

The Vegetative State and the Science of Consciousness

Nicholas Shea & Tim Bayne*

Abstract

Consciousness in experimental subjects is typically inferred from reports and other forms of voluntary behaviour. A wealth of everyday experience confirms that healthy subjects do not ordinarily behave in these ways unless they are conscious. Investigation of consciousness in vegetative state (VS) patients has been based on the search for neural evidence that such broad functional capacities are preserved in some VS patients. We call this the standard approach. To date, the results of the standard approach have suggested that some VS patients might indeed be conscious, although they fall short of being demonstrative. The fact that some VS patients show evidence of consciousness according to the standard approach is remarkable, for the standard approach to consciousness is rather conservative, and leaves open the pressing question of how to ascertain whether patients who fail such tests are conscious or not. We argue for a cluster-based ‘natural kind’ methodology that is adequate to that task, both as a replacement for the approach that currently informs research into the presence or absence of consciousness in VS patients, and as a methodology for the science of consciousness more generally.

Contents

- 1 Introduction
- 2 The Vegetative State
- 3 The Standard Approach
- 4 The Natural Kind Methodology
- 5 Is Consciousness a Special Case?
 - 5.1 Is Consciousness a Natural Kind?
 - 5.2 A Special Obstacle?
- 6 Conclusion

* The paper is fully collaborative and the order of the authors’ names is arbitrary.

1. Introduction

The questions of when and on what basis we should we ascribe consciousness to others are two of the most vexing questions in the philosophy of mind and in the science of consciousness. In this paper we focus on these questions as they manifest themselves in the context of patients in the vegetative state (VS). Such patients have traditionally been assumed to lack consciousness, but a series of recent experiments has put that assumption in doubt. This paper examines the methodological assumptions behind those experiments, with a focus on the question of what it would take to discover evidence of consciousness (or its absence) in VS patients. Current approaches to the study of consciousness in VS—and, indeed, to the study of consciousness in general—are dominated by what we will call the ‘standard approach’. Proponents of this approach look for evidence that behavioural capacities that are pre-theoretically associated with consciousness—such as report and voluntary action—remain intact in VS. In the first half of this paper we draw on the standard approach, arguing that it provides a powerful but less than demonstrative case for thinking that at least some VS patients are conscious. We then turn to the merits of the standard approach itself, arguing that there is a more attractive alternative to it. This alternative approach—which we call the “natural kind methodology”—involves investigating whether consciousness is a natural kind. If it is, then we can study it in ways that go beyond the standard approach.

2. The Vegetative State

After a brain injury caused by trauma or hypoxia, patients may emerge from coma and enter the vegetative state (Bernat [2006]; Jennett [2002]; Plum and Posner [1982]; Schiff [2007]). Unlike coma patients, VS patients show a normal sleep-wake cycle, opening their eyes when awake and making roving eye movements. The crucial clinical signs of VS are negative: there should be no evidence of awareness of self or the environment, no responses to external stimuli of a kind that would suggest volition or purpose (as opposed to reflexes), and no

evidence of language expression or comprehension (Royal College of Physicians [2003], §2.2). A vegetative state is classified as permanent once there is no chance of the patient recovering from it: 12 months after a coma caused by a traumatic brain injury or 6 months after a coma caused by non-traumatic brain injury (e.g. hypoxia). Patients recovering from a brain injury often pass through VS for a short time when emerging from a coma before reaching full consciousness. Other patients remain in the VS for a longer time while retaining some chance of recovery. Although not a recognised diagnostic category, this state is commonly referred to as the ‘persistent vegetative state’.

Might (some) VS patients be conscious? The type of consciousness we have in mind here is phenomenal consciousness: the sort of consciousness that there is something that it is like to enjoy, from the subject’s point of view (Nagel [1974]). One might think that the answer to this question must be ‘no’. After all, VS is commonly said to be a state of ‘wakeful unconsciousness’. Furthermore, clinicians distinguish the vegetative state from the minimally conscious state (MCS),¹ a distinction that seems to presuppose that VS patients are not conscious in any way. If VS were characterised directly by reference to the absence of consciousness, any reason for thinking that an apparently VS patient was conscious would be evidence that the patient was not in fact in VS. That is not, however, how the clinical definition works. Clinicians use a set of agreed guidelines that tie diagnosis to a number of factors, and it remains an open epistemic possibility that patients who meet the guidelines in question might be conscious. We will leave to one side the question of whether patients who meet the current diagnosis of VS but also show signs of consciousness should receive their own diagnostic category (Fins & Schiff [2006]). Our main concern here is not taxonomy but the deeper issue of whether the patients in question—however described—might be conscious.

Of course, not all open possibilities are live possibilities. Are there positive reasons for thinking that some VS patients might actually be conscious? There are—but first some background.

Recent studies show that metabolic activity in the brains of VS patients is typically 40% that of neurologically intact individuals. This is comparable to the activity in the brains of healthy individuals under general anaesthetic, and quite unlike “brain dead” patients who have no metabolic neural activity (measured by PET: positron emission tomography). Other studies have shown preservation of some level of differentiated cortical activity in certain VS patients (e.g. islands of activity in resting state, Schiff [2002]), including in response to a story being read by a relative (PET: de Jong [1997]), the sound of the patient’s own name (fMRI: Staffen [2006], Di [2007]; EEG: Perrin [2006]), pictures of familiar faces (PET: Menon [1998]) and painful stimuli (fMRI: Laureys [2002]).

Although suggestive, these results have little direct bearing on the question of consciousness, for brain activation of this kind is routinely found in the context of unconscious cognitive processing (such as priming). More persuasive are the results reported by Owen et al. ([2005]) and Coleman et al. ([2007]), who found that VS patients show PET and fMRI responses to speech (contrasted with acoustically-matched noise) that correspond to those found in normal subjects, albeit at reduced overall levels of activation. Even more tantalizingly, fMRI studies have found evidence of semantic processing of ambiguous words in VS patients. When normal subjects hear potentially ambiguous sentences,² as compared to low-ambiguity sentences, they show higher activation of a well-characterised language network involving the superior and middle temporal gyrii and the left inferior frontal gyrus (Rodd [2005]). A few VS patients also show activation of this network in response to high-ambiguity but not low-ambiguity sentences.³ Strikingly, this differential response is abolished by even moderate levels of anaesthetic sedation in healthy volunteers (Davis et al. [2007]).

To date, the strongest published evidence for consciousness in the VS derives from an fMRI study of a 23 year old female victim of a car accident who had been in a vegetative state for 5 months (Owen et al. [2006]).⁴ The study involved two kinds of trials. On some trials the patient was played a pre-recorded instruction to imagine playing tennis; on other trials she was instructed to imagine visiting the rooms of her home. In each case, after 30 seconds she was told to relax (“now just relax”) and given a 30 second rest period. In this paradigm, the blood oxygen level-dependent (BOLD) signal from those brain areas preferentially involved in motor imagery and spatial navigation in the two conditions (compared with rest) was indistinguishable from that seen in 34 healthy volunteers: the instruction to imagine playing tennis produced increased activation in the supplementary motor area (SMA) in both control subjects and the patient, and the instruction to imagine walking around the home produced increased activation in the ‘parahippocampal place area’ (PPA)—parahippocampal gyrus, posterior parietal lobe, and the lateral premotor cortex—in both controls and the patient (Boly et al. [2007], Owen et al. [2006]). (This activation is consistent with the results of numerous imaging studies of motor imagery and spatial navigation.) Similar activations have been found in a small number of other VS patients (Owen and Coleman [2008a], [2008b]). The authors draw the following conclusion:

‘These results confirm that ... this patient retained the ability to understand spoken commands and to respond to them through brain activity, rather than through speech or movement. Moreover, her decision to cooperate with the authors by imagining particular tasks when asked to do so represents a clear act of intention, which confirmed beyond any doubt that she was consciously aware of herself and her surroundings.’ (Owen et al. [2006], p. 1402)

One might well take issue with the phrase ‘beyond any doubt’, but these results are certainly striking. Our aim in the first half of the paper is to examine just how strong the case for consciousness in this patient is.

