

*Personal Identity Theory***9.1 Introduction**

In everyday human life, the identification of persons is so commonplace and straightforward that it might seem odd to suggest that personal identity could be the subject of intense scholarly debate. Many social interactions require us to know who the other person in the interaction is. Usually, we have no trouble identifying family members, friends, colleagues, and acquaintances. In our internal psychological lives, each of us presupposes that our memories feature us – not someone else – as their subject and that it is us – not someone else – who will carry out the actions we plan. In moral and legal contexts, we assume that if someone did something culpably wrong at some past time, it is that person who bears primary responsibility for the action. Marriages and other human relationships presuppose the continuing existence of the individuals who occupy the relevant positions in the relationships even if their characteristics change significantly over time. Each person is distinct and persons typically live for many years. We are so familiar with such facts and so adept at identifying and reidentifying persons in everyday life that these phenomena may seem unworthy of investigation.

Yet in some situations personal identity is uncertain. A victim of amnesia may know neither who he is nor the experiences his past contained. An individual in the late stages of progressive dementia may strike others as not being the person they knew before. Meanwhile, the demented individual may feel no connection to her distant past and have no plans for the future. Even in cases that are free of medical pathology, some aspects of our existence might provoke questions about identity. For example, you might wonder when you came into existence. Did you exist as a fetus before it had any mental life? Might you, on the other end of life, continue to exist after losing the capacity for consciousness? The answers to such questions are far from obvious.

As these reflections suggest, questions about personal identity intersect with issues in bioethics. In the bioethics literature, personal identity theory has animated discussions concerning the authority of advance directives, the definition of death, the ethics of human enhancement, and the “nonidentity problem” (as discussed in Chapter 10). Considerations of personal identity have also played a role – sometimes only implicitly – in debates about such topics as abortion and embryo research, preimplantation genetic diagnosis and embryo selection, and prenatal genetic manipulation.

In this chapter, we explore four approaches to personal identity as philosophers have usually understood this concept – that is, in terms of so-called *numerical* identity: person-based accounts, a biological account, a mind-based account, and a social account. In exploring person-based accounts, we introduce the distinct concept of *narrative* identity, which involves a person’s self-conception or self-told story about herself and her life. The chief upshots of our theoretical investigation are, first, that person-based accounts and the social account are implausible accounts of numerical identity and, second, that the mind-based and biological accounts are *both* plausible. This motivates a pluralistic approach to personal identity in the sense that policies and practices should be consistent with both accounts.

With these theoretical resources, we turn to three areas of application. In the first, we neutralize some concerns about human enhancement through biomedical means. In the second, we investigate and ultimately vindicate the authority of advance directives in cases of severe dementia. Finally, we take up the controversy over the definition of death and associated questions about unilateral discontinuation of life support and vital organ procurement. We find that proper resolution of these issues turns primarily on practical considerations other than the nature of death. The overarching lesson of our practical investigations *deflates* the role of personal identity theory in bioethics. Contrary to the claims of numerous bioethicists who have invoked personal identity, after we have narrowed down the theoretical options to genuinely plausible accounts, the latter do not have far-reaching implications in bioethics.

9.2 Person-Based Accounts and Their Difficulties

Dating back at least to the pre-Socratics in the West, philosophers have wondered about the fact that objects of ordinary experience can change over time yet continue to exist. A diary may change gradually – its pages

becoming more yellow, the binding less sturdy – yet common sense suggests that one object, the diary, persists through the changes. On the other hand, truly radical change may end an object's existence, for example, when the diary is burned to ashes. Some changes, apparently, are compatible with the persistence of a particular object, while other changes are not.¹ Identity in this sense is called *numerical identity*.

What are the criteria of identity for beings like us, that is, human persons? In Chapter 7, we conceptualized persons as *beings with narrative identities* – *relatively complex understandings of themselves as persisting over time and as having an implicit life story*. Thus, while many mammals and birds may display a degree of rationality and some awareness of themselves as persisting over time, few if any nonhuman animals possess a narrative identity. We assume that most human beings have acquired a narrative identity and therefore qualify as persons by the age of three or four.

Narrative identity constitutes a second sense of the term “personal identity.” Whereas numerical identity concerns criteria for persistence over time, narrative identity concerns an individual's self-conception: their implicit autobiography, most central values, and identifications with particular persons, activities, and roles. This is the sense of identity that is at risk of falling apart in an identity crisis.² When an adult undergoing a mid-life crisis or a confused adolescent asks, “Who am I?” they are typically not suffering from amnesia or for any other reason trying to work out their numerical identity. Rather, they are trying to get their bearings about how to “define” themselves, what is most important to them, and what self-image should guide them through major life decisions. In this way, narrative identity straightforwardly involves an individual's psychology. It therefore varies with each individual insofar as each has a distinct self-conception.

Our personhood is so central to our lives that we might find it natural to assume that we are *essentially* persons – that we literally cannot exist at any time without being persons at that time. We may call this thesis *person*

¹ Criteria of identity for a particular object *over time* involve persistence conditions. But criteria of identity also determine whether a particular object *would exist in counterfactual situations* – a matter of “trans-world identity.” For example, assuming a particular copy of a book was made from a particular stock of paper, could that very same copy have existed despite being made from a different stock of paper? Could you have come into existence as the result of the fertilization of a different sperm and egg cell than those that were actually involved in your history? For a classic discussion of trans-world identity, see Saul Kripke, *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1970). This chapter focuses on persistence conditions rather than conditions for trans-world identity – except in discussing genetic interventions on gametes or embryos that might affect who comes into being.

² For an excellent exposition of narrative identity, see Marya Schechtman, *The Constitution of Selves* (Ithaca, NY: Cornell University Press, 1996).

essentialism. If person essentialism is true, then our numerical identity consists in the continuing existence of a being with a narrative identity, suggesting that we come into existence as such beings and persist as long we maintain our narrative identities. Reflection suggests that this view cannot be correct. Even though it is debatable precisely when one of us comes into being, it seems obvious that we existed by the time of our birth. Yet person essentialism implies that we did not come into existence until the emergence of a narrative identity at around age three or four. Even if we adopted some other conception of personhood within the Lockean tradition, which invoked a trait such as rationality or self-awareness, person essentialism would implausibly entail that none of us existed as a newborn infant (since newborns clearly lack such traits). The infant would have to be someone else, a predecessor rather than oneself at an early age. This “newborn problem” suggests that person essentialism is untenable.³

Another possible problem with person essentialism concerns the other end of life, when our cognitive capacities may diminish gradually rather than suddenly. If person essentialism is true, then a person who is undergoing the progressive cognitive decline associated with Alzheimer’s disease would go out of existence whenever narrative capacity is lost during the cognitive decline. Yet, at this time, a sentient or conscious human patient continues to live.⁴ Perhaps some will accept the implication that they could go out of existence during the gradual cognitive decline despite the continuing existence of a sentient individual. But many will find this very difficult to believe. They would judge that it would be reasonable for someone in the early stages of progressive dementia to *fear* (for their own sake) reaching the late stages, implying the judgment that they would *still exist* in the late stages of dementia. For those who think along these lines, person essentialism is unacceptable not only due to the newborn problem but also due to the “dementia problem.”

9.3 The Biological Account

According to the biological account, human persons are most fundamentally living human animals (or organisms).⁵ The idea that we are fundamentally animals has the attraction of cohering with a world view

³ David DeGrazia, *Human Identity and Bioethics* (Cambridge: Cambridge University Press, 2005), chap. 2.

⁴ This point is explored in Jeff McMahan, *The Ethics of Killing* (New York: Oxford University Press, 2002), 47–48.

⁵ For the preeminent defense of this position, see Eric Olson, *The Human Animal* (New York: Oxford University Press, 1997).

according to which we are part of the natural world. In contrast to person-based accounts, the biological account implies that our existence and persistence conditions are biological, not psychological. Thus, we come into existence either at conception or shortly thereafter (when integration of embryonic parts makes spontaneous twinning impossible⁶), and we go out of existence when the human animal dies.⁷ Within biological approaches there is debate about what counts as dying. Some argue that total brain failure is sufficient for a human being's death and others that irreversible loss of circulatory-respiratory function is also necessary. We examine these possibilities in a later section.

