Brain Death Without Definitions

Winston Chiong

ABSTRACT: Most of the world now accepts the idea, first proposed four decades ago, that death means "brain death." But the idea has always been open to criticism because it doesn't square with all of our intuitions about death. In fact, none of the possible definitions of death quite works. Death, perhaps surprisingly, eludes definition, and "brain death" can be accepted only as a refinement of what is in fact a fuzzy concept.

Until recently, "brain death" was widely regarded as one of the crowning conceptual achievements of bioethics. After all, less than four decades after the whole-brain criterion of death was first proposed, it has come to supplant the traditional, cardiopulmonary criterion of death throughout much of the world. Not only doctors but also lawmakers and religious authorities have embraced the view that we die when our brains irreversibly cease to function. This revolution in our thinking about human death has had profound practical implications. It has opened the way to vital organ donation and unilateral withdrawal of treatment from patients with beating hearts but no hope of recovering brain function.

In recent years, however, the whole-brain criterion of death has come under increasing criticism, and a growing consensus has developed among bioethicists and philosophers that brain death is actually incoherent. While the proponents of brain death have typically defended it on the grounds that the brain is necessary for the integrated functioning of the organism of the whole, recent findings appear to contradict this claim.[1] Furthermore, extensive study of the "brain dead" has shown that even after the standard battery of diagnostic tests for brain death has been fulfilled (including documentation of coma, the absence of brainstem reflexes, and the absence of respiratory effort), many brain functions persist—including such presumably integrative functions as hormone secretion and thermoregulation.[2]

These apparent inconsistencies have led even one of brain death's most prominent defenders to admit that "Brain death was accepted before it was conceptually sound."[3] I will argue in this paper that, while the whole-brain criterion of death is roughly correct, the conceptual

[1] D.A. Shewmon, "Chronic 'Brain Death': Meta-analysis and Conceptual Consequences," *Neurology* 51 (1998):1538-45.
[2] A. Halevy and B. Brody, "Brain Death: Reconciling Definitions, Criteria, and Tests," *Annals of Internal Medicine* 119 (1993): 519-25; and R.D. Truog, "Is It Time to Abandon Brain Death?" *Hastings Center Report* 27, no. 1 (1997): 29-37.
[3] James Bernat, in G. Greenberg, "As Good As Dead," *New Yorker* August 13, 2001.

framework that its advocates have appealed to is deeply philosophically flawed. In their arguments in support of the whole-brain criterion, the advocates of brain death have appealed to a misguided philosophical model of what is required for the justification of a criterion of death, which their opponents have adopted and turned against them. In particular, this model depends on some claims about language that, while initially plausible, have been seriously undermined by Ludwig Wittgenstein, Saul Kripke, and Hilary Putnam, whose arguments most philosophers of language regard as decisive. Drawing upon their insights, and also upon promising recent work in the philosophy of biology, I propose a new model for our understanding of life and death, which I argue provides a more secure justification for the whole-brain criterion.

The Challenge to the Whole-Brain Criterion

According to the whole-brain criterion of death, death occurs when the whole brain irreversibly ceases to function. Coma, absence of respiratory effort, and absence of brainstem reflexes are the standard tests for the loss of whole-brain function. The two main alternatives to the whole-brain criterion are the higher-brain criterion and the traditional cardiopulmonary criterion. Advocates of the higher-brain criterion claim that it is not the irreversible loss of the functioning of the whole brain, but only of the neocortex—the part of the brain responsible for consciousness, memory, personality, and perception—that is necessary and sufficient for death. This criterion would only require permanent unconsciousness for the declaration of death, dismissing lower-brain functions such as respiratory drive and brainstem reflexes as irrelevant. Advocates of the cardiopulmonary criterion, on the other hand, claim that the irreversible loss of circulatory functioning and the irreversible loss of respiratory functioning together are necessary and sufficient for death. As the cardiopulmonary criterion is usually interpreted, it does not matter whether these functions are carried out spontaneously or via external measures (such as a ventilator or chest compressions). Thus, according to the cardiopulmonary criterion, a patient without brain function and therefore without respiratory drive can be kept alive on a ventilator in the absence of spontaneous breathing, while the whole-brain and higher-brain criteria would class such a patient as dead.

Given the abstract character of debates over the nature of death, how can we hope to make any progress in deciding between these criteria? In an influential early series of articles in support of the whole-brain criterion, James Bernat, Charles Culver, and Bernard Gert offered a framework for these disputes, which I call the *definitions-criteria-tests* model. Bernat set out this framework as follows:

> "This analysis of brain death should be conducted in three sequential phases: (1) the philosophical task of making explicit the *definition* of death that is implicit in our traditional conception of death; (2) the combined philosophical and medical task of identifying the *criterion* of death—that generally determinable standard that shows that the definition is satisfied by being both necessary and sufficient conditions for death; and (3) the medical task of devising a set of bedside *tests* to show that the criterion of death has been fulfilled. Thus, the optimal sequence of argument must proceed from the intangible and conceptual to the tangible and measurable.[4]"

[4] J.L. Bernat, "How Much of the Brain Must Die in Brain Death?" *Journal of Clinical Ethics* (1992): 21-28, at 21-22. See similar passages in J.L. Bernat, C.M. Culver and B. Gert, "On the Definition and Criterion of Death," *Annals of Internal Medicine* 94 (1981): 389-94, at p. 389, and "Defining Death in Theory and in Practice," *Hastings Center Report* 12, no. 1 (1982): 5-9, at 5-6. Culver and Gert have changed their views about the definition of death in ways that resemble the