Before we turn to that task two important points must be made. Firstly, even if this patient is conscious, the nature of her consciousness is likely to be abnormal in fundamental ways. VS patients have suffered serious global brain damage; many fail to regain consciousness at all, and even those who do remain very significantly mentally impaired. (For this reason, and because they have suffered from memory impairments, it is not surprising that there are no accounts from recovered VS patients that report in a verifiable way on what it was like to have been conscious while in the VS.) We should resist the temptation to think of consciousness in the vegetative state on the model of consciousness in the locked-in syndrome. Despite having had a focal brain lesion that leaves them almost incapable of motor activity, locked-in syndrome patients enjoy a rich conscious life that they are able to describe by means of simple motor actions such as eye-blinks. Any consciousness that might occur in VS is likely to be extremely fragmented; not dissimilar, perhaps, to the kind of consciousness seen in delirium and dreaming (Zeman [2000]). Indeed, it is possible that consciousness comes in degrees and that certain patients lie in the penumbra between full consciousness and its absence.

The second point concerns what we might call the problem of error management. Two kinds of error are possible when it comes to the ascription of consciousness. Errors of *commission* involve judging an unconscious organism to be conscious; errors of *omission* involve judging a conscious organism to be unconscious. Of necessity, the task of formulating and applying criteria for the ascription of consciousness must take into account the relative costs of these two forms of error. Those who regard errors of omission as more grave than errors of commission will lean towards a liberal approach to the ascription of

consciousness, whereas those who hold the opposite view will be motivated to adopt a conservative approach to the ascription of consciousness. Arguably, the task of balancing off these two kinds of error against each other is not a purely scientific one but requires attention to complex and contested ethical questions. We will not engage in that task here, but it should be borne in mind in what follows.

3. The Standard Approach

According to what we will call ‘the standard approach’, the ascription of consciousness in contested cases such as VS should be governed solely by those pre-theoretical markers that we use to ascribe consciousness to each other, namely, reportability and various forms of voluntary behaviour. The standard approach requires that the neuroimaging data is interpreted in terms of either reportability or volition in order for it to support the ascription of consciousness. We consider first the argument from reportability before turning to the argument from volition.

In appealing to report as a measure of consciousness most theorists appear to have ‘introspective report’—that is, reports of one’s current experiential states—in mind. It seems clear that the patient did not produce introspective reports. Another form of report that is also—and indeed, more typically—used as a measure of consciousness is ‘environmental report’: a report of one’s current environment (or bodily state). But there is no evidence that the patient produced environmental reports either. So, if we were to take reportability as our ‘gold standard’ for the ascription of consciousness then we would be led to conclude that the patient was not conscious—or at least, that we have no evidence of her being conscious.⁵

But perhaps we have been too hasty. Granted that the patient did not produce any reports, perhaps we have evidence that she was *capable* of producing reports. In a recent experiment, Owen and colleagues showed that healthy volunteers could communicate via

mental imagery that is read by real-time fMRI; in principle, it would be possible to employ this paradigm to communicate with patients (Owen et al. [2006]; Owen et al. [2007b]; Monti, Coleman and Owen [2008]). If a VS patient were able to use this paradigm to communicate reliably, for example by correctly answering a range of biographical questions, then that would show beyond reasonable doubt that they were conscious. The patient in (Owen et al. [2006]) didn't use this paradigm to communicate, but one might argue that the imagery result demonstrates that she *could* have communicated in this way, and that the *reportability* condition could thereby be said to be satisfied.

There is certainly some daylight between 'being able to issue a report here and now' and 'being able to issue reports' in the sense in which the reportability criterion requires. A locked-in syndrome patient whose one good eye muscle has been temporally paralysed won't be able to issue a report right now, but there is still a very natural sense in which such a patient retains the *ability* to produce reports. So, the fact that the patient in Owen's study didn't produce any reports does not demonstrate that she lacked the capacity to produce reports. Nonetheless, we do not have any reason to think that this patient retained the ability to produce reports. Normal subjects can use mental imagery that is read by real-time fMRI to communicate, but we have no evidence that *this patient* also had this ability. Our initial assessment was correct: a reportability-based argument will not support the claim that the patient described in the Owen study is conscious.

Just where this leaves us depends on how the connection between reportability and consciousness is to be understood. Taking reportability as only one ground on which to ascribe consciousness leaves open the case for consciousness in this patient. But if we suppose that our only access to consciousness is via reportability then we will have to conclude that we have no evidence that this patient is conscious—indeed, we might have to conclude that the evidence favours the view that the patient is unconscious.

Is reportability the sole measure of consciousness? A number of theorists have claimed that it is. In a commentary on (Owen et al [2006]) Naccache states that “consciousness is univocally probed in humans through the subject’s report of his or her own mental states.” Naccache went on to claim that the Owen et al. study was “not totally convincing on the issue of consciousness” (Naccache [2006], p. 1396), presumably on the grounds that the patient didn’t produce a report of her own mental states. Naccache is far from being alone in claiming that reportability—in particular, introspective reportability—is the unique measure of consciousness (see Papineau [2002], p. 182; Frith et al. [1999], p. 107; Weiskrantz [1997], p. 84).

The problem with this line of thought is that we routinely ascribe consciousness to subjects who are incapable of producing environmental reports let alone introspective reports. Although there are some who are reluctant to ascribe (phenomenal) consciousness to pre-verbal children and non-linguistic animals, such theorists are clearly in the minority. Pre-verbal creatures might lack *self*-consciousness, but there is little intuitive support for the thought that they are altogether unconscious. Even when it comes to adult, neurologically intact human beings there is reason to think that reportability is not the only measure we have of consciousness. Split-brain patients will usually deny seeing stimuli presented in their left visual field, but they are able to employ representations of the presented items in the service of rational agency. Furthermore, unlike blindsighted subjects, split-brain subjects do not need to be coaxed into using these representations—their use of them is spontaneous (Sperry et al. [1979]; Seymour et al. [1994]; Zaidel et al. [2003]). This fact suggests that these perceptual representations are conscious, even though the patient is not able to draw on their contents for the purposes of report. In sum, one can have good evidence of consciousness even when reportability is not present.

This leads us to a second argument for thinking that the patient studied by Owen and colleagues was conscious—an argument from volition. It does seem plausible to suppose that the neural activity they found is evidence of intentional agency. And, on the face of things, intentional agency seems to be a good marker of consciousness (Dretske [2006]). Indeed, it is precisely the ability to perform intentional actions that leads clinicians to regard MCS patients as conscious, despite the fact that they cannot produce reports of any kind (Giacino et al. [2002]). Let us call this the argument from volition.

One line of objection to the argument from volition takes issue with the claim that this patient was performing a volitional action. Some commentators have objected that the data do not support a volitional interpretation on the grounds that the imagery could be an automatic response to the stimuli (Greenberg [2007]; Nachev & Husain [2007]). Just what to make of this objection depends in no small part on what is meant by ‘automatic’ here. One form of automaticity occurs in the context of semantic priming. Could this account for the data that Owen and colleagues report? That seems highly unlikely, for the priming literature tells us that whole sentences are unable to function as unconscious primes (Greenwald [1993]). Further, Owen and colleagues found that a healthy control showed no increased activity in SMA when he heard the word “tennis” in the context of the sentence “The man played tennis”; nor did he show any increased activity in PPA when he heard the word “house” in the context of the sentence “The man walked around his house” (Owen et al. [2007a]).⁶ A further reason for thinking that the activation cannot be accounted for in terms of semantic priming concerns its time-course. Whereas priming is typically short-lived—in the order of hundreds of milliseconds or less—the activation found in this case persisted for a full 30 seconds (Owen et al [2007a]; although it is not clear that separate analyses were made of different parts of the 30 second blocks). Such sustained activation is very unlikely to be due to automatic processing in healthy subjects, although it cannot be excluded that the temporal

dynamics of the brains of VS patients differ significantly from healthy controls in this respect.⁷ A still further reason for rejecting the automaticity interpretation is that the areas of activation (SMA, PPA) are not associated with the semantic content of the words but with acts of imagery (see Owen et al [2007a]). So, if the neural activity is to be understood in terms of automatic responses, the notion of automaticity in play is much more likely to be that of automatic behaviour rather than that of automatic semantic activation.⁸ (We examine such accounts of this data below.)

