

.....

JEFF MCMAHAN

## **Brain Death: Metaphysics, Morality, and Law**

### **1 Against Brain Death**

In the debate that led to the replacement of the cardiopulmonary criterion of death by the criterion of *brain* death, the main arguments focused on the idea that brain death is known, with no room for doubt, to be sufficient for the irreversible loss of the capacity for consciousness. No one can possibly recover *consciousness* after brain death. The report that played a decisive role in the acceptance of brain death as the criterion of death was not called ‘A New Definition of Death’ but ‘A Definition of Irreversible Coma.’ This suggests that many people took the irreversible loss of the capacity for consciousness to be equivalent to death.

In the notorious cases of Karen Ann Quinlan, Nancy Cruzan, and Terry Schiavo—cases of persistent vegetative state in which the patient had lost the capacity for consciousness—many people said these women were no longer there, that they had ceased to be and that all that remained were their bodies. Yet they were not brain dead.

Cases like these have led some people to propose what has been called a ‘higher-brain’ criterion of death, according to which a person is dead when those areas of the brain in which consciousness is realized are dead or irreversibly nonfunctional. But the problem with the higher-brain criterion is that it declares death in cases in which the brain stem remains intact and functional and the organism breathes

spontaneously and successfully carries out all other somatic functions, apart from the generation of consciousness. It is obviously false to say that such an organism is biologically dead.

But brain death faces a similar problem. A human organism can be maintained almost indefinitely in a comprehensively functional state following brain death, with only minimal external support. There have been several cases, for example, in which the body of a brain dead pregnant woman has been kept sufficiently functional to carry the fetus to term, allowing for a live birth.

If brain death is the correct criterion of death, we have to say that these maternal organisms were biologically dead, despite the fact that, with little external prompting except from a respirator, they were able to maintain their immune responses but restrain them in subtle ways in response to the fetus, to redistribute blood flow to increase the supply to the uterus, to provide the fetus with nutrition, and so on. It is hard to believe that an entity capable of these internally and intricately organized functions is *dead*.

Next consider a human embryo three weeks after conception. By that point the embryonic cells have become differentiated and are performing different but coordinated functions. They clearly constitute an organism and that organism is uncontroversially alive. Yet it has no brain, and therefore no brain functions. The notion of brain death implies that a certain level of brain activity is necessary for a human organism to be alive, and so implies that the embryo cannot be alive. Like the cases of the brain dead pregnant women, this suggests that brain death cannot be equivalent to biological death in an organism.

(One might say that *if* an organism has a brain, that brain must be alive for the organism to be alive. But if an organism does not have a brain, then it can be alive without a living brain. A plant, for example, can be alive without a functional brain. But on this view it would be possible for a human organism from which the brain had been removed to be alive but not for an organism with a dead brain to be alive, even if they were otherwise indistinguishable.)

Our conception of death has to make coherent sense of all these problems. I think the best solution emerges if we draw a distinction between ourselves and our organisms. We can then distinguish two concepts of death: the death of a subject of consciousness, such as yourself, and the death of an organism. *Your* death — your ceasing

to exist — will occur when you irreversibly lose the capacity for consciousness. The death of your organism will occur when its various parts and subsystems irreversibly cease to function together in an internally integrated or coordinated way.

This dualist view offers plausible accounts of the various problem cases. In genuine persistent vegetative state, when an individual has irreversibly lost the capacity for consciousness, the conscious subject has died but the organism remains alive. The same is true in cases in which an individual's brain dies but his organism remains systematically functional as a result of external life support.

An embryo beyond two weeks after conception is a living human organism that does not yet support the existence of a conscious subject such as yourself. On this view, *we* begin to exist in association with our organism when the fetal brain acquires the capacity to generate consciousness, somewhere between 22 and 28 weeks after conception.

Suppose, however, that I am wrong in claiming that we are not human organisms. Even so, it is not plausible to identify the death of a human organism with the death of the brain. Brain death is neither necessary nor sufficient for the biological death of a human organism.

That it is not necessary for the death of an organism is shown by the fact that one's brain could be extracted from one's organism and kept alive, while one's organism is allowed to die. (In a series of macabre experiments on dogs and rhesus monkeys, Robert J. White has shown that a mammalian brain can be kept alive and functional following separation from the body.) Some object to this point that the living brain would be the original organism pared down to its core. But I think there is no more reason to believe that a living, isolated brain is an organism than that a living, isolated heart is. The brain is an organ, not an organism.

