Awake But Not Aware?
Probing for Consciousness in Unresponsive Patients

Gary Williams
3-02-14

Abstract: The standard approach in clinical neurology is to diagnose disorders of consciousness (DOC) on the basis of operationally defined behaviors. Critics of the standard approach argue that it relies on a flawed behaviorist epistemology that methodologically rules out the possibility of covert consciousness existing independently of any observable behavior or overt report. Furthermore, critics point to developments in neuroimaging that use fMRI to “actively probe” for consciousness in UWS using mental imagery tasks (Owen et al. 2006). Critics argue these studies showcase the limitations of the standard approach. The goal of this paper is to defend the standard approach against these objections. My defense comes in two parts: negative and positive. Negatively, I argue that these new “active probe” techniques are inconclusive as demonstrations of consciousness. Positively, I reinterpret these active probes in behavioral terms by arguing they are instances of “brain behaviors”, and thus not counterexamples to the standard approach.

Introduction

When clinicians diagnose patients recovering from disorders of consciousness (DOC) such as coma, unresponsive wakefulness syndrome (UWS),¹ and minimally conscious state (MCS), it is important to ask how these diagnostic decisions are justified. Answering this question is morally significant insofar as such decisions pertain to pain management, therapy, and end-of-life issues. Furthermore, trying to answer the question quickly runs into a philosophical problem analogous to the traditional problem of other minds, i.e. what justifies our beliefs about other minds?

Clinicians typically justify their diagnostic decisions for DOC on the basis of bedside observation using pre-theoretical behavioral criteria such as eye-opening, withdrawal from painful/noxious stimuli, and

¹ UWS is a new term for the condition previously called “vegetative state” (Laureys et al., 2010; Laureys & Boly, 2012). Researchers proposed the name change because until an “active probe” with modern brain-imaging technology is used, it is hasty to infer a patient lacks communicative readiness. In addition to the pejorative connotations of “vegetable” in reference to individuals, they argue that the term “vegetative” is inaccurate as a blanket diagnostic label because it ignores the possibility that they will find residual cognitive function using innovative brain-imaging equipment.
visual pursuit of objects, gaze-following, command following, and fluid verbal expression [Giacino ref]; Shea and Bayne call this the “standard approach” in clinical neurology [2010]. The reliance on behavior is traditionally defended as a methodological necessity. Since consciousness as a subjective phenomena cannot itself be observed, the science of consciousness must rely on some kind of behavioral report of consciousness. For this reason, standard diagnostic criteria for DOCs have been behavioral in nature since the 1960s [Jennett & Plum, 1972; Plum & Posner]. Nevertheless, critics argue that the standard approach in clinical neurology is “seriously flawed” [Overgaard 2009 p. 13] because of its implicit behaviorist epistemology (Kurthen 1991). Shea and Bayne also recommend abandoning the pretheoretical requirement of reportability (2010).

Critics of the standard approach point to recent evidence from neuroimaging that indicates the presence of consciousness in patients diagnosed with UWS according to traditional behavioral assessment. For example, Owen et al. (2006) used a novel brain imaging technique to “actively probe” for consciousness in an UWS patient using mental imagery tasks and functional magnetic resonance imaging (fMRI). Critics of the standard approach argue these findings cast doubt on the reliability of behavioral scales [references].

My aim in this paper is to defend the standard approach in light of these objections. First, I show that the new “active probe” techniques are inconclusive as demonstrations of consciousness. I argue that due to lack of theoretical consensus on which theory of consciousness is correct, we cannot provide principled justifications for choosing a criterion other than behavioral criteria, which have proven to be useful and reliable in the clinical context. Second, I argue that active probes are not at odds with the standard approach; rather, they show we need to expand our concept of “behavior”. I contend via analogy to “inner speech” that these new active probes using fMRI are a means to detect previously unmeasurable behaviors in the same way brain-machine interface (BMI) technology are a means to behave, e.g. using your brain to control a computer cursor.
Accordingly, Owen’s findings do not undermine the “gold standard” approach of using behavioral criteria to diagnosis DOC. To be clear,

2.0 Problems with the “Gold Standard” Approach

2.1 Terminological preliminaries

The term “consciousness” is infamously ambiguous and can refer to different phenomena to different people in different contexts [Zeman, 2001, 2005; Wilkes, 1988]. For now, I will adopt Ned Block’s distinction between phenomenal consciousness and access consciousness [ref: 1995].