Other commentators have objected to a volitional interpretation of the neural activity on less direct grounds. Naccache asks, ‘Why wouldn’t [the patient] be able to engage in intentional motor acts, given that she had not suffered functional or structural lesion of the motor pathway?’ ([2006], p. 1396). The question is a fair one, but it could do with some development. We might put the objection as follows. Assume, for the sake of *reductio*, that the patient intentionally imagined playing tennis and walking around her home. It seems to follow from this that she ought to have been able to form various behavioural intentions, such as the intention to actually play tennis, walk around the home, or simply get out of bed. And if she could form behavioural intentions, then she would have executed them had she been motivated to do so (which she clearly would have been). Since she didn’t execute any behavioural intentions, we must conclude that she couldn’t form any intentions at all, and hence did not intentionally imagine playing tennis.

One response to this objection is to take issue with the assumption that someone who is able to form intentions to perform various acts of mental imagery will also be able to form intentions to perform various motor actions. It is possible that intention has a somewhat modular structure, and that the ability to form intentions can be selectively impaired. Of some relevance to this case is the fact that patients with Parkinson’s disease can construct stimulus-driven intentions but not endogenous intentions ([Jahanshahi & Frith 1998]). A second

account of why the patient might not have been able to engage in intentional motor acts is that she *has* suffered lesions of the motor pathway, *pace* Naccache's assumption. Although we have no information about the motor pathways in this particular patient, independent information about VS patients affords this proposal some plausibility. As Coleman et al. remark, '... A large number of patients progressing to the vegetative stage also suffer complex peripheral nervous system changes. Many have extensive contractures, limited range of movement and muscle wastage preventing sufficient motor output to respond to command.' (Coleman et al. [2007]. p. 36; see also Kinney and Samuels [1994]). So, although Naccache's question is a good one, it can be satisfactorily answered.

This brings us to the relationship between volition and consciousness. In a very interesting discussion of this case, Levy ([2008]) takes issue with the claim that volition is evidence of consciousness. Levy points out that there is a large literature within both clinical and social psychology which suggests that action can occur independently of consciousness. For example, giving subjects stimuli that prime for thoughts of old age leads them to walk more slowly than control subjects (Bargh et al. [1996]; see also Dijksterhuis and Van Knippenberg [1998]). Automatic mirroring, in which subjects unwittingly modulate their actions or the syntax of their speech to match that of others, is also widespread (see Bargh and Chartrand [1999]). To these findings we might add others from cognitive neuroscience, showing that on-line motor control can be guided (Milner and Goodale [2006]) or modified (Lau and Passingham [2007]) by information of which the subject is not conscious. Don't these results cast doubt on the argument from volition?

We think not. Firstly, we should distinguish automatic, over-learned actions from less familiar action types. Even if we can execute highly routinized actions unconsciously, it is doubtful that we can execute relatively novel types of action unconsciously. Are the actions that this patient is taken to have executed routinized or novel? Well, they may not have been

completely novel, but it is unlikely that the patient will have spent much time imagining walking around her house or playing tennis in response to spoken instructions. We very much doubt that we could imagine playing tennis or visiting the rooms in our house without paying attention to what we were doing. William James noted that consciousness tends to depart from where it is not needed, but it would seem to be very much needed here.

Secondly, even when automatic actions are *executed* unconsciously, the *stimuli* that trigger them are typically conscious. One might not be conscious of initiating, guiding or completing the action, but one will usually be aware of the environmental feature to which one is automatically responding. Think of the infamous long-distance truck driver, who navigates ‘on autopilot’. The driver might not be aware of making adjustments to the steering, but she will typically be aware of the features of her environment (the road, the stop-signs, the traffic lights and so on) that motivate such adjustments. (Of course, the awareness of these features might leave few traces on episodic memory.) This point has a direct bearing on experimental work relating to high-level priming in social psychology. Even where the experimental subjects are unaware that their behaviour is affected by the critical stimuli, they are aware of the stimuli themselves.⁹ Relating these thoughts to the present case, one might suggest that even if the patient is not conscious of carrying out certain acts of imagery, she might nonetheless be conscious of the verbal instruction that she has been given, and perhaps also of the images that she is manipulating.

Thirdly, the argument from volition is not meant to be demonstrative. We are not claiming that imagery responses of this kind could never be produced unconsciously; indeed, we aren’t even claiming that it is impossible to produce responses of this kind in unconscious adult human beings. Our claim is only that a response of this kind is *good evidence* of consciousness. In sum, we think that the argument from volition provides fairly strong evidence of consciousness in this VS patient: it is reasonable to take the patient’s BOLD

response as evidence of (non-routine) volition, and it is reasonable to take (non-routine) volition as evidence of consciousness.

Let us summarize the argument of this section. Working within the confines of the standard approach, the fMRI data produced by Owen and colleagues can be interpreted in terms of reportability or in terms of volition. The argument from reportability cannot be sustained, but the argument from volition has some force insofar as the execution of non-routine actions provides us with strong prima facie evidence of consciousness. This prima facie evidence would be defeated if we were to regard reportability as necessary for consciousness, but there is little reason to endorse that position. Even if reportability were the *sine qua non* of consciousness in normal healthy human subjects (which we doubt), VS patients clearly fall outside that category.

We have argued that volition-based considerations suggest that this patient was conscious, but these considerations are clearly not decisive. Further, the fact that some VS patients show covert evidence of voluntary behaviour in the absence of any overt evidence of volition surely gives additional urgency to the question of whether those VS patients who show no sign of volition, or indeed are incapable of volition at all, might also be conscious. This leads us directly to the question of whether there are other forms of evidence that could bear on the question of consciousness in VS patients. The force of this question is not restricted to VS patients, for we might hope to have independent measures of consciousness to verify the bedside measures that are used to determine consciousness in the *MCS* patient. To address these questions we need to go beyond the standard approach. Exactly how to do that is the topic of the next section.

4. The Natural Kind Methodology

In everyday cases we rely on broad functional criteria such as report and volitional activity to determine the distribution of consciousness. When faced with puzzling cases, as with disorders of consciousness, it is tempting to look for evidence of these functional capacities. As we have seen, the results of such an investigation can be illuminating, but they will often be less than decisive—as indeed they are in this case. In light of that, one might wish for additional measures of consciousness. Can we go beyond our pre-theoretical measures of consciousness, and if so, how?¹⁰

According to one influential line of thought, when it comes to ascribing consciousness we are terminally constrained by the measures that we pre-theoretically associate with consciousness (Chalmers [1996], p. 243). These measures need not be taken individually as gold-standards of consciousness, but as a group they limn the marks by means of which consciousness can be ascribed. On this view, the existence of consciousness in any VS patient who was incapable of meeting any of these pre-theoretical criteria would be forever outside of our ken—a mere theoretical possibility.