Proponents of the biological account tend to emphasize the difference between the metaphysical issue of our numerical identity and such value-based questions as what is most important in human life. When we come into existence is one thing; when we have moral status and a life worth continuing is another. Even if each of us lacks moral status as a sentient fetus, as many bioethicists (including us) assert, it would not follow that we did not *exist* as such fetuses. Similarly, when we die is a matter of the termination of biological life. But it might be possible to lose everything that we value in our lives before dying – for example, if consciousness is necessary for the things that we value. From the perspective of the biological account, it appears that psychological accounts of identity – both person-based accounts and the mind-based account to be discussed next – may be motivated, in part, by a conflation of two distinct issues: our identity and, to use Parfit's phrase, "what matters in survival."⁸

The biological account vindicates the claim that we are fundamentally animals while avoiding some questionable implications of other views. First, it avoids the implausible implication of person essentialism that we did not exist as infants. It also avoids the implication that a person would go out of existence during the gradual cognitive decline associated with progressive dementia despite the fact that a conscious patient would remain. Further, by stating that human persons *are* animals, it identifies the relationship between human persons and the animals associated with

⁶ Before significant integration exists among embryonic cells and while twinning remains possible, a particular human organism has yet to be uniquely individuated. One early embryo might yield one being of our kind, a human organism, or it might yield two.

⁷ This assumption is common among personal identity theorists. However, there are dissenters. For example, W. R. Carter argues that we are essentially animals but not necessarily living animals, so we continue to exist as dead animals – corpses – so long as the remains constitute a single object ("Will I Be a Dead Person?," *Philosophy and Phenomenological Research* 59 [1999]: 167–171).

⁸ Derek Parfit, *Reasons and Persons* (Oxford: Clarendon, 1984), Part III.

them more straightforwardly and – some may think – more plausibly than the mind essentialism we discuss in the next section. Some might also find it an advantage of the biological account that it conceptualizes human death in biological terms, whereas mind-based accounts of personal identity understand human death as the loss of some psychological capacity – despite the fact that life and death are, first and foremost, biological phenomena.

Nevertheless, the biological view faces challenges. The most significant challenge, we think, concerns the hypothetical cerebrum-transplant case. In this thought-experiment, Alfred's body remains alive while his cerebrum (the large outer portion of the brain in which the contents of mental life are realized and encoded) is surgically extracted and transplanted, still fully functional, into the head of Bill, whose own cerebrum had been removed while his body remained alive.⁹ There are now both a living, but mindless, human organism associated with Alfred's body and a person associated with Bill's body. That person's mental life is continuous with presurgery Alfred's mental life: the same values and life plans as Alfred had along with his apparent presurgical memories. Following the transplant, where is the individual known as Alfred? Many have the intuition that Alfred has moved, along with his mental life, to Bill's body. Yet the biological view suggests that, because Alfred is essentially a living human animal, he remains alive albeit mindless in his original body – so that the person in Bill's body is *wrong* in believing himself to be Alfred.

While this counterintuitive result challenges the biological view, the challenge is not decisive. For one thing, in the cases that confront us *in everyday life* the biological view has very plausible implications: we track the existence and persistence of human persons in ways that are consistent with the biological understanding of numerical identity. Second, one might argue that the common intuitive reaction that Alfred moves with his cerebrum into Bill's body is a result of overgeneralization from ordinary cases. In all cases with which we are familiar, continuity of a mental life with particular memories, values, and plans coincides with continuity of a particular human person. However, that may be because, in familiar cases, continuity of mental life has always *also* tracked a continuing biological life. In responding to the cerebrum-transplant case, in which the contents of a mental life are separated from the individual's body and realized in a

⁹ If brain parts in addition to the cerebrum are, in fact, implicated in the encoding and realization of the contents of a person's mental life, the thought-experiment may be modified to include those brain parts in the transplant.

different body, we might overgeneralize from familiar cases if we judge that the individual must go with the contents of his earlier mental life. Instead, we might think that Alfred remains with his still-living body, especially if his cerebrum, lodged in an entirely different body, generates a personality and mental life that (contrary to the thought-experiment) are significantly different from those associated with Alfred. In our judgment, the biological approach to personal identity remains a contender.

9.4 A Mind-Based Account

According to the mind-based account of personal identity, human persons are essentially beings with minds.¹⁰ The relevant sense of “mind” here is broad and inclusive; it is sufficient for having a mind in this sense that one has the capacity for any conscious states – any subjective experience – whatsoever. Such *mind essentialism* claims in effect that we are essentially conscious or sentient beings.¹¹ We come into existence whenever the developing human organism first possesses a mind and we continue to exist for as long as that organism has a mind. More specifically, this approach implies that we come into being as late fetuses, whenever sentience emerges (see Chapter 10), and that we go out of existence at the irreversible loss of the capacity for consciousness. So the mind-based account avoids the newborn problem that undermines person essentialism: the implausible implication that we come into existence postnatally. It also avoids the dementia problem associated with person essentialism: the implication that one goes out of existence at some point during the cognitive decline associated with progressive dementia, prior to the onset of irreversible unconsciousness. A further strength of this account is that it avoids the cerebrum-transplant intuition that challenges the biological account by plausibly implying that Alfred would go with his cerebrum – and mind – into Bill’s body. But the mind-based account faces a substantial challenge.

The challenge is to provide a satisfactory explanation of the relationship between the “minded being” – that is, one of us – and the closely associated human organism. According to the mind-based view, a human organism comes into existence at conception (or perhaps shortly afterward,

¹⁰ This account is developed in McMahan, *The Ethics of Killing*, chap. 1.

¹¹ Sentient beings have the capacity to experience feelings such as pain or pleasure. In principle, there could be minded beings that could experience only conscious states that were not feelings, but as far as we know all actual animals, including humans, who have minds experience feelings and are therefore sentient.

as explained earlier), gradually develops in utero, and then, when the fetus becomes sentient, one of us comes into being. At the other end of life, if the human organism enters an irreversibly unconscious state before dying, the minded being goes out of existence although the human organism lives on. So what is the relationship between one of us and the longer-living human organism?

According to the mind-based view, you are not identical to a human animal because the two of you have different persistence conditions: The animal can exist at times when you do not. Thus, the mind-based view implies that, strictly speaking, we are not animals but rather their minds. And what are minds? Most philosophers and scientists today have a naturalistic understanding of minds as inseparable from brains. If we assume, along these lines, that minds are equivalent to brains – or, more precisely, those parts of functioning brains in which conscious states are realized – then, according to the mind-based view, *we* are parts of functioning brains.¹² So the answer to the question of how we relate to the human organisms associated with us is that each of us is *part of* such an organism, which came into existence before we did (since the organism existed before developing a brain) and may outlive us (if our brains irreversibly cease to function while the organism remains alive). Yet this cannot be the whole story: brain parts can continue to exist after irreversibly losing the capacity to function. This suggests that we cannot simply be those brain parts, but must be – or be closely associated with – those brain parts *when they are functional*.

Perhaps, then, we – as minds – are *realized* in functioning brains. But what does this realization consist in? One might hold that we are immaterial substances that are intimately related to our working brains.¹³ But substance dualism is deeply problematic. For example, it leaves the nature of causal interactions between an immaterial mind and the material brain completely mysterious.¹⁴ Another possible suggestion, exploiting an

¹² See McMahan, *The Ethics of Killing*, chap. 1.

¹³ See, especially, Rene Descartes, *Meditations on First Philosophy* (1641).