“Phenomenal consciousness” is difficult to noncircularly define [Goldman, 1993], but it is supposed to capture “experiential properties” such as the painfulness of a toothache or the redness of a sunset. In contrast, “access consciousness” refers to our ability to access and report our experiences. Block and colleagues argue that phenomenal and access consciousness can be dissociated from each other, and thus the absence of report does not indicate an absence of phenomenology; phenomenology “overflows” cognitive access [Block, 2011]. This claim is particularly relevant when researchers evaluate behavioral criteria for the diagnosis of consciousness in UWS patients. On Block’s assertions, the absence of behavior does not entail an absence of phenomenology because someone could be phenomenally conscious without any ability to report it to the outside world.

Not all agree with Block on this issue [Weiskrantz, 1997; Baars & Laureys, 2005; Kouider, 2010; Weisberg, 2011; Cohen & Dennett 2011]. Theorists dispute the overflow hypothesis, arguing that the right kind of access is both necessary and sufficient for consciousness. Since this paper focuses on epistemic criteria for detecting consciousness and not defining consciousness, I will bracket these issues and take researchers at their word when they say they are trying to detect phenomenal consciousness and not

---

2 Footnote on creature consciousness and transitive vs nontransitive concepts of consciousness.
3 Following Block [ref] I am using the term “phenomenology” as synonymous with phenomenal consciousness.
4 “If I am right about the occurrence of the phenomenal NCC [neural correlate of consciousness] without the access NCC, then if some patients who are now classified as vegetative have recurrent processing in sensory cortices, they would have to be re-evaluated.” [Block, 2005, p. 270]
access consciousness.\(^5\)

### 2.2 The Argument from Locked-In Syndrome

Unresponsive wakefulness syndrome is a rare disorder diagnosed on the basis of a clinical team that observe, over a period of time, a decisive lack of behavioral evidence for any awareness of self and environment. In contrast to coma patients (Fig. 2), UWS patients are considered “awake but unaware” given they breathe on their own, open their eyes, go through sleep/wake cycles, and display reflexes ranging from simple twitches to grimacing, crying, and other intentional-looking behavior such as repeating a word [Schiff ref].

Common behavioral indices for diagnosing coma and disorders of consciousness include the Glasgow Coma Scale (GCS), Full Outline of UnResponsiveness (FOUR), Wessex Head Injury Index, 

---

\(^5\) A representative statement comes from cognitive neuroscientist Anil Seth, who writes in the forward to the 2009 text *Coma Science* “Consciousness is the appearance of a world. In its absence there is no self, no environment, no pain, no joy; there is simply nothing at all. Following Thomas Nagel, without consciousness there is ‘nothing like it is to be’ (Nagel 1974)” (p. ix).
Sensory Modality Assessment and Rehabilitation Technique (SMART), JFK Coma Recovery Scale (CRS), JFK Coma Recovery Scale-Revised (CRS-R). By all accounts these behavioral scales are subjective and relative to the skills of the examiner and to the inherent difficulty of interpreting non-verbal behavior. Furthermore, a strict operationalist could use these scales to diagnose recovery from UWS without making any reference to consciousness whatsoever.

Accordingly, there are two main objections to these behavioral scales. First, the case of a patient with Total Locked-In Syndrome (TLIS) is an apparent counterexample to behavioral approaches since by definition a TLIS patient has full preservation of their cognitive capacities without the occurrence of any behaviors that would warrant a diagnosis other than UWS. In contrast to the total locked-in state, other Locked-In Syndrome (LIS) patients can communicate through minimal movement such as blinking their eyelids, nodding their head, or twitching a muscle [references]. Again, by definition, we cannot diagnosis TLIS using bedside behavior, and yet we know that TLIS occurs since some LIS patients retrospectively report having been in a Total LIS state before transitioning into a regular LIS state [reference]. Accordingly, we can make the argument against relying on behavioral criteria:

The Argument from Locked-in Syndrome

1. UWS and TLIS patients display the same behavioral profile.
2. TLIS patients are by definition conscious, whereas UWS are presumably not conscious.
3. Therefore, behavioral criteria for consciousness are inadequate for the diagnosis of consciousness because they will always misdiagnose TLIS as UWS.

