel method fits with a broader set of evidence from multiple, independent sources. Explicating the details of the consilience approach with respect to the science of consciousness can optimize future clinical validation study designs.

Conclusion

Patients with global disorders of consciousness are highly vulnerable. Providing the right diagnosis is imperative, yet diagnostic accuracy of global disorders of consciousness remains one of the most challenging obstacles of modern medicine. Novel methods of assessment may offer a solution. Yet these methods require clinical validation before inclusion in standard diagnostic protocol.

In this article, I have argued that the orthodox approach to clinical validation generates a number of challenges when applied to the science of consciousness. In response, I proposed the consilience approach. This approach estimates the

diagnostic accuracy of a novel method based on its degree of consistency with a variety of evidence from multiple, independent sources. The focus of this article has been on methodological problems in clinical validation of diagnostic tests for global disorders of consciousness. Yet, diagnostic accuracy of other disorders of consciousness may also benefit from reflection on the consistency approach. Study of other transitory and persistent conditions, including absence seizures, toxic encephalopathy, transient global amnesia, pervasive developmental disorders, and a wide range of psychiatric syndromes, is important to our overall understanding of human consciousness. The consistency approach may be fruitful in these domains if similar methodological challenges to clinical validation arise.

The consistency approach raises a number of unique challenges that are in need of further consideration, including how to explain divergent findings, defining the scope of evidence to compare novel methods to, and determining whether consistency is a function of a method's ability to track a patient's true consciousness. Apart from these challenges, the consistency approach also requires formalization: How is clinical evidence aggregated under the consistency approach?; Is the evidential weight different across various methods of assessment?; And, how might consistency apply to other imaging modalities? Reflection on these questions may optimize future clinical validation study designs in the science of consciousness and clarify the precise role of neuroimaging and EEG in differential diagnosis of global disorders of consciousness.

Acknowledgments

I am grateful to Tim Bayne, Charles Weijer, Adrian Owen, Damian Cruse, Lorina Naci, and the neuroethics research group at the Rotman Institute of Philosophy for valuable conversations related to this paper. Earlier versions of this paper were presented at the annual meetings of the Associations for the Scientific Study of Consciousness, in 2015, and the International Neuroethics Society, in 2014. This work is supported by a Vanier Canada Graduate Scholarship and the Rotman Institute of Philosophy.

References

- Andrews K, Murphy L, Munday R, et al. Misdiagnosis of the vegetative state: retrospective study in a rehabilitation unit. *BMJ* 1996;313:13–16.
- Bayne T. Agency as a marker of consciousness. In: Clark, Kiverstein, and Vierkant (eds), *Decomposing the Will*. New York: Oxford University Press, 2013, 160–82.
- Bayne T and Hohwy J. Global disorders of consciousness. *WIREs Cogn Sci* 2014;5:129–38.
- Bayne T and Hohwy J. Modes of consciousness. In: Sinnott-Armstrong (ed), *Consciousness after Severe Brain Damage: Medical, Legal, Ethical, and Philosophical Perspectives*. New York: Oxford University Press, 2015, 57–82.
- Bayne T, Hohwy J and Owen AM. Are there levels of consciousness? *Trends Cogn Sci* 2016;20:405–13.
- Bender A, Jox RJ, Grill E, et al. Persistent vegetative state and minimally conscious state: a systematic review and meta-analysis of diagnostic procedures. *Deutsches Ärzteblatt International* 2015;112:235–42.
- Boly M, Coleman MR, Davis MH, et al. When thoughts become action: an fMRI paradigm to study volitional brain activity in non-communicative brain injured patients. *Neuroimage* 2007;36:979–92.
- Boly M, Tshibanda L, Vanhaudenhuyse A, et al. Functional connectivity in the default network during resting state is preserved in a vegetative but not in a brain dead patient. *Hum Brain Mapp* 2009;30:2393–400.
- Chang H. *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press, 2004.
- Chang H, Cartwright N. Measurement. In: Psillos S. and Curd M. (eds.), *The Routledge Companion to Philosophy of Science*. New York: Routledge, 2008, 367–75.
- Chennu S, Finoia P, Kamau E, et al. Dissociable endogenous and exogenous attention in disorders of consciousness. *NeuroImage: Clin* 2013;3:450–61.
- Childs NL, Mercer WN, Childs HW, et al. Accuracy of diagnosis of persistent vegetative state. *Neurology* 1993;43:1465–467.
- Coleman MR, Davis MH, Rodd JM, et al. Towards the routine use of brain imaging to aid the clinical diagnosis of disorders of consciousness. *Brain* 2009;132:2541–52.
- Coleman MR, Rodd JM, Davis MH, et al. Do vegetative patients retain aspects of language comprehension? Evidence from fMRI. *Brain* 2007;130:2494–507.
- Cruse D, Chennu S, Chatelle C, et al. Bedside detection of awareness in the vegetative state: a cohort study. *The Lancet* 2012;378:2088–94.
- Di H, Boly M, Weng X, et al. Neuroimaging activation studies in the vegetative state: predictors of recovery? *Clin Med* 2008;8:502–07.
- Di Perri C, Bahri MA, Amico E, et al. Neural correlates of consciousness in patients who have emerged from a minimally conscious state: a cross-sectional multimodal imaging study. *Lancet Neurol*; 2016;15:830–42.
- Dikmen SS, Machamer JE, Powell JM, et al. Outcome 3 to 5 years after moderate to severe traumatic brain injury. *Arch Phys Med Rehabil*; 2003;84:1449–57.
- Demertzi A, Soddu A, Laureys S, et al. Consciousness supporting networks. *Curr Opin Neurobiol*; 2013;23:239–44.
- Demertzi A, Gómez F, Crone JS, et al. Multiple fMRI system-level baseline connectivity is disrupted in patients with consciousness alterations. *Cortex* 2014;52:35–46.
- Demertzi A, Antonopoulos G, Heine L, et al. Intrinsic functional connectivity differentiates minimally conscious from unresponsive patients. *Brain* 2015;138:2619–31.
- Fernández-Espejo D, Bekinschtein T, Monti MM, et al. Diffusion weighted imaging distinguishes the vegetative state from the minimally conscious state. *Neuroimage* 2011;54:103–12.
- Fernández-Espejo D, Owen AM. Detecting awareness after severe brain injury. *Nat Rev Neurosci* 2013;14:801–09.
- Fernández-Espejo D, Rossit S, Owen AM. A thalamocortical mechanism for the absence of overt motor behavior in covertly aware patients. *JAMA Neurol* 2015;72:1442–50.
- Fischer DB and Truog RD. What is a reflex? A guide for understanding disorders of consciousness. *Neurology* 2015;85:543–48.
- Forgacs PB, Conte MM, Fridman EA, et al. Preservation of electroencephalographic organization in patients with impaired consciousness and imaging-based evidence of command-following. *Ann Neurol* 2014;76:869–79.
- Giacino JT, Ashwal S, Childs N, et al. The minimally conscious state definition and diagnostic criteria. *Neurology* 2002;58:349–53.
- Giacino JT, Kalmar K, Whyte J. The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. *Arch Phys Med Rehabil* 2004;85:2020–29.
- Giacino JT, Fins JJ, Laureys S, et al. Disorders of consciousness after acquired brain injury: the state of the science. *Nat Rev Neurol* 2014;10:99–114.

