



Neuroethics and the Problem of Other Minds: Implications of Neuroscience for the Moral Status of Brain-Damaged Patients and Nonhuman Animals

Martha J. Farah

Received: 4 January 2008 / Accepted: 7 January 2008

© Springer Science + Business Media B.V. 2008

Abstract Our ethical obligations to another being depend at least in part on that being's capacity for a mental life. Our usual approach to inferring the mental state of another is to reason by analogy: If another being behaves as I do in a circumstance that engenders a certain mental state in me, I conclude that it has engendered the same mental state in him or her. Unfortunately, as philosophers have long noted, this analogy is fallible because behavior and mental states are only contingently related. If the other person is acting, for example, we could draw the wrong conclusion about his or her mental state. In this article I consider another type of analogy that can be drawn between oneself and another to infer the mental state of the other, substituting brain activity for behavior. According to most current views of the mind-body problem, mental states and brain states are non-contingently related, and hence inferences drawn with the new analogy are not susceptible to the alternative interpretations that plague the behavioral analogy. The implications of this approach are explored in two cases for which behavior is particularly unhelpful as a guide to mental status: severely brain-damaged patients who are incapable of inten-

tional communicative behavior, and nonhuman animals whose behavioral repertoires are different from ours and who lack language.

Keywords Philosophy of mind · Brain imaging · Persistent vegetative state · Animal ethics

The Problem of Other Minds

The “problem of other minds” is a central problem in the philosophy of mind. It refers to the difficulty of knowing whether someone or something, other than oneself, has a mind. This general statement of the problem covers a variety of more specific problems, which can be distinguished from one another by the level of skepticism we adopt. In the most skeptical version the problem concerns the difficulty of establishing that there are such things as other minds at all. One could call this the metaphysical problem of other minds.

If we make the assumption that there are minds other than our own in the world, then we encounter the difficulty of determining which entities have minds, and what those minds are like. This is the version we encounter when we ask “How do I know my philosophy professor is not a robot or a mindless zombie?” A less skeptical version of this question grants a mental life resembling one's own to other humans, but notes the difficulty of determining the nature of mental life in other species [32]. These are

M. J. Farah (✉)
Center for Cognitive Neuroscience,
University of Pennsylvania,
3720 Walnut Street,
Philadelphia, PA 19104, USA
e-mail: mfarah@psych.upenn.edu

variants of what could be called the epistemological problem of other minds, in that they concern the difficulty of inferring the existence or nature of a mind from observable evidence. The present article concerns the epistemological problem of other minds. Of course, even the epistemological problem depends heavily on metaphysical assumptions about mind–body relations; the relevance of physical evidence to inferences about mental phenomena depends on one’s view of the relations between the physical and the mental, in ways to which I will return later.

What is the relevance of the problem of other minds to neuroethics? Its relevance to ethics rests on the relation between moral standing and capacity for mental life, particularly the capacity to suffer. If a being is capable of suffering, then it deserves protection from suffering. How and whether we can know about the mental lives of others is therefore an epistemological question with direct relevance to ethics. The relevance of this question to neuroscience rests on the potential value of neuroscience evidence for informing us about a being’s mental life. In this article I will argue that, within the context of a certain class of metaphysical assumptions concerning mind–brain relations, neuroscience evidence is different from the kinds of evidence traditionally used to infer mental life, and that it is in principle more informative.

From Behavior to Mental States: The Argument from Analogy

The problem of other minds is a consequence of mind–body dualism, specifically the idea that there is no necessary relation between physical bodies and their behavior, on the one hand, and mental processes, on the other. Descartes’ famous “I think therefore I am” expresses a basis for certainty concerning the existence of our own mental life. But on what basis can we infer that other people have minds? Descartes invoked the benevolence of God as a reason to trust our inferences regarding other minds. Why would God have given us such a clear and distinct apprehension of other minds if they did not exist [13]?

Non-theological attempts to justify our belief that other people have minds have generally rested on a kind of analogy, also discussed by Descartes and emphasized by Locke [26] and other British empiricists such as J.S. Mill [29]. The analogy uses the

known relation between physical and mental events in oneself to infer the mental events that accompany the observable physical events for someone else. For example, as shown in Fig. 1, when I stub my toe, this causes me to feel pain, which in turn causes me to say “ouch!” When I see Joe stub his toe and say “ouch,” I infer by analogy that he feels pain.

The problem with this analogy is that it begs the question. Why should I assume that same behavioral–mental relations that hold in my case also hold in Joe’s? Joe could be acting and not really feel pain. He could even be a robot without thoughts or sensations at all. The assumption that analogous behavior–mental state relations hold for other people is essentially what the analogy is supposed to help us infer.

