Consciousness, Intention, and Command Following in the Vegetative State

Colin Klein

Abstract

Some vegetative state patients show fMRI responses similar to those of healthy controls when instructed to perform mental imagery tasks. Many authors have argued that this provides evidence that such patients are in fact conscious, as response to commands requires intentional agency. I argue for an alternative reading, on which responsive patients have a deficit similar to that seen in severe forms of akinetic mutism. Akinetic mutism is marked by the inability to form and maintain intentions to act. Responsive patients are likely still conscious. However, the route to this conclusion does not support attributions of intentional agency. I argue that aspects of consciousness, rather than broad diagnostic categories, are the more appropriate target of empirical investigation. Investigating aspects of consciousness provides a better method for investigating profound disorders of consciousness.

1 Introduction

Some vegetative state (VS) patients, upon being instructed to perform mental imagery tasks, show the same patterns of brain activity as healthy controls. Since the first demonstration in a single patient by Adrian Owen and colleagues (in Owen et al. [2006]), this effect has been replicated in a number of patients, in several different labs (Boly et al. [2007]; Monti et al. [2010]; Cruse et al. [2012b]; Owen [2013]). The patients that responded showed no overt behavioral indicators of consciousness. Yet their neural responses to instructions seem like evidence that they are in fact conscious. One of the diagnostic criterion for the Minimally Conscious State (MCS) is 'following simple commands' on a 'reproducible or sustained basis' (Giacino et al. [2002] p. 351). The activation observed in these patients seems like a replicable response to commands, and so evidence of preserved consciousness.

Owen et al. claim that their patient was thus *misdiagnosed* as VS.

> These results confirm that, despite fulfilling the clinical criteria for a diagnosis of vegetative state, this patient retained the ability to understand spoken commands and to respond to them through her 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] 1402)

That conclusion has been both scientifically and philosophically controversial (Naccache [2006]; Levy [2008]; Nachev and Hacker [2010]; Shea and Bayne [2010]; Hohwy and Reutens [2011]). The controversy has mainly focused on *whether* such patients are conscious. This is

an important question, but not the only interesting one. Suppose for the sake of argument that these patients are conscious. A mystery remains: *how* are they conscious? What is it like to be them? An answer, if available, is both scientifically and ethically important.

I will present a tentative story about the awareness that remains. The story will come in two steps. First, I will argue for a particular model of the deficit in these patients. In particular, I claim that responsive patients are best understood as similar to akinetic mute patients. Such patients do not act because they do not intend to act, and they do not intend to act because they have lost the capacity to form and maintain intentions. Thus they display inaction without paralysis.

If this is right, then Owen et al. misdescribe their patient as having made a 'decision to cooperate.' Indeed, most authors treat the positive responses of these patients as evidence of preserved intentional agency. I will explain why my interpretation is preferable. Having done so, I will reflect on what it might like to be such a patient and the best method to investigate such questions. I argue that labels like 'Vegetative State,' while diagnostically useful, don't pick out natural kinds with respect to consciousness or conscious content. Because of this, we ought to instead on what I will call *aspects* of conscious states. Aspects are gradable dimensions along which conscious states can vary. There are potentially many such dimensions. Variation along distinct dimensions can lead to diagnostically indistinguishable states, and variation along the same dimension can put a patient into different diagnostic categories. As such, aspects are the proper carving for scientific investigation.

2 Responses in the Vegetative State

2.1 The Imaging Evidence

I will focus on the fMRI paradigm first used by Owen et al. ([2006]).[1] Owen et al.'s patient was a 23-year old woman with a severe traumatic brain injury, diagnosed as being in a vegetative state VS since her accident. She was instructed to perform two different mental imagery tasks: playing tennis and walking through her house. These instructions were separated by verbally cued periods of rest. The primary data was the observed differential neural activity in each condition versus rest.

The patient showed patterns of brain activity very similar to those of healthy, conscious controls given the same instructions. In the tennis condition, activation compared to rest was seen in the supplementary motor area (SMA); in the spatial navigation condition, differential activation was seen in the parahippocampal gyrus, the posterior parietal cortex, and the lateral premotor cortex. This activation was sustained over the entire 30-second task block; equally importantly, the activation in both groups *disappeared* upon the command to rest (see Figure 2 of Monti et al. ([2010]) for a nice illustration). This is different from the isolated, transient activation in remnant cortex that some VS patients show in response to single words, and which probably does not provide good evidence for consciousness (Plum et al. [1998]). Nor is the activation is linked merely to the *words* 'tennis' and 'house': similar activation is *not* seen in controls presented with either single words or more evocative non-command sentences (Owen et al. [2007]), and similar activation *is* still seen when patients are cued with neutral words (Monti et al. [2010]). In short, the responses seem to be genuine responses to the commands themselves, rather than residual low-level functioning.

The use of fMRI raises some concerns. Vegetative state patients have drastically depressed cerebral metabolism (Plum et al. [1998]; Laureys et al. [2004]). fMRI depends on the incremental change between task conditions and so is insensitive to global metabolic baseline; global baseline, however, is crucial for overall conscious state (Hyde et al. [2002]). Along the same lines, the statistical model used by Owen et al. did not strictly speaking *test* for sustained activity. Rather, it *assumed* sustained activity as part of the model, and showed that the observed activity fit that model with an amplitude significantly larger than zero. So while the scans of both the patient and the controls looked very similar, that similarity might obscure theoretically important differences (a common concern about fMRI. See Roskies [2007]; Klein [2010]).

Having noted these issues, I propose to put them to one side. The fact that both healthy and anesthetized controls did not show activation in a variety of potentially confounding situations is decent evidence that the observed activation in patients is not merely artifactual. Further, relatively few VS patients actually show this activation. If all of them did, one might suspect a mistake. The fact that only a handful do supports the contention that this paradigm really does pick out something different about patients who respond.

This paradigm also avoids many of the standard issues that can arise when interpreting fMRI studies. In particular, it avoids the problems associated with so-called 'reverse inference,' which occur when we make inferences about similar regional activity in distinct task contexts (Poldrack [2006]). Healthy controls perform a task, and regional patterns of activation are observed. Patients are instructed to do the *very same task.* The same activations are observed. We thus infer that patients were doing the same thing as the controls. The Owen et al. paradigm thus doesn't involve changing task contexts, which is arguably at the root of the reverse inference problem (Klein [2012]). Indeed, such inferences don't even need any thesis about how mental states are localized—regional

activation can just be treated as a simple marker of task completion.

2.2 The Need for Models

Call patients who showed a detectible activity in Owen et al's paradigm *responsive* patients. Before we discuss the conscious state of responsive patients, we need a good model of their internal state. That is, we need a story about what is preserved in these patients and what has been damaged.

It may not be obvious that we need a model. One might be tempted to think that Owen et al.'s test is simply a more sensitive indicator of the presence of consciousness than behavioral methods. Functional neuroimaging often promises direct and unmediated access to mental constructs that we otherwise detect only clumsily and indirectly. Whatever you think about that for neuroimaging generally, it's demonstrably false here.

In cases where the sensitivity of Owen et al's paradigm can be assessed, it seems like a relatively *in*sensitive test for residual conscious awareness. Consider Monti et al.'s 2010 study, which used the same procedure as Owen et al. and partially replicated their findings (Monti et al. [2010]). Monti et al. started with 54 patients: 23 diagnosed on behavioral grounds as VS and 21 as MCS. The test gave a positive result for 4 of the VS patients—but only 1 of the MCS patients. So the vast majority of MCS patients are not responsive to the paradigm. Which means that the test does not reliably track consciousness in cases where we have good independent evidence that consciousness is present.

Of course, that is not great evidence *against* the proposition that they are aware, either. For one, the statistical power of fMRI tests is quite low (Yarkoni [2009]; Button et al. [2013]), and so negative results should be interpreted with caution (Monti et al. [2009]).

For another, as these are patients who have been diagnosed as MCS on behavioral grounds, we have independent evidence that these patients are conscious. What it *does* tell against is the proposition that Owen et al's paradigm is simply a better, more sensitive test for MCS than behavioral methods.