- Gibson RM, Fernández-Espejo D, Gonzalez-Lara LE, et al. Multiple tasks and neuroimaging modalities increase the likelihood of detecting covert awareness in patients with disorders of consciousness. *Front Hum Neurosci* 2014;**8**:950.
- Goldfine AM, Victor JD, Conte MM, et al. Determination of awareness in patients with severe brain injury using EEG power spectral analysis. *Clin Neurophysiol* 2011;**122**:2157–68.
- Gould. *The Structure of Evolutionary Theory*. Boston: Harvard University Press, 2002.
- Greenwald BD, Burnett DM, Miller MA, et al. Congenital and acquired brain injury. *Arch Phys Med Rehabil* 2003;**84**:S3–S7.
- Harper W. Consilience and natural kind reasoning. *Boston Stud Philos Sci* 1989;**116**:115–52.
- Kondziella D, Friberg CK, Frokjaer VG, et al. Preserved consciousness in vegetative and minimal conscious states: systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry* 2016;**7**:485–92.
- King JR, Faugeras F, Gramfort A, et al. Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *Neuroimage* 2013;**83**:726–38.
- Klein C and Hohwy J. Variability, convergence, and dimensions of consciousness. In Overgaard M (ed), *Behavioral Methods in Consciousness Research*. New York: Oxford University Press, 2015, 249–63.
- Laureys S, Owen AM, Schiff ND, et al. Brain function in coma, vegetative state, and related disorders. *Lancet Neurol* 2004;**3**:537–46.
- Laureys S, Celesia GG, Cohadon F, et al. Unresponsive wakefulness syndrome: a new name for the vegetative state or apallic syndrome. *BMC Med* 2010;**8**:68.
- Laureys S and Schiff ND. Coma and consciousness: paradigms (re) framed by neuroimaging. *Neuroimage* 2012;**6**:478–91.
- Lloyd EA. Model robustness as a confirmatory virtue: the case of climate science. *Stud Hist Philos Sci* 2015;**49**:58–68.
- Monti MM, Vanhaudenhuyse A, Coleman MR, et al. Willful modulation of brain activity in disorders of consciousness. *New Engl J Med* 2010;**362**:579–89.
- Multi-Society Task Force on PVS. Medical aspects of the persistent vegetative state. *New Engl J Med* 1994;**330**:1499–508.
- Naci L, Monti MM, Cruse D, et al. Brain-computer interfaces for communication with nonresponsive patients. *Ann Neurol* 2012;**72**:312–23.
- Naci L, Cusack R, Jia VZ, et al. The brain's silent messenger: using selective attention to decode human thought for brain-based communication. *J Neurosci* 2013;**33**:9385–93.
- Naci L and Owen AM. Making every word count for nonresponsive patients. *JAMA Neurology* 2013;**70**:1235–41.
- Naci L, Cusack R, Anello M, et al. A common neural code for similar conscious experiences in different individuals. *PNAS* 2014;**111**:14277–82.
- Nakayama N, Okumura A, Shinoda J, et al. Relationship between regional cerebral metabolism and consciousness disturbance in traumatic diffuse brain injury without large focal lesions: an FDG-PET study with statistical parametric mapping analysis. *J Neurol Neurosurg Psychiatry* 2006;**77**:856–62.
- Noirhomme Q, Brecheisen R, Lesenfans D, et al. “Look at my classifier's result”: Disentangling unresponsive from (minimally) conscious patients. *Neuroimage* 2015, first published online Dec 12, 2015, 10.1016/j.neuroimage.2015.12.006.
- Norton L, Hutchinson RM, Young GB, et al. Disruptions of functional connectivity in the default mode network of comatose patients. *Neurology* 2012;**78**:175–81.
