Inverse zombies, anesthesia awareness, and the hard problem of unconsciousness

George A. Mashour a,*, Eric LaRock b

a Department of Anesthesiology and Neurosurgery, University of Michigan Medical School, 1H247 UH/Box 0048, 1500 East Medical Center Drive, Ann Arbor, MI 48109-0048, USA
b Department of Philosophy, Oakland University, Rochester, MI, USA

Abstract

Philosophical (p-) zombies are constructs that possess all of the behavioral features and responses of a sentient human being, yet are not conscious. P-zombies are intimately linked to the hard problem of consciousness and have been invoked as arguments against physicalist approaches. But what if we were to invert the characteristics of p-zombies? Such an inverse (i-) zombie would possess all of the behavioral features and responses of an insensate being, yet would nonetheless be conscious.

While p-zombies are logically possible but naturally improbable, an approximation of i-zombies actually exists: individuals experiencing what is referred to as “anesthesia awareness.” Patients under general anesthesia may be intubated (preventing speech), paralyzed (preventing movement), and narcotized (minimizing response to nociceptive stimuli). Thus, they appear—and typically are—unconscious. In 1–2 cases/1000, however, patients may be aware of intraoperative events, sometimes without any objective indices. Furthermore, a much higher percentage of patients (22% in a recent study) may have the subjective experience of dreaming during general anesthesia.

P-zombies confront us with the hard problem of consciousness—how do we explain the presence of qualia? I-zombies present a more practical problem—how do we detect the presence of qualia? The current investigation compares p-zombies to i-zombies and explores the “hard problem” of unconsciousness with a focus on anesthesia awareness.

© 2008 Elsevier Inc. All rights reserved.

1. Easy problems and the hard problem of consciousness

David Chalmers (Chalmers, 1996, 2000) draws a distinction between the easy problems of consciousness and the hard problem of consciousness. Easy problems are about the performance of cognitive functions and can be solved by discovering the neural or computational mechanisms that perform them. For example, if we could identify the neural mechanism that underlies and performs the focus of attention, we would have a solution to one of the easy problems. Chalmers mentions several other easy problems, such as the deliberate control of behavior, the integration of information by a cognitive system, and the ability to report internal states (Chalmers, 2000). In contrast, the hard problem refers to the problem of experience. What Chalmers means by experience is the subjective aspect of consciousness that normally accompanies the processing of information at higher levels by some cognitive system: ‘When we think and perceive, there is a whir of information-processing, but there is also a subjective aspect’ (Chalmers, 2000). By implication, the mere processing of information does not guarantee experience, even though experience is most likely related to information. The problem of experience is truly a hard
problem because it is not clear that experience could be accounted for in terms of a cognitive system’s functional or neural mechanisms alone. A purely functional explanation of consciousness seems to leave out an explanation of experience itself (see Chalmers, 1996; Chalmers, 2000). This insight motivates the logical possibility of a zombie.

2. Zombies and the irreducibility of consciousness

A zombie is identical to Chalmers (or to any other ostensibly conscious being) in all physical and functional respects, but lacks conscious experience entirely. Demonstrating the logical possibility of a zombie could count as a refutation of physicalism, since it would seem to show that there is no conceptual entailment from functional and physical facts to experiential facts. The central requirement of this type of demonstration is conceptual coherence. One need not show that zombies are naturally possible, only logically possible.

Chalmers talks quite fondly about the logical possibility of his zombie twin (hereafter, ZT). ZT is physically identical to Chalmers from the very lowest levels of physical reality (e.g., quanta) to the higher levels of physical reality (e.g., the neuronal assemblies associated with the performance of cognitive functions). Psychologically speaking, there is no functional difference between ZT and Chalmers. Both can be awake, discriminate between information states, focus attention, categorize, integrate information, and report the contents of internal states to others. The fundamental difference between ZT and Chalmers lies wholly at the experiential level. ZT has no subjective aspect of consciousness, no experiences; all is dark within. Although ZT is physically and functionally identical to Chalmers, there ‘is nothing it is like to be a zombie’ for ZT (Chalmers, 1996).

The zombie-situation just described is conceivable. Hence anyone who claims that ZT is impossible would have to provide a logical counterexample. Zombies, however, are no more logically impossible than mile-high unicycles. None of these terms entails a contradiction. If conceptual analysis cannot be offered to show that a contradiction lurks underneath the terms in question, then ZT is a logical possibility (see Chalmers, 1996).