The question of whether someone is actually in pain or is acting or a robot might seem academic. After all, common sense tells us that there are no robots human-like enough to fool us, and barring very special circumstances there is little reason to suspect anyone of acting. However, there are two cases in which the problem is not purely academic. That is, even with a commonsensical suspension of skepticism concerning robots, actors and the like, in these cases behavioral inferences to mental state seem fraught with uncertainty. In this regard, these cases present us with pragmatic, real-world versions of the problem of other minds. The first such case is humans who have sustained severe brain damage and emerge from coma into a state of wakefulness with little or no behavioral responsiveness to their environment and therefore no ability to communicate behaviorally. The second such case is nonhuman animals whose behavioral repertoires are different from ours and who lack language.

Cognition and Behavior after Severe Brain Damage

Emergence from coma following severe brain damage typically conforms to a pattern whereby arousal

You stub your toe → You feel pain → You say “ouch!”

Joe stubs his toe → *Joe feels pain* → Joe says “ouch!”

Fig. 1 Example of the use of analogy to infer another individual’s mental state (italicized) from observable evidence and one’s own mental state

systems begin to recover first, leading to periods of eyes-open wakefulness, followed in some cases by recovery or awareness, which may be partial and fluctuating or complete [35]. The sequence of possible states through which patients may pass after severe brain damage is illustrated in Fig. 2. The characteristics of these states are summarized in Table 1.

Patients who are arousable but apparently noncognitive are termed “vegetative.” They show a striking dissociation between behaviors indicating arousal and awareness. In addition to opening their eyes, vegetative patients may move their trunks and limbs spontaneously, and have been observed to smile, shed tears, and vocalize with grunts. They may even orient their eyes and heads toward peripheral visual motion or sounds. Yet they do not respond to language or, with the exception of the reflexive orienting responses just mentioned, to other aspects of the environment.

In such cases the problem of other minds weighs heavily on friends and families, who wonder if their loved one is merely incapable of communicating or is truly gone. The bitter dispute that divided Terri Schiavo’s family exemplified this uncertainty. Her parents saw, in their daughter’s behaviors, the presence of a mind that recognized loved ones, enjoyed music, and wanted to live. They attempted to support their view with videotapes of Terri’s shifting eye gaze, facial expressions and other simple but potentially telling behaviors. Her husband, and most of the medical professionals consulted about the case, interpreted the behaviors as the reflexive actions of an unaware being.

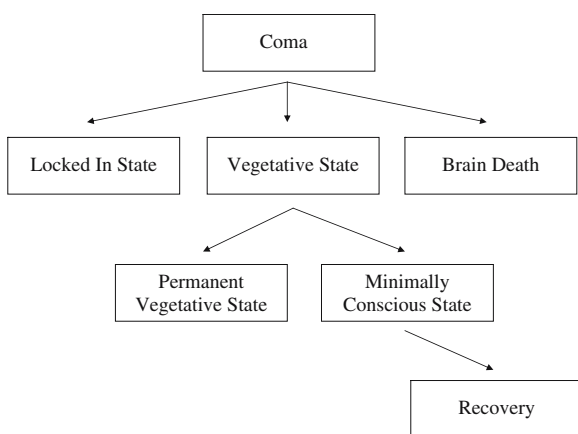


Fig. 2 Sequence of possible states through which patients may pass after severe brain damage

Table 1 Characteristics of possible states through which patients may pass after severe brain damage

Diagnosis	Arousal	Cognition
Coma	Absent	Absent
Vegetative	Present	Absent
Minimally Conscious	Present	Minimal and fluctuating
Locked In	Present	Present

The possibility of mental life in such patients has many societal implications beyond the question of continuing or withdrawing life support, which divided Schiavo’s family. For example, institutions typically provide purely custodial care, in which patients’ bodily functions are sustained without any attention to their experience. To treat a conscious human being as an insensate object for years on end would clearly be inhumane. Additionally, the omission of analgesia for pain, which may be common with vegetative patients [38], would be unconscionable.

The criterial role of mental life for the diagnosis of the vegetative state, as well as the ambiguity of behavioral evidence for inferring mental life, is expressed in the most recent definitive statement on this condition, by the Multisociety Task Force on Persistent Vegetative State [31]. They define it as

“a clinical condition of complete unawareness of the self and the environment, accompanied by sleep–wake cycles, with either complete or partial preservation of hypothalamic and brain-stem autonomic functions (p. 1499)... By definition, patients in a persistent vegetative state are unaware of themselves of their environment. They are noncognitive, nonsentient, and incapable of conscious experience” (p. 1501).

They go on to acknowledge that

“a false positive diagnosis of a persistent vegetative state could occur if it was concluded that a person lacked awareness when, in fact, he or she was aware. Such an error might occur if a patient in a locked-in state (i.e. conscious yet unable to communicate because of severe paralysis) was wrongly judged to be unaware. Thus it is theoretically possible that a patient who appears to be in a persistent vegetative state retains awareness but shows no evidence of it... In the practice of neurology, this possibility is

sufficiently rare that it does not interfere with a clinical diagnosis carefully established by experts” (p. 1501).