This approach to the ascription of consciousness might be appropriate if consciousness were a nominal kind, akin to (say) dirt. There is no underlying nature to dirt, and it makes little sense to suppose that something might have the appearance of dirt without really being dirt. But suppose that consciousness is a *natural* rather than a nominal kind. Suppose, in other words, that being conscious is more like having hepatitis C than being dirty. There is an underlying nature to hepatitis C that goes beyond the superficial properties with which the concept may be associated. In medicine, clinicians gather signs of a disease and then examine whether they are ‘syndromic’—whether they seem to be found together better than chance. If so, the search is on for an underlying pathology that explains the syndrome. What was initially characterised as a blood-borne virus (non-A non-B) that caused

hepatitis is now known to be caused by a specific virus, hepatitis C (Choo et al. [1989]). Importantly, this form of viral hepatitis was diagnosed and treated before there was any sure-fire or gold standard way of telling whether a patient was infected with the hepatitis C virus. Tests of liver inflammation, together with the clinical history and exclusion of other causes gave a good indication, but were not determinative. Now sequencing for viral RNA directly gives a highly accurate diagnosis. However, having strong evidence for the existence of an underlying natural property does not depend on there ever being a sure-fire gold standard test.

In many cases where a variety of everyday signs cluster together there is indeed a natural property that gives rise to the pre-theoretic ways we have of characterising the phenomenon (e.g. being H₂O gives rise to characteristics like being liquid, transparent, etc.). In such cases, it is possible to discover instances of that natural property which are incapable of meeting any of our pre-theoretic criteria (e.g. water in the atmosphere of some distant planet). Finding an apparent cluster of properties does not guarantee that there will be a natural property which explains the clustering (a natural kind property), but when the clustering is best explained by a natural kind property, we thereby have the means to go beyond our pre-theoretic ways of characterising the phenomenon, through picking out the natural kind in new ways. If consciousness were a natural kind it would be very surprising to discover that our access to it was limited to the pre-theoretical measures that we associate with it. In every other case, discovering a natural kind property allows us to go beyond our pre-theoretic measures. We see no good reason why consciousness cannot be investigated in the same way.

We call this the ‘natural kind methodology’. The heart of the method involves collecting a wide a variety of evidence for the target property across a range of different cases and looking for “nomological clusters” in this evidence:

Nomological cluster

A set of evidential properties T_i form a nomological cluster iff:

- (i) they are instantiated together better than chance (given background theory);
- and
- (ii) observing subsets of the cluster supports induction to other elements of the cluster.

The existence of a nomological cluster can be explained by there being a natural property that is responsible, causally or constitutively, for the fact that these measures are positively associated with each other. Not only is the instantiation of a natural property *an* explanation of the cluster, in the kinds of cases we are considering here it is likely to be the *best* explanation. That natural property may be the property of being conscious, or it may be the more determinate property of being conscious in some particular way (e.g. being perceptually conscious; being visually conscious; or being conscious of redness). When there is a reason why the properties over which we induce cluster together we can say that they form a natural kind. The property in virtue of which they so cluster is also called a natural kind or natural kind property.

There are two broad conceptions of natural kinds in philosophy (Mellor [1974]; Schwartz [1980]; Sterelny [1983]). Some restrict the term to a Lockean conception in which some inner intrinsic essence is responsible for the identifiable symptoms (Putnam [1975]). That would exclude cases where the property explaining the cluster was extrinsic. For example the property of being a member of a given biological species is partly an historical property, with similarities in surface properties explained by the fact that conspecific individuals are related by a process of descent, involving conservative copying of features, from the very same individual (i.e., by being members of the same clade). We are adopting a

much broader conception of natural kinds, according to which any property that supports induction as a result of nomological principles or natural laws counts as a natural kind (Hacking [1991], Griffiths [1997], Millikan [2000]).

For a physical property to be a natural kind is a matter of degree, depending upon how broad and various are the properties over which it supports inductions. The natural kind methodology is appropriate no matter where consciousness falls on this spectrum. It is important that it should be a physical property that supports a range of inductions, but it doesn't matter what sort of property is founding those inductions. Some distinguish natural properties, which support inductions for a natural reason, from natural kinds, which support a wide range of inductions (i.e. are causally or constitutively related to a large cluster of properties) (Bird and Tobin [2009]). The property of being positively charged applies to a very large variety of things, from the positron up to the sun,¹¹ and those objects share only a very limited number of other properties in virtue of being positively charged. The more kind-like consciousness is the easier will be the task of finding nomological clusters of properties connected to it. Correlatively, if being conscious supports only a very narrow range of inductions then the methodology we recommend will be correspondingly more difficult, even if being conscious is a perfectly natural property.

The prevailing approach in consciousness research is to look for some property of experimental subjects that is distinctive of consciousness. Many paradigms look for a correlate of being conscious in a particular way (being 'F-ly conscious'), for example of consciously seeing a briefly-presented black ring vs. not. Others look for correlates of being globally conscious vs. being unconscious (which correlates with things like the operation of the reticular activating system). Many of these tests may be carried over, piecemeal, to VS patients. But that would be to miss a crucial part of the natural kind methodology, which is to establish that the tests really do identify a common underlying natural property. To do that,

we need to see if the tests form a nomological cluster: that subjects who are conscious according to some test T_1 are also conscious according to T_2 , etc., and that those who are not conscious according to T_1 are not conscious according to T_2 , and so on with the other putative tests. Since each of these inferences is defeasible we should not expect perfect agreement amongst the tests, but all the inference to the existence of a common natural property requires is that they *are* associated.

Existing consciousness research contains a multitude of tests that could profitably be combined in the search for a nomological cluster.¹² For example, subjects appear to be insensitive to the automatic stem completion effect only when they are conscious of the relevant stimulus (Debner and Jacoby [1994]; Merikle et al. [2001]). The susceptibility of subjects' grip width to a size contrast effect seems also to require consciousness of the object being reached for (Hu and Goodale [2000]). Many other functional tests are candidates: ones which our initial gross functional tests suggest either require consciousness, or are performed in a different way when the subject is conscious, so that the mechanism deployed when performing the task when relying on conscious states has a different functional signature from the non-conscious mechanism for performing the task. As well as such non-obvious functional properties, more direct measures of brain properties are relevant, and portable to VS patients: gamma-band neural synchrony, activation of a global workspace integrating perceptual areas with prefrontal cortex, existence of cortico-thalamic loops, locally recurrent processing within cortex, and so on. None of these tests has yet received widespread acceptance as a good neural correlate of consciousness, but the natural kind methodology will help to choose between them.

There are two crucial features of the natural kind approach that we want to draw attention to. Firstly, it allows us to identify tests for consciousness that *go beyond* the superficial tests with which we begin. As the natural kind method starts to pay off, it will

provide us with new measures of consciousness. Secondly, from the perspective of the natural kind methodology the everyday ('gross') functional roles that are roughly correlated with consciousness lose their privileged epistemic status. If other tests are found to correlate closely with what we take to be conscious vs. unconscious states, and with one another, then these too will be good evidence for consciousness. We will then face the question of how to weigh the evidential force of these new tests against that of the everyday functional tests with which we started. Once it is just a matter of gathering and validating various kinds of evidence, all manner of evidence will, in principle, be admissible: non-obvious functional properties (certain kinds of susceptibility to priming can indicate that the subject was *not* conscious of the stimulus), functional properties that can only be elicited by special means (e.g. transcranial magnetic stimulation), and evidence of non-functional physical properties of the subject, like neural structures, cytology, connectivity and activation.

Although some of these tests may not be applicable in the context of disputed cases, others will be. Here is the sort of test that could be used in VS patients. There are two different ways of forming an association between a tone and a puff of air to the eye so that the tone comes to cause an eye blink: 'delay conditioning' and 'trace conditioning'. In delay conditioning a puff of air to the eye is administered during the occurrence of a tone (after the start of the tone, hence 'delay' conditioning); by contrast, in trace conditioning the air puff occurs shortly after the tone has stopped. Evidence from normals suggests that trace conditioning requires consciousness of the contingency between tone and air puff, whereas delay conditioning does not (Clark et al. [2002]; Perruchet [1985]). If some VS patients show eye-blink trace conditioning, that would suggest that they have become consciously aware of the contingency. Conversely, if delay conditioning is preserved in a VS patient, but trace conditioning is impaired, that would be evidence that the patient was not conscious of the tone, the puff of air and/or the contingency between them.