It is important to note that premise (1) is about total locked-in patients, defined as being fully conscious but lacking in any muscle control by which to communicate with the outside world. Normal locked-in patients with at least one channel of communication such as eye-blinking or head-nodding do display different behavioral profiles than UWS patients. Unlike the reflexive blinking of UWS patients, LIS patients will blink intelligently if a symbol code can be established and systematically tested, e.g. “blink once for yes; blink twice for no”. Morse code can also be used. But total locked-in patients by definition behave
identically to UWS patients. Thus, if clinicians relied on behavioral criteria alone, this would effectively make it impossible to diagnose TLIS patients with consciousness. Of course we cannot dismiss the possibility of a TLIS patient’s consciousness due to lack of muscle capacity; it would be as absurd as stating that a person stops being conscious when they take a muscle relaxer that paralyzes all movement.

The second main objection to behavioral approaches is that they are thought to be fundamentally flawed insofar as they completely ignore the phenomenological components of conscious experience, which for many philosophers is not only the defining feature of consciousness but a morally relevant factor [footnote controversy]. For example, in a review of a recent book defending a behavioral approach to consciousness (ref: Shulman 2013), cognitive neuroscientist Anil Seth writes:

“Almost all active consciousness researchers would agree that any satisfactory definition has to be grounded in the (phenomenal and/or functional) properties of subjective experience and not solely in their overt behavioral accompaniments.”

However, this statement fails to keep separate two distinct issues: theoretical definitions of consciousness vs. operational definitions of consciousness. Even if we agree with Seth that theoretical definitions of consciousness in terms of behavior are unsatisfactory, it’s entirely possible that other considerations will make it desirable to give operational definitions of consciousness exclusively in terms of behavior. I develop this argument further in section 6.

Nevertheless, it’s important to note that Seth’s worry about behavioral definitions is not arbitrary. Indeed, he argues that behavioral definitions are unsatisfactory because Adrien Owen has recently provided evidence based on a neuroimaging case study that purports to demonstrate the existence of “conscious awareness” in a patient that was diagnosed with UWS on purely behavioral grounds. In other words, if this patient were diagnosed with UWS using only behavioral criteria but subsequent neuroimaging

6 Shulman controversially endorses the conjunction of behaviorism and operationalism: “I propose that to explore consciousness we should first define consciousness by observable behaviour and then find physical brain activities that are properties of the person showing that behaviour” (2013, p.22).
Evidence demonstrates they are conscious, then the behavioral criteria are inadequate for diagnosing the presence or absence of consciousness in UWS patients.

However, there are reasons to doubt the conclusiveness of this study. First, just because the patient was diagnosed as unconscious on some behavioral criteria it does not follow that all behavioral criteria will diagnose the patient as unconscious. In other words, the study might indicate that current behavioral criteria are inadequate, not that all possible behavioral criteria are inadequate. Second, the evidential criteria used by Owen and colleagues to demonstrate consciousness could be interpreted as behavioral in nature, albeit “covert” behaviors. Third, and most importantly, any claim about whether patients are conscious because they show capacity X must be justified in terms of a theory of consciousness that says X is necessary for consciousness. Since we lack any such theoretical foundation, such claims are epistemically unconstrained. Before I flesh out these arguments, we need to know more about Owen’s case study.

3.0 A Novel Paradigm for Detecting Consciousness

3.1 Owen’s case study

In July 2005, a 23 year old woman in a car accident suffered severe brain damage. After five months she remained unresponsive, and thus satisfied all behavioral criteria for diagnosis of UWS. However, when Owen and colleagues put the woman in a fMRI machine and presented her with spoken sentences (e.g. “There was milk and sugar in his coffee”) the speech areas of her brain activated similarly to normal controls. Furthermore, when presented with ambiguous spoken sentences (e.g. “The creak came from a beam in the ceiling”) the patient showed additional activation in frontal areas, an indication of deeper semantic processing in normal controls. Owen et al. reasoned that since language understanding is a strong indicator that a person is “consciously aware” there is a possibility this patient has preserved aspects of consciousness that go undetected when using standard behavioral criteria.
However, research on priming undermines Owen’s conclusion. [Reference] strongly suggests that deep semantic processing can occur in the absence of consciousness [references]. So how can we be sure the patient’s brain activity in response to spoken words isn’t an unconscious response? In order to rule out the possibility, Owen and colleagues conducted a second fMRI study with the same patient, using spoken instructions to engage mental imagery. Specifically, the patient was asked to either imagine herself playing a game of tennis or imagine herself walking from room to room in her own house starting from the front door. In normal controls, imagining a motor task reliably activates the supplementary motor area (SMA) [reference].