Without minimizing the wisdom or sincerity of the authors of these words, it must be pointed out that the problem they note here is a problem precisely because we have no agreed upon way of determining whether such false positives are rare.

The diagnostic category of “minimally conscious state” (MCS) was introduced in 1992 for patients who display a limited and possibly intermittent form of responsiveness or communication [15]. Indicative behaviors include the ability to follow simple commands (e.g. “blink your eyes”), to respond to yes/no questions verbally or by gesture, any form of intelligible verbalization, or any purposeful behaviors that are contingent upon and relevant to the external environment.

The differential diagnosis of PVS and MCS is acknowledged to be difficult, particularly given the fluctuating nature of cognition in MCS. Even after several examinations, the likelihood of having missed a patient’s intermittent and unpredictable periods of sentience may be substantial. It is therefore not surprising that a review of patients from one hospital found almost half of the diagnoses of PVS were wrong because the patients did manifest behavioral evidence of cognitive ability consistent with MCS [3]. Of course, these false positive PVS diagnoses were not the result of the independence of mental states and behavior, such that awareness can exist even without indicative behavior. Rather, they were simply the result of insufficient sampling of patients’ behavior.

In contrast, there would appear to be patients for whom cognition and behavior are truly dissociated. It seems likely that, as patients evolve out of vegetative states toward minimally conscious states, even if they do not cross the boundary by manifesting behavioral indications of cognition, some will have periods of mental activity (see [33], discussed later). There is an analogous lag between the acquisition of cognitive abilities and the ability to manifest them behaviorally in infant development. This is attributable to the greater demands placed on the quality of internal representations used to drive external behaviors compared to the demands of purely internal processing [30].

In addition, there is a separate diagnostic category of neurological patients for whom cognition and

behavior are truly dissociated. These individuals continue to experience full awareness of themselves and their surroundings while being unable to indicate their awareness behaviorally. Patients in this condition are said to be “locked in,” a depressingly apt phrase that describes near complete or complete paralysis, the result of interruption of outgoing (efferent) motor connections, most often by stroke. Patients typically emerge from coma to find themselves treated as vegetative, and may try for months or even years to signal their awareness to medical staff and family members [22]. In its most classic form, a degree of preserved voluntary eye movement allows communication, for example answering questions with an upward gaze for “yes” or spelling words by selecting one letter at a time with eye movements. For other patients, the de-efferentation is more complete and no voluntary behavior is possible [5].

Brain Activity in Severely Brain-Damaged Patients

For neurologists examining patients, as for philosophers pondering the problem of other minds, behavior is the most obvious and natural type of evidence to consult for the purpose of inferring the mental life of another being. Yet behavior is clearly inadequate in principle, because it is only contingently related to cognition. Furthermore, certain kinds of brain damage are known to be capable of changing the cognition–behavior relation.

In recent years a different type of evidence has been brought to bear on the study of mental processes, namely functional neuroimaging [36]. These methods have enabled neuroscientists to test hypotheses about cognition in normal [2] and brain-damaged subjects [37]. One of the most exciting applications of this approach has been the assessment of brain activity and brain responses in behaviorally nonresponsive and minimally responsive patients.

Laureys, Owen and Schiff [19] reviewed the literature on brain function after severe brain damage. They report that studies of global brain metabolism at rest show that the brains of locked in patients are almost as active as those of healthy and awake individuals, while the activity measured in comatose, vegetative and minimally conscious patients’ brains is more like that of a sleeping or anesthetized person. Of course, global brain activity is less informative about

mental processes than is activity in specific brain regions associated with cognition, and activity at rest is less informative than activity measured in response to specific meaningful stimuli. Fortunately there is a growing body of literature on these more specific mind-brain correlations.

The specific brain areas associated with awareness of self and environment include the prefrontal and medial parietal cortices. This association is based on measurements of activity in these areas in the normal conscious state and across a variety of states in which conscious awareness is diminished, including general anesthesia, sleep and absence seizures. When resting medial parietal activity is compared across the diagnostic categories discussed here, it is highest for normal control subjects, next highest in locked in patients, lower in minimally conscious patients, and lowest in vegetative patients.

Brain responses to meaningful stimulation can be surprisingly preserved in minimally conscious patients. For example, in a well-known study by Schiff and colleagues [40], subjects underwent fMRI scans while being presented with recordings of a relative telling a personally relevant story and with the same recording played backwards, to approximately control for the auditory characteristics of the stimulus input while varying its meaningfulness. Like the normal control subjects, the subjects in MCS activated a network of language-related areas in response to the meaningful recordings relative to the backwards recordings. The authors concluded that “some MCS patients may retain widely distributed cortical systems with potential for cognitive and sensory function despite their inability to follow simple instructions or communicate reliably” (p. 514).