There are cases where we can accept a new test for a property without a model of how the test links up to the property. However, that requires the results of that test to be cross-calibrated with older, well-established tests (Cronbach and Meehl [1955]). Lacking such cross-calibration, we need a model of the test, the property being tested for, and how the two link up.

A good model of responsive patients will explain two things. First, it will have to explain the observed residual brain activity in response to commands. Second, it will have to explain why responsive patients do *not* make behavioral responses—that is, why they are behaviorally similarly to VS patients, even though their neural responses suggest that they are conscious.[2]

Given that, the space of possible models is constrained. If you consciously intend to act, and you are able to act, you act (all things being equal). Owen et al's patient does not act. So either she is not able to act on her intentions, or she does not have intentions to act in the first place. (A third possibility: she neither intends nor is able to act. But this possibility doesn't explain how she is distinct from other severely disabled VS patients, as the brain responses suggest she is. So I assume that is a nonstarter.) This means that possible models fall into two broad categories: ones that place the deficit at the ability to act, and ones that place the deficit at the intentions themselves. I intend to defend the latter sort of model.

To reiterate, however, what follows is not skeptical about *whether* responsive patients are conscious. Again, let's assume that they are. (Provisionally so; I will return to this question in section 5.1.) While we're at it, let's also assume that behaviorism is false, that being in a certain brain state is sufficient for consciousness, that while we normally get evidence about consciousness and intentions via behavior we might get it from other reliably correlated sources, and that therefore there's nothing in principle odd about getting that sort of information from a brain scan. Even given all that, interesting questions remain.

3 Responsive Patients and Akinetic Mutism

I will argue that responsive patients lack the ability either to form or to maintain intentions. As such, their responses ought not to be considered as intentional actions. The mental imagery that occurs is, as Zoe Drayson suggests, closer to a 'mere happening' than to a full-fledged intentional response (Drayson [2014] pp. 28–9). As such, responses do not indicate consciousness, at least if the route to such indication requires intentional agency.

The route to this conclusion will come in three steps. First, I will introduce the syndrome of *akinetic mutism* (AM). Second, I will argue that akinetic mutism is best understood as a deficit of intention. Third, I will argue that responsive patients ought to be understood as similar to patients with an especially severe form of AM. While this comparison has been suggested in passing (Alkire et al. [2008]), I think the implications of it have not been fully grasped. In particular, I think that this categorization provides a satisfying model both of responsive patients' preserved activity and of their lack of behavior in ordinary circumstances.

A word of caution about this argument. I am not arguing that responsive patients *are* AM.

Such a diagnosis would be impossible without clinical examination. My argument does not in any case require diagnosis. Rather, I focus on AM because it provides an important existence proof. Severely AM patients fail to act at all unless prompted externally. That is very odd. It might seem philosophically questionable if it did not actually happen. The literature on AM patients also makes clear that their deficit is one of intention, rather than (as I will consider later) in their ability to act. I offer akinetic mutism as a plausible model of responders only in the sense that it is plausible that they share the same deficit of intention. Whether responsive patients are actually AM is not relevant to the remainder of the argument; my focus will be on evidence for a shared deficit, rather than on the syndrome as such.

Before I begin, a note on the term "intention." I will use the term as it is typically used in the literature on impaired states of consciousness: that is to say, rather loosely. In most discussions of impaired consciousness, "intention" simply stands for whatever internal motivational state gives rise to a particular action, subject only to the restriction that it is a sufficiently complex state that it cannot occur completely automatically. Intentional action, in this sense, might (for example) stem simply from some occurrent conscious desire. In many philosophical contexts, of course, "intention" and "intentional agency" are relatively demanding contexts, and the subject of considerable debate. The minimal view seems to require at least a particular desire-belief pair (Sinhababu [2013]); more sophisticated accounts require capacities like self-reflection and the ability to make temporally extended plans (e.g. Bratman [2000]; see Holton ([2009]) Ch1 for a nice review).

I take no stand on competing philosophical accounts, and intend what follows to be neutral between them (whether I am successful I leave to the reader to judge). If 'intention' were not so widely entrenched in the scientific literature, I would prefer another term. That said, I think the scientific use is not without merits. For one, we can surely say things

about the intentions of impaired patients without giving a full analysis of the term. For another, treating intention as a (potentially) easy state to fulfill is the less risky choice in the present context. It's very implausible that badly damaged patients have full capacities for self-reflection and temporally extended planning, for example. If we set the bar that high, the game is over before it starts. Conversely, denying that responsive patients have intentional agency even in a (potentially) minimal sense is a relatively more difficult task; if I am successful, then, so much the better.

3.1 Akinetic Mutism

A common symptom of damage to the frontal or cingulate cortex is diminished 'drive' or 'self-activation.' This is termed *abulia*: 'the specific neurological syndrome comprising slowness, decreased responsiveness, apathy, etc.' (Fisher [1983] p. 9). Abulic patients show lack of spontaneous activity (including speech) and flattened affect and apathy without depression. The activity that is present is often slow, minimal, and halting (Fisher [1983] p. 15).

A severe, persistent abulic state has been traditionally termed *akinetic mutism*, following Cairns et al. ([1941]). Akinetic mutism can be caused by damage anywhere to the thalamocortical loop that includes pre-SMA, cingulate cortex, striatum, and the paramedian and intralaminar thalamic nuclei (Ure et al. [1998]; Tekin and Cummings [2002]). Akinetic mutism is a

> wakeful state with prominent apathy, indifference to pain, thirst or hunger; lack of motor and psychic initiative, spontaneous movements, verbalization and response to commands... Abulia is a less severe form of this psychomotor

retardation.

(Tekin and Cummings [2002] p. 651)

There appears, then, to be a gradient of deficit running from abulia down through akinetic mutism, with more severe forms corresponding to a more severe deficit in motivation and drive.

The term 'akinetic mutism' itself has fallen out of favor as a label for several reasons. The term suggests an all-or-nothing condition, rather than a gradable one. The 'mutism' is strictly speaking redundant (Fisher [1983] p. 11). There has been some concern that the term has not been used in a consistent manner (American Neurological Association [1993]). I appreciate the point of this downgrading. However, I suggest that Akinetic Mutism—precisely because it is more narrowly focused—represents a scientifically and philosophically interesting syndrome worth closer inspection.

Left to themselves, akinetic mute patients tend neither to speak nor to act (hence the term). However, AM patients remain capable of action, and often quite complex action, when prompted. Bogousslavsky et al. report an AM patient who became 'passive' after a stroke. He remained in a bed or an armchair all day, initiating movement only to go to the bathroom. He was unresponsive to events in the ward around him. However, he was perfectly capable of complicated, intelligent activity, for 'with constant activation he was able to move and walk normally, he could play cards, answer questions, and read a test and comment on it thereafter; however, these activities would stop immediately if external stimulation disappeared' (Bogousslavsky et al. [1991] p. 308). Similarly Laplane et al. report a patient who did nothing all day unless prompted, but when prompted could 'perform quite correct complex tasks (for example, playing bridge)' (Laplane et al. [1984] p. 377).

Should you find it too implausible that severely damaged patients can understand and follow instructions in the absence of more complex higher function, I note that command-following even in the face of severely compromised higher function has long been documented (Luria [1961] Ch3). Encoding and execution of spoken instructions seems to rely on a network of brain regions that are partially distinct from the regions that can be damaged to cause AM. A study and meta-analysis by Stocco et al. found that the encoding and execution of novel instructions involves posterior parietal cortex, the caudate nucleus, and rostral lateral prefrontal cortex (Stocco et al. [2012]). There is little overlap between these areas and the medial frontal and subcortical structures implicated in AM. I suggest that this provides a plausible explanation for preserved command-following in AM patients in the absence of other motivational states.

Akinetic mutism is a graded condition. Less severe forms respond to questioning, and can thus be quizzed about their mental state. Unresponsive patients also occasionally recover, and so can be quizzed about what it was like. What they say is instructive.