On this model, the primary philosophical problem is to arrive at the correct "definition" of death—presumably as a result of something like a conceptual analysis of our ordinary notion of death. Such a definition, intended to cover all literal biological uses of our English word "death," is supposed to be general and devoid of specific reference to human physiology; for instance, *the permanent cessation of the integrated functioning of the organism as a whole*, or *the departure of the animating or vital principle*. Once we have the proper definition, we may then proceed to a more practical level, arriving at a species-specific "criterion" of death that gives necessary and sufficient conditions for satisfying the definition of death in human beings.[5] Finally, with this criterion in hand, we can devise clinical "tests" that indicate when the criterion has been satisfied.

Bernat and other proponents of the whole-brain criterion have appealed to a definition of death as the permanent cessation of the integrated functioning of the organism as a whole, by which they mean "that set of vital functions of integration, control, and behavior that are greater than the sum of the parts of the organism, and that operate in response to demands from the organism's internal and external milieu to support its life and to maintain its health.[6] With this definition in hand, they have gone on to make the empirical claim that the destruction of the whole brain is necessary and sufficient for satisfying this definition in human beings— because the brain is the "master organ" or "critical system" that integrates the activities of the other organs into a cohesive whole.[7]

However, the empirical claim that the loss of whole-brain function is necessary and sufficient for the cessation of integrated functioning has since been cast into doubt, prompting a revival of interest in the cardiopulmonary criterion. Given advances in critical care, it is now possible to maintain bodies for long periods after they meet clinical tests for brain death, during which they exhibit numerous forms of integration and control that are "greater than the sum of the parts of the organism" and demonstrate functional responsiveness to their physiological milieu. Alan Shewmon has published a report identifying 175 cases in which the bodies of patients reliably diagnosed as fulfilling the whole-brain criterion were maintained for at least one week (and in rare cases years), sometimes with little aggressive intensive care besides mechanical ventilation.[8] These bodies exhibit a "litany of non-brain-mediated somatically integrative functions," including

- homeostasis of a limitless variety of physiological parameters and chemical substances;
- assimilation of nutrients;
- elimination, detoxification, and recycling of cellular wastes;
- energy balance;
- maintenance of body temperature (albeit subnormal);
- wound healing;

---

definition offered by defenders of the brainstem criterion of death (discussed in section 3 of this paper), but remain committed to this model. See B. Gert, C.M. Culver and K.D. Clouser, *Bioethics: A Return to Fundamentals* (Oxford: Oxford University Press, 1997), ch. 11.

[5] I interpret Bernat's claim that criteria are meant to give "necessary and sufficient" conditions for death as a claim about nomological, rather than metaphysical, necessity and sufficiency. After all, the WBC, CPC and HBC must be intended as criteria for death in a given actual species (i.e., *Homo sapiens*)—they could not seriously be proposed as criteria for death in plants or microorganisms, or in possible species with unknown physiologies.

[6] J. L. Bernat, "A Defense of the Whole-Brain Concept of Death," *Hastings Center Report* 28, no. 2 (1998): 14-23, at 17.

[7] See, e.g., J.L. Bernat, "The Biophilosophical Basis of Whole-Brain Death," *Social Philosophy and Policy* 19 (2002): 324-42.

[8] Shewmon, "Chronic 'Brain Death'."

- fighting of infections and foreign bodies;
- development of a febrile response to infection (albeit rarely);
- cardiovascular and hormonal stress responses to incision for organ retrieval;
- successful gestation of a fetus (as in thirteen pregnant women of the prolonged survivors);
- sexual maturation (in two prolonged-surviving children);
- and proportional growth (in three children).[9]

Some defenders of the whole-brain criterion have responded by attempting to narrow the sense of "integrated functioning" so that it refers not to these functions, but only to functions that are in fact carried out by the brain. For instance, Bernat has proposed a defense of the whole-brain criterion that appeals to "critical functions," by which he means functions that are necessary for the continued health and life of the organism.[10] But since it is unclear whether only brain functions satisfy this standard, or why this standard should be preferred to the broader reading of "integrated functioning," so this move has struck many critics as ad hoc.

On these grounds, advocates of the cardiopulmonary criterion have argued that the whole-brain criterion is not sufficient for the permanent cessation of integrated functioning. Instead, Shewmon's cases suggest that the irreversible loss of circulatory and respiratory functioning is necessary and sufficient for satisfying this definition of death. Thus, if the definitions-criteria-tests model is right, and if the correct definition of death is the permanent cessation of integrated functioning, then it seems that the traditional cardiopulmonary criterion is the proper criterion for human death.

Intuitive Grounds for Doubt

Ultimately, however, it is not the whole-brain criterion but the definitions-criteria-tests model that must be given up. At the level of definitions, there neither is nor need be any general feature that defines death, and at the level of criteria, there are no necessary and sufficient conditions for an organism's being dead.

Although a definition of death as "the permanent cessation of integrated functioning" would favor the cardiopulmonary criterion, there are strong intuitive grounds to reject this conclusion: the cardiopulmonary criterion gives no intrinsic consideration to consciousness as a characteristic of biological life, but on a commonsense view an organism's being conscious, in itself, counts very strongly in favor of its being alive. Consider the following example.