In contrast, imaging studies of vegetative patients have yielded little evidence of the kinds of neural processing associated with mental life. In the largest study of its kind, 15 carefully evaluated patients meeting criteria for persistent vegetative state were subjected to painful stimulation while being scanned, and like normal subjects showed activation of midbrain, thalamic, and primary sensory cortical areas. Unlike normal subjects, however, higher cortical areas normally involved in responding to painful stimuli, such as the anterior cingulate cortex (ACC) were not activated [21].

Single case studies of vegetative patients have occasionally shown preserved brain responses to

meaningful stimuli, although few of the cases were unambiguously vegetative at the time of imaging. For example, a patient whose face recognition system responded to photographs of faces [28] was described as either “upper boundary vegetative state or lower boundary minimally conscious state” [19]. The most striking finding to date in this literature comes from a study by Owen and colleagues [33] of a vegetative patient who later recovered, but while meeting diagnostic criteria for the vegetative state showed patterns of brain activation indicative of language comprehension and voluntary mental imagery.

One indication of preserved cognition in this patient was her increased brain activity when presented with sentences containing ambiguous words, in the same region as for normal subjects, consistent with the additional cognitive processing required for resolving the ambiguity of such sentences. In addition, when instructed to perform mental imagery tasks, her brain activity indicated that she understood the instructions and was able to comply. When asked to imagine playing tennis she activated parts of the motor system and when asked to imagine visiting each of the rooms of her home she activated parts of the brain’s spatial navigation system. Furthermore, her patterns of brain activation were indistinguishable from normal subjects’.

In sum, functional neuroimaging provides a new window on the mental status of severely brain-damaged patients. Although there are still relatively few imaging studies on well-characterized patients, it is clear that at least some patients with little or no capacity for purposeful behavior nevertheless show patterns of brain activation consistent with cognition.

Behavior and Brain Activity as Evidence of Mental Life

Why is brain imaging able to provide evidence of mental life when behavior cannot? Is brain activity simply a more sensitive measure of cognitive processing than behavior, but otherwise plays the same role in inferences regarding mental life? Or is it qualitatively different from behavior? Consider the possibility that brain activity and behavior play analogous roles. Figure 3 illustrates this possibility, by replacing the “ouch” behavior of Fig. 1 with

You stub your toe → You feel pain → Your ACC activates

Joe stubs his toe → *Joe feels pain* → Joe's ACC activates

Fig. 3 Example of the use of brain activity to fill the role played by behavior in Fig. 1

activation of the anterior cingulate cortex (ACC), part of the brain's pain network.

The problem with the argument from analogy in Fig. 3 is that it implies that feeling pain causes ACC activation, just as it causes saying "ouch." However, the relations between mental states and behaviors are different in kind from the relations between mental states and brain states. Mental states and behaviors are contingently related. What one means by a term like "feeling pain" is not a behavior, or even a behavioral disposition. Although this possibility was explored in earnest by some behaviorist philosophers several decades ago, for example by Ryle [39], it is not now regarded as a viable approach to the meaning of mental state terms.

For purposes of knowing mental states, behavior is like an indicator light. Indicator lights can be disabled or disconnected, or they can be turned on by other means. Their relation to the thing indicated is contingent on being hooked up a certain way. Inferences based on indicator lights and, analogously, behavior are therefore fallible. In contrast, virtually all contemporary approaches to the mind–body problem regard the relation between mental states and brain states as noncontingent.

The predominant view of the relation between mental states and brain states in cognitive neuroscience and contemporary philosophy of mind is one of identity: mental states *are* brain states. According to one version of this view, "type identity," each type of mental event is a type of physical event [8, 45]. According to a weaker version, "token identity," every instance of a mental event is an instance of a physical event. The most widely accepted version of token identity is based on "functionalism," which identifies the functional role of a physical state, in mediating between the inputs and outputs of the organism, as the determinant of its corresponding mental state [7]. Functionalism has many versions of its own, some of which blur the line between type and token mind–brain identity theory (e.g., [4, 24]).

There is an alternative to mind–brain identity based on the idea that mental states "supervene" on brain states, which avoids substance dualism yet stops short

of equating mental states with brain states [9, 17]. Figure 3 is incompatible with supervenience theories as well as identity theories. This is because, despite the nonidentity of mental and brain states according to supervenience theories, the relation between with the two is stronger than mere causality. According to supervenience, mind–brain relations are noncontingent. In the words of Davidson, "there cannot be two events alike in all physical respects but differing in some mental respects [and] an object cannot alter in some mental respects without altering in some physical respects" (p 214).