Laplane et al.'s patient described his state as 'a blank in my mind' (Laplane et al. [1984] p. 377) Bogousslavsky et al.'s patient 'never mentioned his previous activities and, when asked about his job, he answered that he had no project to go back to work. When asked about his private thoughts he just said 'that's all right', 'I think of nothing', 'I don't want anything'.' (Bogousslavsky et al. [1991] p. 310). Another of their patients 'did not have any projects for the future and did not have any personal thoughts' (Bogousslavsky et al. [1991] 314) Engleborgh et al.'s patient 'rarely spoke spontaneously and took no verbal initiative. When asked about the content of his thoughts, the patient claimed he had none, suggesting a state of mental emptiness.' (Engelborghs et al. [2000] p. 1763).

3.2 AM as a Deficit in Intentions

The primary deficit in AM patients seems to lie in forming intentions. AM patients present with a picture of profound apathy. It appears that they do not act because they do not want to act. That negation should be read with wide scope: the patients have no intentions whatsoever, rather than an intention not to act. Discussing patients like these, Bogousslavsky et al. note that although capable of intellectual activity, 'They seem to have lost not only the willingness to search for satisfactions, but also the very need for mental and affective activities.' (Bogousslavsky et al. [1991] p. 314).

Treating AM as a deficit in intention is partly a theoretical move. AM patients are, in many cases, physically able to act. Laplane et al.'s patient could play bridge when prompted—he was clearly not paralyzed, nor incapable of fine motor planning, nor unable to do the (quite complex) cognitive tasks required by the game. Ordinarily if someone has all of these abilities and yet fails to act, we treat them as lacking intentions—for intention is the only ingredient that's missing.

As AM is a graded condition, the responses of less impaired patients also provide an important clue. Those first-person reports support the hypothesis that what is missing is the will to act. As Watt and Pincus put it, 'Lesser versions of the syndrome seen in more limited cases of bilateral cingulate disease typically show sufficient recovery that patients are later able to report experiencing events but lacking desire or intention.' (Watt and Pincus [2004] p. 102).

This interpretation of AM patients needs further defense, however. Some AM patients still respond to commands. Indeed, it is that preserved response to commands that I will argue sheds light on the preserved activity in responsive patients in Owen et al.'s study.

Command-following is usually taken to be an intentional action, however. Suppose I tell you to get me a lemonade and you do so. You may not have had an antecedent desire to get me a lemonade; you may have had no special pro-attitudes towards lemonade; you may have been especially obedient and so acted without explicit deliberation.[3] Nevertheless, few would deny that you had performed an intentional action.

We may formalize this a bit further. Call *endogenous intentions* intentions that are caused and maintained by an agent's desires or other standing desire-like internal states of mind. Call *stimulus-evoked cognition* any mental state that that is primarily caused or maintained by factors external to an agent. Stimulus-evoked cognition includes complex states that arise from verbal commands. The dividing line between the two is not crisp, and ordinary action is often caused by a mix of both endogenous intentions and stimulus-evoked cognition. I might, for example, decide to drive to the store to buy some milk, and then do it. That action will play out in a way that reflects both my endogenous intentions (my desire for milk, my decision to drive) and features of the world that generate stimulus-evoked cognition and action (the traffic lights, the sirens, the police bullhorn).

Nevertheless, the distinction is useful when it comes to analyzing syndromes like Akinetic Mutism. AM patients appear to show a dissociation between endogenous intention and stimulus-evoked cognition. They appear to more or less lack endogenous intentions altogether: their profound apathy suggests that there is little coming from within to move them. Stimulus-evoked cognition is more complicated. They have the ability to take up and carry out commands, which suggests some preserved cognitive abilities. The more severe the condition, however, the more difficult it is to create or maintain such mental states, and the simpler the actions that can be reliably invoked.

This implies that command-following can occur in the absence of endogenous

intention—that someone can follow a command without having an occurrent intention or desire to follow a command. That is not, I take it, terribly implausible—there are a number of cases where one can simply take up a command and act on it without further deliberation. This is most obvious in cases of motor learning, where close regulation of behavior by a teacher through a stream of verbal instructions is necessary to pick up a skill. It can also occur in sport and military circumstances, where rapid, non-deliberative response to commands is a crucial prerequisite for organized action.

In each of these cases, a command can be taken up and acted upon, I suggest, *without* having to link into a corresponding explicit endogenous intention to follow the particular command. This does not mean that such command-following is *automatic* or *involuntary*, however. Agents typically have conscious veto-power over acting on stimulus-evoked cognitions. So it is usually the case that, were a command to conflict with a subject's endogenous intentions, then the subject would not perform the command (or, at least, would have to deliberate about it).

That last clause is crucial, though. Intentional actions support non-trivial counterfactuals about what *would* have happened had the agent's beliefs, desires, and other motivational states been different. Your trip to the refrigerator counts as intentional, even when command-evoked, because you would not have done so had you not wanted to go, or had you judged that going was a bad idea, or strongly intended to do something else. Non-intentional actions, by contrast, don't have this counterfactual sensitivity: mere automatic responses, we think, are usually insensitive to what we desire or judge is best or otherwise intend to do.

I leave it deliberately vague which counterfactuals need to be supported. It will be a moot point when it comes to talking about AM patients. Whatever the plausible candidates are,

it appears that the command-evoked activities of severely AM patients fail to support any of them. As outlined above, these patients appear to lack all endogenous forms of intention. Most of the counterfactuals suggested above include some sort of endogenous intention (or something that generates an endogenous intention) in their antecedent. Hence severely AM patients cannot make any such counterfactual non-trivially true. We have no reason to think, therefore, that any preserved command-following in severely AM patients is an *intentional* action, rather than a mere happening.

3.3 Arguments for the Link

Akinetic mutism comes in more or less severe forms. Severity tracks the difficulty with which stimulus-evoked cognition can be induced. What, then, would the most severe forms look like? From the outside, I submit that *it would look exactly like a vegetative state.* Patients would not have intentions to act: endogenous intentions would be absent, and stimulus-evoked cognition too difficult to elicit. Lacking both, no action occurs. In certain borderline cases, however, it might be possible to elicit something like mental imagery.

I suggest that responsive patients have a deficit that is comparable to that seen in very severe cases of Akinetic Mutism. I have two arguments. First, responsive patients in the literature have the right sort of damage. Second, treating them having a deficit similar to that seen in AM would account for all of the data, including both the residual neural activity and the the fact that responsive patients don't act even though they are not paralyzed.

3.3.1 The Structural Argument

Inference about the source of damage from published fMRI scans is necessarily uncertain. Nevertheless, many of the responsive patients have gross structural abnormalities that can be easily discerned from the published scans. The original Owen et al. patient suffered severe frontal damage from a car accident and subsequent decompressive craniotomy (Owen et al. [2006] SOM p2). On the published scans, signs of severe damage are apparent in the frontomedial region. Crucially, this damage appears to be mostly anterior to the VCA line. The VCA line is the traditional dividing line between pre-SMA and SMA proper. So there is good evidence, at least in this patient, of severe pre-SMA damage—again, precisely the sort that leads to akinetic mutism. Similar patterns are discernible in Monti et al.'s six responsive patients (Monti et al. [2010] Figure 1). Two of the responsive patients have obvious severe damage to medial frontal cortex (patients 4 and 22), and a third has obvious damage to the cingulate cortex (patient 23). The remaining two responsive VS patients have damage that is more difficult to discern on the published scans; both, however, had suffered traumatic brain injury (Table 1), so similar damage is not ruled out.

The involvement of the pre-SMA is important because of the link to the existing literature on Akinetic Mutism; again, damage to the pre-SMA and more anterior medial frontal regions can cause abulia (Laplane et al. [1977]). It is also important for reasons specific to Owen et al's paradigm.