In many cases of sudden cardiac arrest, the victim remains conscious for several seconds after blood stops flowing to the brain. This would also be true of someone who suffered an *irreversible* cardiac arrest while simultaneously suffering a second injury that irreversibly stopped respiration. On any ordinary understanding of life, this double victim would clearly remain alive as long as he remains conscious; if he managed to mouth a few words or flail around before lapsing into unconsciousness, we would have little inclination to say that these words and actions were produced by a dead organism (a sort of animated corpse?); we would say instead that they were produced by an organism in the process of dying. Yet on the cardiopulmonary criterion he is dead when he suffers the irreversible loss of circulation and respiration, regardless of whether he briefly

[9] D.A. Shewmon, "'Brainstem Death,' 'Brain Death' and Death; A Critical Re-Evaluation of the Purported Equivalence," *Issues in Law and Medicine* 14 (1998): 125-45, at 139-40.

[10] J.L. Bernat, "Refinements in the Criterion of Death," in *The Definition of Death: Contemporary Controversies*, ed. S.J. Youngner, R.M. Arnold, and R. Schapiro (Baltimore, Md.: Johns Hopkins University Press, 1999), 83–92.

retains consciousness.

We should note that this is not only a counterexample to the cardiopulmonary criterion, but also to a definition of death as the permanent loss of integrated functioning. Quite plausibly, an organism that can no longer breathe or circulate blood is no longer functioning in an integrated way. If this is right, then in this case the permanent loss of integrated functioning does not amount to death. Furthermore, this counterexample does not rely upon special assumptions about human consciousness, self-consciousness, or personhood. If mice also retain consciousness for several seconds after sudden cardiac arrest, we would be just as inclined to judge that they remain alive so long as they are conscious.

This suggests that something's being conscious is, in itself, a strong indicator that it is a living organism. Some might then suggest abandoning Bernat's proposed definition of death as the loss of integrated functioning in favor of a definition of death as the permanent loss of consciousness. Such a definition would favor the higher-brain over the whole-brain criterion: the destruction of the whole brain is not necessary for the permanent loss of consciousness, as the destruction of the neocortex is sufficient. However, there are counterexamples to this criterion and definition even more serious than the counterexamples to the cardiopulmonary criterion raised earlier. Brain injuries that destroy the neocortex while sparing the brainstem produce persistent vegetative states (PVS). Those in PVS are not conscious, but often show other classic signs of biological life—including spontaneous breathing and sleep/wake cycles (when "awake" they are not conscious, but are generally more active), and in virtue of brainstem reflexes they may cough when their throats are irritated, blink when their corneas are touched, and swallow food placed in their mouths. I think it intuitively clear that an organism that breathes spontaneously, has circadian rhythms, and exhibits these complex (though non-conscious) responses to stimuli is not dead on any ordinary understanding of life and death—strongly suggesting that the irreversible loss of consciousness also does not amount to death.

Against the Definitions-Criteria-Tests Model

These cases ground strong intuitive objections to the cardiopulmonary and higher-brain criteria, and also to the definitions that would most naturally be taken to justify them. They also undermine an idea implicit in the definitions-criteria-tests model: that there is some special characteristic common to all living or to all dead things, in virtue of which they are alive or dead. In the sudden cardiac arrest case, what seems to guide our judgment about whether or not the victim is alive or dead is the presence of consciousness. But in the PVS case, our judgments about life and death seem to track different characteristics entirely—most notably, the presence of spontaneous respiration, irrespective of the presence of consciousness.

Taken together, these cases suggest that the irreversible loss of consciousness *and* the irreversible loss of spontaneous respiration are each individually necessary for death—neither is sufficient on its own. This finding echoes claims made by defenders of the brainstem criterion of death, which is closely related to the whole-brain criterion and has been adopted in the United Kingdom. (The brainstem criterion can be thought of as the anatomical converse of the higher-brain criterion, with the rationale that while neocortical function is necessary for consciousness, brainstem activation is also required for consciousness as well as for the nonconsciously mediated behavior observed in PVS.) As Christopher Pallis and D.H. Harley write, "We consider human death to be a state in which there is irreversible loss of the capacity for consciousness combined with irreversible loss of the capacity to breathe spontaneously (and hence to maintain a spontaneous heartbeat).

Alone, neither would be sufficient.[11]

Critics of the brainstem criterion have claimed that this rationale for a criterion of death does not satisfy the definitions-criteria-tests model because the definition offered is not theoretically unified. For instance, Shewmon has complained that "apneic coma as a *concept of death* is completely idiosyncratic, pulled out of philosophical thin air."[12] Similarly, Baruch Brody argues that "Neither a purely respiratory criterion nor a combined respiratory/consciousness criterion lends itself to a justifying definition. The former criterion involves only one of the traditional vital "bodily fluids," and it is hard to see why one is to be preferred to the other. The latter criterion comes from two very different definitions, and it is hard to see why the two criteria should be combined."[13]Shewmon and Brody thus insist that a criterion for death appeal to a nonidiosyncratic, unified "justifying definition." But what underlies this demand for a unified definition of death—particularly when this demand seems to conflict with our best intuitions about cases?