In sum, across all these different contemporary metaphysical positions on the mind–body problem, the relationship between mental states and brain states is not contingent, as with the causal relations diagrammed in Fig. 1. For type identity theories as well as functionalist theories, the ACC activation of the example is identical to a pain. For supervenience theories, the ACC activation cannot exist without there being pain. Thus it makes more sense to diagram the inferences from brain activity to mental state as in Fig. 4. The gist of this figure is that, however sure you are of the ACC activation in Joe's brain, you can be that sure that Joe is in pain. In sum, the argument from analogy with brain activity is immune to the alternative interpretations that plague the behavioral analogy.

The Problem of Other, Nonhuman, Minds

Like the severely brain-damaged humans just discussed, nonhuman animals have limited communicative abilities, and this limitation deprives us of the usual methods for learning about their mental states [1, 11]. Although few people today would agree with Descartes' conclusion that animals lack mental states altogether, most of us feel uncertain about the extent and nature of animals' mental lives. On the one hand,

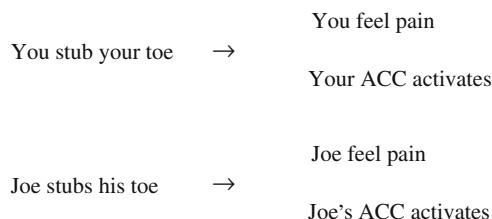


Fig. 4 Example of the use of brain activity to infer mental state

many of us anthropomorphize certain animals, especially our pets, attributing complex thoughts and expectations to them on the basis of what a human in the same situation might think. On the other hand, the mental life of animals is often treated by us as hypothetical, incomparably different from our own, or even nonexistent. How else to explain our acceptance of glue traps for rodents and boiled lobster dinners?

As illustrated in Fig. 5, nonhuman animals present us with a version of the problem of other minds for which the usual problematic analogy is even more problematic because of differences between human and animal behavioral repertoires. Animals cannot talk, and may not even express distress in nonverbal ways that are analogous to ours. For example, they may not vocalize at all, and may freeze rather than struggle when afraid.

Can the neuroscience approach provide traction for exploring the mental life of other species? To a degree it already has, yet according to the present analysis it could provide even more. Ethicists have previously brought physiological data to bear on the question of animal suffering, specifically the similarities between human and animal pain systems. For example, Singer ([42] pp. 12–13) quotes at length from the writings of a pain researcher to the effect that pain processing is a lower level brain function which differs little between humans and other animals. This use of physiological data differs in two ways from the present one.

First, according to the present analysis, physiological data are not simply one more source of evidence about a being's mental life, to be weighed together with behavioral evidence, valuable as they might be in that role. Rather, physiological data can play a qualitatively different and more definitive role because of their noncontingent relation to mental states, as argued in the last section. In terms of the inferences diagrammed earlier, this is the difference between Figs. 3 and 4.

The second difference results from the relatively new ability of cognitive neuroscience to parse brain processes into psychologically and ethically meaningful

You stub your toe → You feel pain → You say "ouch!"
 Bat stubs his toe → *Bat feels pain* → Bat squeaks

Fig. 5 An illustration that nonhuman animals present a version of the problem of other minds

categories. In the present case, it has revealed the neural basis of the distinction between what could be called "mere pain" and suffering. Pains vary along many dimensions, and one dimension of particular ethical relevance is the psychological quality of the pain [10, 12, 16]. Some pain experiences are primarily physical while others are psychologically distressing. The latter, characterized by Dawkins as both unpleasant and intense, warrant the term "suffering." The neural states corresponding to pain states appear to respect this important distinction, demarcating the physical and psychological components of pain experience by the involvement of different brain areas.

Research with animals and humans has revealed a widespread network of brain areas that become active in response to pain-inducing stimuli, including thalamic and somatosensory cortical regions as well as regions further removed from the sensory input such as the insula and anterior cingulate cortex. When the physical intensity of pain is varied, for example in an imaging experiment by having human subjects touch a painfully hot surface that varies in temperature, the level of activity throughout this network varies [6]. Taking advantage of the human ability to report their mental states (and in principle the possibility of first person research in which one introspects on one's own mental states), it is possible to vary independently the physical and psychological dimensions of pain and map the brain states that correspond to each. Morphine, for example, is known to diminish the psychological component of pain. Patients commonly report that they still feel the "physical" pain but that they are less bothered by it. The same is reported by patients whose pain is treated with hypnotic suggestion. Both interventions have their neural effects primarily in the ACC [18, 25, 34]. When people who are not being subjected to pain are empathizing intensely with someone who is, their ACCs become activated in the absence of physical pain [44]. These findings indicate that ACC activation reflects suffering rather than "mere" pain.

Shriver [41] points out that mammals have ACCs and are thus neurally equipped for psychological as well as physical pain. Following Shriver we can substitute an animal for Joe in Fig. 4. However, because brain states can only be as similar as the brains that have them, we must amend the diagram as shown in Fig. 6 to specify a human ACC is one's own case and an animal ACC in the animal's.