Other tests of whether the subject is conscious give rise to direct measures of brain activity. Experiments on the attentional blink have revealed a brain activation signature that correlates with subjects reporting that they see the stimulus (detecting electrical activity on the scalp with fine-grained temporal resolution via ERP: Shapiro et al [1997]; Sergent et al. [2005]). Similarly, the ‘ignition effect’ is potentially portable to the context of VS (stimuli are briefly presented, followed by a mask, and when the presentation is long enough that subjects report the stimuli to be conscious a qualitatively different brain response is observed: Dehaene et al. [2003]). Schnakers et al. ([2008]) recently reported the use of another ERP test directly on VS subjects. Hearing one’s own name produces a large brain response, but in healthy controls what is called the P3 component of the ERP response is larger when subjects have been asked actively to count occurrences of their name than when they listen passively. Schnakers et al. observed a similar difference between active and passive conditions in 14 MCS patients but found no difference in the 8 VS patients that they tested. If such a test were validated by finding that it correlates with other tests of consciousness across a range of different circumstances, then its absence in this group of VS patients would be evidence that this group did not include any patients that were conscious when tested (which would be unsurprising, given that the evidence so far suggests that at best only a small subset of VS patients are conscious).

Certain tests of being globally conscious—of being conscious as such—will also be possible in VS patients. An example is the way that the increased activation of language areas produced by high-ambiguity sentences is abolished by anaesthesia (mentioned above). First we need to ask if this test clusters together with other putative tests of consciousness: e.g. is the ‘ignition effect’ also abolished by anaesthesia and other states of impaired consciousness like sleep? If so, finding such modulations in a VS patient would be further evidence that that patient was conscious. We are not suggesting that it would be necessary to anaesthetise VS

patients, but subjects do sometimes fall asleep in the scanner. It seems likely that the task-specific activations found by Boly et al. ([2007]) in healthy volunteers would be abolished by sleep—it would be useful to check. The same test could be performed in VS patients.¹³ We predict that they, too, would fail to show task-specific activation when asleep. If so, that would provide further evidence that a subject was indeed conscious when awake and exhibiting task-specific activation like the SMA response to the tennis instruction.

To the extent that we already have some reason to believe that current tests of consciousness form a nomological cluster, two features of the fMRI data offer some direct evidence that the patient was conscious: the areas involved, and its time course. Although the patient's pattern of activation in the tennis/house visual imagery task matched those in conscious healthy volunteers, it is not yet clear to what extent the same areas can be activated in unconscious subjects or subliminally in conscious subjects. In particular, unconscious activation of the parahippocampal place area may be possible. But if it turns out that SMA is rarely (if ever) unconsciously activated, then that would be evidence that those VS patients who show matching activation are conscious. Evidence about the areas involved in the semantic ambiguity task would have similar force. The fact that both the imagery and the semantic ambiguity effects are found only in a small number of VS patients lends further support to their utility as tests which could discriminate amongst VS patients, identifying that subset who are conscious.

Secondly, as we have already noted, Owen and colleagues found that the BOLD signal in response to the imagery instructions was sustained throughout the 30 second block, up until the point when the instruction to rest was given.¹⁴ They invoked this data as evidence that the patient was performing a voluntary action (and hence was conscious), but this data can also be employed within the natural kind methodology. Brain activation associated with subliminally-presented stimuli in conscious subjects lasts on the order of hundreds of

milliseconds or less, so finding a sustained pattern of task-specific brain activation in a healthy subject would certainly suggest that the subject was conscious in relation to some function performed by that brain network. The inference is weakened by the fact that we are dealing with patients who have suffered severe brain damage and it is always possible that the brain damage substantially alters the time-course of brain activity, but the fact remains that time-course data is *an* important source of direct evidence that this patient was conscious.

An important component of the natural kind methodology is the task of separating out confounding variables. For example, that a test is abolished by sleep may be due to one of the many other differences between sleep and wakefulness other than those that concern consciousness. Accounts of the functional role of consciousness typically focus on consciousness in the context of normal wakefulness, but we need to somehow separate out the role played by wakefulness—as opposed to that played by consciousness—in such conditions. Similarly, we need to know how putative tests cluster and dissociate in other kinds of pathology where we have good reason to think that consciousness is still present (e.g. Parkinson's, Alzheimer's, epilepsy). The full picture can include a huge range of evidence: developmental data in infants and children, studies in other animals, and so on. We don't need to wait for all this data to be in before we can begin to make better inferences about consciousness and its absence in VS. The first step is just to look at how a small cluster of individually well-validated tests cluster together across a narrow range of cases (awake, asleep, anaesthetised) in healthy controls, in VS patients, and in some clearly conscious patient group with relevant pathology (e.g. recovered VS patients or those VS patients who do achieve communication via BOLD response).

5. Is consciousness a special case?

The natural kind methodology is routine in science, so it is somewhat surprising that consciousness research has not adopted it. What might explain this fact? One possibility is that consciousness is not best thought of as a natural kind (5.1). Another possibility is that although consciousness is in fact a natural kind, it possesses certain features that prevent the natural kind methodology from being usefully applied to it (5.2).

5.1 Is consciousness a natural kind?

Certain theorists have suggested that “consciousness” is a catch all label for a variety of loosely related phenomena as opposed to a natural kind term (Churchland [1983]; Dennett [1988]; Rey [1983]; Wilkes [1988]). We have some sympathy with the claim that the ordinary notion of consciousness picks out a number of different phenomena (phenomenal consciousness, self-consciousness, access consciousness, etc.), but these worries do not undermine the narrower project of investigating *phenomenal* consciousness as a natural kind. In fact, the natural kind methodology requires only that there are natural kinds corresponding to what we pre-theoretically think of as determinates of phenomenal consciousness, such as perceptual experience, visual experience, and so on. It is, of course, an empirical question whether there is a nomological cluster associated with any one of these determinates, but the evidence to date provides the proponent of a natural kind analysis of these notions with reasons for optimism.

A closely-related worry is that consciousness is really a nominal kind: that a full conceptual analysis of “consciousness” reveals that reportability, or voluntary agency, or some combination of the two are constitutive of consciousness. We won’t rehearse here the arguments against this kind of analytic functionalism—we are presupposing a rejection of that position, rather than setting out to refute it. However, our discussion of the standard

approach to studying consciousness in section 3 above does show some of the limitations that derive from tying consciousness tightly to any particular set of pre-theoretical criteria.

Further concerns about the natural kind methodology derive from the thought that consciousness comes in degrees. If this were the case, then the existence of borderline cases is unavoidable, built into the basic ontology of the phenomenon. The answer to this worry is that the natural kind methodology is applicable in any event. The scientific method has proved itself adept at dealing with properties that come in degrees, even managing to quantify them in many cases (e.g. mass). Of course, if consciousness is a graded property then it may support relatively few inductions, which would indeed make it difficult to investigate. At the same time, the fact that a natural property comes in degrees can sharpen the reliability and increase the variety of inductions that it can support, especially if the property in question can be quantified. Whether consciousness can be quantified is also a question for empirical investigation, but it is a question that the natural kind methodology is adequate to address.

Further reasons for doubting that consciousness is a natural kind derive from the thought that consciousness is multiply realisable. Suppose, the critic might say, that consciousness can be realized by different physical structures and in different cognitive architectures. If that is so, then how can we use physical and functional measures with which consciousness happens to be associated with in us—measures that might not be essential to consciousness as such—as a guide to the presence of consciousness?