¹⁴ Another major difficulty with substance dualism, which is rarely noted, involves what happens to consciousness in “split brain” patients. As a treatment for severe epilepsy, certain patients have had their corpus callosa (singular: collosum) – neural connections joining the two cerebral hemispheres – severed. Under experimental conditions it becomes clear that the surgery has caused the conscious life of each patient to divide into two independent spheres of consciousness – in brief, has caused one mind to divide into two (see R. W. Sperry, “Hemisphere Deconnection and Unity in Conscious Awareness,” *American Psychologist* 23 [1968]: 723–733). But, according to substance dualism, the mind is an *immaterial* substance. Why would severing connections between two cerebral hemispheres have the effect of dividing one

analogy with computers, is that we are like the software that permits the hardware of the brain to operate. But this would imply that we are abstract beings that are incapable of existing concretely without an underlying physical substrate – just as a particular shape has no concrete existence without things that possess the shape. The analogy to software also suggests that each of us, as a mind, could enable mental life in an indefinite number of human bodies just as Windows software can enable an indefinite number of computers to operate. The analogy to software seems unpromising.

A more promising variant of mind essentialism asserts that each of us is *constituted* by a human organism. The idea of constitution may be understood with the example of a bronze statue. Once a hunk of bronze is sculpted into the shape of a statue, the statue exists and is constituted by its material substrate, the hunk of bronze. The hunk of bronze exists before the statue does, may outlast the statue if the latter is melted down, and coincides or overlaps with the statue when it has the shape of a statue. In brief, the hunk of bronze constitutes the bronze statue when and only when the hunk has the relevant shape. Similarly, according to the present suggestion, one of us is constituted by a human animal when and only when the animal has the capacity for consciousness. Perhaps the constitution thesis offers a plausible answer to the question of the relationship between one of us and the associated human animal. In any case it is clear that the mind-based account, like the biological account, is a contender among theories of personal identity.

9.5 A Social Account

One further theoretical approach deserves our attention. This account is a variant of person essentialism, but, by construing numerical identity in partly social terms, it may be able to sidestep the newborn problem and the dementia problem we described above.¹⁵

On this social account, the lives of persons feature a cluster of properties and relations that ordinarily support each other. A paradigmatic human life features a characteristic developmental trajectory that can be understood in terms of typical human traits and capacities, activities and social

immaterial substance into two? And is an immaterial substance even a plausible candidate for something that can divide?

¹⁵ For a recent elaboration of this general approach, see Marya Schechtman, *Staying Alive* (Oxford: Oxford University Press, 2014). See also Hilde Lindemann Nelson, *Damaged Identities, Narrative Repair* (Ithaca, NY: Cornell University Press, 2001), chap. 3.

relations, and status within a community of persons.¹⁶ Most human persons come into being as a member of a particular family; develop into individuals who can speak, socially interact, plan, and play; and are recognized as members of a community and various smaller social groups. They continue to mature physically before beginning a gradual physical decline, which continues until they die in relatively old age. While these general features are *characteristic* of the lives of human persons, on this account, they are not all strictly *necessary* for a human to qualify as a person. Someone might, for example, suffer from cognitive disabilities that prevent some human-typical forms of psychological and social life. Nevertheless, if she has several of the other person-characteristic traits and is an *identifiable locus of social interaction within a community of persons*, then she is a person.¹⁷ Thus, on this account, our numerical identity is a function of having and sustaining the life of a person, which means existing and persisting as an identifiable locus of social interaction.

This account integrates narrative identity into its account of numerical identity. It does so by way of two assertions: (1) that we are essentially persons and that persons are *characteristically*, though not necessarily, individuals who possess self-narratives and (2) that the relevant sorts of narratives are not only *internal* to an individual person – as suggested by the term “self-narrative” – but also *social* in that other persons’ understandings of one’s life (e.g., “This boy is the son of so-and-so”) are part of one’s broader personal narrative. An important implication is that your existence begins long before you are able to construct a self-narrative. In our society it begins no later than birth, when one becomes a locus of social interaction and the subject of a personal narrative (commenced by other individuals). And your existence may continue after you have lost narrative capacity, say, due to advanced dementia. The narrating work, in this case, would be continued by family members, hospital staff, and the state, all of whom continue to regard you as a living member of a family (someone to visit and talk to, for example) and of the broader community (someone who maintains the legal status of a person). According to the social account, then, the severely demented patient persists as a person until the time of what is socially recognized as death.

The social account avoids the implication of standard versions of person essentialism that we never existed as newborns. It also plausibly allows that we could survive into a dementia that destroyed our narrative capacity and ability to plan. Moreover, it offers a possible response to the problem of

¹⁶ Schechtman, *Staying Alive*, chap. 5. ¹⁷ *Ibid.*

nonparadigm humans (as discussed in Chapter 7): by implying that newborns and the severely demented are persons, it implies that they have full moral status.

Despite such attractions we find this novel account of personal identity inadequate. Consider its basis for assigning moral status. In the case of an infant or deeply demented human being, who lacks their own narrative capacity, the view permits other individuals to recognize the individual as a person, to do the “narrating work” for them so to speak, thereby socially conferring personhood on the individual. We do not believe that moral status can depend on social recognition or special relationships. As discussed in Chapter 7, the concept of moral status attributes a type of inherent value to individuals such that moral agents have obligations regarding their treatment of those individuals. But inherent value must be *inherent*, resting on the valuable entity’s characteristics. By contrast, relationships are “external” to an individual rather than inherent. So relationships cannot be a basis for moral status. This judgment fits with the plausible idea that a being’s moral status gives moral reasons to *all moral agents* to treat that being in particular ways, whereas special relationships would provide moral reasons only to those who stand within the relationships. In addition to this conceptual point, there is a substantive moral concern: social recognition, by depending on the perceptions and perhaps choices of others, seems too unstable and contingent to be a plausible basis for moral status.

In addition, social recognition is implausibly unstable as a basis for numerical identity. Consider a society that did *not* regard newborn infants as persons and members of the community. On the conception of numerical (and narrative) identity that understands it as constituted by one’s self-conception and others’ social recognition there would be no grounding for newborns’ numerical identity. This implies, implausibly, that in the imagined society those newborns who survive and later “grow into” persons with narrative capacity did not actually exist as newborns. That is, in such a society, it would be true of children who are persons that they had not existed as newborns, who were instead numerically distinct predecessors. So the newborn problem would apply to human beings in such a society.

Personhood, numerical identity, and moral status must rest on the properties of individuals, not on their relationships to others or on social recognition more generally. The social account of numerical identity fails to meet this requirement. So, despite its apparent advantages, we do not consider it a viable contender among accounts of numerical identity.

9.6 Our Approach to Personal Identity

As previously noted, we reject person-based accounts of numerical identity. Standard person-based accounts have deeply counterintuitive implications about our identity as found in the newborn problem and the dementia problem. We also reject the social account, a variant of the person-based approach, since it makes some individuals' personhood and moral status – and, in some possible societies, even their existence as newborns – depend on social recognition and special relationships.

Our discussion of the biological view, according to which we are essentially living human animals, noted several advantages over person-based accounts and no fatal weaknesses. We reached a similar verdict for the mind-based approach, according to which human persons are essentially beings with the capacity for consciousness. Therefore, as far as numerical identity is concerned, our approach is pluralistic: we regard both the mind-based account and the biological account as reasonable options, so our policies and practices should be as consistent as possible with each. On either account, we are essentially living beings – assuming, as we do, that we must be alive in order to have a mind.¹⁸ Meanwhile, we regard narrative identity as closely connected with much of what human persons care about in their lives. That is because we care about – even if not *only* about – our existence as persons; and in our view persons are beings with narrative capacity and an (internal) narrative identity.

9.7 Personal Identity and Human Enhancement

Background

Human enhancement through biomedical means has received ample scholarly attention in recent decades. There is a substantial literature on such topics as cosmetic surgery, the pharmacological enhancement of cognition and mood, doping in athletics, and genetic enhancement.¹⁹ There is even a literature on “moral bioenhancement,” the aim of which is to improve, through biomedical means, moral capacities such as those

¹⁸ This may not be true for advanced forms of artificial intelligence, but “we” here refers to human persons.