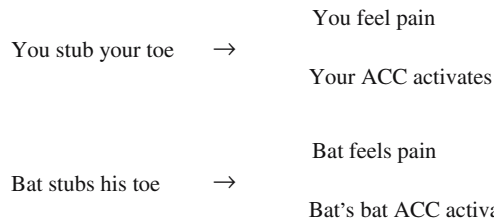


Fig. 6 Example of the use brain activity to infer mental state in a different species

This raises the question of how could one determine whether mind–brain relations established with one species' brain generalize to other species. Behind this question is a more fundamental one about how degrees and types of variation in brain states correspond to degrees and types of variation in mental states, a question that will arise even within a given species because no two brains are identical. In principle, one could manipulate human brains (including one's own) to systematically vary all the different biophysical characteristics by which brains differ, in order to discover what the relevant aspects of the brain state are for determining the mental state. Of course in practice this is not even remotely possible.

At best we can suppose that similarity of psychological state will fall off as similarity of brain state falls off, without knowing which aspects of brain state similarity are relevant or how sharply the one falls off relative to the other. Edelman, Baars and Seth [14] provide an example of the attempt to identify functional similarities in brain architecture across species, including nonmammalian species. Shriver [41] attempts to address the problem of generalization from human to nonhuman in the case of pain by citing evidence that the ACC plays a similar role in rat and human pain experience (although this evidence is admittedly based on behavior, which the present appeal to brain evidence was intended to replace): LaGraize and colleagues [20] compared the behavior of rats with and without lesions of the ACC when forced to choose between staying in the dark, which rats generally prefer, and avoiding electric shocks to their feet. All of the rats reacted similarly when shocked, by withdrawing the shocked foot, thus indicating preserved perception of pain. However, the lesioned rats were more willing to experience the shocks for the sake of staying in a dark region of the experimental apparatus. Like patients on morphine, they appeared to be less distressed by the pain. This

implies that rat ACCs play a role similar to human ACCs.

Am I suggesting that neuroscience can tell us what it is like to be a bat? Yes and no. When Nagel [32] framed this question, he chose the bat as his nonhuman animal because bats use echolocation to perceive the world, a sense which humans lack. Knowing what it is like to perceive the world with a sense we lack remains a problem, even with the help of neuroscience, because the neural systems that perform echolocation in bats have no obvious homolog in the human brain. However, given that we do share the same general pain physiology with bats, including an ACC, we can know certain things about what it's like to be one of them. Specifically, we can infer that to be a bat with an injured toe is more like being a human with an injured toe and no pain relief than it is like being a human with an injured toe who has been given morphine.

The problem of animal minds has not thus far figured prominently in the field of neuroethics. One reason may be that neuroethics is young and has yet to engage all of the subject matter that will eventually comprise the field. Another reason may be that the personal and political rancor associated with animal ethics has discouraged scholars from approaching this topic. Given the real-world importance of animal ethics, and the special role that neuroscience evidence can play in this endeavor, the study of animal neuroethics would seem to hold great promise.

Assumptions and Conclusions

The idea that neuroscience can reveal ethically relevant information about severely brain-damaged patients and nonhuman animals rests on a number of assumptions. One assumption that has not been examined in this article is that our ethical obligations toward a being depend at least in part on the mental life of that being. Although this assumption hardly needs defending, there is much more that could be said about which specific aspects of mental life have which specific ethical implications. Perhaps the most important further clarification concerns those aspects of mental life that obligate us to prevent suffering, and those that obligate us to protect life.

The present article has focused on the question of whether another being has the capacity for relatively

simple mental states, those with some consciously experienced content and affective valence. This mental capacity has more limited ethical implications than the mental capacity to conceive of oneself and one's life and have an explicit preference to continue living [23, 43]. The neuroscience evidence discussed so far pertains only to the capacity of patients and animals to experience the former kind of mental state, and the relevant ethical implications are therefore confined to preventing suffering rather than protecting life. However, this is not an in-principle limitation of neuroscience data. Given the appropriate research program, there is no reason why we could not identify the neural systems, and states thereof, corresponding to the self-concept and the desire to continue living. This knowledge would have implications for many aspects of end-of-life decision making, and might even obligate us to refrain from killing certain animals.

Another assumption that deserves explicit discussion concerns the relation between cognitive processing, of the kind that cognitive neuroscientists correlate with brain activation in imaging experiments, and consciousness. This is an important assumption in the present context because our ethical concern is with conscious mental life, and conscious suffering in particular, rather than with unconscious information processing. Based on most of the views of mind–brain relations reviewed earlier, certain types or instances of neural processing are identical to, or are necessarily associated with, certain mental states, including conscious mental states. Therefore the problem is one of determining empirically which brain states correspond to which conscious mental states. This is not a trivial problem, but it is in principle solvable. Indeed, if one is willing to accept other normal humans' reports of conscious experience as evidence, we are on our way to solving it in practice. (Skeptics unwilling to accept others' reports of conscious experience would have to be scanned themselves, which could be done to verify specific findings but not be feasible as a means of verifying all cognitive neuroscience knowledge.)