The question is a good one, but it needs to be handled with some care. Worries about multiple realisability have most force when we are dealing with creatures that are very different from us, such as Martians, robots or even octopuses. In such cases, the natural kind methodology will be extremely difficult to apply ([Block 2002]). But such worries have far less force in this context, for VS patients are fellow human beings with whom we share physiological and fine-grained functional properties. There is good reason to think that

functional and non-functional properties which cluster together in some members of the species (in oneself, for example), also do so in other members of the species, given our shared relation of descent with conservative copying from a single common ancestor in the relatively recent past (Sober [2000]). Consciousness may be multiply *realisable*, but when it comes to the members of a single species it is unlikely to be multiply *realised*.

Of course, it is *possible* that consciousness is realised in a *sui generis* way in VS, but it seems highly unlikely that the physical-functional basis of consciousness in VS should stand apart from that in all other normal and pathological brains, including those of recovered VS patients. The natural kind approach would also be in difficulty if putative tests of consciousness clustered in healthy subjects but not in VS patients. Then we would need to have recourse to a much larger research project, investigating whether the tests cluster in a range of other pathologies, and in infants, and in other animals, and so on. That would take us beyond the scope of the present paper. We are not claiming that the natural kind approach is guaranteed to deliver a correct and determinate answer to the question of whether VS patients are conscious, only that it has good prospects of doing so.

A rather different worry with the natural kind conception of consciousness is that it appears to be at odds with the epistemic authority we possess concerning our own states of consciousness. A phenomenally conscious state is simply a state that there is ‘something it is like’ to be in. But ‘what it’s likeness’ is, one might think, a ‘self-revealing’ or ‘luminous’ property, in the sense that it’s presence or absence is directly ascertainable from the first-person perspective. Natural kind properties, by contrast, have an underlying structure that is not immediately apparent. In short, the worry is that the natural kind conception of consciousness appears to threaten the appealing thought that we are authorities when it comes to our own states of consciousness.

We certainly have no wish to jettison first-person authority with respect to consciousness. Luckily, we have no need to do so, for the apparent tension between that authority and the natural kind conception is merely apparent. The objection goes wrong in assuming that being a natural kind need be a matter of microstructure. Consider another natural kind property, such as being a polar bear. If two animals both possess polar bear-ness (i.e. the property of being a polar bear), then there is indeed no further question of whether they are polar bears—irrespective of their underlying microstructure. Since ‘what it’s likeness’ is the phenomenon we are setting out to investigate, it would be a mistake to assume that it must be a superficial property, separate from the natural kind which is consciousness.

This section we have canvassed a number of objections to the natural kind conception of consciousness, and have argued that none of these objections is decisive. However, even if consciousness is a natural kind, one might worry that standard natural kind methodology cannot be applied when the investigation must start with the evidence furnished by subjects’ reports. We turn now to such worries.

5.2 A special obstacle?

In order to generate the set of tests that form our putative nomological cluster, we have to *start* somewhere. The science of consciousness starts with subjects’ reports: introspective reports on their own experiential states, and environmental reports in contexts where asking the subjects afterwards (and being subjects themselves) makes the researchers fairly sure that such environmental reports reflect the relevant phenomenal states. Given this starting point, additional tests of consciousness—for example, neural signatures of various kinds—look to be evidence of consciousness only insofar as they are evidence of the ability to report. So the challenge is this: given that the whole method is parasitic on report, wouldn’t it be paradoxical to arrive at a test that could tell us that a subject was phenomenally conscious, although we thought their phenomenal states were no longer reportable? If further tests are

only ever evidence of consciousness because they are evidence of reportability, which is itself evidence of consciousness, then we cannot ever really have measures of consciousness that are independent of the pre-theoretical criteria with which we begin.

The challenge would be legitimate if phenomenal consciousness were a nominal kind, defined at least in part in terms of reportability. But once our objector has come with us to the point of accepting that phenomenal consciousness may be a *natural* kind then this challenge loses its force: if there is a property that explains the cluster of tests, including reportability, then there is no reason to think of tests in the cluster as being indirect tests of consciousness. There is indeed an inference, along the lines we discussed in section 3, from a correlative test, via reportability, to consciousness: that they are *evidence of* evidence of consciousness. Once we have evidence that there is indeed a natural kind property underlying the cases that we pre-theoretically took to be instances of consciousness, the fact that those initial instances happened to be generated in a particular way (e.g. by report) presents no obstacle to what we can subsequently discover about them. If one accepts that phenomenal consciousness is a natural kind, there is no reason to think that the natural kind methodology poses any *sui generis* obstacle to the investigation of consciousness by the third-person methods of science.

Deep puzzles remain, which arise out of the fact that we are each subjects of conscious experience. In this paper we don't take ourselves to be answering the hard problem or bridging the explanatory gap between first-person and third-person data about consciousness. Indeed, the existence of this explanatory gap may be part of what explains the reluctance of those engaged in the science of consciousness to dive in with their standard methods. The fact that we have direct first-personal acquaintance with phenomenal consciousness might itself contribute to the widespread reluctance to treat consciousness as suitable target of the natural kind methodology. Forceful as those psychological motivations undoubtedly are, they lose their force when the project is to improve our third-personal

theories of the phenomenon, and to reason about its presence and absence in others in the ordinary abductive way.

6. Conclusion

According to the standard approach to the study of consciousness, consciousness is uniquely identified by high-level measures, such as reportability and voluntary behaviour. Commitment to the standard approach leads theorists to assume that evidence of consciousness in disorders of consciousness such as VS must take the indirect form of finding neural evidence that the criteria reportability and volition have indeed been met, albeit it in a covert or private form. As we have seen, this approach does suggest that certain VS patients might be conscious, although the evidence here is far from conclusive.

Our central aim here has not been to assess consciousness in VS by reference to the standard approach, but to argue that the standard approach is by no means compulsory. The natural kind methodology that we have outlined provides a superior framework within which consciousness can be studied, and this framework promises to prove particularly useful when studying members of our own species. In line with the sciences of other biological entities (such as diseases), the science of consciousness should proceed by looking for clusters of properties—both functional and physical—that support induction between elements of the cluster. This approach allows us to count as direct evidence of consciousness neural and functional properties that are not pre-theoretically associated with consciousness. There is no guarantee that this method will deliver definitive answers to the question of consciousness in VS patients, but on the reasonable assumption that (phenomenal) consciousness is indeed a natural property it does offer some reason for optimism on that score.

Funding

NS: OUP John Fell Research Fund, James Martin 21st Century School, Wellcome Trust (Oxford Centre for Neuroethics) and Somerville College (Mary Somerville Junior Research Fellowship). TB: Australian Research Council Discovery grant DP0984572 "Conscious States in Conscious Creatures".

Acknowledgements

The authors would like to thank Stephen Laureys, Adrian Owen, Adam Zeman, and audiences in Oxford and Canberra for helpful discussion of this material, and two referees for *British Journal for the Philosophy of Science* for constructive criticism of an earlier version of this manuscript.

Nicholas Shea

Faculty of Philosophy and Somerville College

University of Oxford

Oxford OX1 4JJ

Email: nicholas <dot> shea <a> philosophy <dot> ox <dot> ac <dot> uk

Tim Bayne

Faculty of Philosophy and St. Catherine's College

University of Oxford

Oxford OX1 3UJ

Email: tim <dot> bayne <a> gmail <dot> com

References

Bargh, J. A., Chen, M., Burrows, L. [1996]: 'Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action', *Journal of Personality and Social Psychology*, **71**, pp. 230–44.

Bargh, J.L. and Chartrand, T.L. [1999]: 'The unbearable automaticity of being', *American Psychologist*, **54/7**, pp. 462-79.

Bernat, J.L. [2006]: 'Chronic disorders of consciousness', *Lancet*, **367**, pp. 1181-92.

Bird, A. and Tobin, E. [2009]: 'Natural kinds', in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2009 Edition),
<plato.stanford.edu/archives/spr2009/entries/natural-kinds/>

Block, N. [2002]. 'The harder problem of consciousness', *The Journal of Philosophy*, **99**, pp. 1-35.