¹⁹ Helpful introductions to biomedical enhancement include Erik Parens (ed.), *Enhancing Human Traits* (Washington, DC: Georgetown University Press, 1998); and Julian Savulescu and Nick Bostrom (eds.), *Human Enhancement* (Oxford: Oxford University Press, 2009).

for sympathy and fairness.²⁰ The focus in this section is a particular set of concerns about biomedical enhancements *in relation to personal identity*.

Before turning to those concerns, it is worth noting the range of ways in which enhancement is understood. Many scholarly discussions of biomedical enhancement define it by way of contrast with medical treatment or therapy: enhancements are interventions designed to improve human form or function without responding to genuine medical need.²¹ So treatment endeavors to restore health or normal functioning, in keeping with the traditional objectives of medicine, whereas enhancement aims to take an individual beyond what is needed for health or normal functioning. The medication modafinil, for example, is used to treat the sleep disorder narcolepsy, but it can also be used to enhance academic performance by reducing the (normal) need for sleep and improving cognitive focus. Defining enhancement by contrasting it with treatment invites concerns about whether enhancement lies beyond the proper scope of medicine.

Some authors, however, doubt that the distinction between promoting health and normal functioning and seeking to transcend these standards has any fundamental importance. Accordingly, they define enhancement independently of any contrast with therapy. For example, Allen Buchanan conceptualizes enhancement as any deliberate intervention that aims to improve an existing capacity or create a new capacity in a human being.²² This broad definition includes such nonbiomedical enhancements as education and athletic training. Our purposes do not require deciding between traditional and nontraditional definitions of enhancement, for the interventions that are likely to provoke concerns about personal identity – concerns about a kind of death or loss of oneself – are also likely to satisfy any reasonable definition of enhancement.

Concerns about Numerical Identity

Substantial, deliberate endeavors to improve oneself – which we may call *enhancement projects* – appear to be connected with personal identity because they concern who someone is or who they want to become. The fact that Dasha's greatest dream is to become a highly accomplished writer

²⁰ For a helpful introduction, see Thomas Douglas, "The Morality of Moral Neuroenhancement," in Jens Clausen and Neil Levy (eds.), *Handbook of Neuroethics* (Dordrecht: Springer, 2015), 1227–1249.

²¹ See, e.g., Eric Juengst, "What Does 'Enhancement' Mean?," in Parens (ed.), *Enhancing Human Traits*, 25–43.

²² Allen Buchanan, *Beyond Humanity?* (New York: Oxford University Press, 2011), 23.

says much about who she is – that is, what sort of person – and concerns the sort of individual she wants to become. In some sense, then, Dasha's enhancement project involves her identity. But what sense of identity is at issue? Some scholars suggest that numerical identity is sometimes implicated. Walter Glannon, for example, expresses this thesis in the context of discussing genetic therapy:

[G]ene therapy designed to correct or treat a cognitive or affective disorder would be more likely [than gene therapy with no direct effects on mental life] to alter one's identity. The manipulation of the relevant neurotransmitters or regions of the brain that generate and support mental life would directly affect the very nature of the mental states definitive of personhood and personal identity through time.²³

Glannon applies this contention about identity to already-existing persons and not merely to fetuses and newborns. His reasoning would extend from therapy to enhancements with similarly far-reaching effects on one's mental life.

But the contention that enhancing an already-existing person's mental characteristics might result in a numerically distinct person is deeply implausible. Consider the implications of various contending accounts of numerical identity. If the biological account is correct, then the contention is obviously false, because a single human animal persists through the improvements of mental life. The contention is also false according to the mind-based account because a single embodied capacity for consciousness survives the changes associated with the enhancement project. Indeed, the contention is dubious even if one shares Glannon's assumption of person essentialism.²⁴ For the post-enhancement person will presumably *remember* pre-enhancement life, suggesting continuity of one and the same person-constituting mental life. Furthermore, most of the individual's attitudes and intentions are likely to survive the enhancement. In sum, there is no reason to think that enhancing an existing person's mental abilities would affect numerical identity. Apparently, Glannon has reasoned as follows: he intuitively appreciates that a substantial enhancement of mental characteristics would occasion a significant change in *narrative* identity, the person's self-conception; but, failing to distinguish the two types of identity, he fallaciously infers a change in numerical identity.²⁵

²³ Walter Glannon, *Genes and Future People* (Boulder, CO: Westview, 2001), 81–82.

²⁴ *Ibid.*, 25.

²⁵ In later work Glannon avoids this conflation (see *Bioethics and the Brain* [New York: Oxford University Press, 2007], 113–114).

Although a *person's* enhancement project will not alter their numerical identity, the enhancement of a human organism or of its constituent biological material *at a sufficiently early stage* could affect numerical identity. The timing would be sufficiently early if it occurred prior to the origination (the coming into being) of one of us. When – at what stage of development – one of us originates depends on what we essentially are. To focus our investigation, let's consider deliberately induced genetic changes with the intention of improving the cognitive capacities of the resulting individuals. If the mind-based account of numerical identity is correct, then we are essentially minded beings and come into being when the capacity for consciousness emerges – that is, sometime in the later months of pregnancy when the human organism becomes sentient. In this case, genetic interventions for the purpose of cognitive enhancement would affect identity if they occurred before the onset of sentience and the emergence of a minded being.²⁶ It seems plausible to judge that identity would be affected: an individual would originate who is numerically distinct from the one who would have originated had the genetic enhancement not taken place. Rather than affecting numerical identity by ending the existence of one of us, this sort of enhancement would determine that a different being comes into existence. Now, if the biological account is correct, then in order to affect numerical identity, a genetic enhancement would have to occur at an earlier stage (so that it precedes the origination of one of us). If we come into existence at conception, the critical changes would have to occur in the gametes prior to fertilization; if we come into being somewhat later – say, when the primitive streak forms and spontaneous twinning is precluded – then the critical changes could occur in the early embryo.

These reflections suggest that certain types of enhancement could affect numerical identity so long as they introduce a significant change – such as a nontrivial change in cognitive capacities – that would alter the “nature” of the individual who comes into existence. Does this observation suggest that some enhancements are morally troubling for reasons pertaining to identity? Emphasizing that we are still focusing on numerical identity, we answer negatively. That is because determining *which* particular human individual comes into existence is not a morally weighty matter (setting

²⁶ Here we make the simplifying assumption that the genetic intervention succeeds in making a change to someone's genome that will, in fact, have some significant effect on their mental life. It is debatable whether making a genetic (pre-origination) change that induced no significant change should count as affecting the numerical identity of the individual who comes into being.

aside cases in which such a choice is expected to have a substantial negative impact on quality of life).²⁷ One might think otherwise, holding that an individual who would have come into being were it not for the genetic enhancement of the pre-origination gametes, embryo, or fetus would be “cheated” out of existence and thereby wronged. But, surely, prior to the origination of one of us, there is no being with moral status. No such being can be cheated in any significant sense any more than potential persons who are never conceived – say, because two prospective parents decide not to attempt pregnancy – can be cheated. In the present set of cases, interventions that bring about enhancements affect numerical identity in a morally innocuous way.²⁸

Concerns about Narrative Identity

Despite the concerns expressed by various commentators,²⁹ biomedical enhancements either do not affect numerical identity or do so in a morally innocuous way. At the same time, enhancements are likely to affect narrative identity because an agent embarking on an enhancement project intends to change themselves and their life in a way that matters to them. Perhaps some such changes raise significant moral issues. If so, they apply only to persons, who have narrative identities, and not to embryos, fetuses, or infants, who do not.