Sure enough, when the patient was asked to imagine playing tennis, significant activity was observed in the SMA. In contrast, when the patient was asked to imagine herself walking through her home, significant activity was observed in the parahippocampal gyrus, the posterior parietal cortex, and the lateral premotor cortex. Crucially, the patient’s activations in both conditions were indistinguishable from the activation of normal controls given the same task. In conclusion, Owen et al. argue, “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” [2006, p. 1402].

3.2 The skeptical challenge and the argument from persistence

Despite their confidence in this conclusion (“beyond any doubt”), there are prima facie reasons to doubt the strength of this inference. First of all, how do we determine whether the patient’s brain activity isn’t unconscious activation in response to the words “tennis” and “house”? Owen responds to this challenge by explaining that the pattern of activation “was not transient, but persisted for the full 30 seconds of each imagery task, that is far longer than would be expected, even given the haemodynamics of the fMRI response” [Owen, Assessment, Neurology of Consciousness, p. 170]. What is the relevance of transience? Owen et al claim, “Sustained activity in these regions of the brain is impossible to
explain in terms of unconscious responses to either single 'key' words or to short sentences containing those words.” Their evidence for this claim is based on a supplementary study [Owen et al 2007 Response to comments] that showed no sustained activity in response to single key words in healthy volunteers.

In sum, Owen et al. claims “Such [persistent] responses are impossible to explain in terms of automatic brain processes” [reference]. What does Owen mean by “automatic” in this context? Earlier he says, “Although it is well documented that some words, can, under certain circumstances, elicit wholly automatic neural responses in the absence of conscious awareness, such responses are typically transient (i.e., lasting for a few seconds) and, unsurprisingly, occur in regions of the brain that are associated with word processing” (p. 170). Owen's argument runs as follows:

The Argument from Persistence

1. Unconscious brain activity in response to stimuli are typically transient, lasting less than 30 seconds.
2. The patient's brain activity in response to stimuli persisted for 30 seconds.
3. Therefore, the patient's brain activity indicates conscious awareness.

There is reason to be skeptical about the strength of this argument. First, it’s important to note this argument is probabilistic in nature because the first premise says unconscious responses are typically transient but it doesn't say unconscious responses are always transient. How do we know that brain damage patients are typical in this respect? Furthermore, even if we strengthened premise (1) to say unconscious activity is always transient, there are possible counterexamples. Take the case of David Armstrong’s long-distance truck driver [reference]. The driver at the end of his shift feels tired, but he still pays attention to the road. Suddenly, the driver “comes to” and realizes he has no recollection of the past five minutes, yet he managed to drive safely during that time. Armstrong and others argue that the best interpretation of the case is that he lost consciousness while “blanked out”.

Far from mere speculation, a real-life case of “homicidal somnambulism” has been reported. [Paragraph on case]
The upshot is that if certain theories of consciousness are true, it’s not at all “impossible” for complex unconscious processing to persist for longer than 30 seconds. Indeed, if John Bargh is right that unconscious processing makes up 99% of all mental activity [Bargh ref; James quote] then it seems plausible much unconscious brain activity will persist longer than 30 seconds. Of course, some theorists think the best interpretation of this case is that the driver was consciously aware of the road the entire time but his introspective machinery was unable to access this awareness [references; Searle? Block]. I don’t purport to resolve the debate between First-order and Higher-order theories of consciousness in this paper. My point, however, is that whether or not it’s “impossible” for unconscious processes to persist longer than 30 seconds depends on which theory of consciousness is true out of dozens of contested options.

3.3 Voluntary command-following

In addition to the Persistence Argument, Owen also seems to make the following argument:

**The Argument from Voluntary Command-following**

1. The patient was able to understand spoken commands and voluntarily follow instructions.
2. Voluntarily following instructions is something that cannot happen unconsciously.
3. Therefore, the patient's response indicates conscious awareness.