- Oreskes N. The scientific consensus on climate change: How do we know we're not wrong? In: DiMento JFC and Doughman P (eds), *Climate Change: What it Means for us, our Children, and our Grandchildren*. Boston: MIT Press, 2007, 65–99.
- Owen AM, Menon DK, Johnsrude IS, et al. Detecting residual cognitive function in persistent vegetative state. *Neurocase* 2002;**8**:394–403.
- Owen AM, Coleman MR, Boly M, et al. Detecting awareness in the vegetative state. *Science* 2006;**313**:1402–402.
- Owen AM. Detecting consciousness: a unique role for neuroimaging. *Ann Rev Psychol* 2013;**64**:109–33.
- Owen AM. Disorders of consciousness: diagnostic accuracy of brain imaging in the vegetative state. *Nat Rev Neurol* 2014;**10**:370–71.
- Peterson A, Cruse D, Naci L, et al. Risk, diagnostic error, and the clinical science of consciousness. *NeuroImage: Clin* 2015;**7**:588–97.
- Plum F and Posner JB. *The Diagnosis of Stupor and Coma*. Vol. 19. New York: Oxford University Press, 1982.
- Schnakers C, Vanhaudenhuyse A, Giacino J, et al. Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. *BMC Neurol* 2009;**9**:35.
- Seel RT, Sherer M, Whyte J, et al. Assessment scales for disorders of consciousness: evidence-based recommendations for clinical practice and research. *Arch Phys Med Rehabil* 2010;**91**:1795–813.
- Sleigh J and Warnaby CE. Functional brain imaging: gatecrashing the clinical party? *Lancet* 2014;**384**:476–77.
- Shea N and Bayne T. The vegetative state and the science of consciousness. *Brit J Phil Sci* 2010;**61**:459–84.
- Schiff ND. Global disorders of consciousness. In: Velmans M, Schneider S. (eds.), *The Blackwell Companion to Consciousness*, Malden: Blackwell, 2007, 589–604.
- Sitt JD, King JR, El Karoui, et al. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain* 2014;**137**:2258–70.
- Stender J, Gosseries O, Bruno MA, et al. Diagnostic precision of PET imaging and functional MRI in disorders of consciousness: a clinical validation study. *The Lancet* 2014;**384**:514–22.
- Stender J, Kupers R, Rodell A, et al. Quantitative rates of brain glucose metabolism distinguish minimally conscious from vegetative state patients. *J Cereb Blood Flow Metab* 2015;**35**:58–65.
- Stender J, Mortensen KN, Thibaut A, et al. The minimal energetic requirements of sustained awareness after brain injury. *Curr Biol* 2016;**26**:1494–99.
- Thibaut A, Bruno MA, Chatelle C, et al. Metabolic activity in external and internal awareness networks in severely brain-damaged patients. *J Rehabil Med* 2012;**44**:487–94.
- Thurman D and Guerro J. Trends in hospitalization associated with traumatic brain injury. *JAMA* 1999;**282**:954–57.
- Tomaiuolo F, Cecchetti L, Gibson RM, et al. Progression from vegetative to minimally conscious state is associated with changes in brain neural response to passive tasks: a longitudinal single-case functional MRI study. *J Int Neuropsychol Soc* 2016;**22**:620–30.
- Vanhaudenhuyse A, Noirhomme Q, Tshibanda LJ, et al. Default network connectivity reflects the level of consciousness in non-communicative brain-damaged patients. *Brain* 2009;**133**:161–71.
- Vezer MA. Aggregating evidence in climate science: consilience, robustness and the wisdom of multiple models. *UWO Electronic Thesis and Dissertation Repository* 2015, Paper 2837.
- Weijer C, Bruni T, Gofton T, et al. Ethical considerations in functional magnetic resonance imaging research in acutely comatose patients. *Brain* 2016;**139**:292–99.
- Young GB. Coma. *Ann N Y Acad Sci* 2009;**1157**:32–47.