In ordinary cases, it would seem that self-chosen enhancements that affect one’s sense of self and thereby narrative identity would be morally unproblematic – especially when we limit consideration to identity-related moral concerns – so long as the enhancements are genuinely *self*-chosen or pursued autonomously. For example, a young woman who is a talented but uncultivated athlete might decide (autonomously) to lift weights, improve her diet, and train intensely in her sport in order to become an outstanding athlete. Success in this enhancement project is likely to change

²⁷ See our discussion of the “nonidentity problem” in Chapter 10.

²⁸ An issue we have not considered here (because it does not focus on identity) is the concern of some disability advocates that preimplantation and prenatal genetic manipulations and genetic diagnosis – whether with therapeutic or enhancement-based intent – may express a devaluation of persons with disabilities. See, e.g., Adrienne Asch, “Reproductive Technology and Disability,” in Sherrill Cohen and Nadine Taub (eds.), *Reproductive Laws for the 1990s* (Clifton, NJ: Humana, 1989), 69–124. We believe that use of such genetic technologies need not express any disrespectful message. For elaboration, see David DeGrazia, *Creation Ethics* (New York: Oxford University Press, 2012), 102–106.

²⁹ For further examples of such concerns, see Carl Elliott, *A Philosophical Disease* (New York: Routledge, 1999), 28–29; and President’s Council on Bioethics, *Beyond Therapy* (Washington, DC: PCB, 2003), 300.

her sense of herself in a way that is important to her. But this feature of her endeavor does not make it morally problematic. Suppose her enhancement project used some sort of genetic intervention to bring about a greater capacity for growing muscle. Although the genetic means to her end would introduce some important ethical considerations – including concerns about safety and fair play – it would not introduce new identity-related concerns. Whether her means are traditional ones or exotic genetic means, her enhancement project will affect her narrative identity, but there seems to be nothing inherently objectionable about doing so.

In a provocative discussion, Farah Focquaert and Maartje Schermer argue that some types of moral enhancement – the attempt to improve an individual's moral capacities – could raise serious issues in connection with narrative identity.³⁰ The concerns they raise could also apply to other types of enhancement projects and to therapeutic interventions that use similar means to their respective enhancements. Among the means to moral enhancement, Focquaert and Schermer distinguish (1) direct interventions, which affect the brain directly and thereby indirectly affect the individual's way of thinking (e.g., deep brain stimulation) and (2) indirect interventions, which directly affect the agent's way of thinking, thereby indirectly affecting her brain (e.g., cognitive psychotherapy). They suggest that the direct/indirect distinction is important insofar as it tracks a distinction between relatively active and relatively passive roles of the agent. Someone who participates in psychotherapy, for example, is active in processing ideas and changing ways of thinking, whereas the recipient of deep brain stimulation assumes a considerably more passive role. Direct means and a correspondingly passive role for the individual might induce abrupt, profound psychological changes with little connection to her life story, threatening the coherence of her narrative identity. Additionally, the authors argue, there is the greater possibility of major changes in personality that go unnoticed by the enhanced individual, resulting in a kind of self-blindness and inauthenticity. Such disruptions in narrative identity or self-awareness, the authors imply, are likely to diminish the well-being of the individuals in question.

We agree with Focquaert and Schermer that moral enhancement might sometimes lead to harmful disruptions in narrative identity. At the same time, we find their response to the possibilities of such outcomes

³⁰ Farah Focquaert and Maartje Schermer, "Moral Enhancement: Do Means Matter Morally?," *Neuroethics* 8 (2015): 139–151.

satisfactory.³¹ As the authors propose, where direct interventions aimed at enhancement are involved, the individuals in question should undergo a very robust informed consent process that explores the possibilities of disrupted narrative identity or self-blindness. Moreover, during the intervention (if it is undertaken in steps and communication between the steps is feasible) and following it, the individuals should receive counseling that serves to minimize these possibilities. Such autonomy-promoting measures, we think, reduce the reasons for concern to tolerable levels – both by lowering the risks of narrative-identity-related harm and by increasing awareness of any residual risks so that individuals may accept or reject them in a well-informed manner.

9.8 The Authority of Advance Directives in Cases of Severe Dementia

Background

Advance directives are exceptionally important instruments for medical decision-making. They permit competent adult patients to provide instructions regarding their medical care in the event that they lose decision-making capacity in the future.³² In this way, advance directives allow the patient's proxy decision-makers and medical personnel to respect the patient's autonomy even though the patient lacks the capacity for autonomous choice at the time decisions are needed. It is sometimes said that advance directives facilitate respect for *precedent* autonomy – the autonomy one had during a prior period of time.

One concern about the employment of advance directives is that in certain cases in which a patient undergoes massive psychological change, the individual who exists following the change might be a different person than the one who earlier completed the directive. If there is in fact a disruption of numerical identity, there is substantial reason to challenge the authority of the directive in question, since an advance directive applies only to the individual who authored it. This is the “someone else problem” of advance directives.³³

³¹ *Ibid.*, 147–148.

³² They also permit such patients to designate proxy decision-makers in the event of losing decision-making capacity. But a proxy decision-maker needs some basis for deciding on behalf of the incompetent patient. Instructional advance directives supply such a basis, so our discussion focuses on this type of advance directive.

³³ David DeGrazia, “Advance Directives, Dementia, and ‘the Someone Else Problem,’” *Bioethics* 13 (1999): 373–391. See also DeGrazia, *Human Identity and Bioethics*, chap. 5.

Dementia comes in various forms, including Alzheimer's disease, vascular dementia, substance-induced persisting dementia, HIV-related dementia, and dementias caused by head trauma, Parkinson's disease, and Huntington's disease. The various forms of dementia share the following features:

1. Memory impairment
2. One or more of the following: (a) language disturbance, (b) impaired ability to carry out motor activities despite intact motor function, (c) failure to recognize objects despite intact sensory functioning, (d) disturbed executive functioning (planning, organizing, sequencing, etc.)
3. As a result of the above symptoms, significant impairment in social or occupational functioning that represents a substantial decline from the patient's baseline of functioning.³⁴

Here we focus on Alzheimer's disease, which features a fairly characteristic pattern of decline. In mild or early-stage Alzheimer's, symptoms are relatively subtle. For example, a patient may experience mild memory loss and feel unusually disorganized at times. Moderate or middle-stage Alzheimer's is characterized by more obvious symptoms such as frequent lapses in memory, losing valued objects, and more pronounced difficulty in performing tasks like bill-paying and planning a trip. Severe or late-stage Alzheimer's, by contrast, features profound loss of capacity. A patient might be nearly or entirely unable to speak, to perform activities of daily living such as getting dressed and self-feeding, and to recognize loved ones.

The Someone Else Problem and Numerical Identity

Suppose that Mary, at age sixty, is familiar with several cases of Alzheimer's disease in her extended family. After reading up on the disease, she thinks carefully about the possibility of succumbing to it and how she would want her life to go. Her conclusion is that, in view of the value she puts on intellectual pursuits and personal independence, she would not want to live for very long in a state of conscious obliviousness. Mary completes an advance directive that calls for withholding life-extending medical

³⁴ This summary draws from the American Psychological Association, *Diagnostic Criteria from DSM-IV* (Washington, DC: APA, 1994), 85–93. In the current, fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*, the term “dementia” is deemphasized while the various dementias are clustered – together with delirium, amnesia, and various other conditions – under the umbrella of “Major Neurocognitive Impairment” (Washington, DC: APA, 2013).

interventions, including artificial nutrition and hydration as well as antibiotics, in the event that she becomes incapable of (1) performing any activities of daily living (dressing, bathing, walking, toileting, etc.) and (2) recognizing loved ones. After a few years of “mild cognitive impairment,” Mary is diagnosed with early-stage Alzheimer’s disease at the age of sixty-five. The dementia progresses over the next ten years with the result that at age seventy-five Mary is severely demented and resides in a nursing home. She is now unable to perform independently any activities of daily living and shows no indications of recognizing loved ones. Nevertheless, she does not appear to be suffering. Indeed, she seems to enjoy being dressed, fed, wheeled around in her wheelchair, and spoken to by staff and visitors.