Boly, M., Coleman, M.R., Davis, M.H. et al. [2007]: 'When thoughts become action: an fMRI paradigm to study volitional brain activity in noncommunicative brain injured patients', *Neuroimage*, **36**, pp. 979-92.

Chalmers, D. [1996]: *The Conscious Mind*, Oxford: Oxford University Press.

Choo Q.L., Kuo G., Weiner A.J., Overby L.R., Bradley D.W., Houghton M. [1989]: 'Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome', *Science*, **244**, pp. 359-62.

Churchland, P.S. [1983]: 'Consciousness: The transmutation of a concept', *Pacific Philosophical Quarterly*, **64**, pp. 80-95.

Clark, R.E., Manns, J.R., Squire, L.R. [2002]: 'Classical conditioning, awareness, and brain systems', *Trends in Cognitive Sciences*, **6/12**. pp. 524-31.

Coleman, M. R., Owen, A. M., Pickard, J. D. [2007]: 'Functional imaging and the vegetative state,' *Advances in Clinical Neuroscience and Rehabilitation*, **7(2)**, pp. 35-6.

Davis, M.H., Coleman, M.R., Absalom, A.R., Rodd, J.R., Johnsrude, I.S., Matta, B. F., Owen, A.M., Menon, D.K. [2007]: 'Dissociating speech perception and comprehension at reduced levels of awareness', *Proceedings of the National Academy of Sciences U.S.A.*, **104**, pp. 16032-7.

Debner, J. A. and Jacoby, L.L. [1994]: 'Unconscious perception: attention, awareness, and control', *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20(2)**, pp. 304-17.

de Jong, B., Willemsen, A.T., Paans, A.M. [1997]: 'Regional cerebral blood flow changes related to affective speech presentation in persistent vegetative state', *Clinical Neurology and Neurosurgery*, **99**, pp. 213-16.

Dehaene, S., Sergent, C., & Changeux, J. P. [2003]: 'A neuronal network model linking subjective reports and objective physiological data during conscious perception', *Proceedings of the National Academy of Sciences U.S.A.*, **100**, pp. 8520-5.

Dennett, D. [1988]: 'Quining qualia', in A. Marcel & E. Bisiach (eds), *Consciousness in Contemporary Science*. Oxford: Oxford University Press.

Di, H.B., Yu, S.M., Weng, X.C., et al. [2007]: 'Cerebral response to patient's own name in the vegetative and minimally conscious states', *Neurology*, **68**, pp. 895-9.

Dijksterhuis A, and Van Knippenberg A. [1998]: 'The relation between perception and behavior, or how to win a game of trivial pursuit', *Journal of Personality and Social Psychology*, **74**, pp. 865-77.

Dretske, F. [2006]: 'Perception without awareness', in J. Hawthorne & T. Gendler (eds), *Perceptual Experience*, Oxford: Oxford University Press.

- Farah, M. [2008]: 'That little matter of consciousness', *The American Journal of Bioethics*, **8(9)**, pp. 17-8.
- Fins, J.J. and Schiff, N.D. [2006]: 'Shades of gray: new insights from the vegetative state', *Hastings Center Report*, **36**, p. 8.
- Frith, C., Perry, R., and Lumer, E. [1999]: 'The neural correlates of conscious experience: an experimental framework', *Trends in Cognitive Sciences*, **3**, pp. 105-14.
- Giacino, J.T., Ashwal, S., Childs, N., Cranford, R., Jennet, B., Katz, D.I., et al. [2002]: 'The minimally conscious state: definition and diagnostic criteria', *Neurology*, **58**, pp. 349-53.
- Greenberg, D.L. [2007]: 'Comment on "Detecting awareness in the vegetative state"', *Science*, **315**, p. 1221.
- Greenwald, A.G. [1992]: 'New Look 3: Reclaiming unconscious cognition', *American Psychologist*, **47**, pp. 766-79.
- Griffiths, P. [1997]: *What Emotions Really Are*. Chicago, IL: University of Chicago Press.
- Hacking, I. [1991]: 'A tradition of natural kinds', *Philosophical Studies*, **61**, pp. 109-26.
- Hu, Y. and Goodale, M.A. [2000]: 'Grasping after a delay shifts size-scaling from absolute to relative metrics', *Journal of Cognitive Neuroscience*, **12(5)**, pp. 856-68.
- Jahanshahi, M. and Frith, C.D. [1998]: 'Willed action and its impairments', *Cognitive Neuropsychology*, **15**, pp. 483-533 .
- Jennett, B. [2002]: *The Vegetative State*, Cambridge: Cambridge University Press.
- Kinney, H.C. and Samuels, M.A. [1994]: 'Neuropathology of the persistent vegetative state: a review', *Journal of Neuropathology and Experimental Neurology*, **53**, pp. 548-58.
- Lau, H. C. and Passingham, R. E. [2007]: 'Unconscious activation of the cognitive control system in the human prefrontal cortex', *Journal of Neuroscience*, **27**, pp. 5805-11.

Laureys, S., Faymonville, M.E., Peigneux, P., Damas, P., Lambermont, B., Del Fiore, G., Degueldre, C., Aerts, J., Luxen, A., Franck, G. et al. [2002]: ‘Cortical processing of noxious somatosensory stimuli in the persistent vegetative state’, *Neuroimage*, **17**, pp. 732-41.

Laureys, S., LeMaire, C., Maquet, P., Phillips, C., and Franck, G. [1999]: ‘Cerebral metabolism during vegetative state after recovery to consciousness’, *Journal of Neurology, Neurosurgery and Psychiatry*, **67**, p. 121.

Levy, N. [2008]: ‘Going beyond the evidence’, *The American Journal of Bioethics*, **8/9**, pp. 19-21.

Mellor, D. N. [1974]: ‘Natural kinds’, *British Journal for the Philosophy of Science*, **28**, pp. 299–312.

Menon, D.K., Owen, AM., Williams, E.J., et al. [1998]: ‘Cortical processing in persistent vegetative state’, *Lancet*, **352**, p. 200.

Merikle, P. M., D. Smilek and J. D. Eastwood [2001]: ‘Perception without awareness: perspectives from cognitive psychology’, *Cognition*, **79**, pp. 115-34.

Milner, A.D. and Goodale, M.A. [2006]: *The Visual Brain in Action*, Oxford/New York: Oxford University Press.

Monti, M.M., Coleman, M.R., Owen, A.M. [2008], ‘fMRI and the Vegetative State: Solving the behavioral dilemma?’, *Annals of the New York Academy of Sciences Series 2008: Disorders of Consciousness*.

Naccache, L. [2006], ‘Is she conscious?’, *Science*, **313**, pp. 1395-6.

Nachev, P., and Husain, M. [2007]: ‘Comment on “Detecting awareness in the vegetative state”’, *Science*, **315**, p. 1221.

Nagel, T. [1974]: ‘What is it like to be a bat?’, *Philosophical Review*, **83**, pp. 435-50.