That brain death is not sufficient for the death of an organism is shown by the cases in which human organisms have been kept alive for extensive periods following brain death. Of course, those who claim that brain death is death have to claim that these organisms merely *seem* to be alive. But I think we should be reductionists about this matter. We know various facts about the state of such an organism — for example, that its respiration is generated externally and that the heart then beats in response to the artificially induced

respiration, setting off a cascade of other complexly interrelated functions, which occur without any further external prompting. I doubt that whether the organism is alive is a further fact. Whether it is alive is a matter of whether the claim that it is alive coheres better with our beliefs about life and death in paradigm cases.

The orthodox view, which I accept, is that life in an organism is a matter of whether the various parts function together in an internally regulated way to maintain homeostasis. No one suggests that people who are conscious but require a respirator cannot be alive because of the external life support. So why should a brain dead organism whose functions are sustained by a respirator not also be alive?

Defenders of brain death try to explain the difference by saying that people on respirators who are conscious are alive because their brains continue to regulate the various functions of their bodies, while the functions that can be maintained in brain dead organisms are externally rather than internally regulated. But there is in fact very little difference in the degree of internal organization in the two cases. In brain dead organisms whose functions are sustained by external ventilation, the various organs and systems function in response to signals they receive from each other. Coordination and integration are *internal* but *decentralized*.

The common claim that the brain is necessary for internal integration, and thus for life in a human organism, is therefore false.

Suppose now that I am right and that we are not organisms but are instead conscious subjects (that is, beings for whom the capacity for consciousness is an essential property). In that case, brain death is *sufficient* for our ceasing to exist but *not necessary*. I would cease to exist if my cerebral hemispheres were completely destroyed while my brain stem remained intact and functional.

It is true that brain death affords *maximal certainty* about the irreversible loss of the capacity for consciousness. The possibility of residual consciousness in the brain stem of a patient in a genuine persistent vegetative state cannot be ruled out with absolute certainty. This is a problem about the limits of our knowledge. If people in what we now refer to as a persistent vegetative state in fact retain the capacity for consciousness, then on my view they remain alive. In that case, brain death might be the right criterion of death *for us*, even if we are not organisms.

But the evidence overwhelmingly suggests that the mind, or the capacity for consciousness, can cease to exist when some functions in the brain stem remain. If that is right, then brain death is not equivalent to the death of a human organism *or* to the death or ceasing to exist of a person, or conscious subject.

Still, if there is significant *uncertainty* in determining when the capacity for consciousness is lost *irreversibly*, then it might make sense to adopt brain death as the criterion of death for *pragmatic* reasons. But we could also reasonably expect that as our knowledge advances, we will be able to replace brain death with a more precise criterion of the irreversible loss of the capacity for consciousness, and thus for when *we* cease to exist.

Having now summarized what I think is the most plausible view about human death, I will offer a few thoughts about the arguments in Professor Dieter Birnbacher's stimulating and challenging paper.

## 2 Comments on Dieter Birnbacher's Paper

I agree with a great deal of what Professor Birnbacher says. He distinguishes three types of account of death: reconstruction, explication, and pragmatic account. He argues that we should seek an explication rather than a reconstruction. Here I completely agree. What most people believe about when our lives begin and end is largely irrelevant. It may be the product of an implausible system of religious beliefs. Or it may be the result of confusion, as in the case of those cited by Birnbacher who believe that human life begins with the *implantation* of the embryo, which cannot be right because life must be a function of intrinsic rather than relational properties.

I also agree with Birnbacher that while we should seek an explication, the best explication may not be translatable directly into policy or law. But we should aim to be open about the ways in which our pragmatic proposal diverges from the best metaphysical account of death.

Birnbacher suggests four desiderata of a satisfactory explication of the concept of death: univocity, exhaustiveness, inclusiveness, and symmetry with our concept of the beginning of life. I will consider each of these in turn.

My observations will suggest that coherence and adequacy are

connected, and that we can satisfy Birnbacher’s desiderata only by sacrificing coherence. If we are to achieve coherence, we will have to reject some of the desiderata.