This requires a bit of reconstruction, but I think the best available rationale for this demand, and indeed for the entire definitions-criteria-tests model, relies upon a natural but flawed picture of rigorous theoretical investigation. The picture goes something like this: if we're going to investigate some phenomenon X (such as death), we must begin with a definition of the term "X" that serves both metaphysical and semantic purposes. Metaphysically, the definition of "X" is supposed to give us a unified account of *what it is* for something to be X (rather than, say, Y or Z) and thus an account of the truth-conditions of the claim that one or another particular thing is X. Otherwise, we wouldn't think that such a definition would help us to find the characteristic or characteristics that all Xs have in common that is necessary and sufficient for being X. Semantically, the definition of "X" is supposed to explain how, when we use the term "X", we succeed in referring to things that are X rather than things that are not X—the idea being that there is some implicit concept or mental content associated with the term "X" that picks out the common characteristic that is definitive of Xs. Otherwise, we wouldn't expect conceptual investigation of our traditional understanding of X to be of much use in revealing the nature of X.

This *descriptivist* picture of how terms like "death" work was similarly dominant within analytic philosophy of language and philosophy of science until the 1970s. One early challenge to this picture concerns the metaphysical role that these definitions are supposed to play. When theorists present "definitions" of death like *the permanent cessation of the integrated functioning of the organism*, or *the irreversible loss of consciousness*, they are attempting to tell us what is common to all dead things in virtue of which they are dead—we might say, they are attempting to state the *essence* of death. However, as Wittgenstein famously argued, in natural languages we find many terms for which there is no essential characteristic that determines whether the term applies in a given case. Consider his refusal to offer a unified definition of a "language-game":

> "Instead of producing something common to all that we call language, I am saying that these phenomena have no one thing in common which makes us use the same word for all,—but that they are *related* to one another in many different ways. And it is because of this relationship, or these relationships, that we call them all "language". I will try to explain this.

---

[11]C. Pallis and D.H. Harley, *ABC of Brainstem Death* (London, U.K.: BMJ Publishing Group, 1993), 28. See also Gert, Culver and Clouser, *Bioethics*.

[12] Shewmon, "'Brainstem Death,' 'Brain Death' and Death," 132.

[13] B. Brody, "How Much of the Brain Must Be Dead?" in *The Definition of Death: Contemporary Controversies*, 71-82, at 78.

> Consider for example the proceedings that we call "games". I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all?—Don't say: "There *must* be something common, or they would not be called 'games'"—but *look and see* whether there is anything common to all.—For if you look at them you will not see something that is common to *all*, but similarities, relationships, and a whole series of them at that. To repeat: don't think, but look!—Look for example at board-games, with their multifarious relationships. Now pass to card-games; here you find many correspondences with the first group, but many common features drop out, and others appear. When we pass next to ball-games, much that is common is retained, but much is lost."[14]

These points are directed against philosophers who would demand a unified definition of what it is for something to be a game, and can be extended to bioethicists who, appealing to the definitions-criteria-tests model, demand a unified definition of death or life. The terms we use don't always require any such shared, essential characteristic to do their work.

Still more serious objections to this style of descriptivism were later presented by Saul Kripke and Hilary Putnam, who attacked the semantic role that such definitions had been thought to play. On the descriptivist model, the reference of terms like "life" and "death" is fixed by our implicit mental association of these terms with definitions that specify essential features of their referents. In other words: on this view, what enables me to use the word "death" to make claims about dead things is my grasp of some description that tells me what it *means* for something to be dead. Here Kripke argued that the mental contents associated with certain terms, particularly proper names and natural kinds, may be merely contingent or even entirely mistaken. To cite Kripke's most famous example: it could be that all of the descriptions that we use to identify Gödel are in fact not true of Gödel but instead of some more obscure mathematician; but even if this were so, in using the term "Gödel" we would still be talking about Gödel rather than his colleague.[15]

Putnam extended Kripke's suggestion that the reference of terms does not depend solely on the mental contents of the speaker, but also on contingent causal connections between the speaker and the world, the nature of which may even be inaccessible to the speaker.[16] On this view, philosophical analysis of the mental contents or concepts associated with a term may be quite irrelevant, in itself, to discovering necessary and sufficient conditions for the term's application. Instead, Putnam suggested that these mental contents or concepts might merely provide "operational definitions" rather than essential ones: ways of picking out the referents of our terms that appeal to contingent rather than necessary and sufficient characteristics of their referents. So an English speaker before the advent of chemistry, who would have been ignorant of the fact that what *makes* something water is having the chemical composition H_2O, could still refer to water by conceiving of it under some contingent description— as the primary constituent of the lakes and rivers of the actual world, for example, or as the primary constituent of some paradigmatic sample of water. Furthermore, an analysis of this speaker's "definition" of water wouldn't, in itself, get us very far in trying to find out what makes something water (that is, being H_2O): this is a matter for empirical rather than conceptual investigation.

---

[14] L. Wittgenstein, *Philosophical Investigations*, 3rd ed., trans. G.E.M. Anscombe (New York: MacMillan, 1958), §§65-66.
[15] S. Kripke, *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1972).
[16] H. Putnam, "Meaning and Reference," *Journal of Philosophy* 70 (1973): 699-711.