Boly et al. performed a study designed to test the discriminability of various imagery tasks in healthy controls, in order to validate the use of Owen et al.'s paradigm in experimental contexts ([2007]). The goal of the paper was to contrast the patterns of activation associated with different kinds of mental imagery. The study was thus primarily concerned with the differences between imagery conditions. However, Boly et al. also performed a

conjunction analysis to see what was common to the performance of distinct imagery tasks.

That analysis found two areas: the pre-Supplementary Motor Area and a portion of dorsal premotor cortex. They note that pre-SMA is often implicated in the endogenous generation of intentions and conclude that

> Though our study did not investigate the neural mechanisms of volition itself, pre-SMA and dorsal premotor cortex activation in our volunteers during an active task compared to rest, regardless of the nature of the task, could be understood as reflecting a certain form of cognitive control, or the subject's intention to perform a task. (Boly et al. [2007] p. 989)

This seems like a plausible linkage: the association between pre-SMA and voluntary, endogenously initiated action has been found in numerous studies (Eagleman [2004]; Nachev et al. [2008]).

Crucially for present purposes, we have pre-SMA activity as a plausible marker for endogenous intention *within the very same task context*. By the logic suggested in section 2.1, it is thus reasonable to use that activation as a marker for endogenous intention. The absence of such activity—and, more importantly, the fact that damage *precludes* such activity—thus tells us something about the absence of endogenous intention in these patients.

Of course, this structural argument is not knock-down. Data from many patients is lacking, and other damage might also play a causal role. Remember, though, that the argument is not that these patients have suffered thus-and-such type of damage. Nor is it that they ought to be classified as AM. Rather, the argument is that they ought to be considered as having the same sort of deficit as AM patients have. Finding that some

responsive patients have the sort of damage that is associated with that deficit is thus evidence in favor of this hypothesis.

3.3.2 The Abductive Argument

If responsive patients are relevantly similar to AM patients, then we also have a straightforward model of our original target: the preserved patterns of brain responses alongside the lack of behavioral response. Responsive VS patients lack the capacity to maintain and generate endogenous intentions. They've also almost completely lost the ability to take up stimulus-evoked cognition as well. What ability remains explains the response to command detected by the scans. What's missing explains the lack of overt behavior.

The lack of behavior is explained simply by the lack, in almost all circumstances, of sufficient intention to act. The lack of endogenous intentions means that responsive patients won't show spontaneous, self-initiated activity. The almost complete lack of ability to take up stimulus-evoked cognition means that ordinary behavioral assessment methods will not be sufficient to cause activity either. Only commands that involve no actual motion at all end up preserved. And those, of course, will be the hardest to detect behaviorally.

That might occur because mental imagery is relatively undemanding: physical action requires the coordination of muscle groups, while imagery doesn't. Or, more plausibly, it might occur because stimulus-evoked cognition is modular, and so mental imagery might be preserved in the absence of other functions that can be evoked. For example, Drayson argues that there is an important difference between physical and mental actions ([2014]). If so, then the differential demands of the two types of actions might lead to differential impairment following damage.

Owen et al.'s paradigm thus detects remnant responses to imagery tasks. That explains the patterns of brain responses as preserved flickers of command-following. In such patients, mental imagery is the last of the functions available to be evoked by command (again, either because it is easier or because of the modular structure of cognition), and so the last to be extinguished.

Note in this regard that the tasks the patients have been instructed to perform do not require anything like introspection. Even in Monti et al's question-answering paradigm, responses are contingent upon the patient's *semantic* knowledge about themselves, rather than on their present conscious state. Nor would it require anything like choice between non-task-related objectives. The patients need not, therefore, be described as hearing a command and then deciding to act—a description that implies preserved endogenous intentions. Rather, they are simply commanded to act in a certain way. If that action can be done, they do it. By thus treating responsive patients as examples of AM, we can account for exactly the pattern of data observed.

3.4 Interim Conclusion

I argued that responsive patients lack a capacity to form and maintain intentions. They have the same deficit that is seen in relatively severe forms of Akinetic Mutism, a condition marked precisely by such a deficit in intentions.

Note a non-obvious consequence of my position. Syndromes like Akinetic Mutism *cross-cut* categories like VS and MCS. Patients with mild forms will be diagnosed as disabled, those with more severe forms as MCS, and the most severe as VS. Conversely, there are many ways to become disabled to the point of being diagnosed as MCS or VS. So there are many

VS/MCS patients who are not cases of AM. So we have two non-overlapping ways to categorize disorders of consciousness. I will return to the implications of this in section 5.2.

My interpretation is not the only possible one. To complete my defense, I will present considerations against two alternative models.

4 Other Models

4.1 A Deficit in Ability?

I argued in section 2.2 that there are two possible families of models for responsive patients: ones that place the deficit at intentions, and ones that place it at the ability to act on intentions. I have defended the former sort of model. The latter sort is extremely common, however, and so warrants discussion.

Many authors have suggested that responsive patients have the intention to act, but lack the ability to do so. The fMRI scanner thus acts as a sort of prosthetic, allowing intentions to be expressed and registered despite the disability. Owen, for example, claims that his paradigm is 'the key to unlocking signs of covert consciousness in situations where all forms of physical response have been rendered unavailable' (Owen [2013] p. 111). This suggests that responders, like healthy controls, retain certain endogenous intentions to act but lack is the *ability* to make behavioral indications of those intentions. Responders thus have a problem with acting, not with thinking — or, more generally, if we think of a path from perception to conscious awareness to action, their deficit lies either on the path from consciousness to action or on the ability to act itself. Responsive patients thus intend to act in various ways—including complying with the instructions that constitute the

experimental task—but that intention is normally thwarted by a lack of ability to act.

The presence of such intentions is the crucial link between the brain data and attributions of consciousness. Recall that Owen et al. claimed of their patient that she was *cooperative*, and that this '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). Owen and Coleman similarly claim that the most 'parsimonious explanation' is that 'this patient was consciously aware and willfully following the instructions given to her, despite her diagnosis of vegetative state' (Owen and Coleman [2008] pp. 136–7). Shea and Bayne, endorsing this argument, similarly note that

> 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... 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. (Shea and Bayne [2010] p. 465)

So these patients intend to participate in the experiment. They presumably also intend to do other sorts of things, but are blocked from doing so.

Call such models *no-ability* models. If a no-ability model is correct, the closest comparison to responsive patients would be with with Locked-In Syndrome (LIS). LIS patients are conscious. Damage to motor efferents renders them unable to express their conscious state in nearly every way. Yet there is no doubt that they have intentions to act. The preservation of cranial nerves controlling the eyes allows them to express those intentions using blinks or eye movements. Further, it is plausible that LIS can become total: that is,

the remaining connections to the cranial nerves can become severed, producing a patient who is conscious but completely unable to interact with the world (see Bauer et al. [1979] case 7). If so, we would find patients with intentions but no ability whatsoever to act. We might be able to find out about those abilities via a suitably sensitive brain scan, but otherwise they would be cut off from the world. Inattentive or uninformed doctors might mistakenly diagnose them as VS.

The no-ability model paints a similar picture of responsive patients. Of course, responsive patients are unlikely to be just like LIS patients. Classic LIS patients do not have higher cortical damage. So their conscious state is mostly untouched. Responsive patients, as we have seen, have rather serious cortical damage. That surely alters their conscious state, a disanalogy with LIS. Following Owen's description, however, responsive patients at least have conscious intentions to act. These intentions are what are detected in paradigms like Owen et al's. The presence of these intentions is the crucial feature that warrants diagnosis as MCS. And these intentions are, arguably, one of the things that make responsive patients worth serious ethical attention.

The no-ability model has a pleasing simplicity. There is, however, a rather serious problem with it. It appears that at least some responsive patients *can* act. They just don't.