## 2.1 Univocity

Our concept of death *is not* univocal. When Jesus is quoted as saying that “whosoever liveth and believeth in me shall never die,” he does not mean that the organisms of believers will remain functionally integrated forever. He is not making a biological claim or even asserting anything about organisms. He means that believers won’t ever cease to exist. So there is, in the language, a clear sense in which death — particularly in the case of a human being — just *means* ceasing to exist. But there is also a biological sense in which death is only contingently connected with ceasing to exist. When an organism dies, it usually does not cease to exist but continues to exist as a corpse. And at least some living organisms can cease to exist without dying in this biological sense. This happens, for example, when an ameba divides to form two new amebas.

I claimed earlier that the best solution to puzzles involving persistent vegetative state and brain dead but functional organisms is to distinguish two concepts of death: the death of the conscious subject and the biological death of the organism. Reflection on Jesus’ words and on cases of biological fission or twinning shows that my proposal does not require any conceptual revision. We *already have* the two required concepts of death. Both are clearly present in the language.

It seems, then, that our concept of death cannot be univocal in the way that Birnbacher would like. What do we lose if we cannot have univocity? Birnbacher says that a univocal concept “dissolves the ambiguities inherent in our everyday thinking about life and death and simplifies things” (p. 136). But the ambiguities and complexities are there for a reason. We can achieve the sort of simplicity that univocity offers only by abandoning certain concepts. But why should we try to restrict our range of concepts if the unruly complexity of reality demands the range we currently have? Unnecessary conceptual amputations are no more desirable than unnecessary physical ones. So I suggest that a satisfactory explication will not demand univocity.

## 2.2 Exhaustiveness

This is the requirement that life and death be exhaustive, that there be no third possibility. The virtue of this requirement is that it “is a further step towards the reduction of complexity and ambiguity” (p. 136). Yet Birnbacher himself seems to recognize a third possibility, which is that an entity may be neither alive nor dead (since being dead presupposes having been alive), yet be such that it can, unlike a rock, *become* alive. This is an interesting suggestion. It seems that it must have happened at least once that nonliving stuff was converted to a very primitive form of life.

Of course, it may be that in this kind of case the living entity is not identical with the nonliving entity from which it was created. If that is so, then this is not an instance of the third state that a being that is at some time alive can be in. But if the living thing is identical with the nonliving thing from which it developed, then Birnbacher has supplied his own counterexample.

I think there is another possibility intermediate between life and death: suspended animation, or cryogenic suspension. But I do not think there is a serious issue here. If we are reductionists about life and death, we know all the facts about an organism that is frozen. There is no further fact about whether it is alive or not. What we should say about such an individual is just an issue of conceptual coherence. If we say that a living human being who is frozen remains alive, that requires that we have what Birnbacher calls a dispositional or capacity-based account of life. For there is no *actual* or occurrent mental life or integrated functioning while the individual is in a frozen state. In a case in which a frozen individual is revived, his being alive while in the frozen state must have *consisted in* his retention of the capacity for mental activity or integrated physical functioning rather than actual activity or functioning. Yet Birnbacher rejects a dispositional account of life in favor of what he calls an actualistic account.

But suppose we say that a living human being who is then frozen is dead. This conflicts with Birnbacher’s next desideratum — the requirement of inclusiveness, which is that death is necessarily permanent. For if an entity cannot die and then live again, and if a frozen body is dead, then it cannot be the same life that is revived when the body is thawed. Thawing, on this pair of assumptions, in-

volves the creation of a new life. But there is no reason to believe that.

### 2.3 Inclusiveness

According to the requirement of inclusiveness, life cannot be discontinuous. An individual cannot die and then live again. So death is necessarily permanent. But again our ordinary concept of death *does not* make it a necessary truth that death is irreversible. If that were true, the claim that Lazarus was raised from the dead, or that Jesus was resurrected, would be *incoherent*. These claims are false; but if it were a conceptual truth that death is irreversible, they would not be false but *nonsensical*.

I do think, however, that there is something important in the idea that death as a *biological* phenomenon is irreversible. It may be a conceptual truth that an organism can be revived from death only by a violation of the laws of nature — that is, only by a literal miracle of the sort that Jesus is thought by some to have performed. For in cases not involving miracles, if an organism that was thought to be dead is restored to integrated functioning, our tendency is to conclude that we were mistaken in assuming that it was dead. Some people, of course, will say that the organism was dead but was non-miraculously restored to life. To make this claim acceptable, they need to offer good reasons for thinking the organism was dead, given that it is now alive.