Similarly, armchair theoretical investigation of our ordinary concepts of death is unlikely to help us decide among possible criteria of death. Linguistic terms can be successfully used to refer even when the mental contents associated with them only contingently pick out their referents, and sometimes when these contents do not even truly apply to their referents. For instance, even if there is no such thing as an immaterial soul, someone whose concept of death is *the departure of the immaterial soul from the physical body* might nonetheless succeed in referring to the dead rather than the living—perhaps by deferring to authorities in her community, or by having false beliefs about the relationship of the soul to the physiological features that actually are involved in death, or by deferring to authorities with such beliefs.

A further advantage of this approach to the semantics of "life" and "death" is that it allows that people might have different concepts or mental contents associated with the terms "life" and "death" and yet still be talking about the same things. It's quite plausible to me that people in different cultures and times, and often in the same culture and at the same time, have different concepts of life and death. But we still want to be able to say that (for example) a vitalist and a materialist about life are actually *disagreeing* about the nature of life, not merely talking about two different things.

Taken together, I believe that the arguments of Wittgenstein, Kripke, and Putnam together show the definitions-criteria-tests model, and the associated demand for a unified justifying "definition," to be deeply philosophically flawed. The right way to argue for a given criterion of death is not to argue that it fits some singular "definition" of life or death that results from any conceptual analysis of these terms. Metaphysically, it may be that there is no shared characteristic common to all dead things in virtue of which they are dead; and semantically, the concepts or mental contents that we associate with death may merely be operational definitions that appeal to accidental features rather than necessary and sufficient conditions for death. We ought, therefore, to seek some alternative theoretical model that can make sense of our best intuitive responses to cases—such as the cardiac arrest and PVS cases.

Life and Death as Cluster Kinds

When we give up the idea that criteria for death must be backed by a unified definition of death, we can abandon the claim that there need be any single characteristic, or even any conjunction of characteristics, that is both necessary and sufficient for an organism to be alive or dead. Indeed, cases like the cardiac arrest and PVS cases lead me to suspect that no *single* characteristic works this way; instead, the property of being alive (like the property of being a language or a game) involves a cluster of characteristics—none of which is in itself necessary and sufficient for an organism to be alive, but all of which contribute to an organism's being alive and tend to reinforce one another in paradigm cases. To list some important examples:

1. Consciousness;
2. what might be called, at the risk of circularity, spontaneous vital functions, which may vary from species to species (where "vital" means those functions that are necessary for the persistence of the other functions of the organism and "spontaneous" means that these functions are regulated and maintained by activities that are internal rather than external to the organism);[17]

[17] I take the terms "spontaneous," "internal," and "external," as used here, to admit of borderline cases—see the discussion of indeterminacy in the next section.

3. behavior—that is, functional responsiveness to environmental stimuli, regardless of the presence of consciousness;
4. integrated and coordinated functioning of multiple subsystems—a certain degree of organizational complexity and coherence;
5. the ability to resist decay and putrefaction;
6. the capacity to reproduce; and
7. the capacity to grow via the assimilation of nutrients.

No doubt this list is incomplete, but I take it as a start. None of these characteristics appears necessary for life: the cardiac arrest victim might only exhibit the first feature, while those in PVS lack it. But at the same time, at least in paradigm cases of living things, these characteristics are related: they tend to be mutually supporting and reinforcing. For instance, in higher animals (we suspect), consciousness contributes to an organism's functional responsiveness to its environment, which helps it to respond to situations in ways that support its spontaneous vital functions and the coordination of its subsystems, which in turn support the structures necessary for consciousness and behavior.

In this proposal I appeal to Richard Boyd's influential recent work in the philosophy of biology, and more specifically to the distinction between natural kinds (or "real kinds") and nominal kinds. Natural kinds involve categories that occur in nature and independently of human interests (e.such as "water" or "tigers"), while nominal kinds involve categories that we take to be useful or otherwise answer to human interests ("chair" or "salad fork"). Traditionally, realists about natural kinds have held that the members of natural kinds, unlike those of nominal kinds, are unified in virtue of having in common some intrinsic characteristic or set of characteristics that is necessary and sufficient for membership.[18]

If this traditional understanding of natural kinds is correct, then the following objection could be raised to the cluster theory of life and death that I have proposed: "Languages and games are human inventions while life and death are not human inventions. It's acceptable that languages and games don't involve shared, necessary, and sufficient membership conditions because there's no *objective*, mind-and-language-independent fact of the matter whether something is a language or game anyway. However, life and death are matters of biological fact, independent of human purposes and intentions, and therefore there must be shared, necessary, and sufficient conditions for something's being alive or dead."

However, this traditional conception of natural kinds has also been recently undermined. It turns out that almost none of the categories investigated in biology, nor in most of the other special sciences— such as psychology, meteorology, astronomy, economics, or linguistics— involve shared intrinsic characteristics that are necessary and sufficient for membership. The most-discussed controversy is over species. Michael Ghiselin and David Hull have argued on evolutionary grounds that species do not involve this sort of necessary and sufficient membership condition—for instance, there needn't be any special characteristic that tigers all have in common that makes it the case that they're all tigers—and therefore are not categories or *kinds*, but in fact are massively

[18] 'Intrinsic characteristics' is a term of art for properties that are non-disjunctive and non-relational. Much traditional thinking about natural kinds is put in terms of intrinsic characteristics—see Robert A. Wilson's discussion of the role of "intrinsic properties" in "traditional scientific realism" in "Promiscuous Realism," *British Journal for the Philosophy of Science*, 47 (1996): 303-16, at 304-05. See also a related discussion of the reality of cluster concepts as "class terms" in D.L. Hull, "A Matter of Individuality," *Philosophy of Science* 45 (1978): 335-60, at 355.

spatiotemporally extended *individuals* composed of numerous organisms.[19] But this proposal faces serious objections of its own, and it is difficult to see how it might be generalized to other categories in the special sciences that are also taken to be natural kinds but are not spatiotemporally continuous in the way that species may be.