First, there is a straightforward story about why *locked-in* patients cannot act: damage to the ventral pons or the surrounding area severs motor efferents heading to the spinal cord (Patterson and Grabois [1986]). LIS patients are thus paralyzed for the same reason as quadriplegics are: the neural commands issued by cortex simply can't reach the muscles. In general, we want our models of responders to explain and justify our claims about the residual deficits. We would want a similar story about responsive patients. That's not easy. The damage they have suffered is not the sort that causes paralysis. (As noted in section

3.3.1, there need not even be a single focus of damage) As Lionel Naccache wonders in a commentary, 'If this patient is actually conscious, why wouldn't she be able to engage in intentional motor acts, given that she had not suffered functional or structural lesion of the motor pathways?' (Naccache [2006] p. 1396) That she doesn't suggests an explanatory gap on the no-ability account: it postulates a deficit in the ability to act, but gives neither explanation nor evidence for that deficit.

Of course, if the no-ability account was the only possible explanation for the observed phenomena, we might be justified in simply inferring the requisite lack of ability. (I suspect that this is why most authors simply assume a lack of ability, rather than actually arguing for it). But it is not the only possible explanation: responsive patients might lack intention rather than ability. More defense is needed.

In their 2010 review, Nicholas Shea and Timothy Bayne argue that it remains an open question whether Owen et al.'s patient '*has* suffered lesions of the motor pathway, *pace* Naccache's assumption.' (Shea and Bayne [2010] p. 467). They note in support of this that many VS patients present with contractures and muscle wastage. Contractures and muscle wastage are a nonspecific sequela of severe brain damage, however, occurring in both VS and MCS patients (Elliott and Walker [2005]); it's not obvious that they can play a specific explanatory role in these cases.

More importantly, there is positive evidence that Owen et al.'s patient *can* move. Owen et al. note that that 'she had preserved, but inconsistent, reflexive behaviour (startle, noxious, threat, tactile, olfactory).' (Owen et al. [2006] SOM p2) They note that such responses are difficult to elicit, and are never elaborated to more complex actions—but general paralysis does not seem to be an issue. Similarly, the clinical description of one of Monti et al.'s patients notes that he 'showed reproducible, but inconsistent, response to some commands

(i.e., 'move your leg')' (Monti et al. [2010] SOM p4). In short, at least some of these patients appear capable of motor activity at the gross level, and even (in the case of the latter patient) of at least occasional motor responses to command. So the lacuna for no-ability models cannot be dismissed so easily.

This argument is not, to be sure, entirely conclusive. Brain damage is messy, and patients diagnosed as VS are likely have a cluster of deficits. Damage to the supplementary motor cortex, or to the frontal cortex more generally can produce a variety of motor problems (Nachev et al. [2008]), including frontal apraxias that consist in difficulties in ordering and executing movements (Luria [1980] p. 233ff). One might also model the deficit of responders as a conduction deficit, affecting the linkage between intention and an otherwise intact motor system.

Note, however, that an acceptable story will be strongly constrained by other features of the case. Residual activity in the SMA is one of the markers in Owen et al.'s paradigm, so a story would have to be consistent with preserved functionality in the more posterior areas of motor cortex (lest the test itself be called into question). Any story would also have to explain why mental imagery in response to command is preserved. A general disconnection between intention and intentional action won't explain that.

There may well be such a story to be told. Absent one, however, the no-ability model is far less attractive than the Akinetic Mutism model.

4.2 Modular Intentions?

Shea and Bayne, in their response to Naccache, also offer a slightly different diagnosis of responsive patients' inability to act. They suggest that intention might have 'a somewhat

modular structure and that the ability to form intentions can be selectively impaired.' (Shea and Bayne [2010] p. 467). Responsive patients, then, might lack endogenous intentions, but still preserve enough of some other sort of intentions to allow response during Owen et al.'s paradigm. The modular structure of intentions postulates that intentions to form mental imagery are preserved (explaining responses to the Owen et al. paradigm), alongside the lack of intentions necessary for other sorts of command-following (explaining the general lack of response elsewhere).

This is, I suspect, the most promising way to preserve the claim that responsive patients have intentional agency. I agree that intention is likely to have a modular structure. Further, I agree that such an account would fit the surface phenomena. Nevertheless, the only evidence for preserved intentions in responsive patients is the fact that they respond. The impaired-intention account must claim that mental imagery in response to command is *inevitably* an intentional action. Whether this is so turns in part on what we mean by 'intentional,' and in part on the philosophical work to which we want to put the term.

As I have used it, intention is primarily a *personal-level* concept. Whole agents act intentionally, and whether a particular act counts as intentional depends, in part, on other personal-level features of the agent. Modular explanations, by contrast, are sub-personal: they explain certain patterns of personal-level behaviors and deficits by positing separable subcomponents with differential specializations in the cognitive machinery. Failure of a module might explain why an otherwise ordinary agent is impaired on certain tasks. But if enough is lacking at the *personal* level, we ought to become wary about claims of intention. That's true even if some modules are preserved, the activity of which would, in ordinary cases, result in intentional activity.

This is, of course, a refinement of the argument offered in section 3.2. There, I claimed that

intentional actions must support appropriate counterfactuals. That counterfactual structure is lacking in severely impaired patients. It is the personal-level integration between command-evoked dispositions and the rest of our intentional life that permits us to call command-evoked activity intentional in ordinary cases. In drastically impaired patients, the question becomes more complex. On both impaired-intention accounts as well as my own, this background structure is missing. As evoked responses to commands don't seem to have many of the markers of intentional action—no role for deliberation, not even links with antecedent intentions and desires, and so on—it is hard to maintain that responders must have preserved intentional agency. The akinetic mutism account, by contrast, gives a much more straightforward story. Not all stimulus-evoked activity need be intentional, even if it is reasonably complex. Thus there is no bar to some activity being preserved even in the absence of *all* endogenous intentions.

Finally, I note that the difference between the impaired intention model and the modular intention model is open to empirical test. My model crucially relies on the distinction between endogenous intentions and command-evoked activity. It should be possible to design more complex paradigms that would require endogenous intentions to correctly perform. One might, for example, present commands in two distinct voices, instructing the subject to choose one and only respond to it. If successful, this would provide evidence of preserved endogenous intentions: a *choice* would be necessary, and clearly distinguishable from a mere failure to find a response. I know of no experiment that has actually been run along these lines, but the fact that one could suggests that the debate is not *just* about the semantics of 'intention.' Similarly, I have argued that responsive patients do not possess veto-power over the preserved aspects of stimulus-evoked cognition. While it might difficult to distinguish true vetoing from mere failure to respond, should such a task be developed that would provide further test of the theory.

That said, I suspect that the present concern is partly orthogonal to Shea and Bayne's true target. Shea and Bayne are concerned with whether responsive patients are conscious. They have a much stronger argument here: even in cases of non-intentional responses to mental imagery commands, it's unquestionable that ordinary subjects are aware of both the command and the resulting mental state (Shea and Bayne [2010] p. 468). I do not doubt that part of the story. What I do doubt is that response to command is a marker of intentional activity. That makes a difference to *how* such patients are conscious of the world. And it is to that question that I now turn.

5 Consciousness and Method

5.1 Are they conscious?

Return to the question with which I began. Assume that I am right, and that responsive patients have a severe deficit in intentions. What is it *like to be* one of the responsive patients? Here, it's fruitful to return to the literature on Akinetic Mutism, which offers our best-studied model of the severely abulic state.

First, I think, there should be no doubt that there is *something* it's like to be a responsive patient. Note first that the less severe forms of AM are capable of making self-reports about their conscious state. Those reports suggest an impoverished, curiously empty state of mind. Nevertheless, they make introspective reports about that impoverished, empty state of mind—and that is extremely good evidence that responsive patients are conscious.

So lack of intention *per se* does not seem to eliminate consciousness. One might nevertheless claim that *severely* AM patients are not conscious. Antonio Damasio has

suggested this, though without much argument (Damasio [2000] pp. 121ff). One might think that behavior is necessary for consciousness, or that consciousness depends on intentional action and so is abolished if intentions are completely absent. Watt and Pincus suggest that the 'emptying out' of consciousness characteristic of AM may lead to the lack of phenomenal content in severe cases, because stimuli must 'have at least some potential affective significance in order to gain access to the conscious workspace' (Watt and Pincus [2004] p. 102) One might, I suppose, also think that consciousness arises in AM patients only when stimulus-evoked cognition kicks in, and so insofar as AM patients are conscious it is fleeting and stimulus-bound (Damasio also suggests this interpretation, at ([2000] p. 91)).