Suppose we agree with Birnbacher that the concepts of life and death must be univocal and that the appropriate concepts are organismic rather than mentalistic. The requirement of inclusiveness entails that it is a conceptual impossibility for an organism to die and then be revived. Biological death is irreversible. Birnbacher says that a “[d]eath that is followed by resurrection . . . is not truly death” (p. 138). But suppose that some person is annihilated by a nuclear bomb. Immediately afterward, God gathers all the atoms that composed that person’s body and reassembles them. There is now a person who is qualitatively identical with the person who was just annihilated. If we accept the requirement of inclusiveness, we have to say that this is a different person, one newly created.

This seems implausible. Consider a parallel case. Suppose I have



an old wind-up watch. I disassemble it and place each cog, spring, etc. in a separate envelope. I mail the various parts to different people in various distant locations with instructions to send them back to me in a couple of weeks. While the parts are dispersed over the globe, there is no watch. It has ceased to exist. But when the parts are returned to me and reassembled, I think what I have is the same old watch. Disassembly is not the same as permanent destruction. Nothing really hinges on the stipulation that the parts of the watch are for a while geographically separated. If a watch repair person had disassembled my watch in his workshop to lubricate the parts and then put it back together, I would not accuse him of having destroyed my watch and given me only a cleverly disguised replica.

In short, I think the existence of things can be discontinuous. That is true in principle of biological organisms as well. And I think it is true in practice that *life* can be discontinuous, as in cases of cryogenic preservation. So I think we should abandon not only the requirement of univocity, but the requirements of exhaustiveness and inclusiveness as well.

The requirement of inclusiveness also seems incompatible with the rejection of a dispositional account of life and the acceptance of an actualist account. Birnbacher says that

the organismic life of a human individual continues even during periods in which it temporarily ceases to function (as in some forms of freezing) or in which it temporarily loses the capacity to function (as in deep-freezing immediately after the cessation of heartbeat) on the condition that physical functioning, or the capacity of physical functioning, is regained at a future time. (p. 138)

There are several problems with this claim. One is that the reference to freezing immediately after death is incompatible with his stipulation that death cannot be followed by the resumption of life—that is, the inclusiveness requirement—unless he accepts that the new conscious life is a different life. But if the individual continues to be alive while frozen, a dispositional as opposed to an actualist account of life must be correct. Yet Birnbacher insists that an *actualist* account is correct.

To illustrate this claim, he assumes, for the sake of argument, a

mentalist account of life. He then asks us to consider two patients, A and B. A becomes unconscious at  $t_1$  but retains the *capacity* for consciousness. B becomes unconscious at  $t_1$  and at the same time loses the capacity for consciousness irreversibly. Suppose that they are both blown up in an explosion at  $t_3$  without A having regained consciousness. Birnbacher says that their mental lives ended at the same time.

In one sense, that is uncontroversial. After  $t_1$ , neither of them experienced consciousness or mental activity. But the relevant question is not whether there is any ‘mental life’ after  $t_1$ . It is instead whether, if a mentalist account of life is correct, either A or B *continues to exist* after  $t_1$ . That is, the relevant question is whether there is still *someone there* after  $t_1$ . I claim that A continued to exist after  $t_1$  but that B did not. Here is why A did not cease to exist at  $t_1$ , though his mental activity ceased at that point.

Suppose that at some time,  $t_2$ , between  $t_1$  and  $t_3$ , I *could* have caused A to regain consciousness. If I had, then according to the requirement of inclusiveness, he must have been alive between  $t_1$  and  $t_2$ , since on this requirement a person’s life, or existence, cannot be discontinuous. If he had been alive at  $t_2$ , he must have been alive at all times between  $t_2$  and the earlier time when he began to exist. (Although I reject the requirement of inclusiveness as inherent in the *concept* of death, I think that *in fact* there are no instances in which people cease *exist* and then begin to exist again.)

But notice that in fact I *did not* wake A up. On the actualist account that Birnbacher says is the only reasonable view, it follows that A was not alive between  $t_1$  and  $t_2$ . The actualist account therefore implies that whether A was alive or dead between  $t_1$  and  $t_2$  depends entirely on whether I waked him up at  $t_2$ . I did not wake him up, so he was in fact dead during that time. But if I had waked him up, he would have been alive during that time.