A final assumption concerns the accuracy and completeness of cognitive neuroscience. For purposes of exploring the in-principle prospects and limitations of neuroscience evidence as a solution to the problem of other minds, I have written as if we know the brain states associated with specific mental states.

Unfortunately, this is not true. Although cognitive neuroscience has made tremendous progress in the last few decades, the current state of our knowledge is far from complete. For many mental states, including suffering, we have good working hypotheses about the brain regions that are relevant, but future research will undoubtedly call for the revision of some of these hypotheses. In addition, we know little about the specific mechanisms by which these regions implement the relevant mental states. "Activation" observed in brain imaging studies is closely related to neural activity measured at the single cell level, but does not map perfectly onto a specific aspect of neuronal behavior such as action potentials [27]. Furthermore, any single measure of brain activity, be it single cell or aggregate, electrical or chemical, will omit potentially important features of neuronal function. It is possible that activation as measured by our current methods is not diagnostic of the relevant neuronal activity and that under some circumstances it will be misleading. Knowing more about the specific computations performed by neurons in the brain regions implicated by brain imaging, including their interactions with neurons in other regions, will be particularly important as we attempt to evaluate cross-species homologies.

Acknowledgements I thank Liz Camp, Neil Levy, Adrienne Martin, and Susan Schneider for their helpful comments on an earlier draft of the paper. The writing of this paper was supported by NIH grants R21-DA01586, R01-HD043078, R01-DA14129 and R01-DA18913, and by the MacArthur Project on Neuroscience and the Law.

References

1. Allen, C. 2004. Animal pain. *Nous* 38: 617–643.
2. Aguirre, G.K. 2005. Functional imaging in cognitive neuroscience I: Basic principles. In *Patient-Based Approaches to Cognitive Neuroscience*, eds. M.J. Farah, and T.E. Feinberg 2nd Cambridge: MIT Press.
3. Andrews, K., L. Murphy, C. Munday, and C. Littlewood. 1996. Misdiagnosis of the vegetative state: Retrospective study in a rehabilitation unit. *British Medical Journal* 313: 13–16.
4. Armstrong, D.M. 1968. *A Materialistic Theory of the Mind*. London: RKP.
5. Bauer, G., F. Gerstenbrand, and E. Rimpl. 1979. Varieties of the locked-in syndrome. *Journal of Neurology* 221: 77–91.
6. Becerra, L.R., H.C. Breiter, R. Wise, R.G. Gonzalez, and D. Borsook. 2001. Reward circuitry activation by noxious thermal stimuli. *Neuron* 32: 927–946.