When Mary contracts pneumonia, which would ordinarily call for antibiotics, the nursing home staff consult her advance directive and revisit its instructions. A dispute erupts among the staff. Some believe that the directive straightforwardly applies and should be followed: no antibiotics, with the predictable result that Mary will die within days. Other staff members stress that the patient, who appears to have a relatively pleasant existence, is so psychologically cut off from the time of authoring the advance directive that it should no longer apply. She seems to have independent interests that favor life-extending antibiotic treatment. Who is right?

An advance directive is supposed to apply to decisions regarding medical care for the individual who completed the directive. Various commentators have suggested that a severely demented patient is not the individual who, while competent, expressed their medical preferences through the directive. Consequently, the advance directive should not apply. Some of these authors claim that the severely demented individual is *a different person* from the person who completed the directive.³⁵ Others, who regard persons as beings with the capacity for *relatively complex* forms of consciousness, have judged that someone with (sufficiently) severe dementia is not a person at all.³⁶ A and B cannot be the same person if B is not even a person. Given our analysis of personhood, we agree that a severely demented individual like Mary-at-75 is not a person and so cannot be a different person. But perhaps the nonperson is, in virtue of that fact, “someone else” from the competent individual who preceded her.

³⁵ See, e.g., Rebecca Dresser, “Life, Death, and Incompetent Patients: Conceptual Infirmities and Hidden Values in the Law,” *Arizona Law Review* 28 (1986): 373–405.

³⁶ See, e.g., Allen Buchanan, “Advance Directives and the Personal Identity Problem,” *Philosophy and Public Affairs* 17 (1988): 277–302.

The reasoning, applied to the present case, may be reconstructed this way:

1. Mary-at-60 (who is competent) is a person.
2. Mary-at-75 (who is very severely demented) is not a person.
Therefore:
3. Mary-at-75 is *numerically distinct from* Mary-at-60.

This reasoning is fallacious. The conclusion about numerical distinctness does not follow from premises (1) and (2). In order to infer the conclusion, one needs another premise: that Mary is *essentially* a person, so that the individual who completed the advance directive cannot exist at any time without being a person at that time. But the only reason to assume that Mary is essentially a person would be the general thesis of person essentialism, which earlier we found to be deeply implausible. What *is* plausible is that Mary existed for many years as a person but may survive for a time as a nonperson after losing traits such as narrative capacity.

Our pluralistic approach to numerical identity regards both the mind-based account and the biological account as plausible. Both vindicate the central metaphysical assumption underlying the presumed authority of advance directives in cases of severe dementia: that the author of the advance directive is numerically identical to – is the same individual as – the severely demented patient to whom the directive apparently applies. The mind-based account reaches the verdict because Mary-at-60 retains her basic capacity for consciousness throughout the cognitive decline leading to the condition of Mary-at-75. The biological account reaches this result because Mary-at-60 and Mary-at-75 are one and the same living animal. So the earlier advance directive applies to the older Mary. If there is any significant reason for doubting its authority, the reason will not concern numerical identity.

What about Narrative Identity?

In considering the possibility that personal identity theory challenges the authority of advance directives in certain cases, we have focused on numerical identity. With this focus, it appears that the someone else problem is misconceived. But *narrative* identity can challenge the authority of advance directives in a distinct way.

Consider Mary. While her numerical identity is maintained throughout the ravages of dementia, her narrative identity is disrupted. Mary's sense of herself as the protagonist of her self-narrative is entirely absent at age seventy-five. She is at most very weakly psychologically connected to the

life of the person who completed the directive, a past she does not remember at all. Although, strictly speaking, Mary has persisted into her present state, she has in an important sense become someone other than the earlier Mary. While the relevant sense of distinctness here is qualitative, not numerical, it captures a profound sort of qualitative change: the termination of one's self-narrative. Moreover, the loss of narrative capacity not only entails that she is no longer a person, as we have conceptualized personhood; it is precisely what destroys her awareness of being the same individual as Mary from the past. Such disruption of narrative identity, one might argue, threatens the authority of the earlier advance directive *just as if* the earlier person and present demented individual were numerically distinct. For all practical purposes, they might as well be distinct because the present patient is psychologically cut off from her past. We should not assume that the interests of deeply demented Mary are the same as those of the younger, competent Mary; nor should we assume that the younger person had authority to make decisions for the later demented woman who neither remembers nor cares about the values and priorities of the woman who completed the directive. Mary at age seventy-five has her own, relatively simple interests in states and activities that give her enjoyment – and these interests call for staying alive.

Despite some initial intuitive appeal, the narrative identity-based challenge to the authority of advance directives is problematic. In claiming disruption of narrative identity, the argument views identity *retrospectively* from the standpoint of the demented or irreversibly unconscious individual, rather than prospectively from the standpoint of the earlier person. The prospective standpoint is no less important. Individuals who actively contemplate the possibility of becoming severely demented or irreversibly unconscious and take these possibilities into account in completing advance directives presumably *identify*, in some sense, with such future possible stages of themselves and care about their fate during those stages. In the case of Mary, it is clear that the patient identifies in a relevant way with her possible future, deeply demented self because she realizes she might *become – transform into – such an individual* and decides that she does not want to live for long if that transformation occurs.

One might challenge the idea that the person who completes an advance directive and anticipates the possibility of severe dementia identifies with the later individual in any sense of “identification” richer than bare acknowledgment of numerical identity. One might, accordingly, deny that this very weak identification is enough to give the person authority over her later self. In response, we suggest distinguishing and examining the implications of two senses of narrative identity.

Bearing in mind that any sort of narrative identity presupposes numerical identity, we may distinguish strong and weak narrative identity.³⁷ *Strong* narrative identity involves one's continued existence with significant psychological unity and the capacity for appreciating one's unfolding narrative. Given our conception of personhood, it also involves one's continuing existence *as a person*. Strong narrative identity does not obtain between Mary-at-60 and Mary-at-75, because the latter clearly lacks narrative capacity. Now consider *weak* narrative identity: the relationship between a person at a particular time and possible future stages of herself that she (at that earlier time) consciously acknowledges as part of her life and cares about prudentially. Weak narrative identity obtains between Mary-at-60 and Mary-at-75. For Mary, at age sixty, completed an advance directive after carefully considering the possibility of becoming deeply demented and decided that she would not prefer to persist for long in such a state.

We contend that it would be unreasonable to require strong narrative identity as a condition for recognizing an advance directive's authority. After all, while Mary-at-60 is only weakly narratively identical to Mary-at-75, no other individual has a stronger claim of being identical to Mary-at-75.³⁸ Moreover, the very purpose of advance directives is to permit individuals to authorize, prospectively, the provision or withholding of certain kinds of medical treatment at a later time when they are incapacitated. A requirement of strong narrative identity would invalidate all use of advance directives when patients have lost narrative capacity. Such a radical revision to current practice and restriction in the scope of individual autonomy rights would require a powerful rationale. We do not believe any such rationale has been advanced.

We conclude that, contrary to some authors,³⁹ considerations of personal identity pose no significant threat to the presumptive authority of advance directives in cases featuring severe dementia.

³⁷ We borrow the distinction from DeGrazia, *Human Identity and Bioethics*, chap. 5.

³⁸ Thanks to David Benatar for suggesting this argument.

³⁹ See, e.g., Allen Buchanan and Dan Brock, *Deciding for Others* (Cambridge: Cambridge University Press, 1989), chap. 3; Rebecca Dresser, "Advance Directives, Self-Determination, and Personal Identity," in Chris Hacker, Ray Mosely, and Dorothy Vawter (eds.), *Advance Directives in Medicine* (New York: Praeger, 1989), 155–170; Jeffrey Blustein, "Choosing for Others as Continuing a Life Story: The Problem of Personal Identity Revisited," *Journal of Law, Medicine, and Ethics* 27 (1999): 20–31; and Mark Kuczewski, "Whose Will Is It Anyway? A Discussion of Advance Directives, Personal Identity, and Consensus in Medical Ethics," *Bioethics* 8 (1994): 27–48. Interestingly, while these authors incorrectly assert that cases of severe dementia feature a disruption of numerical identity, each ultimately vindicates the authority of advance directives.