In a recent series of papers, Richard Boyd has proposed an alternative understanding of natural kinds that does not involve necessary and sufficient membership conditions, but instead appeals to what he calls "homeostatic property clusters."[20] These are Wittgensteinian families of properties that tend to be nonaccidentally coinstantiated, in that something's possessing some of the properties in the cluster makes it more likely that it will also possess the other properties in the cluster. Boyd has argued that a number of biological categories (not only biological species, but also the higher taxa) involve homeostatic property clusters, as do many of the categories studied in economics and geology.

This proposal shows that cluster kinds can be natural kinds: that categories can occur in nature prior to our classificatory schemes without any intrinsic characteristic that all members of the category have in common. Thus, against Ghiselin and Hull's presuppositions, two animals might both belong to the natural kind "tiger" even though there is no special characteristic that both possess that all nontigers lack—because some of the characteristics that they do possess have the natural higher-order property of being members of a homeostatic property cluster. Another such natural kind, I think, is "living organism." Even though there is no single special characteristic that a recent victim of cardiac arrest and a PVS patient possess and that all nonliving things lack, both of them might belong to the natural kind" living organism" because some of the characteristics that they do possess (for the cardiac arrest victim, consciousness, and for the PVS patient, the capacity for spontaneous respiration, integrated functioning, and so on) have the higher-order property of being members of a homeostatic property cluster. At least in humans and higher animals, these characteristics are coinstantiated in the vast majority of cases and function to sustain one another.

The Indeterminacy of Life and Death

While the properties in a Wittgensteinian cluster tend to be coinstantiated, they are not always coinstantiated. In some cases, an individual will possess some but not all of the properties in the cluster. When some property is central to the cluster—as I've argued consciousness is—then possessing only this one property may be sufficient for membership in the natural kind. However, merely possessing one or several properties that are peripheral to the cluster may not be sufficient for membership. Consider organizational complexity and behavior: some robots are organizationally complex and functionally responsive, though intuitively not alive. In between clear cases of membership and exclusion from the natural kind there will be borderline cases, in which it is indeterminate whether something is a member of the kind.[21]

Intuitively, viruses strike me as a borderline case: there seems to be no determinate answer to

[19] See Hull, "A Matter of Individuality."

[20] R. Boyd, "Homeostasis, Species, and Higher Taxa," in *Species: New Interdisciplinary Essays*, ed. R.A. Wilson (Cambridge, Mass.: MIT Press, 1999), 141-85, at 143-144. See also "What Realism Implies and What it Does Not," *Dialectica* 43 (1989): 5-29; and "Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds," *Philosophical Studies* 61 (1991): 127-48.

[21] Boyd: "Moreover, there will be many cases of extensional indeterminacy, which are not resolvable even given all the relevant facts and all the true theories. There will be things that display some but not all of the properties in [the cluster] such that no rational considerations dictate whether or not they are to be classed under [the natural kind term], assuming that a dichotomous choice is to be made." "Homeostasis, Species, and Higher Taxa," p. 144.

the question of whether or not viruses are alive. Of the various characteristics I have identified as relevant to life, viruses exhibit only the capacity for reproduction. While consciousness may be determinately sufficient for something to be alive, I take it to be indeterminate whether reproductive capacity is by itself sufficient, which is why the question of whether viruses are alive admits no determinate answer.

It is also very plausible to think that, in the course of dying, many people pass through a borderline state in between being determinately alive and determinately dead: think of a comatose patient in multisystem organ failure, losing one bodily function or capacity after another. Before the advent of critical care and mechanical ventilation, this progression might have advanced so quickly as to escape notice—but now we often face patients in whom this process may be arrested indefinitely. How should we deal with these indeterminate cases?

One response might be to refuse to treat such people as either alive or dead, perhaps devising some third category for the indeterminate cases. But this approach faces serious practical and theoretical problems. One practical problem is that there is as yet no consensus on how to treat those who are "neither alive nor dead," which would likely be troubling and confusing for their families. A theoretical problem is that this approach would only succeed in pushing off the basic problem to a higher level. This approach does not account for the phenomenon of higher-order indeterminacy. Just as there are borderline cases between determinate life and death, there are also borderline cases between determinately determinate life and the determinately indeterminate cases. So even on this approach, at some point a discontinuous boundary must be introduced.

A different approach, commonly applied to indeterminacy elsewhere, is to sharpen an originally indeterminate distinction by introducing an artificially defined cutoff that can be used to sort the borderline cases determinately into different categories.[22] Consider adulthood, which also often admits of borderline cases. When for legal or social purposes it is important to make a dichotomous distinction between adults and children, we may do so by introducing a cutoff at eighteen years. This cutoff is not entirely "arbitrary"—it clearly fits better with the original category than a cutoff at six years or thirty—but it is no more consistent with the original category than many other cutoffs, such as a cutoff at seventeen-and-a-half years. So some ways of sharpening an indeterminate distinction are admissible, while others are not: a cutoff at seventeen-and-a-half years is admissible, while a cutoff at six or thirty is not. For a cutoff to be admissible it must agree with the original distinction in the determinate cases, and a seven-year-old is very definitely not an adult.