7. Block, N., and J. Fodor. 1972. What psychological states are not. *Philosophical Review* 81: 159–181.
8. Churchland, P.M. 1988. *Matter and Consciousness*. 2nd Cambridge: MIT Press.
9. Davidson, D. 1970. *Mental events. Reprinted in essays on actions and events*. Oxford: Clarendon Press.
10. Dawkins, M.S. 1985. The scientific basis for assessing suffering in animals. In *In Defense of Animals*, ed. P. Singer, 27–40. New York: Basil Blackwell.
11. DeGrazia, D. 1996 *Taking Animals Seriously*. Mental Life and Moral Status. Cambridge University Press.
12. Dennett, D.C. 1996. *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.
13. Descartes, R. 1968. *Discourse on Method and the Meditations* (trans. F. E. Sutcliffe). Penguin.
14. Edelman, D.B., B.J. Baars, and A.K. Seth. 2005. Identifying hallmarks of consciousness in non-mammalian species. *Consciousness and Cognition* 14: 169–187.
15. Giacino, J.T. et al. 2002. The minimally conscious state: definition and diagnostic criteria. *Neurology* 58: 349–353.
16. Hardcastle, V.G. 1999. *The Myth of Pain*. Cambridge, Massachusetts: MIT Press.
17. Kim, J. 2005. *Physicalism, or Something Near Enough*. Princeton, NJ: Princeton University Press.
18. Kupers, R., M.E. Faymonville, and S. Laureys. 2005. The cognitive modulation of pain: hypnosis- and placebo-induced analgesia. *Progress in Brain Research* 150: 251–269.
19. Laureys, S., Owen, A., and Schiff, N. 2004. Brain function in coma, vegetative state, and related disorders. *The Lancet Neurology* 3: 537–546.
20. LaGraize, S.C., C.J. LaBuda, M.A. Rutledge, R.L. Jackson, and P.N. Fuchs. 2004. Differential effect of anterior cingulate cortex lesion on mechanical hyperalgesia and escape/avoidance behavior in an animal model of neuropathic pain. *Experimental Neurology* 188: 139–148.
21. Laureys, S., M.E. Faymonville, P. Peigneux, P. Damas, B. Lambermont, G. Del Fiore, C. Degueldre, J. Aerts, A. Luxen, G. Franck, M. Lamy, G. Moonen, and P. Maquet. 2002. Cortical processing of noxious somatosensory stimuli in the persistent vegetative state. *Neuroimage* 17: 732–741.
22. Laureys, S., F. Pellas, P. Van Eeckhout, S. Ghorbel, C. Schnakers, F. Perrin, J. Berre, M.E. Faymonville, K.H. Pantke, F. Damas, M. Lamy, G. Moonen, and S. Goldman. 2005. The locked-in syndrome: what is it like to be conscious but paralyzed and voiceless? *Progress in Brain Research* 150: 495–511.
23. Levy, N. 2008. *Cognition and Consciousness*. in press.
24. Lewis, David. 1980. Mad pain and Martian pain. In *Readings in the Philosophy of Psychology*, Vol. I., ed. N. Block, 216–222. Harvard University Press.
25. Lidstone, S.C., and A.J. Stoessl. 2007. Understanding the placebo effect: contributions from neuroimaging. *Molecular Imaging & Biology* 9: 176–185.
26. Locke, J. 1959. *Essay Concerning Human Understanding*, ed. A.C. Fraser. Dover.
27. Logothetis, N.K., and B.A. Wandell. 2004. Interpreting the BOLD signal. *Annual Review of Physiology* 66: 735–769.
28. Menon, D.K., A.M. Owen, E.J. Williams et al. 1998. Cortical processing in persistent vegetative state. *Lancet* 352: 200.
29. Mill, J.S. 1979. The collected works of John Stuart Mill, volume IX—An examination of William Hamilton’s philosophy and of the principal philosophical questions discussed in his writings, ed. John M. Robson, Introduction by Alan Ryan (Toronto: University of Toronto Press, London: Routledge and Kegan Paul, 1979). Chapter: CHAPTER XII: The psychological theory of the belief in matter, How far applicable to mind. Accessed from <http://oll.libertyfund.org/title/240/40871> on 2007-12-13
30. Munakata, Y. 2001. Graded representations in behavioral dissociations. *Trends in Cognitive Sciences* 5: 309–315.
31. Multi-Society Task Force on PVS. 1994. Medical aspects of the persistent vegetative state (1). *New England Journal of Medicine* 330(21): 1499–1508.
32. Nagel, T. 1974. What is it like to be a bat? *Philosophical Review* 83: 435–450.
33. Owen, A.M., M.R. Coleman, M. Boly, M.H. Davis, S. Laureys, and J.D. Pickard. 2006. Detecting awareness in the vegetative state. *Science* 313: 1402.
34. Petrovic, P., E. Kalso, K.M. Pettersson, and M. Ingvar. 2002. Placebo and opioid analgesia—imaging a shared neuronal network. *Science* 295: 1737–1740.
35. Plum, F., and J.B. Posner. 1980. *The diagnosis of stupor and coma*. 3rd Philadelphia: FA Davis Co.
36. Posner, M.I., and M. Raichle. 1997. *Images of mind*. New York: WH Freeman.
37. Price, C.J., and K.J. Friston. 2005. Functional imaging in cognitive neuroscience II: Imaging patients. In *Patient-based approaches to cognitive neuroscience*, eds. M.J. Farah, and T.E. Feinberg 2nd Cambridge: MIT Press.
38. Rifkinson-Mann, S. 2003. Legal consequences and ethical dilemmas of pain perception in persistent vegetative states. *Journal of Health Law* 364: 523–548.
39. Ryle, G. 1949. *The concept of mind*. Chicago: University of Chicago Press.
40. Schiff, N.D., D. Rodriguez-Moreno, A. Kamal, K.H.S. Kim, J.T. Giacino, F. Plum, and J. Hirsch. 2005. fMRI reveals large-scale network of activation in minimally conscious patients. *Neurology* 64: 514–523.
41. Shriver, A. 2006. Minding mammals. *Philosophical Psychology* 19: 443–442.
42. Singer, P. 1990. *Animal Liberation*, revised edition. New York: Random House.
43. Singer, P. 1999. *Practical ethics*, 2nd ed. Cambridge University Press.
44. Singer, T., B. Seymour, J. O’Doherty, H. Kaube, R.J. Dolan, and C.D. Frith. 2004. Empathy for pain involves the affective but not sensory components of pain. *Science* 303: 1157–1162.
45. Smart, J.J.C. 1959. Sensations and Brain Processes. *Philosophical Review* 68: 141–156.