9.9 The Definition and Determination of Human Death

Background

For the most part, questions about the definition and determination of human death did not get much public attention until well into the twentieth century.⁴⁰ Sufficient destruction of the brain, including the brain stem, ensured respiratory failure, which led quickly to terminal cardiac arrest. Conversely, prolonged cardiopulmonary failure led inevitably to total, irreversible loss of brain function. After the invention of mechanical respirators in the 1950s, however, it became possible for a previously lethal extent of brain damage to coexist with cardiopulmonary function, which sustained the functioning of other organs. Was a patient in this condition alive or dead? In the 1960s widespread dissemination of such technologies as respirators and defibrillators to restore cardiac function underscored the possibility of separating cardiopulmonary and neurological functioning. While these developments cleared conceptual space for a debate over the nature of human death, other developments motivated the debate in more practical terms.

Two developments are especially noteworthy. First, soaring medical expenditures and a limited supply of beds in intensive care units (ICUs) provoked concerns about prolonged, possibly futile treatment of patients who presented some but not all of the traditionally recognized markers of death. If the patient were dead, it would be permissible to discontinue life support and remove them from an ICU bed. Second, the evolving techniques of organ transplantation motivated physicians not to delay unnecessarily in determining that a patient had died. Moreover, if physicians were permitted to remove organs from patients whose cardiopulmonary function was sustained by respirators yet who were judged to be dead by neurological criteria, the well-perfused organs would be more likely to benefit recipients. Removing vital organs as quickly as possible, and with organs in the best condition possible, would improve the prospect of saving other patients' lives. But removing vital organs from living patients would cause them to die, violating laws against homicide as well as the commonly accepted moral rule prohibiting the intentional killing of innocent human beings (see Chapter 4).

⁴⁰ For a historical discussion, see M. Pernick, "Brain Death in a Cultural Context: The Reconstruction of Death, 1967–1981," in Stuart Youngner, Robert Arnold, and Renie Shapiro (eds.), *The Definition of Death* (Baltimore, MD: Johns Hopkins University Press, 1999), 3–33.

These developments set the stage for a landmark 1968 Harvard Medical School committee report, which supported a whole-brain standard of death.⁴¹ A 1981 US President's Commission report, *Defining Death*, recommended the Uniform Definition of Death Act (UDDA), which incorporated the whole-brain standard of death alongside the traditional cardiopulmonary standard.⁴² The UDDA states that "an individual who has sustained either (1) irreversible cessation of circulatory and respiratory functions or (2) irreversible cessation of all functions of the entire brain, including the brainstem, is dead."⁴³ Current law in the United States incorporates both standards of death, with most states adopting the UDDA, while others have embraced similar language. The United Kingdom and Canada both use a brainstem standard, which has virtually the same practical consequences as a whole-brain standard but requires fewer clinical tests.⁴⁴ Most countries around the world have accepted the whole-brain standard or the brainstem standard either by themselves or alongside the cardiopulmonary standard.⁴⁵

This legislative consensus, however, has not prevented physicians, lawyers, philosophers, and others from defending a variety of significantly different views about the nature of human death. These views often invoke, explicitly or implicitly, accounts of our numerical identity. Understanding and contributing to the ongoing debate requires understanding the implications of these competing accounts.

Implications of the Biological Account

The biological account of our numerical identity asserts that human persons are essentially living human animals and *not* essentially persons or even minded beings. Therefore, just as we came into existence prior to becoming conscious, we might continue to exist after irreversibly losing our mental lives. According to the organismic definition of death, which coheres with the biological account of our numerical identity, death

⁴¹ Ad Hoc Committee of the Harvard Medical School, "A Definition of Irreversible Coma – Report of the Ad Hoc Committee of Harvard Medical School to Examine the Definition of Brain Death," *JAMA* 205 (1968): 337–340.

⁴² President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research, *Defining Death* (Washington, DC: Government Printing Office, 1981).

⁴³ *Ibid.*, 119.

⁴⁴ Andrew McGee and Dale Gardiner, "Differences in the Definition of Brain Death and Their Legal Impact on Intensive Care Practice," *Anaesthesia* 74 (2019): 569–572. For elaboration on the brainstem standard, see Christopher Pallis, "On the Brainstem Criterion of Death," in Youngner, Arnold, and Shapiro (eds.), *The Definition of Death*, 93–100.

⁴⁵ Elco Wijdicks, "Brain Death Worldwide: Accepted Fact but No Global Consensus on Diagnostic Criteria," *Neurology* 58 (2002): 20–25.

involves the termination of biological life. Like other organisms, we die when the organism stops functioning as a more or less integrated unit. Whereas a living creature is a dynamic entity that extracts energy from the environment to maintain its own structure and functioning, death yields an inert piece (or pieces) of matter subject to disintegration and decay. In the case of humans, like other animals and plants, death involves the irreversible loss of integrated bodily functioning. While this organismic definition of death is relatively clear, its proper translation into a standard for the determination of human death is disputed.

According to the traditional *cardiopulmonary standard*, human death is the irreversible cessation of circulatory and respiratory function. According to the *whole-brain standard*, human death is the irreversible cessation of functioning of the entire brain, including the brain stem. Proponents of each standard generally maintain that only *it* correctly specifies an organismic definition of death for the case of human beings.⁴⁶

The difference between assisted and unassisted respiration is crucial in distinguishing these two standards. A mechanical respirator can enable a “brain-dead” patient – a patient whose entire brain is nonfunctional – to breathe and thereby sustain cardiac function. Yet such a patient is incapable of unassisted or spontaneous respiration. On the cardiopulmonary standard, such a patient counts as alive as long as breathing and circulation occur, no matter how they occur. But such a patient is dead according to the whole-brain standard. At the same time, this standard judges that a patient who is irreversibly unconscious yet retains some measure of brainstem function is alive.

Reviewing a few points of neurology may help to clarify the two standards’ distinct clinical implications. The human brain includes what might be called the “higher brain” – consisting of the cerebrum, the primary vehicle of consciousness, and the cerebellum, which enables the control and coordination of voluntary muscle movements – and the “lower brain” or brainstem. The brainstem includes the medulla, which controls spontaneous respiration, and the reticular activating system, a sort of on/off switch that makes consciousness possible without affecting its contents. Whole-brain death involves loss of function in both the higher brain and brainstem. In what is called a *persistent vegetative state* (PVS), while damage

⁴⁶ See, e.g., the arguments of Lawrence Becker (“Human Being: The Boundaries of the Concept,” *Philosophy and Public Affairs* 4 [1975]: 334–359), who defends the traditional standard, and those of James Bernat, Charles Culver, and Bernard Gert (“On the Definition and Criterion of Death,” *Annals of Internal Medicine* [94]: 398–394), who defend the whole-brain standard.

to the higher brain causes irreversible unconsciousness, a largely functional brainstem permits some or all of the following: spontaneous respiration and heartbeat, sleep and wake cycles (but without consciousness), eye movements and pupillary reactions to light, and such reflexes as swallowing and coughing. A patient in *irreversible coma*, by contrast, never appears to be awake due to a damaged reticular activating system but remains capable of unassisted breathing. The ability to breathe spontaneously in either PVS or irreversible coma – as well as the sleep/wake cycles and reflexes present in PVS – presents the impression that patients in these conditions are alive. Both the whole-brain and the cardiopulmonary standards affirm this judgment.