I think that if there is one desideratum that it is hard to deny, it is that whether or not a person is alive at a given time depends on that person’s intrinsic properties at that time and not on what may or may not happen later. Whether A was alive *between*  $t_1$  and  $t_2$  cannot depend on whether or not I restore him to consciousness *at*  $t_2$ . I think, therefore, that whether we accept a mentalist or organismic account of life, that account has to be dispositional rather than actualist.

## 2.4 Symmetry

One reason that Birnbacher gives for thinking that an actualist account of life is preferable is that it enables us to have symmetrical accounts of the beginning and end of life—his fourth desideratum. If, on a mentalist account, we begin to exist only when conscious experience begins, then symmetry requires that we cease to exist when consciousness permanently ceases.

Although I think we should reject the requirements of univocity, exhaustiveness, and inclusiveness, I think the requirement of symmetry is right. (I can hardly deny that since Birnbacher kindly says that I have shown that it is correct.) But I think we should preserve symmetry in a different way. I think we are driven, for the reason just given, to accept that we cease to exist only when the *capacity* for consciousness is lost. I therefore accept that we begin to exist when the fetal brain first acquires the capacity for consciousness.

## 3 Reflections on Method

It is hard to give an account of the method I think we should follow in trying to understand death. I agree with Birnbacher that a reconstruction is largely useless. But I think that the requirement of coherence in developing an explication forces us to be respectful of the concepts we already have.

Explication is, I think, largely a matter of metaphysics. But it appeals to our intuitions about what kinds of change we could survive and what kinds we could not survive. It appeals, in other words, to intuitions about egoistic concern. Although I have not said anything about this so far, I follow Derek Parfit in thinking that identity is not the basis of rational egoistic concern. This conclusion comes from reflection on hypothetical cases of human fission, or division. There is obviously not time for me to go into all this, but what it means is that if I could divide like an ameba into two identical people who would both be physically continuous with me, it would be rational for me to be egoistically concerned about both of them even though I would not actually *be* either of them.

This kind of division is, I think, in principle possible but not practically possible, at least at present. So there are no actual cases in

which it might be rational for a person to be egoistically concerned about a person with whom he would not be identical. But I think there *are* cases in which it is rational for a person to have virtually no egoistic concern about himself in the future.

Suppose, for example, that you were in the early stages of progressive dementia and that it was predictable that some years from now your dementia will have progressed so far that you will be barely conscious, with no memory of the past, but will experience great pain. I think you *now* have little reason to fear that future pain. I accept, in other words, that there can be cases in which it can be rational to have little concern about oneself in the future. This exposes my view to an objection concerning methodology. I began by noting that people frequently say of a patient in a persistent vegetative state that ‘she is not there anymore.’ Or they imagine themselves suffering the same sort of injury and they think: ‘That body would not be me. I would not be there any longer.’ But perhaps the explanation of these attitudes is not that people cease to exist when a persistent vegetative state occurs, but that once they are in a persistent vegetative state, it no longer matters, *for them*, what happens to them. In short, the intuition that one would no longer exist in a persistent vegetative state might be just a confused interpretation of one’s sense that at that point one would be beyond benefit or harm, that it would no longer matter what happened to one.

So the appeals to intuition that have a role in metaphysical argument do not necessarily support my view that we cease to exist with the loss of the capacity for consciousness over the rival view that there is no reason to care now about what will happen to us after we lose the capacity for consciousness.

#### 4 Law

One possibility that Birnbacher notes is that the best metaphysical account of death may not be workable as a matter of policy or law. Pragmatic considerations must guide the formulation of the law and if the best metaphysical view is one that people cannot understand or is inconsistent with their beliefs about our nature or identity, then the law may have to be based on a metaphysical view that people can accept even if it is not the view that is best supported by the

arguments. So the challenge to my view is to consider whether it is too bizarre to be acceptable to most people.

Many philosophers have told me that the dualism of person and organism that I defend is deeply counterintuitive. (They usually express the point differently. They say that it is *crazy*.) If they are right about that, then it might be impossible to build the law around my view. People are naturally reluctant to accept a law that has to be defended by appeal to metaphysical claims that are crazy. But the view that we are not organisms is not counterintuitive at all, at least to most people. Most people — at least in the US and other countries where most people retain religious beliefs — do not believe that they are organisms. Anyone who believes that after death he will continue to exist as a disembodied soul cannot believe that he is essentially an organism. So, perhaps surprisingly, I think my view turns out to be reasonably well adapted to the rather primitive metaphysical beliefs that most people actually hold.