I think these readings, while tempting, are undermined by the clinical literature. Some severely AM patients recover, or at least have the severity of their condition lessened. Sometimes they are amnestic for the most severe periods—but sometimes they aren't. Damasio and van Hoesen report on an AM patient who was initially completely unresponsive. Though not entirely clear from the clinical report, her total lack of response would seem to qualify her for a diagnosis of VS. She later recovered to a relatively less severe AM state, and was asked about her experience. They report that:

> Asked if she ever suffered anguish for being apparently unable to communicate she answered negatively. There was no anxiety, she reported. She didn't talk because she had 'nothing to say.' Her mind was 'empty'. Nothing 'mattered'. She apparently was able to follow our conversations even during the early period of the illness, but felt no 'will' to reply to our questions. (Damasio and Van Hoesen [1983] 98-99).

This suggests that the patient was conscious even during the period of severe, VS-like

inactivity. Further, it suggests that she was aware of things like the conversations going on around her—that is, things other than direct commands and their consequences. So it seems that even severely AM patients are conscious of their surroundings. That suggests that responsive patients are as well.

Of course, the mental state of AM patients might still seem mysterious. What is it like to lack endogenous intentions? It is probably not understandable as a merely quantitative change—that is, we cannot understand the AM patient as like us, but with fewer and fewer intentions. What we intend shapes what we are aware of, and the removal of those intentions should produce a striking qualitative change in our awareness.

Here, I ask you to permit a bit of speculation. Ordinarily we are conscious when we follow a set of directions, and conscious of those directions, because of *other* intentions we have that might interfere with executing our plan. Consciousness thus marks the need for selection amongst intentions. At the subpersonal level, one is conscious of the tasks that require focused processing resources because they are novel or complex and so are potentially interfered with by other intentions you have (Shallice [1988] Ch14). At the personal level, consciousness is associated with the choice to follow some set of directions *rather than* some other intention one might have. Conversely, some tasks do not require full conscious engagement. Over-learned motor skills and navigation in familiar environments, say, can fade into the background. One might still be conscious of stimuli relevant to the tasks involved, and become more fully aware of stimuli that signal something unusual or otherwise relevant to task performance. But under ordinary circumstances, those take only a fraction of awareness. Note that this claim is carefully hedged. You might think that that selection is a function (or even *the* function) of consciousness awareness. Or you might just think that this is the function of attention, and that consciousness goes along with attention. The link may be even more indirect, so long

as consciousness and selection for action are associated.

Following Uriah Kriegel, call these two modes of awareness *focal* versus *peripheral* consciousness (Kriegel [2004]). I suggest that responsive patients may have preserved peripheral consciousness in the absence of focal consciousness. Since they have no endogenous intentions, they have no intentions that need to take up focal consciousness. Nor do they have any intentions with which current plans might conflict. They still monitor the world for potentially relevant information—but in the absence of intentions, no information becomes actually relevant. External commands can, under the right circumstances, cause focal consciousness to be directed by some stimulus-evoked cognition. But in the absence of such a driving command, there is only periphery without focus.

Purely peripheral consciousness is difficult to imagine. It is as close to a purely passive state of consciousness as you can get. As Watt and Pincus put it, AM patients must 'live in a kind of strange, virtually unfathomable netherworld close to the border of a persistent vegetative state' (Watt and Pincus [2004] p. 102). Nevertheless, I argue, it fits quite well with both behavioral data and self-reports of AM patients. It would also make sense of Owen et al's patient and her curious inaction. We might get some sense of what it's like if we consider the way in which we are aware of our surroundings while we are engaged in some other task—the dim awareness of the road as we talk while driving, or our awareness of the trees outside the window as we type. Save that in our case, our intentions might change in such a way that those peripheral things capture our attention, either because our desires change or because the world itself grabs us. Responsive patients do not have the same structure of intentions, and so most of the time they have nothing to elicit focal awareness. Only when a sufficiently strong command grabs them might there be focal awareness. Focal awareness might last just as long as there is some intention in which it is of service, and then ebb away again.

Responsive patients, then, are conscious and aware of their surroundings. That awareness is not like a faded or fragmented version of our awareness, though. We always have *some* endogenous intentions, even we we are half-asleep or drunk or bored. The change in responsive patients is more profound: not just a lack of conscious intentions, but a corresponding change in the very structure of conscious contents themselves. A striking change, perhaps—but not, I think, an unintelligible one.

5.2 Aspects Versus Levels

Astute readers will note that, despite arguing against Owen et al., I have come to the very same conclusion they have, via a more circuitous route. That might seem like the sort of improvement that only a philosopher could love. Note, however, that the methodological presuppositions of my argument are really quite different. I conclude by elaborating the virtues of my approach.

Owen et al., focused on different *levels* or *modes* of consciousness. 'Vegetative State', 'Minimally Conscious State', and similar labels are broad groupings that depend on a few simple diagnostic criteria and group patients into a rough hierarchy of levels of awareness. Most of the debate about responsive patients so far has been over whether they meet the criterion to be classed in MCS rather than VS. My argument, in contrast, focused on a particular syndrome, akinetic mutism. In section 3.4, I suggested that syndromes and levels cross-cut one another: AM patients might belong to any of several levels depending on severity. Conversely, each level includes patients with a variety of different syndromes.

Thinking about levels of awareness is important for many purposes. MCS patients might be more likely to recover, for example, and so be entitled to a larger share of a finite pool

of resources (Giacino [2005]). So too with responsive patients (Di et al. [2008]). MCS patients, because they are definitely conscious, might also deserve better treatment. These propositions aren't uncontroversial—see (Wilkinson et al. [2009]) for a critical review—but they do show the sort of thing that broad levels of awareness might be useful for determining.

On the other hand, if we care about the details of an individual's conscious state—its structure, contents, relations, and so forth—then we ought to carve things differently. Why? For starters, two patients can meet the criteria for being at a level for completely different neurological reasons. Levels arguably thus don't form projectable natural kinds. Discovering that one MCS patient has some mental property shouldn't give us much reason to believe that another MCS patient has the same property (outside of properties that track the definitional criteria themselves, of course). Similar remarks apply to, for example, the use of propofol as a matched control condition when examining unconscious processing (Davis et al. [2007]). Propofol globally depresses firing; the resulting state is behaviorally comparable to VS. But unless you assume that this behavioral similarity picks out a natural kind, it's not at all obvious whether discoveries about the one have anything to do with the other.

Instead, I suggest, we ought to focus on *aspects* of consciousness. An aspect, as I will use the term, is a set of functionally similar conscious capacities that vary together in a more or less gradable manner. The preceding has argued for one important aspect of consciousness: the capacity to form and maintain intentions. AM patients give similar first-person reports. Most appeared to be aware of the world, yet with a 'blank' state of mind, a lack of interest in action, and a lack of intention to act.

The capacity to form and maintain intentions is only one of several potential aspects of

conscious awareness. Most researchers recognize that at least two dimensions—awareness and wakefulness/arousal—are necessary to capture the variability amongst different disorders (Monti et al. [2009] Fig. 1). Expansions to this minimum have been proposed, however. Watt and Pincus suggest that at least six aspects are relevant to disorders of Consciousness: Arousal, Intention, Emotion, Intention, Short-Term Memory, and Working Memory capacity (Watt and Pincus [2004] Table 3.1). Bayne and Hohwy similarly consider a whole range of possible aspects that might be impaired in disorders of consciousness. These include attentional capacity, different kinds of conscious content, global features like bandwidth and accessibility of contents, and so on. Importantly, they suggest that the overall level of consciousness might be aspect-dependent—that is, that a patient might count as VS if tested on one aspect and MCS if tested on a different one (Bayne and Hohwy, forthcoming).