3. Inverse zombies

Although the possibility of p-zombies may be logically coherent, their existence is naturally improbable. As such, they are not appealing constructs to those adopting an empirical approach to the problem of subjectivity. A related concept that is grounded in the probable—rather than merely the possible—would be an advance for the investigation of consciousness. By inverting the properties of p-zombies, we may develop a more fruitful philosophical and neuroscientific approach to the problem.

What would an inverse (i-) zombie look like? Since the p-zombie is a creature that behaves and responds as if it were conscious when in fact it is unconscious, we posit an i-zombie to be a creature that appears to be unconscious when in fact it is conscious. Any query of a p-zombie elicits a response to indicate consciousness; thus, any query of an i-zombie should thus elicit a response (or lack thereof) to indicate the absence of consciousness. Characteristics of the unconscious appearance of an i-zombie could be unresponsiveness to verbal commands, absence of spontaneous or evoked vocalization or speech, absence of spontaneous or evoked movement, and unresponsiveness to noxious stimulus. Like the p-zombie, the concept of the i-zombie entails no logical contradiction and hence can be considered both conceivable and possible. Unlike the p-zombie, however, i-zombies are naturally probable. We argue that a subset of patients experiencing awareness during general anesthesia, or “anesthesia awareness,” may fall into the category of i-zombie.

Having looked at some differences, we might also consider some similarities between p-zombies and i-zombies. It would seem that whatever solution we find for the problem of detecting consciousness in the case of i-zombies would be equally applicable to p-zombies in some important sense. What sense do we have in mind? In i-zombie cases, some type of consciousness detector could be used to confirm or disconfirm the hypothesis that anesthetized (or possibly even comatose) patients are conscious. In p-zombie cases, we could also use some type of consciousness detector to confirm or reject the same hypothesis with respect to infants, humans, animals, or aliens, which behave and function as if they are conscious. A consciousness detector of some sort would have to be able to distinguish between the presence and absence of consciousness in any possible creature and would therefore apply in detecting both p-zombies and i-zombies. Below we explore potential solutions to this consciousness detection problem (see Section 5).

4. Anesthesia awareness and anesthetic depth

Although the terms “awareness” and “explicit recall” are distinct and dissociable cognitive processes, in the clinical practice of anesthesiology “anesthesia awareness” denotes both awareness and subsequent explicit recall of intraoperative events. Anesthesia awareness is a problem receiving increased attention by clinicians, patients, and the general public. A multi-center American study estimated incidence of awareness with explicit recall of approximately 0.13% (Sebel et al., 2004), a rate consistent with large European studies demonstrating awareness in 1-2/1000 cases (Sandin, Enlund, Samuelsson, & Lennmarken, 2000). A proportion of patients experiencing awareness may subsequently develop serious psychological sequelae, including post-traumatic stress disorder (Osterman, Hopper, Heran, Keane, & van der Kolk, 2001).
There are a number of subjective states that are associated with general anesthesia. In a recent study, dreaming has been reported in 22% of patients undergoing elective surgery (Leslie, Skrzypek, Paech, Kurowski, & Whybrow, 2007). Awareness itself can vary from the transient perception of conversations in the operating room to the sensation of being awake, paralyzed, and in pain (Sebel et al., 2004). The condition of anesthesia awareness is truly a clinical “problem of consciousness.” This can also occur in patients with neurologic injury leading to vegetative states or locked-in syndromes (Laureys, Perrin, & Bredart, 2007). How can such undetected consciousness occur? The case of anesthesia awareness requires a perspective of how anesthesiologists assess anesthetic depth.

In the early days of anesthesia, inhalational agents such as ether or chloroform were employed to achieve the four therapeutic goals of general anesthesia: hypnosis (suppressed consciousness), amnesia, analgesia, and immobility. In order to assess adequate dosing, there have been three major historical conceptions of anesthetic depth: (1) anesthetic stages, (2) minimum alveolar concentration (MAC), and (3) techniques based on electroencephalography (EEG). These techniques are focused primarily on determining anesthetic-induced unconsciousness.

Shortly after the discovery of ether anesthesia, various stages and planes of anesthesia were defined based on physiologic endpoints such as respiratory rate, muscle tone, as well as the size and function of the pupils. Although useful clinically, these qualitative stages were particular to ether and were therefore not useful in comparisons of different anesthetics. In the 1960s, Eger and colleagues developed the conception of MAC as a gauge of anesthetic depth (Eger, Saidman, & Brandstätter, 1965). MAC is the minimum alveolar concentration required to prevent movement in response to a noxious stimulus in 50% of the population. It has become clear, however, that there is a neuroanatomical segregation of anesthetic effects, with immobility mediated primarily in the spinal cord (Antognini & Schwartz, 1993). Thus, since movement is the endpoint of MAC, it is primarily a reflection of spinal cord function rather than the brain. As such, it does not tell us anything directly about the state of consciousness.