What considerations might favor one standard over the other? To begin, why did proponents of the whole-brain standard propose a departure from the traditional cardiopulmonary standard? There were two main reasons. The first was straightforwardly practical: the whole-brain standard would permit extraction of vital organs while a patient maintained cardiopulmonary function with mechanical assistance, meaning more viable organs would be available for transplant. Although it may seem strange that a criterion of human death could be motivated by the good consequences that would result from its acceptance, the seminal report by the Harvard Medical School committee contained no philosophical or conceptual justification at all for its recommendation of the whole-brain standard. By contrast, the later President's Commission report did offer such a justification – one that connected the organismic definition of death to the whole-brain standard by way of the thesis that the human brain is the *primary integrator* of overall bodily functioning. Accordingly, the argument goes, only loss of function of the entire brain is necessary and sufficient for human death. Life involves the integrated functioning of the whole organism. While circulation and respiration are centrally important, so are hormonal regulation, maintenance of body temperature, and various other functions – including consciousness in humans and many other animals. A central integrator – the brain – permits the integration of these vital functions.

Although the whole-brain standard has been recognized in jurisdictions around the world, it has always had detractors and has recently faced especially vigorous counterarguments. Some traditionalists never accepted the whole-brain standard, rejecting its practical advantages as irrelevant to the nature of death and maintaining that cardiopulmonary function, no matter how it is sustained, entails the continuation of life.⁴⁷ Most of the

⁴⁷ See, e.g., Hans Jonas, "Against the Stream: Comments on the Definition and Redefinition of Death," in Hans Jonas, *Philosophical Essays* (Chicago: University of Chicago Press, 1974), 132–140.

intense recent pushback against the standard, by contrast, has appealed to empirical evidence that lack of brain function is not sufficient for the irreversible collapse of integrated bodily functioning. Many of the human body's integrated functions, according to this challenge, are not mediated by the brain and can persist in individuals who satisfy whole-brain criteria for death by standard clinical tests. These functions include homeostasis, detoxification of cellular waste, assimilation of nutrients, wound healing, fighting infections, and hormonal stress responses to unanesthetized incisions; in a few cases, brain-dead bodies have even matured sexually, grown, or gestated a living fetus.⁴⁸

Champions of the whole-brain standard have replied to this and other challenges.⁴⁹ Rather than entering into the details of those replies or adjudicating between the two sides, we simply note that we find both the cardiopulmonary standard and the whole-brain standard reasonable specifications of the organismic definition of death – which, as stated earlier, is consonant with the biological account of our numerical identity. We therefore find the UDDA, which states that a human being has died upon satisfying either standard (whichever applies first), a sensible compromise from the standpoint of a biological understanding of human life, identity, and death. But consideration of a more progressive standard, discussed in the next section, will take us in a more radically pluralistic direction.

Implications of the Mind-Based Account

The mind-based account of identity asserts that human persons are essentially beings with the capacity for consciousness. This view implies that each of us came into existence when the developing fetus first became sentient and began to have subjective experiences such as tactile and auditory experiences. On the other side of human life, the mind-based view implies that we die – and go out of existence – upon the irreversible loss of the capacity for consciousness. This implication of the mind-based view of identity is known as the *higher-brain standard* of human death. No jurisdiction in the world has yet adopted this standard.

⁴⁸ D. A. Shewmon, "The Brain and Somatic Integration: Insights into the Standard Biological Rationale for Equating 'Brain Death' with Death," *Journal of Medicine and Philosophy* 26 (2001): 457–478.

⁴⁹ See, e.g., James Bernat, "A Defense of the Whole-Brain Concept of Death," *Hastings Center Report* 28 (1998): 14–23; and President's Council on Bioethics (PCB), *Controversies in the Determination of Death* (Washington, DC: PCB, 2008).

One common argument against the higher-brain standard appeals to the nearly irresistible intuition that a human body that can breathe unassisted is alive. Surely, the argument goes, if someone falls into an irreversible unconscious state but continues to breathe spontaneously, that individual cannot be dead. Indeed, it would seem grotesque to cremate or bury a breathing body even if it were irreversibly unconscious.

In response to this challenge, the mind-based account can assert that in cases of this sort there are in fact two deaths: the death of the person or minded being, who dies upon irreversibly losing their capacity for consciousness, and the death of the human organism that constituted them. In cases of the present sort, the person dies before the organism does. The organism, though still breathing, is insentient and lacks moral status. For this reason, it would not wrong the person whose body it was to cremate or bury the breathing body. Of course, since doing so would probably strike most people as grotesque, that would be sufficient reason to ensure that the body died too before cremation or burial.

Just as we find the mind-based account of our numerical identity fairly plausible, as discussed earlier, we believe that the higher-brain standard of death deserves a place next to the cardiopulmonary and whole-brain standards as among those criteria for human death that it is reasonable to believe.

Living with Pluralism about Standards of Death

Our view of numerical identity, which finds the biological and mind-based accounts about equally reasonable, leaves us with full-fledged pluralism regarding the three leading standards of human death.

Where does our pluralism leave us in terms of practical guidance? We believe it invites a reconsideration of the presumed practical significance of death. Earlier we noted that the modern debate over the nature of human death was driven both by technological advances that made it possible to separate cardiopulmonary and neurological function and by certain practical concerns. Let us revisit the latter. Clarity on the appropriate standard for death was supposed to help physicians decide whether it would be permissible (1) to discontinue life support for patients whose cardiopulmonary function was maintained artificially and (2) to harvest vital organs from the same class of patients (given their prior consent). We believe that affirmative answers to these questions may be justified *independently* of whether the patients are dead. So we recommend detaching the ethics of universal discontinuation of life support and of organ transplantation from

the determination of death. Both practices can be justified without any assumption that the patients in question have died.

Unilateral withdrawal of medical life support from patients who have entered an irreversible unconscious state (IUS) is justified on the basis of resource constraints and the near-impossibility of benefiting them in such a state. Patients who will never again return to consciousness cannot enjoy life or derive satisfaction from anything in it. The only way they can derive benefit is if some aim of theirs is satisfied despite their lack of awareness. Very few important aims can be satisfied by being kept on life support. We therefore think that it would be excessively demanding to expect hospitals and nursing homes to continue to dedicate their resources to preserving the lives of individuals who are irreversibly unconscious. In our view, it does not matter whether or not the patients have died according to the traditional or whole-brain standard of death (although they will necessarily have satisfied the higher-brain standard). Other considerations justify unilateral withdrawal of life support.

In parallel fashion, the harvesting of vital organs for transplant is justified by ethical considerations other than whether or not the patients have died. It is justified by the prior valid consent of the patient to donate upon reaching a particular medical condition (e.g., an IUS, whole-brain death) and by the fact that, from the standpoint of this patient's values, continuing existence is less important than donating vital organs. In brief, donation is justified by respect for autonomy rights and is not prohibited by nonmaleficence.⁵⁰

What is the upshot of our approach for laws concerning the determination of death? Two broad possibilities suggest themselves. One possibility is to leave current laws for determining death as they are, even if they fail to acknowledge the higher-brain standard as a reasonable option. If these laws remain in place, then it is certain practices that have been traditionally associated with death – unilateral withdrawal of life support, organ donation, and perhaps others – that should be extricated from the need for determinations of death. For example, the extraction of vital organs from still-living patients causes them to die, violating laws against homicide. Open practice of such organ procurement may therefore require modification of these laws. A second, more radical possibility would be to change laws about death in a pluralistic direction. For example, one standard – say, the whole-brain criterion – might serve as a legal default,

⁵⁰ Cf. Franklin Miller and Robert Truog, "Rethinking the Ethics of Vital Organ Donations," *Hastings Center Report* 38 (6) (2008): 38–46.

from which competent individuals could depart in selecting a different standard for themselves or for their legal dependents. In effect, the standard for your death would become a matter of your choice (within the reasonable options we have discussed). While moving in this radical, pluralistic direction might have unique advantages insofar as people might consider what constitutes their own death a highly personal matter, one worth troubling over, this move would also be costly in terms of the multiplicity of laws that would have to change and would no doubt prove confusing to many people. Given the recognition that death has less practical importance than we have traditionally assumed, it might be preferable to limit our changes to modifying practices such as those we have discussed.