A focus on aspects, I suggest, would be more scientifically fruitful than a focus on levels, for several reasons. First, aspects are more tightly defined. As I have used the term, an aspect collects up functionally similar properties of the conscious state. Hence we ought to expect that patients who have differing levels of the same aspect have something in common. In the case of Akinetic Mutism, for example, each patient has lost (to a greater or lesser degree) the capacity to form and maintain endogenous intentions.

That in turn provides a more consistent basis for inference about the effects of variation along an aspect. We saw that the verbal reports of AM patients also had many similarities. As aspects are graded, we ought to be able to infer from less severe forms of a syndrome to more severe ones. That is, we can take cases where we *do* have evidence for some particular conscious content, and try to run that through to the limit to see what it would be like to be such a patient.

The argument in the preceding sections thus represents a different method for getting at states of awareness in the profoundly disabled. Owen et al. focused on a small set of diagnostic criteria that were sufficient to place patients in a relatively broad category. One may then infer that patients have the other properties characteristic of other patients at that level, including the possibility for recovery. That is useful in medical contexts, where chance of recovery is something we would very much like to predict. On the other hand, such broad labels are less sensitive to facts about conscious state. If I am right, some vegetative state patients are in fact conscious, but in a way that is very unlike the way in which we are conscious.

Further, and more importantly, patients diagnosed as minimally conscious might be conscious in ways that are radically different from one another. My account, again, only applies to a subset of minimally conscious patients: the ones who have a relatively severe deficit in forming intentions. We can say something about their conscious state, but that probably won't apply to patients who are MCS for different reasons. Other MCS patients, for example, might have preserved endogenous intentions but other profound cognitive disabilities that limit their ability to act in situationally appropriate ways. Those differences matter if we want to theorize about consciousness. They might also have profound ethical implications, for only patients that have endogenous intentions can have intentions that are unfulfilled, frustrated, and thwarted.[4]

In sum, we have good theoretical and ethical reasons to group together patients whose conscious states are similar to one another. I have argued that aspects of consciousness form a better basis for such inference, and are more likely to form natural kinds. These kinds cross-cut diagnostic categories, which suggests that diagnostic categories aren't a suitable basis for scientific and philosophical theorizing about conscious states.

That, I suggest, is the important consequence that follows from the groundbreaking work of Owen et al. As many have noted, it is not terribly surprising that we might get evidence for conscious state in the absence of the typical behavioral indicators of conscious state. What is more surprising is that, within otherwise unresponsive patients, we might nevertheless find a great diversity of conscious states. We are only now beginning to tease apart that diversity.

Funding

Australian Research Council (DP1097264 to David Chalmers; FT140100422 to C.K.).

Acknowledgements

Thanks to Derek Baker, Tim Bayne, David Chalmers, Jakob Hohwy, to two Anonymous Reviewers, and to audiences at the Australian National University and Macquarie University for helpful feedback on earlier drafts. Research was made possible by support from the Centre for Consciousness at the ANU.

Department of Philosophy
Macquarie University NSW 2109
Australia
colin.klein@mq.edu.au

Notes

[1]EEG-based methods have subsequently been developed Cruse et al. [2012a;c]. My story ought to apply to those as well, though the lack of structural scans will make some points necessarily less certain in particular cases.

[2]I will assume that this similarity is not simply a matter of *overlooking* behavioral cues. That is a serious possibility: many MCS patients are undoubtedly misdiagnosed as VS because the behavioral signs of MCS are subtle and easy to overlook (Gill-Thwaites [2006]). However, despite occasional reports of responsive patients showing residual behavioral signs of MCS upon re-examination, most accounts do not suggest that this is what's going on. In many cases, responsive patients otherwise continue to meet the behavioral criteria for VS.

[3]Though discussions of this are rare in the philosophy of mind literature, many political philosophers note that authoritative commands are meant to *preclude* deliberation and to be effective regardless of an agent's intentional state. See for example Raz's discussion of exclusionary reasons in (Raz [1975]).

[4]Note that this is not the *only* thing that affects ethical status. The ability to feel pain might be preserved, for example, which would also have obvious ethical implications. The point is merely that when we turn to ethical implications of a disordered state, the details of the preserved conscious content might make a difference to permissible and impermissible activities.

References

Alkire, M. T., Hudetz, A. G., and Tononi, G. [2008]: 'Consciousness and anesthesia', *Science*, **322**(5903), pp. 876–80.

American Neurological Association [1993]: 'Persistent vegetative state: Report of the American Neurological Association committee on ethical affairs', *Annals of Neurology*, **33**, pp. 386–90.

Bauer, G., Gerstenbrand, F., and Rumpl, E. [1979]: 'Varieties of the locked-in syndrome', *Journal of Neurology*, **221**(2), pp. 77–91.

Bayne, T. and Hohwy, J. [Forthcoming]: 'Modes of consciousness', . In Sinnott-Armstrong, W., *(ed.)* Forthcoming, *Consciousness after Severe Brain Damage: Medical, Legal, Ethical, and Philosophical Perspectives*. Oxford University Press: New York.

Bogousslavsky, J., Regli, F., Delaloye, B., Delaloye-Bischof, A., Assal, G., and Uske, A. [1991]: 'Loss of psychic self-activation with bithalamic infarction', *Acta Neurologica Scandinavica*, **83**(5), pp. 309–16.

Boly, M., Coleman, M., Davis, M., Hampshire, A., Bor, D., Moonen, G., Maquet, P., Pickard, J., Laureys, S., and Owen, A. [2007]: 'When thoughts become action: An fMRI paradigm to study volitional brain activity in non-communicative brain injured patients', *Neuroimage*, **36**(3), pp. 979–92.

Bratman, M. E. [2000]: 'Reflection, planning, and temporally extended agency', *Philosophical Review*, pp. 35–61.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., and Munafò, M. R. [2013]: 'Power failure: Why small sample size undermines the reliability of neuroscience', *Nature Reviews Neuroscience*, **14**, pp. 365–76.

Cairns, H., Oldfield, R., Pennybacker, J., and Whitteridge, D. [1941]: 'Akinetic mutism with an epidermoid cyst of the 3rd ventricle', *Brain*, **64**(4), pp. 273–90.

Cronbach, L. and Meehl, P. [1955]: 'Construct validity in psychological tests.', *Psychological bulletin*, **52**(4), pp. 281–302.

Cruse, D., Chennu, S., Chatelle, C., Bekinschtein, T. A., Fernández-Espejo, D., Pickard, J. D., Laureys, S., and Owen, A. M. [2012a]: 'Bedside detection of awareness in the vegetative state: A cohort study', *The Lancet*, **378**(9809), pp. 2088–94.

Cruse, D., Chennu, S., Chatelle, C., Fernández-Espejo, D., Bekinschtein, T. A., Pickard, J. D., Laureys, S., and Owen, A. M. [2012b]: 'Relationship between etiology and covert cognition in the minimally conscious state', *Neurology*, **78**(11), pp. 816–22.

Cruse, D., Chennu, S., Fernández-Espejo, D., Payne, W. L., Young, G. B., and Owen, A. M. [2012c]: 'Detecting awareness in the vegetative state: Electroencephalographic evidence for attempted movements to command', *PloS one*, **7**(11), pp. 1–9.

Damasio, A. [2000]: *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, Harcourt, Brace, and World, Inc.: New York.

Damasio, A. and Van Hoesen, G. [1983]: 'Emotional disturbances associated with focal lesions of the limbic frontal lobe', . In Satz, P. and Heilman, K. M., (*eds*) 1983, *Neuropsychology of Human Emotion*. Guilford Press: New York, pp. 85–110.

Davis, M., Coleman, M., Absalom, A., Rodd, J., Johnsrude, I., Matta, B., Owen, A., and Menon, D. [2007]: 'Dissociating speech perception and comprehension at reduced levels of awareness', *Proceedings of the National Academy of Sciences*, **104**(41), pp. 16032–37.