These shortcomings led to the development of EEG techniques to assess anesthetic depth and detect consciousness. In the 1930s, it was demonstrated that the EEG was sensitive to the effects of anesthetics (Gibbs, Gibbs, & Lennox, 1937). There is not, however, a unique electrical signature that is common to all agents. Furthermore, the apparatus is bulky, labor intensive, and requires a dedicated observer in the operating room. Due to these limitations, processed EEG modules that often rely on Fourier transformation have been developed. Such “awareness monitors” include the Bispectral Index, Narcotrend, Patient State Index, A-line, and others (Mashour, 2006). In general, these modules collect raw EEG and/or electromyographic data, subject them to Fourier transform, and then analyze parameters that are thought to best represent a state of hypnosis. The output is often a dimensionless number, usually on a scale of 100 (wide awake) to 0 (isoelectric EEG). One such monitor has been shown to reduce the incidence of awareness in a high-risk population (Myles, Leslie, McNeil, Forbes, & Chan, 2004), although the results of this study have recently come into question (Avidan et al., 2008). It is important to note that although “awareness monitors” may hold promise as measures of hypnosis, they do not address the other cardinal features of a general anesthetic, viz., amnesia, analgesia, and immobility.

Such EEG-based monitors, although promising, also have limitations (Dahaba, 2005). Many of these modules are insensitive to well-known anesthetics such as nitrous oxide, ketamine, and xenon. These agents may be pharmacologically similar in their effect on the N-methyl-D-aspartate glutamate receptor. Conversely, EEG monitors can be sensitive to agents that do not suppress consciousness, such as B-adrenergic blockers or neuromuscular blockers. There are other ways by which such “awareness monitors” can be confounded, such as individuals who have a congenitally low-voltage EEG, as well as patients who are hypothermic or hypoglycemic. Finally, such monitors are subject to artifact from other electrical equipment in the operating room.

The current limitations of assessing anesthetic depth entail that we have no completely reliable way to ensure the absence of consciousness in a patient undergoing anesthesia and thus there is a class of individuals who may appear completely unconscious and yet who are nonetheless conscious. Furthermore, despite advances in demonstrating intentionality in patients with persistent vegetative states (Owen et al., 2006), neuroimaging techniques are not practical or even possible for real-time intraoperative monitoring. In short, for all practical purposes, i-zombies are not simply possible or probable—they are known to exist.

5. Philosophical implications of i-zombies

Could i-zombies rule out the feasibility of certain theories of mind? We might consider behaviorism as such a theory. B.F. Skinner, for example, held that mind is reducible to behavior: “we may take feeling to be simply responding to stimuli” (Skinner, 2000). Standard philosophical criticisms of behaviorism are built around conceptual considerations alone and sometimes appeal to intuitions that behaviorists would find question-begging. By contrast, the existence of an i-zombie implies a compelling, empirically based counterexample to behaviorism. An i-zombie is not only real, but has feelings without the possibility of behaviorally responding to stimuli. Therefore, feeling is not simply responding to stimuli. As we observed, eliminating the possibility of responding to stimuli through the administration of certain anesthetic agents does not always guarantee the elimination of painful feelings. During some intraoperative events anesthetized patients have painful feelings associated with the sensation of being cut, but cannot manifest characteristic behaviors that often correlate with such painful feelings. When this occurs we have an instance of an i-zombie. I-zombies, therefore, demonstrate that feeling is not reducible to behavior.
A plausible alternative to behaviorism is functionalism. Functionalism arose on the philosophical scene in response to the shortcomings of behaviorism and type-type identity theory. Functionalism holds that mental states are interdefined in terms of causal relations: the defining characteristic of any mental state \( P \) is the set of causal relations that \( P \) has with respect to inputs, internal mental processes, and behavioral outputs (Fodor, 2000; see also Churchland, 1996). Instead of characterizing the mind simply in behavioral terms, functionalists argue for the causal efficacy of mental states. For example, my belief that a tidal wave is about to form is caused in me by my perception of wave patterns characteristic of tidal waves; and in relation to my desire to preserve my life, the fear of a potential tidal wave will cause me to seek shelter. In contrast to type-type identity theory, functionalists do not hold that mental states can be identified exclusively with a single type of matter (e.g., the neural stuff that composes our brains), but instead maintain that mental states can be realized in any suitably organized system.