There are several problems with this argument. First, there is evidence that people can follow complex verbal instructions *without* the full accompaniment of conscious awareness. For example, consider the anesthetic technique called “procedural sedation”--sometimes called “conscious sedation”-- used when doctors need a patient to communicate and follow verbal instructions during painful or uncomfortable procedures. To do this, conscious sedation involves a combination of two drugs: (1) a powerful painkiller such as fentanyl, and (2) a short term “amnesic” drug such as propofol, which binds to GABA receptors in the brain [reference]. I experienced this myself in a medical setting after dislocating a shoulder. I can remember receiving fentanyl, and I can remember suddenly “coming to”-- shocked to realize my arm was
readjusted in a cast and that fifteen minutes elapsed since taking the amnesic. In other words, there is a distinct “gap” or “blind spot” in the memory, which accords with William James’s observation that, “If the consciousness is not aware of [time gaps], it cannot feel them as interruptions” (Chap IX, p. X).

Was I conscious during the procedure? Maybe I was conscious the whole time, but I didn’t lay down any memories of my consciousness. Or maybe memory formation in a “global workspace” is essential to consciousness [Baars, X]. Either way, the doctors later told me I was following directions and communicating, albeit in a somewhat dissociated fashion. Of course, the truth of whether people are really conscious during procedural sedation hinges on which theory of consciousness we deem correct. Regardless of which theory of consciousness is most accurate, the relevant point is that according to some strongly supported theories consciousness can clearly be dissociated from command following.

Another counterexample to the argument from command-following is that on some theories of consciousness dogs are not conscious [Carruthers reference] yet are clearly able to “willfully” obey spoken commands with persistent responses. One might disagree with this theory of consciousness and argue that any theory which denies consciousness to dogs is implicitly wrong. Given that contemporary philosophers argue that bacteria are conscious [Thompson], rocks are conscious [Strawson], and that consciousness is an intrinsic property of the universe akin to mass, spin, and electric charge [Chalmers], we cannot sustain appeals to common sense, intuition, or “apparent” assumptions over consciousness when such appeals carry no theoretical leverage.

Furthermore, consider the phenomenon of hypnosis. [paragraph on hypnosis]

Lastly, consider the case of “automatic writing”. In Principles of Psychology, William James wrote: [quote and paragraph on automatic writing]

3.4 Methodological refinements

Some people worry that Owen’s command-following paradigm is too cognitively demanding because the patients must hold two planned responses in their memory (“Tennis for yes” and “Walking for
no”) in addition to listening to instructions and generating imagery. The consideration is whether Owen’s task is too complicated for patients with brain damage who might be able to complete simpler tasks. Accordingly, researchers modified Owen’s paradigm and developed a hierarchical “single-response” paradigm where the patient generates a single type of mental image rather than choosing between generating two types of mental images. This study focused on six brain-damaged patients with different diagnoses including minimally conscious state (MCS), emerged from MCS, and locked-in state.

In the first level of the hierarchy, patients were asked to follow the command, “Imagine swimming now...stop”. At the second level, the patients were asked binary yes or no questions such as, “If your mother’s name is Norma, imagine swimming now...stop.” For the third level, the patients were first shown a playing card and asked to respond to either the suit (“heart, spade, etc.”) or face-value (“Jack, Queen, etc.”). Then the patients were given four multiple choice questions: “If the face of the card is a Jack, imagine swimming now...stop” and so on. All of these probes were validated with healthy controls.

The results were inconclusive. Three out of six patients failed to show any significant activity for the first command-following level of the task, with two out of these three meeting diagnostic criteria for MCS, and the third recently emerged from MCS. Of the three that passed the first level, two were in MCS and one was in LIS. A particularly striking feature of these results is that one patient who had emerged from MCS displayed fluid and expressive language but showed no significant activation on these imagery tasks, even though she confirmed she had attempted to perform the task. A dissociation between fluid verbal behavior and success on the imagery tasks was also seen for the binary and multiple-choice tasks in two subjects.

Therefore, even if these mental imagery paradigms are capable of detecting willful command-following in some brain damaged patients, it would be rash to use neuroimaging data alone to
make significant ethical decisions, due to the risk of false negatives. In its current form, neuroimaging technology is just one more diagnostic tool available for clinicians in addition to traditional bedside measures.