- Owen, A.M., Coleman, M.R., Menon, D.K. et al [2005]: ‘Residual auditory function in persistent vegetative state: a combined PET and fMRI study’, *Neuropsychological Rehabilitation*, **15**, pp. 290-306.
- Owen, A.M., Coleman, M.R., Boly, M., Davis, M.H., Laureys, S., Pickard, J.D. [2006]: ‘Detecting Awareness in the Vegetative State’, *Science*, **313**, p. 1402.
- Owen, A.M., Coleman, M.R., Boly, M., Davis, M. H., Laureys, S., Jolles, D., Pickard, J.D. [2007a]: ‘Response to comments on “Detecting Awareness in the Vegetative State”’, *Science*, **315**, p. 1221c.
- Owen, A.M., Coleman, M.R., Boly, M., Davis, M.H., Laureys, S., Pickard, J.D. [2007b]: ‘Using functional magnetic resonance imaging to detect covert awareness in the vegetative state’, *Archives of Neurology*, **64**, pp. 1098–1102.
- Owen, A.M. and Coleman, M.R. [2008a]: ‘Functional neuroimaging of the vegetative state’, *Nature Reviews Neuroscience*, **9**, pp. 235-43.
- Owen, A.M. & Coleman, M.R. [2008b]: ‘Detecting awareness in the vegetative state’, *Annals of the New York Academy of Sciences*, **1129**, pp. 130-8.
- Papineau, D. [2002]: *Thinking about Consciousness*, Oxford/New York: Oxford University Press.
- Peigneux P., Orban P., Balteau, E., Degueldre, C., Luxen, A., Laureys, S., and Maquet, P. [2006]: ‘Offline persistence of memory-related cerebral activity during active wakefulness’, *PLoS Biology*, **4(4)**, p. e100.
- Perrin F., Schnakers, C., Schabus, M., Degueldre, C., Goldman, S., Bredart, S., Faymonville, M.E., Lamy, M., Moonen, G., Luxen, A., Maquet, P., and Laureys, S. [2006]: ‘Brain response to one's own name in vegetative state, minimally conscious state, and locked-in syndrome’, *Archives of Neurology*, **63(4)**, pp. 562-9.

- Perruchet, P. [1985]: 'Expectancy for airpuff and conditioned eyeblinks in humans', *Acts Psychologica*, **58**, pp. 31-44.
- Plum, F. & Posner, J. [1982]: *Diagnosis of Stupor and Coma*, New York: F. A. Davis.
- Putnam, H. [1975]: 'The Meaning of "Meaning"', in his *Mind, Language and Reality: Philosophical Papers, Volume 2*, Cambridge: Cambridge University Press, pp. 215-71.
- Royal College of Physicians [2003]: *The Vegetative State: Guidance on Diagnosis and Management*, London: Royal College of Physicians.
- Rey, G. [1983]: 'A reason for doubting the existence of consciousness', in R. Davidson, G. Schwartz and D. Shapiro (eds), *Consciousness and Self-Regulation, Vol 3*, New York: Plenum, pp. 1-39.
- Rodd, J.M., Davis, M.H., and Johnsrude, I.S. [2005]: 'The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity', *Cerebral Cortex*, **15**, pp. 1261-9.
- Schiff, N., Ribary, U., Moreno, D., Beattie, B., Kronberg, E., Blasberg, R., et al. [2002]: 'Residual cerebral activity and behavioural fragments in the persistent vegetative state', *Brain*, **125**, pp. 1210-34.
- Schiff, N. D. [2007]: 'Global disorders of consciousness', in M. Velmans and S. Schneider (eds), *The Blackwell Companion to Consciousness*, Oxford: Blackwell.
- Schnakers, C., Perrin, F., Schabus, M. et al. [2008]: 'Voluntary brain processing in disorders of consciousness', *Neurology*, **71**, pp. 1614-20.
- Schwartz, S. P. [1980]: 'Natural kinds and nominal kinds', *Mind*, **89**, pp. 182-95.
- Sergent, C., Baillet, S., and Dehaene, S. [2005]: 'Timing of the brain events underlying access to consciousness during the attentional blink', *Nature Neuroscience*, **8(10)**, pp. 1391-1400.

Seymour, S.E., Reuter-Lorenz, P.A., and Gazzaniga, M.S. [1994]: 'The disconnection syndrome: Basic findings reaffirmed', *Brain*, **117**, pp. 105-15.

Shapiro, K. L., Arnell, K. and Raymond, J. E. [1997]: 'The attentional blink: A view on attention and a glimpse on consciousness', *Trends in Cognitive Science*, **1(8)**, pp. 291-5.

Sober, E. [2000]: 'Evolution and the problem of other minds', *Journal of Philosophy*, **97**, pp. 365-87.

Sperry, R.W., Zaidel, E. and Zaidel, D. [1979]: 'Self-recognition and social awareness in the disconnected minor hemisphere', *Neuropsychologia*, **17**, pp. 153-66.

Staffen, W. Kronbichler, M., Aichhorn, M. et al. [2006]: 'Selective brain activity in response to one's own name in the persistent vegetative state', *Journal of Neurology, Neurosurgery and Psychiatry*, **77**, pp. 1383-4.

Sterelny, K. [1983]: 'Natural kind terms', *Pacific Philosophical Quarterly*, **35**, pp. 9-20.

Weiskrantz, L. [1997]: *Consciousness Lost and Found: A Neuropsychological Exploration*, Oxford: Oxford University Press.

Wilkes, K. [1988]: 'Yishi, duh, um and consciousness', in A. Marcel and E. Bisiach (eds), *Consciousness in Contemporary Science*, Oxford: Oxford University Press.

Zaidel, E., Iacoboni, M., Zaidel, D. W., and Bogen, J. [2003]: 'The callosal syndromes', in K. H. Heilman and E. Valenstein (eds), *Clinical Neuropsychology*, Oxford: Oxford University Press, pp. 347-403.

Zeman, A. [2002]: 'The persistent vegetative state: Conscious of nothing?', *Practical Neurology*, **2**, pp. 214-7.

Footnotes

¹ Royal College of Physicians [2003], § 2.5.

² E.g. ‘The creek/creak came from a beam in the ceiling’. Typical subjects assume the first meaning, *creek*, until they reach ‘beam’, leading them to re-interpret the term as *creak*.

³ Out of 17 VS patients, three showed greater activation of this network to high ambiguity sentences (Owen et al. [2005], Owen et al. [2006], Coleman et al. [2007], Owen & Coleman [2008a]).

⁴ The patient in question was in a *persistent* vegetative state, and it is not yet known whether similar results can be found in *permanent* vegetative state patients.

⁵ Here we depart from Owen et al. ([2006]) who suggest that the patient did produce reports.

⁶ It would be useful to give similar tests to VS patients: to record their response to the instructions, indicative sentences involving the same words, the words “tennis” and “house” alone, etc.

⁷ Some evidence against this challenge is furnished by the fact that the brain responses of VS patients to other kinds of stimuli (sounds, words, semantic ambiguity, own name, etc.) have temporal dynamics of the same order as the responses in normals, and certainly do not show the sustained activation found in response to the imagery instruction.

⁸ Voluntary agency typically produces transient activation of ventro-lateral prefrontal cortex that can be detected by event-related fMRI (making a choice, performing a mental action, changing task rules). Although Owen et al. did not analyse the fMRI data from their patient to see if this ‘signature of volition’ was present, doing so would have been very interesting, as it could have a significant bearing on the volitional interpretation of the patient’s response.

⁹ These data suggest that a useful further test would be to see if VS patients can follow a two stage instruction. For example, they could be asked, “when you hear this tone ____, switch back and forth between imagining playing tennis and imagining walking around your home”. Even if the reaction to the instruction to “imagine playing tennis” and to “now just relax” could be voluntary but unconscious, it is much less plausible that following a two-stage instruction triggered by a neutral tone could be an act of voluntary unconsciousness. Although more demanding than the imagery test alone, such two-stage instructions would probably be considerably less demanding (and so a less conservative test) than using imagery to communicate in the way that Owen and colleagues have suggested.

¹⁰ Farah ([2008]) also discusses this issue.

¹¹ To use an example of Alexander Bird’s.

¹² The tests needn’t be simply conjunctive, for example, it could be that T_1 is a good test of the presence of consciousness when T_2 is also satisfied, but not when it is not—all sorts of logical relations are possible.

¹³ Behavioural, physiological and even EEG markers could be used to tell when subjects are asleep (subject to technical difficulties of concurrent measurement with fMRI, especially of the fMRI-EEG combination).

¹⁴ As we noted above, the latter part of the 30s block has not been analysed separately. If this were done, it might provide stronger support for the claim that the patient was performing a voluntary action.