An explanatory advantage of functionalism is that it affirms the mental as the source of behavior causation by insisting that mind is defined in terms of function, or by what it does—an interdefined web of causal relations between inputs, inner processes and outputs. An explanatory weakness, however, is that by defining mind in terms of causal relations, functionalism is logically compatible with the absence of experience itself (Armstrong, 2000, p. 142; see also Chalmers, 1996; Churchland, 1996; LaRock, 2007). But does this explanatory weakness imply that the functionalist framework could not be utilized to address the problem of i-zombies? Not necessarily. Our primary concern is practical in nature, for it fundamentally involves detecting consciousness in i-zombie cases. Within this context, we think a plausible functionalist approach would be informed by findings in anesthesiology that could actually begin to address the problem of detecting consciousness in i-zombie cases. In order to motivate functionalism within this practical context, we need to answer a basic question, such as: Where is consciousness caused in the brain?

Answering the “where” question of consciousness in functionalist terms returns us to our discussion of anesthetic depth. In order to localize the neurophysiologic endpoints of anesthesia such as loss of consciousness, we should not use structural space but rather functional or phase space. Phase space, fractal geometry, and strange attractors are now being employed to characterize states of consciousness and anesthesia. In the late 1980s, Watt & Hameroff (1988) demonstrated that phase space analysis of EEG reveals distinct attractors and dimensions for the waking state, anesthesia, and burst suppression. More recent work from van den Broek, van Rijn, van Egmond, Coenen, & Booij, (2006) confirms fractal dimensionality as a measure of anesthetic depth. Walling and Hicks (Walling & Hicks, 2006) found that emergence from sevoflurane was associated with a transition from an ordered attractor of unconsciousness to a chaotic strange attractor of wakefulness. Data from Lee, Kim, Noh, and Choi (2007) using phase space reconstruction of multi-channel EEG demonstrate that the information integration capacity of conscious subjects diminishes after induction of anesthesia. Furthermore, they demonstrate statistically significant differences in fractal dimensionality of the conscious and unconscious states.

Taken together, one answer to the “where” question of consciousness and anesthesia is “phase space.” This form of explanation is consistent with functionalism as it does not attach itself to a specific neural process or location, but rather considers the overall dynamic or “functional” properties of the system. Furthermore, because it can be applied to EEG analysis of the anesthetized patient, it also holds promise in the detection of i-zombies in the clinical realm. Further work on phase space EEG technologies is warranted.

6. The hard problem of unconsciousness

As discussed above, the hard problem of consciousness relates not to the explanation of cognitive processing per se, but rather subjective experience or qualia. This hard problem is more clearly manifested by the logical possibility of p-zombies. I-zombies, on the other hand, bring into relief another kind of “hard problem:” how do we detect the presence of qualia? It should be clear immediately that the hard problem of unconsciousness is fundamentally practical or clinical. The fact that there is no uniformly reliable method to identify or predict intraoperative awareness leaves us with a situation in which consciousness is truly a problem. Assuming 30,000,000 general anesthetics delivered every year in the U.S. alone, with an incidence of anesthesia awareness of approximately 0.15%, we are left with 45,000 patients each year who have not had the adequate suppression of qualia. If we include patients who dream during general anesthesia, the number of potential i-zombies increases dramatically.

This problem is not limited to the operating room; it is becoming clear that patients who carry a clinical diagnosis of persistent vegetative state are capable of “responding” (as assessed by functional imaging) in a way that indicates both comprehension and conscious intentionality (Owen et al., 2006). This hard problem of unconsciousness—detecting the presence of qualia—is again relevant. Decisions of continued life support, as in the highly publicized case of Terry Schiavo, are often made on the assumption of an absence of qualia. A way of accurately “ruling out” the presence of qualia is of paramount importance in making such determinations. This is also true in the case of arguments for “higher brain death”
criteria, where the absence of consciousness rather than the absence of all brain activity becomes the benchmark for the legal determination of “death.”