We may then look upon competing criteria for death not as attempts to state necessary and sufficient conditions for death, but instead as proposals for sharpening the distinction between life and death. Presumably there is no uniquely admissible cutoff; however, some proposals can be ruled out as inadmissible, while some admissible cutoffs may be preferred to others on practical grounds (how easily and reliably they can be clinically determined, and their degree of fit with longstanding cultural traditions).

A proposed sharpening is inadmissible if it disagrees with the original distinction in the

[22] In the philosophical literature on indeterminacy, this sort of sharpening is called precisification. Let me note here that, while the notion of precisification is usually associated with supervaluationist approaches to indeterminacy, I don't think that this sort of practical application of precisifications commits one to supervaluationism. For instance, an epistemicist who thinks there is a discontinuous but unknowable boundary at which adulthood begins can still admit the option of precisifying adulthood by introducing a cutoff at age eighteen for the purposes of epistemically challenged creatures like ourselves.

determinate cases. Recall that something might lack some of the properties in a cluster and yet nonetheless belong determinately to the relevant kind: for instance, mules cannot reproduce, but healthy mules are nonetheless determinately alive. Similarly, I have claimed recent, still-conscious victims of cardiac arrest and people in PVS are determinately alive. Thus, the cardiopulmonary and higher-brain criteria are not admissible cutoffs: adopting one of these would be tantamount to revising, rather than sharpening, the boundaries of life and death.

A similar case could be made against the whole-brain criterion if it made some determinately living individual count as determinately dead— or some determinately dead individual count as determinately alive. To my knowledge, the best potential counterexamples are the cases of "chronic brain death" documented by Shewmon, in which individuals have retained numerous integrative functions over long periods when maintained on mechanical ventilation. Clearly these cases undermine the traditional rationale for the whole-brain criterion, which definies death in terms of integrated functioning. But since I've argued that demands for such analytic definitions are misguided, the relevant question for our present purposes is whether these intuitively are determinate cases of life, determinate cases of death, or borderline cases in between.

Shewmon's original reaction to these cases—which I suspect he would now disclaim—was that these individuals are determinately alive.[23] I would claim, however, that these are borderline cases, in contrast with those in PVS, whom I regard as determinately living. Think, after all, of the many significant functions that those in PVS exhibit and that Shewmon's patients lack, such as spontaneous breathing, sleep-wake cycles, and complicated functional responses to stimuli such as protective coughing, protective blinking, and swallowing. By contrast, I think that Shewmon's cases of "chronic brain death" are so dependent on the external provision of such paradigmatically vital functions as breathing that they cannot be considered clear, determinate cases of life. And if there are no better potential counterexamples, then the whole-brain criterion is an admissible way of sharpening the boundary between life and death.

This defense of the whole-brain criterion as an admissible sharpening of death, rather than as a necessary and sufficient condition for death, is much more modest than the defenses that others have offered for this criterion. As such, it avoids another serious objection raised against the whole-brain criterion. Critics have noted that, as this criterion is usually employed, it does not literally require the irreversible loss of all brain functions. This is to say that, even after the clinical tests for brain death have been met, other untested functions often persist—for instance, hormone secretion and thermoregulation.[24] Yet few advocates of brain death have proposed that we should adopt a more extensive clinical examination for determining brain death, which would test for *all* brain functions that could persist—in part because such testing might prove both difficult and costly, and might thereby delay organ procurement or the process of grieving. But on what principled ground can one maintain, for example, that although brainstem reflexes are relevant to the determination of death, neurohormonal regulation is not?

Treating the whole-brain criterion as an admissible cutoff helps to defuse this objection. On the view I've proposed, there are many admissible cutoffs—including some that require the loss of neurohormonal regulation, and others that do not (such as the whole-brain and brainstem criteria).

---

[23] For his original reaction to these cases see his comments about "T.K." in "'Brainstem Death,' 'Brain Death,' and Death," 136. In recent articles he has expressed doubts about his earlier views, on grounds similar to those presented in this paper: see D.A. Shewmon and E.S. Shewmon, "The Semiotics of Death and its Medical Implications," in *Brain Death and Disorders of Consciousness*, ed. C. Machado and D.A. Shewmon (New York: Kluwer Academic, 2004), 89-114.

[24] See again Halevy and Brody, "Brain Death," and Truog, "Is It Time to Abandon Brain Death?".

Choosing between these is like choosing between cutoffs for legal adulthood at seventeen-and-a-half, eighteen, and eighteen-and-a-half years—all agree with one another (and the original distinction) in the originally determinate cases, though they disagree about the borderline cases. The choice between them must then be settled on practical rather than purely biological grounds—for instance, by how easily and reliably they can be clinically confirmed. (A major consideration favoring the cardiopulmonary criterion in earlier times was that the relevant tests could be reliably performed by any doctor with a stethoscope.)