4.0 The Behaving Brain

In this section I argue that Owen's case study does not undermine the reliability of behavioral criteria if we broaden our understanding of what it means to behave. Throughout most of the 19th and 20th centuries, the only behaviors a clinician could observe were overt muscle movements that resulted in speech or limb movement, or subtle muscle movement like a twitch, blink, or head-nod. However, I contend that advancement in neuroimaging and brain-interfacing technology allows us to expand the concept of “observable behavior” to include what I call “brain behaviors”. For many readers, the idea of a “brain behavior” comes across as an oxymoron, but the idea can be illuminated by examples.

---

7 “It is generally agreed that the present state of imaging technologies cannot provide alternative markers of awareness” - Giacino chap 14 The neurology of consciousness p. 180
The simplest definition of a “brain behavior” is any behavior that is performed using the brain itself as a means for accomplishing a task. Although it’s true that all behaviors ultimately have their causal origin in the nervous system, most behaviors are carried out using our muscles, limbs, and fingers as tools for accomplishing tasks, e.g. using your hand to pick up a beer. However, some tasks can be accomplished without the use of “external” muscles. With biofeedback technology we can use our heart muscles to make decisions, e.g. “For ‘yes’, speed up your heartbeat; for ‘no’, slow down your heartbeat.” In the biofeedback example, the heart is the instrument by which the task is accomplished. Similarly, a brain behavior is a behavior that uses the brain itself as the instrument. Neuroscientists have developed BCIs that enable novel brain behaviors using a variety of techniques (Fig. 3).

Fig. (3) From Sorger et al. 2009

5.2 Types of Brain Behaviors

In the context of Owen’s case study, one might worry that the concept of a “brain behavior” will fail to make sense of more “cognitive” activities such as imagining playing tennis or walking through one’s home. Can these imaginative acts also be construed as types of brain behavior? I think they can. Consider the well-known concept of “inner speech”, which is commonly defined as “silently talking to oneself”. While there are reasons to think there are individual differences in the incidence and vividness of “inner speech”, I believe the phenomenon is familiar to most people. Furthermore, I contend that the boundary between verbal speech (a behavior) and inner speech (an “inner” behavior) is a matter of degree that admits no sharp boundaries. If you begin by repeating a word aloud and slowly lesson your volume to totally “inner” speech, I challenge you to find a sharp cut-off point to mark the moment intending speech behavior stops “feeling” the same as intending inner speech behaviors, e.g. choosing to “yell” or “scream”

---

8 Julian Jaynes poetically described inner speech as an “introcosm” whereby we converse with ourselves in a "secret theater of speechless monologue and prevenient counsel” [p.1].
9 Quote from William James, ref
in inner speech. Accordingly, I propose that phenomena such as inner speech are “analog behaviors”. An analog is a type of model generated at every point by the thing it is a model of (e.g. a map is an analog of a particular environment). Our inner speech is thus an analog of outer speech.

The Russian psychologist Lev Vygotsky is well-known for proposing an “outside-inside” model of inner speech development in childhood [reference, Thought and Language]. He argues that when a toddler first begins to speak there is no inner speech, and that when toddlers are presented with a problem to solve, they “talk through” the steps just as you would count on your hands to “off-load” computational demand: “When circumstances force [the child] to stop and think, he is likely to think aloud” (thought and language p. 35). As children become older, the frequency of their overt speech during problem solving decreases, eventually disappearing altogether. But crucially, Vygotsky argued that the self-regulative function of outer speech was preserved yet fully internalized.

Using inner speech as a paradigm, I contend that we can provide a behavioral interpretation of Owen’s case study. In my view, when Owen and colleagues asked the woman to imagine playing tennis, they were asking her to perform an analog behavior. Analog behaviors can function as brain behaviors if we are hooked up to BCIs. If Owen’s research team had done a localizer scan for inner speech, they could have just as easily asked the patient to imagine herself reciting a poem in her head. This possibility has already been demonstrated in a 2004 study by Yoo et al. (2004). Using fMRI they had subjects perform four different tasks: “right-hand motor imagery”, “left-hand motor imagery”, “mental calculation”, and “inner speech”. Each task was assigned to a direction: right, left, up, down- respectively. Using their brain alone, the subjects were able to use an fMRI device as a controller to navigate a 2D maze on a computer screen. This paradigm case demonstrates analog behavior as brain behavior.