Di, H., Boly, M., Weng, X., Ledoux, D., and Laureys, S. [2008]: 'Neuroimaging activation studies in the vegetative state: Predictors of recovery?', *Clinical Medicine*, **8**(5), pp. 502–7.

Drayson, Z. [2014]: 'Intentional action and the post-coma patient', *Topoi*, **23**, pp. 23–31.

Eagleman, D. M. [2004]: 'The where and when of intention', *Science*, **303**(5661), pp. 1144–6.

Elliott, L. and Walker, L. [2005]: 'Rehabilitation interventions for vegetative and minimally conscious patients', *Neuropsychological Rehabilitation*, **15**(3-4), pp. 480–93.

Engelborghs, S., Marien, P., Pickut, B. A., Verstraeten, S., and Deyn, P. P. D. [2000]: 'Loss of psychic self-activation after paramedian bithalamic infarction', *Stroke*, **31**, pp. 1762–5.

Fisher, C. [1983]: 'Honored guest presentation: Abulia minor *vs.* agitated behavior', *Clinical Neurosurgery*, **31**(9), pp. 9–31.

Giacino, J., Ashwal, S., Childs, N., Cranford, R., Jennett, B., Katz, D., Kelly, J., Rosenberg, J., Whyte, J., Zafonte, R., et al. [2002]: 'The minimally conscious state: Definition and diagnostic criteria', *Neurology*, **58**(3), pp. 349–53.

Giacino, J. T. [2005]: 'The minimally conscious state: Defining the borders of consciousness', *Progress in brain research*, **150**, pp. 381–95.

Gill-Thwaites, H. [2006]: 'Lotteries, loopholes and luck: Misdiagnosis in the vegetative state patient', *Brain Injury*, **20**(13-14), pp. 1321–28.

Hohwy, J. and Reutens, D. [2011]: 'A case for increased caution in end of life decisions for disorders of consciousness', *Monash Bioethics Review*, **28**(2), pp. 14.1–12.

Holton, R. [2009]: *Willing, Wanting, Waiting*, Oxford University Press: Oxford.

Hyde, F., Rothman, D. L., and Shulman, R. G. [2002]: 'Total neuroenergetics support localized brain activity: Implications for the interpretation of fMRI', *Proceedings of the National Academy of Sciences*, **99**(16), pp. 10771–6.

Klein, C. [2010]: 'Images are not the evidence of neuroimaging', *British Journal for the Philosophy of Science*, **61**, pp. 265–78.

Klein, C. [2012]: 'Cognitive ontology and region- versus network-oriented analyses', *Philosophy of Science*, **79**(5), pp. 952–60.

Kriegel, U. [2004]: 'The functional role of consciousness: A phenomenological approach', *Phenomenology and the Cognitive Sciences*, **3**(2), pp. 171–93.

Laplane, D., Baulac, M., Widlöcher, D., and Dubois, B. [1984]: 'Pure psychic akinesia with bilateral lesions of basal ganglia.', *Journal of Neurology, Neurosurgery & Psychiatry*, **47**(4), pp. 377–85.

Laplane, D., Talairach, J., Meininger, V., Bancaud, J., and Orgogozo, J. [1977]: 'Clinical consequences of corticectomies involving the supplementary motor area in man', *Journal of the Neurological Sciences*, **34**(3), pp. 301–14.

Laureys, S., Owen, A. M., and Schiff, N. D. [2004]: 'Brain function in coma, vegetative state, and related disorders', *The Lancet Neurology*, **3**(9), pp. 537–46.

Levy, N. [2008]: 'Going beyond the evidence', *The American Journal of Bioethics*, **8**(9), pp. 19–21.

Luria, A. [1961]: *Speech and the Regulation of Behaviour*, Pergamon Press: London.

Luria, A. [1980]: *Higher Cortical Functions in Man.*, Basic Books: New York, 2nd edition.

Monti, M. M., Coleman, M. R., and Owen, A. M. [2009]: 'Neuroimaging and the vegetative state', *Annals of the New York Academy of Sciences*, **1157**(1), pp. 81–9.

Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., Owen, A. M., and Laureys, S. [2010]: 'Willful modulation of brain activity in disorders of consciousness', *New England Journal of Medicine*, **362**(7), pp. 579–89.

Naccache, L. [2006]: 'Is she conscious?', *Science*, **313**(5792), pp. 1395–6.

Nachev, P. and Hacker, P. [2010]: 'Covert cognition in the persistent vegetative state', *Progress in neurobiology*, **91**(1), pp. 68–76.

Nachev, P., Kennard, C., and Husain, M. [2008]: 'Functional role of the supplementary and pre-supplementary motor areas', *Nature Reviews Neuroscience*, **9**(11), pp. 856–69.

Owen, A. and Coleman, M. [2008]: 'Detecting awareness in the vegetative state', *Annals of the New York Academy of Sciences*, **1129**(1), pp. 130–38.

Owen, A., Coleman, M., Boly, M., Davis, M., Laureys, S., and Pickard, J. [2006]: 'Detecting awareness in the vegetative state', *Science*, **313**(5792), p. 1402.

Owen, A. M. [2013]: 'Detecting consciousness: A unique role for neuroimaging', *Annual review of psychology*, **64**, pp. 109–33.

Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., Jolles, D., and Pickard, J. D. [2007]: 'Response to comments on "Detecting awareness in the vegetative state"', *Science*, **315**(5816), p. 1221c.

Patterson, J. and Grabois, M. [1986]: 'Locked-in syndrome: A review of 139 cases.', *Stroke*, **17**(4), pp. 758–64.

Plum, F., Schiff, N., Ribary, U., and Llinás, R. [1998]: 'Coordinated expression in chronically unconscious persons.', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **353**(1377), pp. 1929–33.

Poldrack, R. A. [2006]: 'Can cognitive processes be inferred from neuroimaging data?', *Trends in Cognitive Sciences*, **10**(2), pp. 59–63.

Raz, J. [1975]: 'Reasons for action, decisions and norms', *Mind*, **84**(1), pp. 481–99.

Roskies, A. L. [2007]: 'Are neuroimages like photographs of the brain?', *Philosophy of Science*, **74**, pp. 860–72.

Shallice, T. [1988]: *From Neuropsychology to Mental Structure*, Cambridge University Press: Cambridge.

Shea, N. and Bayne, T. [2010]: 'The vegetative state and the science of consciousness', *The British Journal for the Philosophy of Science*, **61**(3), pp. 459–84.

Sinhababu, N. [2013]: 'The desire-belief account of intention explains everything', *Noûs*, **47**(4), pp. 680–96.

Stocco, A., Lebiere, C., OReilly, R. C., and Anderson, J. R. [2012]: 'Distinct contributions of the caudate nucleus, rostral prefrontal cortex, and parietal cortex to the execution of instructed tasks', *Cognitive, Affective, & Behavioral Neuroscience*, **12**(4), pp. 611–28.

Tekin, S. and Cummings, J. [2002]: 'Frontal-subcortical neuronal circuits and clinical neuropsychiatry: An update.', *Journal of Psychosomatic Research*, **53**, pp. 647–54.

Ure, J., Faccio, E., Videla, H., Caccuri, R., Giudice, F., Ollari, J., and Diez, M. [1998]: 'Akinetic mutism: A report of three cases', *Acta Neurologica Scandinavica*, **98**(6), pp. 439–44.

Watt, D. F. and Pincus, D. I. [2004]: 'Neural substrates of consciousness: Implications for clinical psychiatry', . In Panksepp, J., *(ed.)* 2004, *Textbook of biological psychiatry*. Wiley-Liss Incorporated: Hoboken, chapter 3, pp. 75–110.

Wilkinson, D. J., Kahane, G., Horne, M., and Savulescu, J. [2009]: 'Functional neuroimaging and withdrawal of life-sustaining treatment from vegetative patients', *Journal of Medical Ethics*, **35**(8), pp. 508–11.

Yarkoni, T. [2009]: 'Big correlations in little studies: Inflated fMRI correlations reflect low statistical power', *Perspectives on Psychological Science*, **4**(3), pp. 294–98.