A Non-relativistic Pluralism

The account of life and death presented here is pluralistic, in that it admits numerous ways of sharpening the indeterminate boundary between life and death. If the whole-brain criterion is to preferred to these other potential cutoffs, this must be on practical grounds rather than biological or metaphysical grounds. Yet while this position is pluralistic, it is not relativistic: many other proposed sharpenings of this distinction are objectively inadmissible—as I have argued in the case of the cardiopulmonary and higher-brain criteria. Ultimately, these proposals must answer to a mind-and-language-independent standard: agreement with the original boundary between life and death in determinate cases. Thus, any proposal that would treat conscious people as dead would be ruled out, while any proposal that would treat spontaneously breathing people as dead would most likely also be ruled out.

On such a pluralistic view, the adoption of different criteria for death in different societies is unproblematic in itself, so long as these different criteria all represent admissible cutoffs. If the brainstem and whole-brain criteria are both admissible ways of sharpening the distinction between life and death, then we need not be troubled by the fact that the brainstem criterion is accepted in the United Kingdom, while the whole-brain criterion is accepted in the United States. Note, by contrast, that if we held to Bernat's claim that criteria must give necessary and sufficient conditions for being dead, then we would have to regard at least one of these different criteria as objectively wrong.

Must a uniform standard be imposed in all contexts within a given society? I believe there is a default presumption in favor of simplicity and universality; however, the account of life and death presented here is compatible with a limited degree of context-dependence. For instance, some states have adopted "conscience clauses" to accommodate the convictions of people whose cultural or religious traditions do not accord with the whole-brain criterion, such as Orthodox Jews, Japanese, and some Native Americans. In essence, such clauses allow individuals (or their families) considerable discretion in choosing which standard of death will be applied to them. If the different candidate criteria for death made available by such a conscience clause are all admissible ways of sharpening the original distinction between life and death, then such a statute need not be inconsistent with the account of life and death defended here. Whether such conscience clauses make sense as a matter of policy is thus not a matter to be settled solely by attending to the nature of life and death, but also to various practical considerations at stake in balancing the social benefit of unanimity against due respect for the adherents of admissible minority views.

An even more controversial example of context-dependence can be seen in the efforts of some organ procurement agencies to expand the pool of potential organ donors by applying different criteria for death in different circumstances. These efforts make use of the fact that the Uniform Determination of Death Act proposed by the President's Commission in 1981 calls for the declaration of death given *either* the irreversible loss of cardiopulmonary function or the irreversible

loss of whole-brain function (although the Commission's report suggests that cardiopulmonary arrest was recognized only as an indicator of the loss of brain function).[25] Thus, some organ procurement organizations employ two different protocols for vital organ donation: one for donors declared dead on the basis of neurological testing, and another for donors who are declared dead after a two-to-five minute interval following cardiopulmonary arrest ("non-heart-beating donors," or NHBDs). These NHBD protocols are used to facilitate organ procurement from people with neurological injuries that impair only some crucial brain functions—for instance, in conscious people who can no longer breathe spontaneously and wish to be organ donors. In these cases, mechanical ventilation is withdrawn in accordance with the patient's wishes to discontinue life-sustaining treatment, thereby inducing a hypoxic cardiac arrest; following a two-to-five minute interval (depending on the site), the organs are removed quickly to minimize ischemic injury.

Most NHBDs in whom mechanical ventilation is withdrawn in this way would not meet the whole-brain criterion of death—while five minutes of cerebral ischemia likely would result in the permanent destruction of cortical structures required for consciousness, some brainstem structures could survive for many more minutes. (Nancy Cruzan was estimated to have suffered twelve to fourteen minutes of cerebral ischemia, which left her persistently vegetative rather than brain-dead.) However, such a protocol might represent a different admissible cutoff between life and death—recall that the protocol is implemented in people who have already lost the capacity for spontaneous respiration, and then lose integrated functioning and the capacity for consciousness following the withdrawal of mechanical ventilation. If so, then applying the whole-brain criterion to one group of potential donors while applying the NHBD protocol to another group of potential donors would represent another instance of a context-dependent application of different admissible cutoffs for the boundary between life and death.

What remains potentially troubling about this practice, of course, is the idea of tailoring of our standards of death in different circumstances to meet the purpose of facilitating organ procurement. Depending on how these disjunctive policies are implemented and applied, they could easily invite confusion and mistrust about organ transplantation among patients; there is also a danger that the ventilator-dependent may be subtly coerced into assenting to withdrawals of treatment that they would not otherwise have chosen for themselves. In general, I suspect that the potential benefits of such policies as a means of meeting the demand for organs have been overstated. Instead, their advocates might do well to emphasize the potential that these policies present for enhancing the autonomy of the terminally ill and ventilator-dependent in determining the character and circumstances of their deaths. On such a view, NHBD protocols might have more in common with "conscience clauses" than is generally recognized.

Acknowledgments

I would like to thank William Ruddick, Derek Parfit, John Richardson, Wade Smith, James Bernat, and Alan Shewmon for many valuable conversations about the material in this paper. I would also like to thank the editors and reviewers for their input and suggestions for improvement.

[25] President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research, *Defining Death: Medical, Legal and Ethical Issues in the Determination of Death*, (Washington, D.C.: U.S. Government Printing Office, 1981). This sort of disjunctive legal standard is also employed in many states that did not adopt the Uniform Determination of Death Act.