In sum, active probes using imagery tasks do not constitute a counterexample to the use of behavioral criteria for diagnosing DOCs. Instead, active probes enable a novel form of communication by

---

10 [footnote Sellars, Antipodians].
co-opting the capacity for performing analog behaviors such as inner speech and motor imagery.

6.0 What Do Active Probes Actually Probe?

6.1 The virtues and vices of behavioral criteria

The purpose of diagnostic criteria is to help determine when we have landed on an accurate diagnosis. But within coma science, there is an open debate about what we should be looking for. Many researchers in the field cite Thomas Nagel (197x) and claim we should be looking for “what-it’s-likeness”, “qualia”, or the “appearance of a world” [reference]. However, there are several problems with framing the goal of coma science in terms of studying “appearances”.

First, a focus on the “appearance of a world” raises questions about which creatures have a world appear to them. For example, the 19th century biologist Jakob von Uexküll pioneered the study of the “Umwelt” or “lifeworld” of animals but did so in order to study the life worlds of simple creatures like ticks and flies. In the context of coma science, the problem is that it seems highly unlikely that Owen’s mental imagery probes are designed to detect a property we have in common with insects.

Furthermore, even if our best theories of consciousness told us that there is something-it-is-like to be a tick, it’s doubtful whether neurologists would still care about probing for consciousness in UWS patients. As Levy argues (2009), the presence or absence of what-it’s-likeness in a creature is arguably relevant to its status as a moral patient but not necessarily as a moral agent. A moral patient is an entity whose well-being matters, particularly with respect to the capacity for experiencing pain or pleasure.11 If UWS are conscious in this sense, then they are clearly moral patients, but the problem is that such a diagnosis would not morally distinguish them from simple creatures that have a relatively low moral status.

Moreover, even if we were trying to detect qualia in coma patients, there is no consensus on which theory we should use to develop a set of measurable criteria for the diagnosis of consciousness in UWS. For example, higher-order theorists will recommend different criteria than first-order theorists

---

11 “The question is not, Can they reason? nor, Can they talk? but, Can they suffer?” (Bentham 1789).
because these theories make different predictions about the minimally sufficient conditions for consciousness, e.g. whether or not attention is a necessary or sufficient condition for consciousness. Without a principled way to decide between these rival theories, we cannot currently evaluate the soundness of the argument from persistence or the argument from command-following, both of which are crucial to rebutting skeptical challenges to the use of new “active probe” methodologies. In sum, until we have a validated theory of consciousness, I am skeptical that any conclusions about consciousness straightforwardly follow from neuroscientific evidence, let alone conclusions that are “beyond any doubt”.

Behavioral criteria do not suffer from either of these problems. There is no deep epistemic problem in determining whether simple creatures are capable of operationalized behaviors. Standard assessment scales like the JFK Coma Recovery Scale have been validated on inter-rater and test-retest reliability (Giacino 2009 fools good). If we are interested in verbal expressiveness, we don’t have to worry about whether there could be verbal expressiveness in the absence of observable behavior precisely because it is an observable behavior. Some philosophers might object this as naive operationalism that defines away the consciousness of patients with locked-in syndrome or Guillain Barre syndrome- an inflammatory disorder of the nerves that can temporarily block all voluntary muscle action. This objection is unwarranted partly because advances in neuroimaging technology enable clinicians to observe analog behaviors that wouldn’t have been observable thirty years prior. Undoubtedly, future technological developments will expand our ability to measure analog behaviors in greater detail and establish high-bandwidth links to patients with motor problems.

6.2 What are active probes actually probing for?

Are active probes probing for the “appearance of a world” or the capacity to imagine and follow verbal directions? It does not seem that we are interested in whether or not an UWS patient can imagine themselves playing tennis because it is an indicator for something else we are interested in: “qualia”. Rather, it seems we are interested in whether a UWS patient can imagine playing tennis because we place
a high value on the capacity for imagination itself.

Consider what it involves to truly follow the directions “imagine yourself playing tennis”. First, one must understand that the word “tennis” refers to a recreational game played with certain equipment in accordance with a set of rules. Thus, an understanding of “tennis” indicates a sophisticated cognitive grasp of human language, culture, and sport. Regardless of whether consciousness is necessary for complex cultural cognition [see Sleutals, *greek zombies*], such capacities are highly valued by the family members of UWS patients.

**Conclusion**

**References**