The foregoing examples highlight the ethical dimension of the hard problem of unconsciousness. The demonstrated natural possibility of i-zombies has implications for our treatment of individuals presumed unconscious. How should clinicians behave in the operating room given the demonstrated incidence of 1–2 individuals/1000 that may still experience qualia during a surgery? Should we comport ourselves acknowledging that the patient has the capacity for suffering? Should we at least ensure that if qualia cannot be extinguished that suffering is minimized with adequate analgesia? Should we restrict our speech to that which is respectful to all patients, conscious or “unconscious”? These ethical implications seem to readily fall out of the possibility of i-zombies: that a proportion of patients can actually be experiencing intraoperative events dictates that we should not assume the Cartesian res cogitans has necessarily been sifted out of the res extensa by general anesthesia. Indeed, part of the goal of general anesthesia is the elimination of subjectivity such that the body can be readily manipulated as a surgical object. The mere possibility that such subjectivity may persist mandates respectful behavior and the withholding of statements that would not be made if the patient were unequivocally known to be awake.

A more controversial ethical question relates to life support for patients with a diagnosis of persistent vegetative state. Given recent data suggesting that these patients may somehow covertly experience undetected qualia, what are the implications? Do we need to further consider the possibility that patients with even more dire diagnoses such as coma or brain death could potentially be i-zombies? The ethical exploration of this question is beyond the scope of this essay and would have important implications for end-of-life decision processes in critical care medicine, as well as organ donation.

It should be noted, however, that there will likely need to be a reconsideration of “responsiveness”, given that patients in persistent vegetative states can be demonstrated to activate brain areas in response to questions assessing their capacity for intentionality. Furthermore, there can still be brain activation during general anesthesia. For example, primary and feed-forward visual processing persists during general anesthesia, while higher order processing is interrupted (Imas, Ropella, Ward, et al., 2005a, 2005b). A study of auditory processing under propofol anesthesia has reached a similar conclusion (Plourde, Belin, Chartrand, et al., 2006). These findings further emphasize the need to assess which brain states are associated with qualia. Mere activation or arousal of the brain does not necessitate consciousness and may still be a feature of an unconscious being. Indeed, this question touches not simply on the detection but on the very definition of i-zombies.

The hard problem of unconsciousness as evidenced by i-zombies may be more practical and less epistemological than the hard problem of consciousness. The former is primarily a problem of detecting consciousness, while the latter is primarily a problem of explaining consciousness. The two problems are nonetheless related. The search for a method to adequately and reliably detect the presence or absence of qualia in real time would be an important step to at least establishing the neural correlates of consciousness. Furthermore, does the fact that an individual can be conscious without any objective indices (i.e., an i-zombie) have the same consequences for physicalism that a p-zombie does? What is a more compelling argument against physicalism, the artificial yet theoretically pure construct of the p-zombie or the practical reality of the i-zombie? Although it obviously can be argued that the inability to detect qualia in a uniformly reliable way is merely a limitation of current technology, it is nonetheless provocative that there is still no method that can unequivocally distinguish consciousness and unconsciousness in a patient who is paralyzed.

7. Conclusion

In summary, we define i-zombies as a class of individuals that appear unconscious but are in fact conscious. We demonstrate that i-zombies are not only logically possible, but also naturally possible and that some patients experiencing anesthesia awareness may fall into this category. Like p-zombies, i-zombies are shown to be linked to a “hard problem” of unconsciousness, a solution of which would require the ability to detect the presence or ensure the absence of qualia. Although we have focused considerable effort on the practical side of this problem, it would be a mistake to conclude that there are no theoretical benefits to be gleaned from our analysis. Since our analysis addresses the where question of consciousness, it would seem that it could begin to pass the ‘dream catcher’ test articulated by Revonsuo (2005). The anesthesiologist is in a unique position to discover ways to pass this test on grounds that i-zombies present an objective means to determine the presence or absence of consciousness. Because it is often not possible for the anesthesiologist to use subjective reports or behaviors of anesthetized patients to detect whether consciousness is present or absent, any advances made on the phenomenon of i-zombies would also count as making positive steps toward passing the dream catcher test.

Finally, previous attempts to make progress on the hard problem of consciousness have typically focused on conceptual or theoretical analyses alone. We find it interesting that a practical approach to solving the consciousness detection problem in the case of i-zombies might bring us closer to a deeper understanding of the relationship between consciousness and the brain. Thus, the empirical reality of i-zombies might provide a practical window through which we might begin to see how to address the purely conceptually based problem of p-zombies that, for many, has been ruled intractable. More research on the clinical, ethical, and philosophical implications of i-zombies and the hard problem of unconsciousness is warranted.

References
